



# Critical Issues in Mycobiota Analysis

**Bettina Halwachs**<sup>1,2,3\*</sup>, **Nandhitha Madhusudhan**<sup>1,2</sup>, **Robert Krause**<sup>4</sup>, **R. Henrik Nilsson**<sup>5</sup>, **Christine Moissl-Eichinger**<sup>3,4</sup>, **Christoph Högenauer**<sup>2,3,6</sup>, **Gerhard G. Thallinger**<sup>3,7</sup> and **Gregor Gorkiewicz**<sup>1,2,3\*</sup>

<sup>1</sup> Institute of Pathology, Medical University of Graz, Graz, Austria, <sup>2</sup> Theodor Escherich Laboratory for Medical Microbiome Research, Medical University of Graz, Graz, Austria, <sup>3</sup> BioTechMed-Graz, Interuniversity Cooperation, Graz, Austria, <sup>4</sup> Section of Infectious Diseases and Tropical Medicine, Department of Internal Medicine, Medical University of Graz, Graz, Austria, <sup>5</sup> Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden, <sup>6</sup> Division of Gastroenterology and Hepatology, Department of Internal Medicine, Medical University of Graz, Graz, Austria, <sup>7</sup> Institute of Molecular Biotechnology, Graz University of Technology, Graz, Austria

## OPEN ACCESS

### Edited by:

David Berry,  
University of Vienna, Austria

### Reviewed by:

Carlotta De Filippo,  
National Research Council, Italy  
Micah Egge Dunthorn,  
Kaiserslautern University of  
Technology, Germany

### \*Correspondence:

Bettina Halwachs  
bettina.halwachs@medunigraz.at  
Gregor Gorkiewicz  
gregor.gorkiewicz@medunigraz.at

### Specialty section:

This article was submitted to  
Microbial Symbioses,  
a section of the journal  
Frontiers in Microbiology

**Received:** 20 May 2016

**Accepted:** 24 January 2017

**Published:** 14 February 2017

### Citation:

Halwachs B, Madhusudhan N,  
Krause R, Nilsson RH,  
Moissl-Eichinger C, Högenauer C,  
Thallinger GG and Gorkiewicz G  
(2017) Critical Issues in Mycobiota  
Analysis. *Front. Microbiol.* 8:180.  
doi: 10.3389/fmicb.2017.00180

Fungi constitute an important part of the human microbiota and they play a significant role for health and disease development. Advancements made in the culture-independent analysis of microbial communities have broadened our understanding of the mycobiota, however, microbiota analysis tools have been mainly developed for bacteria (e.g., targeting the 16S rRNA gene) and they often fall short if applied to fungal marker-gene based investigations (i.e., internal transcribed spacers, ITS). In the current paper we discuss all major steps of a fungal amplicon analysis starting with DNA extraction from specimens up to bioinformatics analyses of next-generation sequencing data. Specific points are discussed at each step and special emphasis is placed on the bioinformatics challenges emerging during operational taxonomic unit (OTU) picking, a critical step in mycobiota analysis. By using an *in silico* ITS1 mock community we demonstrate that standard analysis pipelines fall short if used with default settings showing erroneous fungal community representations. We highlight that switching OTU picking to a closed reference approach greatly enhances performance. Finally, recommendations are given on how to perform ITS based mycobiota analysis with the currently available measures.

**Keywords:** microbiota, mycobiota, internal transcribed spacer (ITS), 16S rRNA gene, multiple sequence alignment (MSA), OTU picking, formalin-fixed paraffin-embedded tissue (FFPE), DNA isolation

## INTRODUCTION

It is now well-established that the microbiota contributes significantly to human health and disease. So far, microbiota investigations have been mainly focused on bacteria, but also archaea, viruses, and micro-eukaryotes such as protozoa and fungi are part of human-associated microbial communities. Fungi are prevalent in all microbially colonized body habitats including skin, the gastrointestinal (GI)-, urogenital-, and respiratory tract (Charlson et al., 2012; Findley et al., 2013; Hallen-Adams et al., 2015). Up to now more than 390 fungal species have been described in humans (Oever and Netea, 2014; Gouba and Drancourt, 2015). Depending on the habitat the abundance of fungal cells varies from <0.1% of microorganisms in the GI tract to up to 10% on skin (Belkaid and Naik, 2013). An average fungal cell is about 100-fold larger than an average bacterial cell, which translates into a significant fungal biomass, providing abundant bioactive molecules to the host and shaping its physiology (Underhill and Iliev, 2014). The GI mycobiota actively interacts with the immune system, for instance through the human innate immune receptor Dectin-1 able to dampen GI inflammation (Iliev et al., 2012). A balanced mycobiota prevents from hyperinflammation of the

GI tract and alterations in fungal community composition due to antifungal drugs exacerbate colitis in mice (Wheeler et al., 2016). In humans genetic defects in certain immune-regulatory genes (e.g., *STAT1*, *CARD9*, etc.) or Il-17 and Il-22 signaling pathways lead to severe fungal syndromic infections, such as chronic mucocutaneous candidiasis or the APECED (Autoimmune Polyendocrinopathy, Candidiasis, Ectodermal Dystrophy) syndrome (Oh et al., 2013; Underhill and Iliev, 2014). Compositional mycobiota shifts are reported in various diseases (Cui et al., 2013) and also interdependencies between the fungal and bacterial component of the microbiota exist. They are exemplified by disease-specific inter-kingdom alterations, reported for instance in inflammatory bowel disease (IBD, Ott et al., 2008; Hoarau et al., 2016; Sokol et al., 2016) or in the lung microbiome of cystic fibrosis patients (Kim et al., 2015). Importantly, fungi contribute significantly to human infections, especially in immune-compromised, chronically ill and intensive care patients wherein the respiratory or GI tract are often the origins of fungal systemic infections (Brown et al., 2012; Krause et al., 2016).

## INTERNAL TRANSCRIBED SPACERS (ITS) AS FUNGAL MOLECULAR BARCODES

Currently, amplicon-based next generation sequencing is the standard measure for the culture-independent assessment of the mycobiota. Also metagenomic approaches are increasingly used, providing functional insights into the mycobiota. However, their broad application is still too costly due to the required sequencing effort to capture the relatively rare fungal biosphere and the special needs for bioinformatics analysis paired with underdeveloped fungal reference genome databases make metagenomics approaches still cumbersome (Tang et al., 2015). Early culture-independent mycobiota investigations used the eukaryotic 18S ribosomal RNA gene, in analogy to the prokaryotic 16S rRNA gene, as molecular target enabling PCR amplification of fungal DNA and subsequent taxonomic profiling via sequence analysis (Simon et al., 1992; Kappe et al., 1996; Smit et al., 1999; Hunt et al., 2004). The 18S rRNA gene, however, is less discriminatory for fungi compared to its prokaryotic equivalent often failing to discriminate fungi at lower taxonomic levels, such as genus or species (Hartmann et al., 2010; Lindahl et al., 2013).

The prokaryotic and the eukaryotic rRNA operons exhibit different genetic architectures (**Figures 1A,B**). The eukaryotic rRNA cistron consists of the 18S (small subunit, SSU), 5.8S, and 28S (large subunit, LSU) rRNA genes transcribed as a unit by RNA polymerase I, including two internal transcribed spacer regions, ITS1 and ITS2, flanking the 5.8S rRNA gene. The two ITS regions are post-transcriptionally removed and are absent in the mature ribosome. Since they are dispensable for ribosome function, they experience a lower evolutionary pressure leading to higher sequence variability (**Figures 1C–E**). The increased level of sequence variability enables discrimination of even closely related taxa (e.g., at species level). In addition ITS sequences seem to represent superior molecular targets for fungal PCR amplification compared to SSU and LSU sequences,

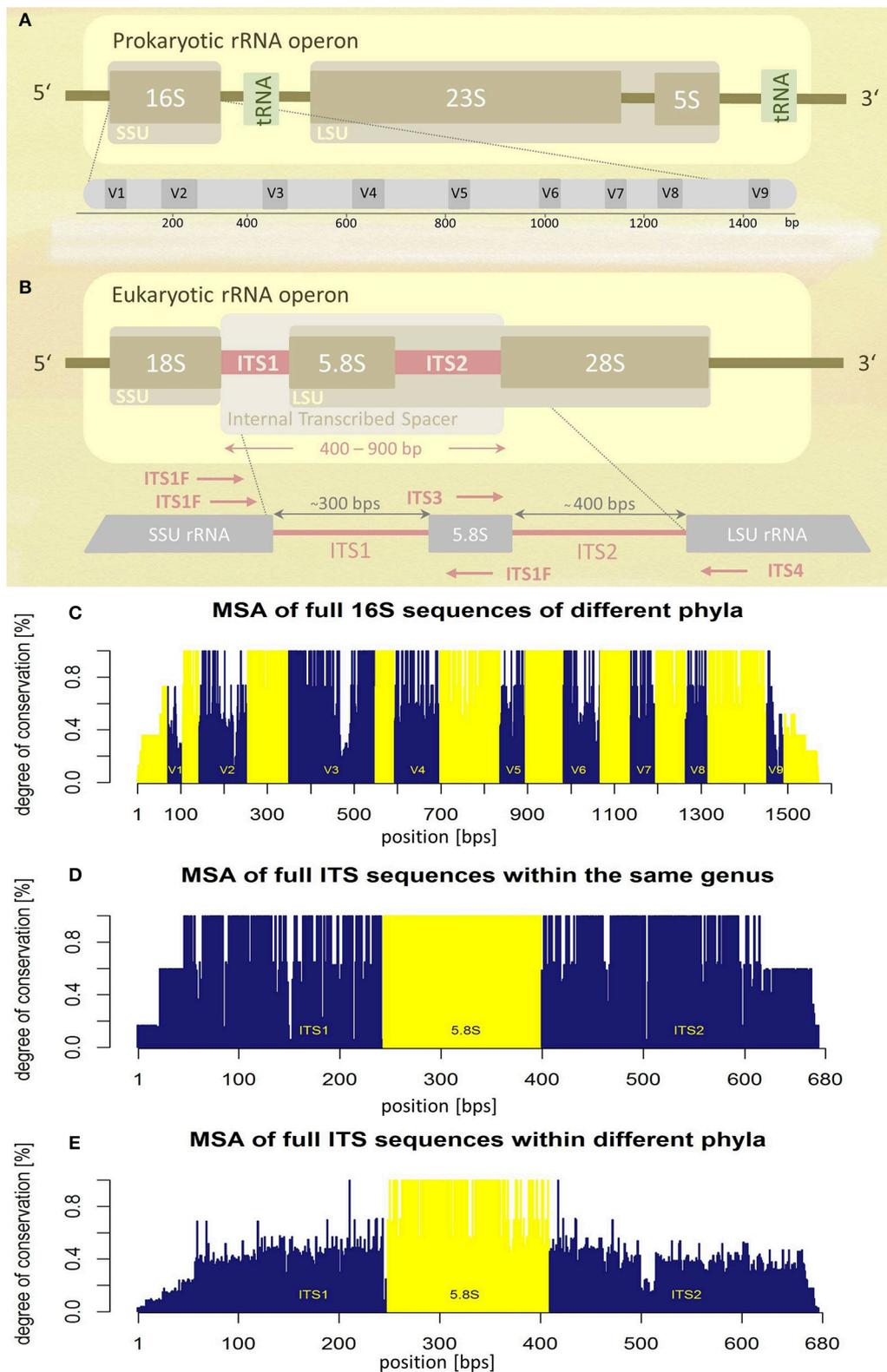
signified by higher positive PCR amplification rates (Schoch et al., 2012). Based on these observations, the Fungal Barcoding Consortium recently denoted the ITS region as the universal barcode for fungi superior to other molecular markers (Schoch et al., 2012).

In the following sections, we discuss the main steps of amplicon-based mycobiota analyses with special emphasis on the bioinformatics challenges emerging if standard bioinformatics analysis pipelines such as mothur, QIIME, or MICCA are employed (Schloss et al., 2009; Caporaso et al., 2010; Albanese et al., 2015).

## FUNGAL DNA ISOLATION

A variety of studies have shown that DNA isolation methods and oligonucleotide primer choice significantly influence the outcome of molecular phylogenetic surveys (Gorkiewicz et al., 2013; Tedersoo et al., 2014; Hallen-Adams et al., 2015). Numerous protocols and kits are available for isolation of fungal DNA and they follow similar basic principles with slight modifications dependent on the specimen type used (Paulino et al., 2006; Ghannoum et al., 2010; Findley et al., 2013; Lindahl et al., 2013; Gosiewski et al., 2014; Oh et al., 2014). The basic protocol involves mechanical cell disruption using bead beating, followed by enzymatic cell lysis. Especially the addition of lyticase, and endoglucanase hydrolyzing the covalent bonds between  $\beta$ -(1-3)-D-glucose molecules in the fungal cell-wall glycan, is an essential step to enable complete fungal cell lysis (Muñoz-Cadavid et al., 2010; Goldschmidt et al., 2014). The final DNA purification step is often performed by using membrane-based procedures (van Burik et al., 1998; Lindahl et al., 2013).

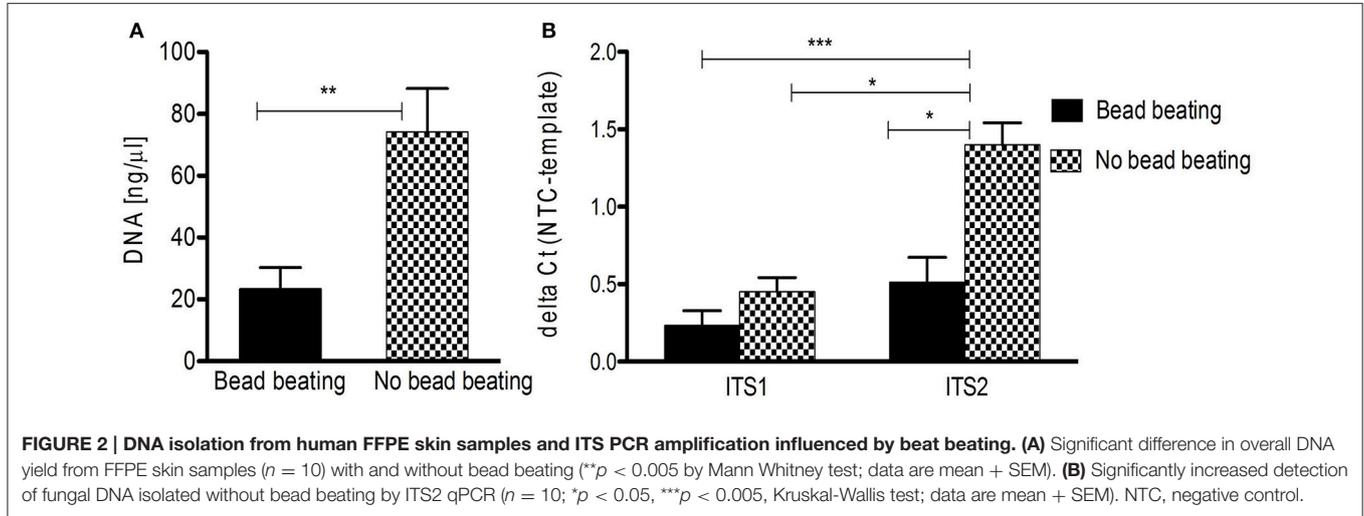
Aside of typically sampled native material (e.g., swabs, etc.) also other resources for mycobiota investigations exist. Formalin-fixed paraffin-embedded (FFPE) tissue samples play an important role in the clinical context. Biopsies or surgically removed tissues are typically fixed in formalin (10%) immediately after they are collected from the patient, thus they represent a well-preserved resource for the analysis of biomolecules including nucleic acids (Sangoi et al., 2009; Kocjan et al., 2015). FFPE specimens are typically used for diagnostic purposes (e.g., histopathology) but are also amenable for molecular scientific investigations. Their prevalence in biological repositories such as biobanks make them ideal specimens to study the mycobiota in the context of human disease (Yuille et al., 2008). About 70 commercially kits are available for DNA extraction out of FFPE material (Kocjan et al., 2015), however, nucleic acid isolation from FFPE material is challenging. Biomolecules are typically cross-linked and fragmented due to formalin, and factors such as the pH of the fixative, duration of fixation, and importantly the DNA extraction method applied greatly influence the quality of the extracted DNA (Bonin and Stanta, 2013; Kocjan et al., 2015). Factors such as residual formalin inhibiting proteinase K activity and omitting complete cell lysis, as well as the presence of PCR inhibitors in the DNA extract might altogether interfere with successful fungal DNA amplification (Coura et al., 2005; Muñoz-Cadavid et al., 2010).



**FIGURE 1 | Schematic representations of rRNA operons and their variability assessed by multiple sequence alignments (MSA). (A)** Prokaryotic and **(B)** eukaryotic rRNA operons. Position and orientation of oligonucleotide primers used for ITS amplification are schematically indicated (for sequence information see **Table 1**). SSU, small subunit; LSU, large subunit; tRNA, transfer RNA; V1-V9, variable regions; ITS, internal transcribed spacer; bps, base-pairs. **(C)** Multiple  
(Continued)

**FIGURE 1 | Continued**

sequence alignment (MSA) of the entire 16S rRNA operon of five different bacterial species (encompassing five different phyla). Variable regions (V1–V9) are highlighted in blue, conserved regions in yellow, positions according to the *E. coli* 16S rRNA (GenBank acc. no.: J01695.2). **(D)** MSA of the complete internal transcribed spacer region of five different fungal species of the same genus (*Hydnum* sp.). **(E)** MSA of the complete ITS region of seven fungal taxa representing different phyla. Information about sequences used for MSA generation **(C,D)** is given as Supplementary Tables S3–S5.



**FIGURE 2 | DNA isolation from human FFPE skin samples and ITS PCR amplification influenced by beat beating. (A)** Significant difference in overall DNA yield from FFPE skin samples ( $n = 10$ ) with and without bead beating (\*\* $p < 0.005$  by Mann Whitney test; data are mean + SEM). **(B)** Significantly increased detection of fungal DNA isolated without bead beating by ITS2 qPCR ( $n = 10$ ; \* $p < 0.05$ , \*\*\* $p < 0.005$ , Kruskal-Wallis test; data are mean + SEM). NTC, negative control.

**TABLE 1 | Overview of commonly used ITS1 and ITS2 oligonucleotide primer pairs.**

Region	Name	Sequence (Forward)	Name	Sequence (Reverse)	Length (bp)	T <sub>m</sub> (°C)	References
ITS1	ITS1	TCCGTAGGTGAACCTGCGG	ITS2	GCTGCGTTCATCATCGATGC	~290	65	White et al., 1990; Muñoz-Cadavid et al., 2010; Schoch et al., 2012
	ITS5	GGAAGTAAAAGTCGTAACAAGG	ITS2	GCTGCGTTCATCATCGATGC	~315	63	White et al., 1990
	ITS1F	CTTGGTCATTTAGAGGAAGTAA	ITS2	GCTGCGTTCATCATCGATGC	~350	51	Mello et al., 2011
	ITS1-F_KYO2	TAGAGGAAGTAAAAGTCGTAA	ITS2_KYO2	TTYRCTRCGTTCATCATC	~300–400	47	Toju et al., 2012
	18S-F	GTAAAAGTCGTAACAAGGTTTC	5.8S-1R	GTTCAAAGAYTCGATGATTCAC	~300–400	*ns	Findley et al., 2013
ITS2	ITS3	GCATCGATGAAGAACGCAGC	ITS4	TCCTCCGCTTATTGATATGC	~330	62	White et al., 1990, Muñoz-Cadavid et al., 2010; Mello et al., 2011; Flury et al., 2014
	ITS3_KYO2	GATGAAGAACGYAGYRAA	ITS4	TCCTCCGCTTATTGATATGC	~400	47	Toju et al., 2012
	fITS9	GAACGCAGCRAAIIGYGA	ITS4	TCCTCCGCTTATTGATATGC	~390	55	Ihrmark et al., 2012
	fITS7	GTGAR TC ATC GAATC TTTG	ITS4	TCCTCCGCTTATTGATATGC	~340	57	Ihrmark et al., 2012
	gITS7	GTGARTCATCGARTCTTTG	ITS4	TCCTCCGCTTATTGATATGC	~340	56	Ihrmark et al., 2012
	5.8S-F	GTGAATCATCGARTCTTTGAAC	28S1-R	ATGCTTAAGTTCAGCGGGTA	~300	*ns	Findley et al., 2013
ITS1-2 incl. 5.8S rRNA gene	ITS5	GGAAGTAAAAGTCGTAACAAGG	ITS4	TCCTCCGCTTATTGATATGC	~641	58	White et al., 1990; Muñoz-Cadavid et al., 2010; Flury et al., 2014
	ITS1F	CTTGGTCATTTAGAGGAAGTAA	ITS4-B	CAGGAGACTTGACACGGTCCAG	~600	55	Gardes and Bruns, 1993
	ITS1-F_KYO2	TAGAGGAAGTAAAAGTCGTAA	ITS4	TCCTCCGCTTATTGATATGC	~700	47	Toju et al., 2012

\*ns, not specified.

These difficulties make a thorough review of the (pre-) analytical process of mycobiota studies mandatory. To highlight the influence of pre-analytics on ITS based mycobiota

investigations we assessed the performance of DNA extraction from human skin FFPE samples (see Supplementary Table S1 for sample information) with a commercially available kit (QIAamp

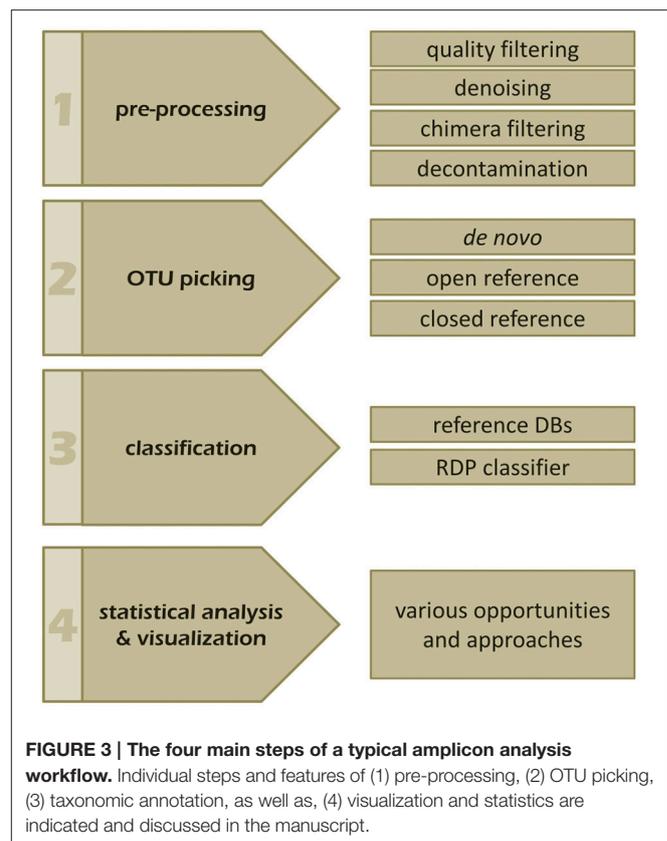
DNA FFPE tissue kit, Qiagen) reported to be efficient for fungal DNA extraction out of FFPE material (Muñoz-Cadavid et al., 2010). We added a mechanical cell disruption step (bead-beating) to the procedure (MagnaLyser, Roche), since this step was shown to be crucial for complete lysis of microbial cells in specimens, significantly influencing correct community representation (de Boer et al., 2010; Reck et al., 2015). A detailed description of the applied method is given in the Data Sheet S1. Interestingly, we observed that bead-beating significantly lead to lower DNA yields and a significantly decreased signal-to-noise ratio in ITS PCR, impairing efficient fungal PCR amplification (Figure 2). Thus, mechanical lysis of specimens could also counteract reliable mycobiota investigations especially if low-biomass samples such as skin are used.

## ITS AMPLIFICATION VIA PCR

For amplification of fungal DNA various primers have been designed targeting different regions of the rRNA operon or other marker genes encoding translation elongation factor 1- $\alpha$ , RNA polymerase II,  $\beta$ -tubulin, and the minichromosome maintenance complex component 7 (MCM7) protein (White et al., 1990; Tanabe et al., 2002; McLaughlin et al., 2009; O'Donnell et al., 2010; Schoch et al., 2012; Toju et al., 2012; Lindahl et al., 2013). Of these, the ITS regions are considered the formal barcode for fungal taxonomy (Schoch et al., 2012; Lindahl et al., 2013). As noted above, ITS1 and ITS2 sequences are highly variable and can be used to discriminate fungi even down to species level (Martin and Rygiewicz, 2005; Porrás-Alfaro et al., 2014). However, each ITS primer combination fails to amplify certain species, a situation similar to bacterial 16S rRNA gene based analysis (Bellemain et al., 2010). Thus the use of multiple primer combinations and/or primers with degenerated nucleotide positions is recommended to capture the entire fungal community (Ihrmark et al., 2012; Toju et al., 2012). **Table 1** summarizes commonly used ITS1 and ITS2 oligonucleotide primers. Of note, the ITS2 region was reported to perform better for fungal DNA amplification out of FFPE material (Muñoz-Cadavid et al., 2010; Flury et al., 2014). We also observed increased PCR performance using ITS2 primers and human skin FFPE samples (Figure 2B). However, other reports obtained similar amplification rates with ITS1 and ITS2 oligonucleotides (Mello et al., 2011; Bazzicalupo et al., 2013; Blaaid et al., 2013; Lindahl et al., 2013).

## BIOINFORMATICS CHALLENGES IN MYCOBIOTA ANALYSES

The bioinformatics analysis workflow of amplicon data can be summarized into four main steps: (i) pre-processing, (ii) OTU picking, (iii) taxonomic classification, and (iv) visualization and statistical analysis (Figure 3; Kuczynski et al., 2012). So far dedicated bioinformatics tools for mycobiota analyses are sparse. Measures originally developed for 16S rRNA gene data, like QIIME (Caporaso et al., 2010) and mothur (Schloss et al., 2009) are often employed to investigate ITS amplicons. However, these



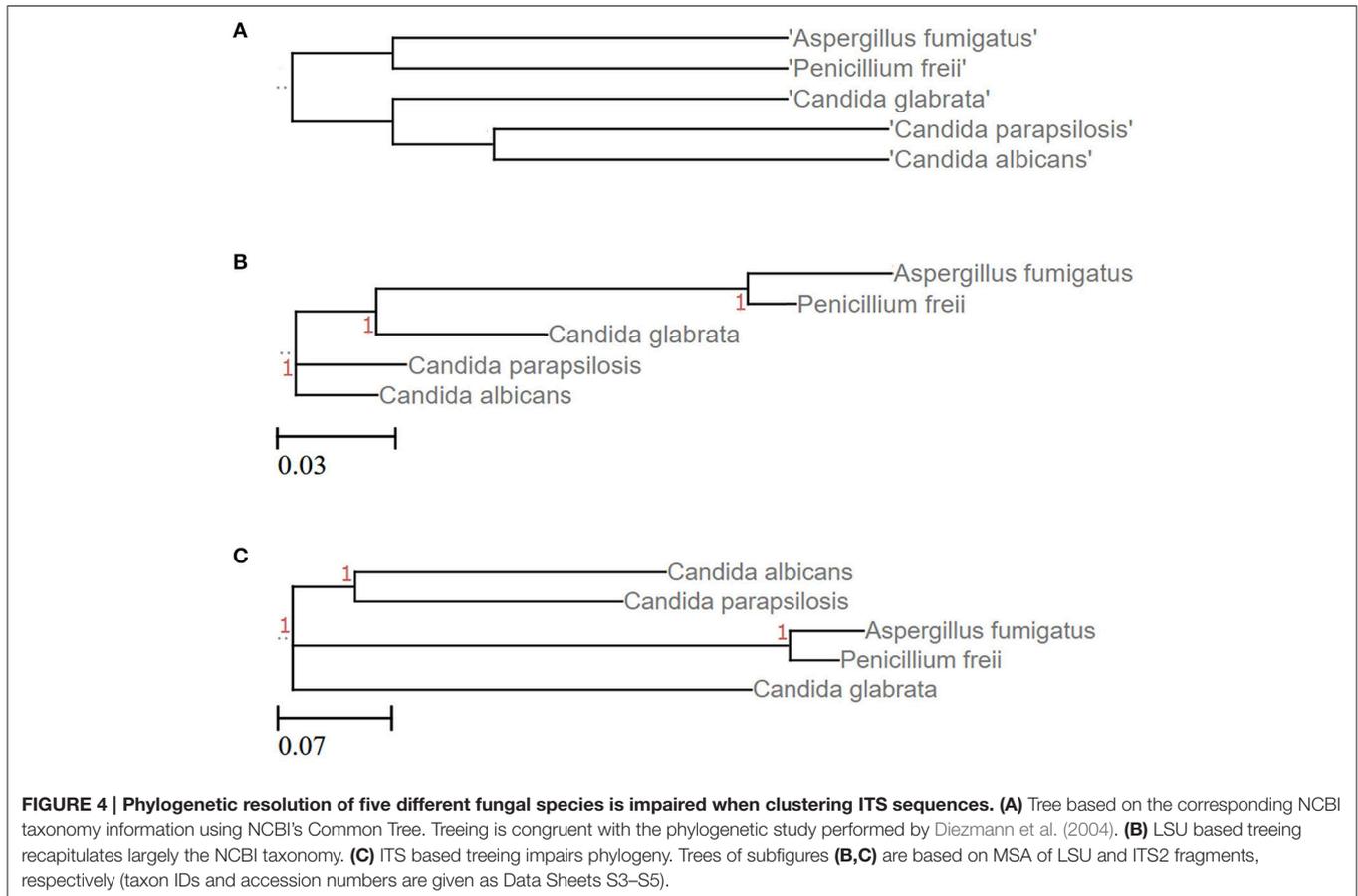
tools pose several shortcomings when applied to ITS sequences, especially when standard protocols are used. In the following the main analytical steps and potential hurdles of ITS based amplicon data analyses are discussed with special emphasis on OTU clustering (OTU picking) and classification. We also highlight the effect of different OTU picking strategies on taxonomic classification of ITS data by comparative analysis of an ITS1 *in silico* mock community.

## PRE-PROCESSING OF AMPLICON RAW DATA

Current pre-processing recommendations include rigorous length filtering of reads, noise reduction (detection, correction, and removal of sequencing errors and artifacts), quality filtering (removal of reads with quality scores below a defined threshold; average > 25), chimera removal (detection and removal of artificially created reads, produced different targets during PCR), as well as removal of singletons/doubletons (Bokulich et al., 2013). The latter could emerge due to sequencing errors (e.g., within homopolymers) leading to OTU inflation of data, which is dependent also on the sequencing technology used (Schirmer et al., 2015). Choice of pre-processing methods and used parameters heavily influence the number of created OTUs, which could lead to underestimation of species diversity if too stringent filtering is applied (Flynn et al., 2015; Kopylova et al., 2016). However, adequate pre-processing of raw reads is mandatory

**TABLE 2 | List of commonly used clustering algorithms.**

Algorithm name	Algorithm type	Multiple sequence alignment required	Integrated in	References
Mothur	Hierarchical	YES	Mothur	Schloss et al., 2009
UCLUST	Greedy	NO	QIIME	Edgar, 2010
UPARSE	Greedy	NO	QIIME	Caporaso et al., 2010; Edgar, 2013; Albanese et al., 2015
SWARM	Agglomerative	NO	QIIME, MICCA	Caporaso et al., 2010; Mahé et al., 2014; Albanese et al., 2015
OTUCLUST	Greedy	NO	MICCA	Albanese et al., 2015

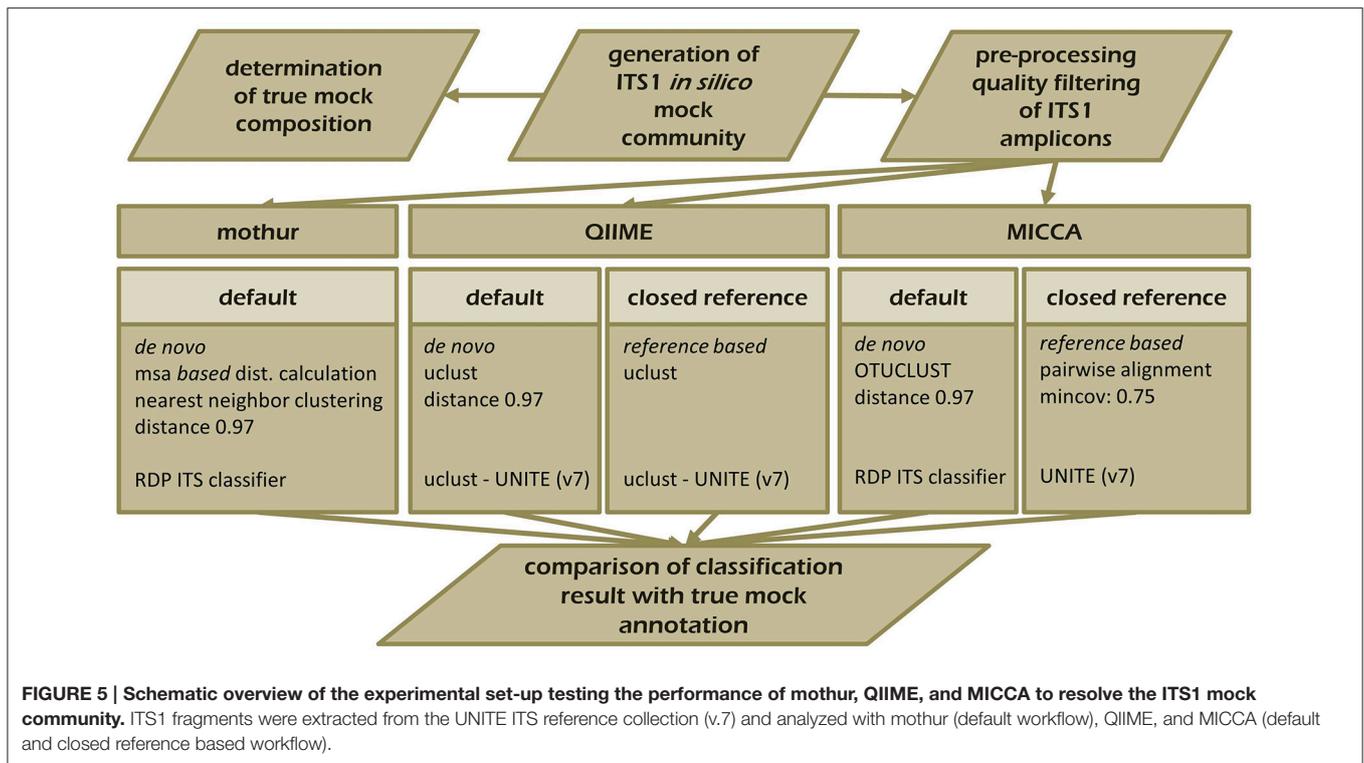


independent of the used marker gene, leading to a reduced number of assigned OTUs and less noise in the data. Basically we refer to the suggestions of Schloss et al. (2011), but as there are no general rules for pre-processing we strongly recommend looking carefully into what is happening during filtering rather than just applying default parameters.

## OTU PICKING—CLUSTERING INTO OPERATIONAL TAXONOMIC UNITS (OTUs)

Numerous approaches and tools are available for clustering sequences into OTUs. Current algorithms developed primarily for 16S rRNA gene amplicons are summarized in Table 2. In general OTU clustering and annotation could be achieved by using three different strategies (i) *de novo*-, (ii) closed

reference-, and (iii) open reference-based clustering. Briefly, a closed reference approach calculates for each input sequence the best pairwise alignment to a pre-defined reference database collection. Sequences with the same best match are binned into the same cluster (i.e., OTU). In contrast, *de novo* based strategies cluster sequences within a pre-defined distance (commonly 3%). For each of these clusters a representative sequence is selected and taxonomically classified. Open-reference OTU picking is a mixture of both. Reads are first clustered using a closed reference approach and all reads which fail in this first step are subsequently clustered using a *de novo* strategy (Rideout et al., 2014; Westcott and Schloss, 2015). A recent comparison of the three different clustering strategies revealed the *de novo* approach based on a global distance matrix (implemented by default by mothur) as the optimal method for clustering 16S rRNA gene sequences into OTUs (Westcott and Schloss,



2015). Such benchmark comparisons are unfortunately missing for ITS amplicons. Importantly, the use of multiple sequence alignments (MSA) for clustering ITS sequences in a *de novo* approach poses a significant problem. ITS sequences show a high degree of intraspecific variation (Figures 1D,E), which leads to the introduction of gaps during the alignment process and subsequently to erroneous multiple sequence alignments exhibiting wrong phylogenetic resolution (Figure 4). In addition, there is no commonly accepted genus or species level cut-off for the formation of ITS clusters, such as 5% variation for genus- and 3% for species-level clustering applied to 16S rRNA gene data (Stackebrandt and Goebel, 1994). Often 3% variation is used and this cut-off seems to perform reasonable for fungal ITS sequences, although taxonomic resolution is clearly impaired within certain taxa. Both, ITS1 and ITS2, show a highly congruent fungal taxonomic resolution (Blaalid et al., 2013).

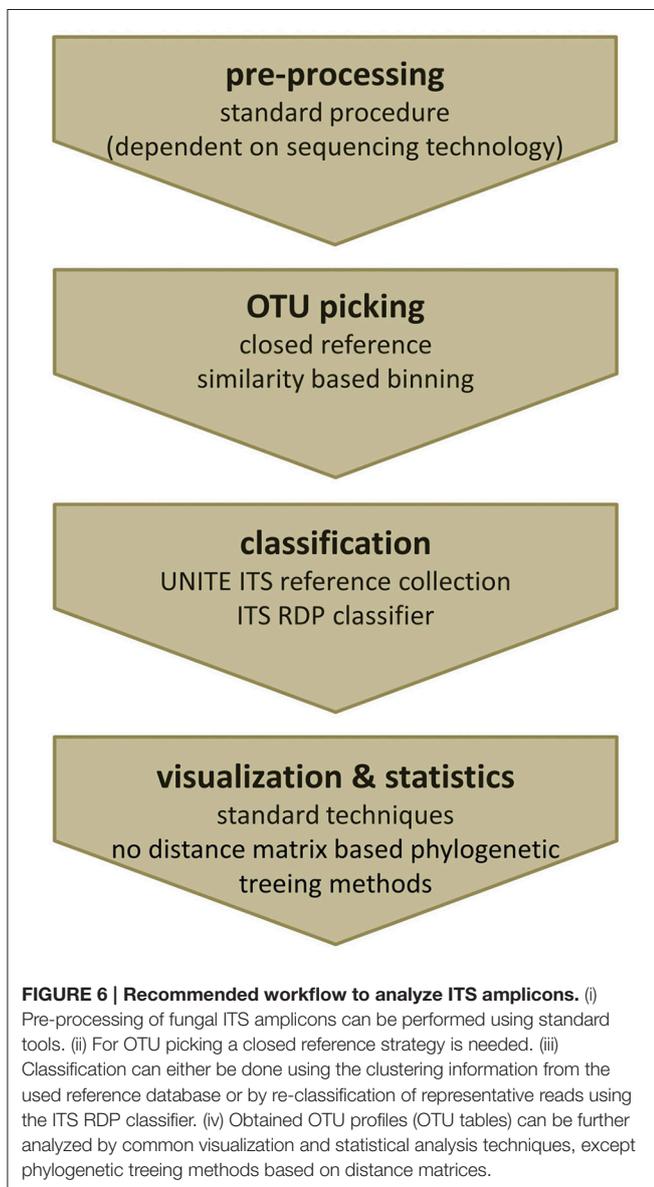
## TAXONOMIC CLASSIFICATION OF OTUs

If a closed reference-based approach is used, taxonomic classification is achieved already during the OTU picking step, wherein OTUs represent clusters of identical matches to the reference database. If a *de novo* strategy is employed a proxy sequence from each cluster is chosen and taxonomically classified either by calculating sequence similarities between the proxy sequence and a reference database or by estimating the classification confidence using a pre-trained classifier, such as the RDP classifier (Wang et al., 2007). The latter one offers training sets for ITS (Porrás-Alfaro et al., 2014) as well as for LSU (Liu et al., 2012) sequences. Accurate taxonomic

classification of sequences requires reference databases of high quality. The UNITE (Unified system for the DNA based fungal species linked to the classification, <https://unite.ut.ee>) database for ITS fragments represents a curated full-length ITS sequence repository devoid of ambiguous sequences (Nilsson et al., 2014). Several factors lead to misannotated ITS sequences in repositories, such as GenBank, EMBL, or DDJB. For instance many fungi have sexual (teleomorph) and asexual (anamorph) forms and they are often classified as different taxa assigned even to different families (Mahé et al., 2012; Underhill and Iliev, 2014). UNITE represents currently the most comprehensive taxonomic ITS classification resource, providing ready-to-use application files for mothur, QIIME, and MICCA. Although still some fungal lineages are uncovered it comprises 536,881 sequence entries (as of January 2016, UNITE version 7.0). Recently, the hand curated ISHAM-ITS reference DNA barcoding database, with 3,200 sequences covering about 415 fungal species (as of December 2015) maintained by the Society for Human and Animal Mycology (ISHAM) was incorporated into UNITE (Irinnyi et al., 2015). Noteworthy is UNITE's key concept, the so-called *species hypotheses* (SH). A SH represents an operational taxonomic unit at approximately species level (Kõljalg et al., 2013). Each SH is represented by the most homologous high quality sequence within a respective sequence cluster linked to a unique, permanent digital object identifier (DOI), which allows for unambiguous identification even in absence of a full formal taxonomic name or when a fungal OTU remains taxonomically unassigned. Of note, the global fungome is estimated to comprise 1.5–6 million different species (Hawksworth, 1991; Blackwell, 2011; Taylor et al.,

**TABLE 3 | Correct classification of the *in silico* ITS1 mock community with different analysis pipelines and OTU picking strategies (% in parenthesis).**

	No. of classified sequences	No. of OTUs	Taxonomic classification					
			Phylum	Class	Order	Family	Genus	Species
<i>In silico</i> ITS1 mock	56,451	11,336	6	45	143	409	1,931	11,336
Mothur	36,255	26,965	35,675 (98.40)	33,686 (92.91)	32,949 (90.88)	29,739 (82.02)	24,000 (66.20)	11,233 (30.98)
QIIME default	56,451	19,779	18,930 (33.53)	14,316 (25.36)	4,865 (8.62)	669 (1.18)	372 (0.65)	37 (0.07)
QIIME closed reference	31,676	8,764	31,504 (99.47)	31,284 (98.77)	30,832 (97.34)	29,211 (92.23)	26,650 (84.14)	19,246 (71.62)
MICCA default	56,446	20,878	54,698 (96.90)	51,095 (90.52)	49,454 (87.61)	44,511 (78.56)	36,549 (64.75)	19,246 (29.73)
MICCA closed reference	52,475	9,942	52,129 (99.34)	49,670 (94.65)	49,434 (94.20)	45,784 (87.25)	41,206 (78.52)	26,400 (50.31)



the “completeness” of the current fungal reference databases (Tedersoo et al., 2014).

## THE EFFECT OF DIFFERENT OTU PICKING STRATEGIES ON TAXONOMIC CLASSIFICATION OF ITS DATA

To demonstrate the influence of different OTU picking strategies on phylogenetic resolution of fungal communities we compared three commonly used analysis pipelines mothur, QIIME, and MICCA, employing an *in silico* created fungal ITS1 mock community. Therefore, 582,779 ITS1 fragments were extracted by ITSx (Bengtsson-Palme et al., 2013) from the public UNITE sequence collection (version 7, comprising 656,899 sequences). Amplicons were filtered for ambiguous lineage definitions, resulting in 345,201 sequences. These amplicons were quality filtered yielding finally 56,451 unique ITS1 fragments (accession numbers and taxonomic annotations are given in Supplementary Table S2). ITS1 fragments were subsequently clustered into OTUs by the default *de novo* strategies employed by mothur, QIIME, and MICCA, according to the standard protocol of each pipeline (for details see Data Sheet S1). For analyses with QIIME and MICCA, sequences were additionally binned into OTUs according to their taxonomic classification using a closed reference OTU picking strategy employing the UNITE database (version 7, 22.08.2016). The database was used for classification of representative sequences either directly for similarity-based comparisons or indirectly for training the RDP classifier. Finally, the assigned taxonomic classifications were compared to the true annotation of the ITS1 mock community. A scheme highlighting the experimental design and used parameters for comparison of pipelines is shown in Figure 5. Table 3 summarizes the comparison results, which clearly indicates that choice of the OTU picking strategy severely impacts the phylogenetic resolution of the ITS mock community. All pipelines used with default parameters failed to accurately classify the mock community down to species level. All approaches classified ITS1 reads with a reasonable accuracy only to the order level (range 87.61–97.34% correct assignment), except QIIME with default settings (*de novo*), which behaved poor (classifying only 33.53% of sequences correctly at phylum level and 0.07% at species level). A high number of singletons emerged by using all three *de novo*

2014), wherein currently 130,000 species are represented in the public sequence repositories (<http://www.speciesfungorum.org/>, accessed March 2016). These counts give already an idea about

approaches, leading to OTU inflation, and wrongly clustered OTUs. Importantly, changing the default OTU picking approach of QIIME (*de novo*) to a closed reference approach increased the amount of correctly classified species to 71.62% (Table 3). Taken together these data indicate that closed reference based strategies should be preferred if ITS amplicons are analyzed. Nevertheless, a relatively large fraction of wrongly annotated OTUs might still persist, thus manual correction of taxonomic assignments (i.e., by individual blast analysis of sequences) might still improve classification (Iliev et al., 2012).

## VISUALIZATION AND STATISTICAL ANALYSIS OF ITS DATA

Visualization and statistical analyses of mycobiota data typically enable measures for community structure, such as alpha-diversity metrics (e.g., richness, evenness, Shannon index), as well as taxonomic turnover (i.e., changes in microbial composition between conditions or groups) called beta-diversity, which can be calculated with different distance measurements (Bray Curtis, Andernberg, UniFrac, etc.). Principle coordinates analysis (PCoA) plots based on these distance matrices enable simplified visualization of the structural resemblance of mycobiota profiles. Statistical identification of differential abundant taxa between groups could be achieved using tools such as LEfSe (Segata et al., 2011) or linear modeling approaches, such as DESeq (Paulino et al., 2006) or edgeR (Robinson et al., 2010). Measures for alpha- and beta-diversity are readily provided by tools such as mothur and QIIME and operate on the created OTU tables. Caution must be taken if measures derive phylogenetic information based on diversity matrices emerging from MSAs of ITS reads, such as UniFrac (Lozupone et al., 2011). Such methods lead to erroneous results because of the bad performance of aligning ITS reads as shown above (Figure 4).

## CONCLUSION

Fungal amplicon studies benefit greatly from the advancements made in the analysis of bacterial communities, nonetheless, many hurdles need still to be solved and standards are waiting to be defined. Although numerous protocols and kits are available for fungal DNA isolation out of complex specimens such as human tissue, protocols need to be adapted to the special study needs. Recommendations on how to perform ITS analyses using mothur and QIIME with non-phylogenetic diversity metrics have been recently released (e.g., [https://mothur.org/wiki/Analysis\\_examples#Sanger\\_16S-ITS\\_rRNA\\_sequence\\_analysis](https://mothur.org/wiki/Analysis_examples#Sanger_16S-ITS_rRNA_sequence_analysis),

accessed February 2017, [http://qiime.org/1.7.0/tutorials/fungal\\_its\\_analysis.html](http://qiime.org/1.7.0/tutorials/fungal_its_analysis.html), accessed April 2016). Based on our experience, pre-processing, and quality filtering of ITS sequencing data, as well as chimera filtering could be done with standard 16S rRNA gene based procedures. We use the default workflow of mothur for ITS data pre-processing, assembling of paired reads, length-, quality-, and chimera filtering, as well as noise reduction as described in the MiSeq 16S SOP of Kuczynski et al. (2012, accessed May 2016). Since mothur employs pairwise distance matrices, which require the creation of multiple sequence alignments, we recommend switching to tools such as QIIME or MIICA for further analyses, which allow for closed reference-based approaches. Subsequently QIIME can be used for visualization of mycobiota data. The crucial step within QIIME is to suppress tree generation within the OTU picking step and to use closed reference OTU picking instead of the default *de novo* strategy. The pre-formatted version of the UNITE ITS reference database which is provided directly by UNITE works perfectly with one of the reference-based OTU picking scripts of QIIME and MICCA. Alternatively sequences can be also classified and binned based on the information gained by the RDP classifier trained for ITS fragments or simply by an individual blast approach. A final summary of the recommended analysis steps for ITS based mycobiota analysis is given in Figure 6.

## AUTHOR CONTRIBUTIONS

Conceptualization: BH, RN, GT, and GG. Data analysis: BH and NM. Manuscript draft: BH and GG. Final manuscript and approval: All authors.

## FUNDING

This work was supported by BioTechMed-Graz and the Austrian Science Fund (FWF W1241-B18).

## ACKNOWLEDGMENTS

Jürgen C. Becker, Karl Kashofer, and Andrea Thüringer are acknowledged for their input regarding mycobiota analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00180/full#supplementary-material>

## REFERENCES

- Albanese, D., Fontana, P., De Filippo, C., Cavalieri, D., and Donati, C. (2015). MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. *Sci. Rep.* 5:9743. doi: 10.1038/srep09743
- Bazzicalupo, A. L., Bálint, M., and Schmitt, I. (2013). Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecol.* 6, 102–109. doi: 10.1016/j.funeco.2012.09.003
- Belkaid, Y., and Naik, S. (2013). Compartmentalized and systemic control of tissue immunity by commensals. *Nat. Immunol.* 14, 646–653. doi: 10.1038/ni.2604
- Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., and Kausserud, H. (2010). ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* 10:189. doi: 10.1186/1471-2180-10-189
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., et al. (2013). Improved software detection and extraction of ITS1 and

- ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* 4, 914–919. doi: 10.1111/2041-210x.12073
- Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P. M., and Kausserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol. Ecol. Resour.* 13, 218–224. doi: 10.1111/1755-0998.12065
- Blackwell, M. (2011). The fungi: 1, 2, 3... 5.1 million species? *Am. J. Bot.* 98, 426–438. doi: 10.3732/ajb.1000298
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276
- Bonin, S., and Stanta, G. (2013). Nucleic acid extraction methods from fixed and paraffin-embedded tissues in cancer diagnostics. *Expert Rev. Mol. Diagn.* 13, 271–282. doi: 10.1586/erm.13.14
- Brown, G. D., Denning, D. W., Gow, N. A., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden killers: human fungal infections. *Sci. Transl. Med.* 4:165rv13. doi: 10.1126/scitranslmed.3004404
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Charlson, E. S., Diamond, J. M., Bittinger, K., Fitzgerald, A. S., Yadav, A., Haas, A. R., et al. (2012). Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. *Am. J. Respir. Crit. Care Med.* 186, 536–545. doi: 10.1164/rccm.201204-0693OC
- Coura, R., Prolla, J. C., Meurer, L., and Ashton-Prolla, P. (2005). An alternative protocol for DNA extraction from formalin fixed and paraffin wax embedded tissue. *J. Clin. Pathol.* 58, 894–895. doi: 10.1136/jcp.2004.021352
- Cui, L., Morris, A., and Ghedin, E. (2013). The human mycobiome in health and disease. *Genome Med.* 5:63. doi: 10.1186/gm467
- de Boer, R., Peters, R., Gierveld, S., Schuurman, T., Kooistra-Smid, M., and Savelkoul, P. (2010). Improved detection of microbial DNA after bead-beating before DNA isolation. *J. Microbiol. Methods* 80, 209–211. doi: 10.1016/j.mimet.2009.11.009
- Diezmann, S., Cox, C. J., Schonian, G., Vilgalys, R. J., and Mitchell, T. G. (2004). Phylogeny and evolution of medical species of *Candida* and related taxa: a multigenic analysis. *J. Clin. Microbiol.* 42, 5624–5635. doi: 10.1128/JCM.42.12.5624-5635.2004
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604
- Findley, K., Oh, J., Yang, J., Conlan, S., Deming, C., Meyer, J. A., et al. (2013). Topographic diversity of fungal and bacterial communities in human skin. *Nature* 498, 367–370. doi: 10.1038/nature12171
- Flury, B., Weisser, M., Prince, S. S., Bubendorf, L., Battegay, M., Frei, R., et al. (2014). Performances of two different panfungal PCRs to detect mould DNA in formalin-fixed paraffin-embedded tissue: what are the limiting factors? *BMC Infect. Dis.* 14:692. doi: 10.1186/s12879-014-0692-z
- Flynn, J. M., Brown, E. A., Chain, F. J., MacIsaac, H. J., and Cristescu, M. E. (2015). Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. *Ecol. Evol.* 5, 2252–2266. doi: 10.1002/ece3.1497
- Gardes, M., and Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2, 113–118. doi: 10.1111/j.1365-294X.1993.tb00005.x
- Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., et al. (2010). Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.* 6:e1000713. doi: 10.1371/journal.ppat.1000713
- Goldschmidt, P., Degorge, S., Merabet, L., and Chaumeil, C. (2014). Enzymatic treatment of specimens before DNA extraction directly influences molecular detection of infectious agents. *PLoS ONE* 9:e94886. doi: 10.1371/journal.pone.0094886
- Gorkiewicz, G., Thallinger, G. G., Trajanoski, S., Lackner, S., Stocker, G., Hinterleitner, T., et al. (2013). Alterations in the colonic microbiota in response to osmotic diarrhea. *PLoS ONE* 8:e55817. doi: 10.1371/journal.pone.0055817
- Gosiewski, T., Salamon, D., Szopa, M., Sroka, A., Malecki, M. T., and Bulanda, M. (2014). Quantitative evaluation of fungi of the genus *Candida* in the feces of adult patients with type 1 and 2 diabetes - a pilot study. *Gut Pathog.* 6:43. doi: 10.1186/s13099-014-0043-z
- Gouba, N., and Drancourt, M. (2015). Digestive tract mycobiota: a source of infection. *Med. Mal. Infect.* 45, 9–16. doi: 10.1016/j.medmal.2015.01.007
- Hallen-Adams, H. E., Kachman, S. D., Kim, J., Legge, R. M., and Martínez, I. (2015). Fungi inhabiting the healthy human gastrointestinal tract: a diverse and dynamic community. *Fungal Ecol.* 15, 9–17. doi: 10.1016/j.funeco.2015.01.006
- Hartmann, M., Howes, C. G., Abarenkov, K., Mohn, W. W., and Nilsson, R. H. (2010). V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. *J. Microbiol. Methods* 83, 250–253. doi: 10.1016/j.mimet.2010.08.008
- Hawksworth, D. (1991). The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycol. Res.* 95, 641–655. doi: 10.1016/S0953-7562(09)80810-1
- Hoarau, G., Mukherjee, P. K., Gower-Rousseau, C., Hager, C., Chandra, J., Retuerto, M. A., et al. (2016). Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. *MBio* 7, e01250–e01216. doi: 10.1128/mBio.01250-16
- Hunt, J., Boddy, L., Randerson, P. F., and Rogers, H. J. (2004). An evaluation of 18S rDNA approaches for the study of fungal diversity in grassland soils. *Microb. Ecol.* 47, 385–395. doi: 10.1007/s00248-003-2018-3
- Ihrmark, K., Bödeker, I. T., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., et al. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* 82, 666–677. doi: 10.1111/j.1574-6941.2012.01437.x
- Iliev, I. D., Funari, V. A., Taylor, K. D., Nguyen, Q., Reyes, C. N., Strom, S. P., et al. (2012). Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science* 336, 1314–1317. doi: 10.1126/science.1221789
- Irinyi, L., Serena, C., Garcia-Hermoso, D., Arabatzis, M., Desnos-Ollivier, M., Vu, D., et al. (2015). International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database - the quality controlled standard tool for routine identification of human and animal pathogenic fungi. *Med. Mycol.* 53, 313–337. doi: 10.1093/mmy/myv008
- Kappe, R., Fauser, C., Okeke, C. N., and Maiwald, M. (1996). Universal fungus-specific primer systems and group-specific hybridization oligonucleotides for 18S rDNA. *Mycoses* 39, 25–30. doi: 10.1111/j.1439-0507.1996.tb00079.x
- Kim, S. H., Clark, S. T., Surendra, A., Copeland, J. K., Wang, P. W., Ammar, R., et al. (2015). Global analysis of the fungal microbiome in cystic fibrosis patients reveals loss of function of the transcriptional repressor *nrg1* as a mechanism of pathogen adaptation. *PLoS Pathog.* 11:e1005308. doi: 10.1371/journal.ppat.1005308
- Kocjan, B. J., Hosnjak, L., and Poljak, M. (2015). Commercially available kits for manual and automatic extraction of nucleic acids from formalin-fixed, paraffin-embedded (FFPE) tissues. *Acta Dermatovenerol. Alp. Panonica. Adriat.* 24, 47–53. doi: 10.15570/actaapa.2015.12
- Köljal, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. doi: 10.1111/mec.12481
- Kopylova, E., Navas-Molina, J. A., Mercier, C. Á., Xu, Z. Z., Mahé, F., He, Y., et al. (2016). Open-source sequence clustering methods improve the state of the art. *mSystems* 1, e00003–e00015. doi: 10.1128/mSystems.00003-15
- Krause, R., Halwachs, B., Thallinger, G. G., Klymiuk, I., Gorkiewicz, G., Hoenigl, M., et al. (2016). Characterisation of *Candida* within the mycobiome/microbiome of the lower respiratory tract of ICU patients. *PLoS ONE* 11:e0155033. doi: 10.1371/journal.pone.0155033
- Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., et al. (2012). Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* 13, 47–58. doi: 10.1038/nrg3129
- Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjöller, R., et al. (2013). Fungal community analysis by high-throughput

- sequencing of amplified markers - a user's guide. *New Phytol.* 199, 288–299. doi: 10.1111/nph.12243
- Liu, K. L., Porras-Alfaro, A., Kuske, C. R., Eichorst, S. A., and Xie, G. (2012). Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Appl. Environ. Microbiol.* 78, 1523–1533. doi: 10.1128/AEM.06826-11
- Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133
- Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* 2:e593. doi: 10.7717/peerj.593
- Mahé, S., Duhamel, M., Le Calvez, T., Guillot, L., Sarbu, L., Bretaudeau, A., et al. (2012). PHYMYCO-DB: a curated database for analyses of fungal diversity and evolution. *PLoS ONE* 7:e43117. doi: 10.1371/journal.pone.0043117
- Martin, K. J., and Rygiel, P. T. (2005). Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiol.* 5:28. doi: 10.1186/1471-2180-5-28
- McLaughlin, D. J., Hibbett, D. S., Lutzoni, F., Spatafora, J. W., and Vilgalys, R. (2009). The search for the fungal tree of life. *Trends Microbiol.* 17, 488–497. doi: 10.1016/j.tim.2009.08.001
- Mello, A., Napoli, C., Murat, C., Morin, E., Marceddu, G., and Bonfante, P. (2011). ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds. *Mycologia* 103, 1184–1193. doi: 10.3852/11-027
- Muñoz-Cadavid, C., Rudd, S., Zaki, S. R., Patel, M., Moser, S. A., Brandt, M. E., et al. (2010). Improving molecular detection of fungal DNA in formalin-fixed paraffin-embedded tissues: comparison of five tissue DNA extraction methods using panfungal PCR. *J. Clin. Microbiol.* 48, 2147–2153. doi: 10.1128/JCM.00459-10
- Nilsson, R. H., Hyde, K., Pawlowska, J., Ryberg, M., Tedersoo, L., Aas, A. B., et al. (2014). Improving ITS sequence data for identification of plant pathogenic fungi. *Fungal Divers.* 67, 11–19. doi: 10.1007/s13225-014-0291-8
- O'Donnell, K., Sutton, D. A., Rinaldi, M. G., Sarver, B. A., Balajee, S. A., Schroers, H. J., et al. (2010). Internet-accessible DNA sequence database for identifying fusaria from human and animal infections. *J. Clin. Microbiol.* 48, 3708–3718. doi: 10.1128/JCM.00989-10
- Oever, J. T., and Netea, M. G. (2014). The bacteriome-mycobiome interaction and antifungal host defense. *Eur. J. Immunol.* 44, 3182–3191. doi: 10.1002/eji.201344405
- Oh, J., Byrd, A. L., Deming, C., Conlan, S., Kong, H. H., and Segre, J. A. (2014). Biogeography and individuality shape function in the human skin metagenome. *Nature* 514, 59–64. doi: 10.1038/nature13786
- Oh, J., Freeman, A. F., Park, M., Sokolic, R., Candotti, F., Holland, S. M., et al. (2013). The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. *Genome Res.* 23, 2103–2114. doi: 10.1101/gr.159467.113
- Ott, S. J., Kühbacher, T., Musfeldt, M., Rosenstiel, P., Hellmig, S., Rehman, A., et al. (2008). Fungi and inflammatory bowel diseases: alterations of composition and diversity. *Scand. J. Gastroenterol.* 43, 831–841. doi: 10.1080/00365520801935434
- Paulino, L. C., Tseng, C. H., Strober, B. E., and Blaser, M. J. (2006). Molecular analysis of fungal microbiota in samples from healthy human skin and psoriatic lesions. *J. Clin. Microbiol.* 44, 2933–2941. doi: 10.1128/JCM.00785-06
- Porras-Alfaro, A., Liu, K. L., Kuske, C. R., and Xie, G. (2014). From genus to phylum: large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. *Appl. Environ. Microbiol.* 80, 829–840. doi: 10.1128/AEM.02894-13
- Reck, M., Tomasch, J., Deng, Z., Jarek, M., Husemann, P., and Wagner-Dobler, I. (2015). Stool metatranscriptomics: a technical guideline for mRNA stabilisation and isolation. *BMC Genomics* 16:494. doi: 10.1186/s12864-015-1694-y
- Rideout, J. R., He, Y., Navas-Molina, J. A., Walters, W. A., Ursell, L. K., Gibbons, S. M., et al. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2:e545. doi: 10.7717/peerj.545
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Sangoi, A. R., Rogers, W. M., Longacre, T. A., Montoya, J. G., Baron, E. J., and Banaei, N. (2009). Challenges and pitfalls of morphologic identification of fungal infections in histologic and cytologic specimens: a ten-year retrospective review at a single institution. *Am. J. Clin. Pathol.* 131, 364–375. doi: 10.1309/AJCP99OOZSNISCSZ
- Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* 43:e37. doi: 10.1093/nar/gku1341
- Schloss, P. D., Gevers, D., and Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 6:e27310. doi: 10.1371/journal.pone.0027310
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (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. doi: 10.1073/pnas.1117018109
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60
- Simon, L., Lalonde, M., and Bruns, T. D. (1992). Specific amplification of 18S fungal ribosomal genes from vesicular-arbuscular endomycorrhizal fungi colonizing roots. *Appl. Environ. Microbiol.* 58, 291–295.
- Smit, E., Leeflang, P., Glandorf, B., van Elsas, J. D., and Wernars, K. (1999). Analysis of fungal diversity in the wheat rhizosphere by sequencing of cloned PCR-amplified genes encoding 18S rRNA and temperature gradient gel electrophoresis. *Appl. Environ. Microbiol.* 65, 2614–2621.
- Sokol, H., Leducq, V., Aschard, H., Pham, H. P., Jegou, S., Landman, C., et al. (2016). Fungal microbiota dysbiosis in IBD. *Gut*. doi: 10.1136/gutjnl-2015-310746. [Epub ahead of print].
- Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 44, 846–849. doi: 10.1099/00207713-44-4-846
- Tanabe, Y., Watanabe, M. M., and Sugiyama, J. (2002). Are Microsporidia really related to Fungi? a reappraisal based on additional gene sequences from basal fungi. *Mycol. Res.* 106, 1380–1391. doi: 10.1017/S09537562000686X
- Tang, J., Iliev, I. D., Brown, J., Underhill, D. M., and Funari, V. A. (2015). Mycobiome: approaches to analysis of intestinal fungi. *J. Immunol. Methods* 421, 112–121. doi: 10.1016/j.jim.2015.04.004
- Taylor, D. L., Hollingsworth, T. N., McFarland, J. W., Lennon, N. J., Nusbaum, C., and Ruess, R. W. (2014). A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecol. Monogr.* 84, 3–20. doi: 10.1890/12-1693.1
- Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N. S., Wijesundera, R., et al. (2014). Fungal biogeography: Global diversity and geography of soil fungi. *Science* 346:1256688. doi: 10.1126/science.1256688
- Toju, H., Tanabe, A. S., Yamamoto, S., and Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* 7:e40863. doi: 10.1371/journal.pone.0040863
- Underhill, D. M., and Iliev, I. D. (2014). The mycobiota: interactions between commensal fungi and the host immune system. *Nat. Rev. Immunol.* 14, 405–416. doi: 10.1038/nri3684
- van Burik, J. A., Schreckhise, R. W., White, T. C., Bowden, R. A., and Myerson, D. (1998). Comparison of six extraction techniques for isolation of DNA from filamentous fungi. *Med. Mycol.* 36, 299–303. doi: 10.1080/02681219880000471
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Westcott, S. L., and Schloss, P. D. (2015). *De novo* clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ* 3:e1487. doi: 10.7717/peerj.1487

- Wheeler, M. L., Limon, J. J., Bar, A. S., Leal, C. A., Gargus, M., Tang, J., et al. (2016). Immunological consequences of intestinal fungal dysbiosis. *Cell Host Microbe* 19, 865–873. doi: 10.1016/j.chom.2016.05.003
- White, T., Bruns, T., Lee, S., and Taylor, J. (1990). “Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics,” in *PCR Protocols: A Guide to Methods and Applications*, eds M. Innis, D. Gelfand, J. Sninsky, and T. White (Orlando, FL: Academic Press), 315–322.
- Yuille, M., van Ommen, G. J., Brechot, C., Cambon-Thomsen, A., Dagher, G., Landegren, U., et al. (2008). Biobanking for Europe. *Brief. Bioinform.* 9, 14–24. doi: 10.1093/bib/bbm050

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Halwachs, Madhusudhan, Krause, Nilsson, Moissl-Eichinger, Högenauer, Thallinger and Gorkiewicz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.