Metataxonomic comparison between internal transcribed spacer and 26S ribosomal large subunit (LSU) rDNA gene

Authors: Jatziri Mota-Gutierrez, Ilario Ferrocino, Kalliopi Rantsiou, Luca Cocolin*

Affiliations:

Department of Agricultural, Forest, and Food Science, University of Turin, Largo Paolo Braccini 2, 10095, Grugliasco, Torino, Italy

Corresponding Author:

Luca Cocolin, Ph.D., Department of Agricultural, Forest and Food Science of the University of Turin (DISAFA), University of Turin, Grugliasco, Torino, 10095, Italy

E-mail: lucasimone.cocolin@unito.it

Highlights

- Primer selection plays critical roles in the sensitivity and tracking fungal communities to assess reliable and accurate ecological populations
- The 26S target region exploited in rRNA sequencing demonstrated greater taxonomical depth for fungal communities
- Preferential amplification phenomenon contributes to underestimations and overestimations of fungal species
- The limited availability of updated databases to assess ecological populations.

Abstract

Next-generation sequencing has been used to strengthen knowledge about taxonomic diversity and ecology of fungi within food ecosystems. However, primer amplification and identification bias could edge our understanding into the fungal ecology. The aim of this study is to compare the performance of two primer pairs over two nuclear ribosomal RNA (rRNA) regions of the fungal kingdom, namely the ITS2 and 26S regions. Fermented cocoa beans were employed as biological material and the fungal ecology during fermentation was studied using amplicon-based sequencing tools, making use of a manually curated 26S database constructed in this study, and validated with SILVA database. To explore potential biases introduced by PCR amplification of fungal communities, a mock community of known composition was prepared and tested. The relative abundances observed for ITS2 suggest that species with longer amplification fragments are underestimated and concurrently species that render shorter amplification fragments are overestimated. However, this correlation between amplicon length and estimation is not valid for all the species analysed. Variability in the amplification lengths contributed to the preferential amplification phenomenon. DNA extracted from twenty fermented cocoa bean samples were used to assess the performance of the two target regions. Overall, the metataxonomic data set recovered similar taxonomic composition and provided consistent results in OTU richness among biological samples. However, 26S region provided higher alpha diversity index and greater fungal rRNA taxonomic depth and robustness results compared with ITS2. Based on the results of this study we suggest the use of the 26S region for targeting fungi. Furthermore, this study showed the efficacy of the manually curated reference database optimized for annotation of mycobiota by using the 26S as a gene target.

Keywords: Amplicon sequencing; fungal ecology; primer bias; Illumina; fungal database.

1. Introduction

Fungi are eukaryotic microorganisms which belong to one of the most diverse kingdoms on Earth (Blackwell, 2011). They play an important role in the safety, quality, and stability of all foodstuff to some degree, whether they are required during processing or whether they have a negative impact during shelf life. Therefore, tracking fungal communities of food systems has been a concern in food research. To date, most recent studies on the microbial diversity of fermented food such as vegetable, seafood, beverages, cheese, olives and spontaneously fermented American cool ship ale fermentations have employed amplicon sequencing approaches (Bokulich et al., 2012; Cocolin et al., 2013; Ercolini et al., 2012; Li et al., 2011; Roh et al., 2010).

Illumina sequencing platform has been currently providing a sensitive description of the microbial dynamics within food ecosystems. Some of the advantages of using this technology is that it yields greater sequencing coverage and increased sample throughput at lower cost *per* sample compared to other platforms (Caporaso et al., 2011; Quail et al., 2012). The sensitivity of this approach relies on the high coverage and accurate taxonomic resolution of short amplicon length (Quail et al., 2012). Recent advances in the microbial diversity using next-generation sequencing technologies (NGS) have underlined the importance of the reliability of PCR primers targeting a specific genetic marker (Bokulich and Mills, 2013). In spite of the importance of the amplification of shorter fragments amplified by PCR in NGS, recent studies described a more reliable community of fungi using shorter Internal Transcribed Spacer amplicons (ITS) of the nuclear ribosomal RNA (rRNA) (Bokulich and Mills, 2013; Ihrmark et al., 2012).

The ITS region is considered the universal barcode for identification of fungi and includes the ITS1 and ITS2 regions, separated by the 5.8S gene. These two regions (ITS1-2) are characterized by high evolutionary rates and are edged by highly conserved regions with suitable target sites for universal primers (Begerow et al., 2010). However, the complete ITS region located between the 18S and 28S genes in the nuclear ribosomal RNA is considered too long for 454

sequencing or other NGS (Bellemain and Carlsen, 2010). Therefore, various primers are used to amplify parts of the ITS region. In this study, we selected the primer ITS3ngs that targets a site in the ITS1 and the degenerate reverse primer ITS4 which targets an ITS-flanking site in the ribosomal large subunit (LSU) encoding regions (White TJ, Bruns T, Lee S, 1990) based on their ability to amplify fungal species through *in silico* analysis (Bellemain and Carlsen, 2010; Tedersoo et al., 2015). Nevertheless, the nuclear rRNA large subunit (LSU/28S/26S) and small subunit (SSU/18S) genes have also been often used to address fungal diversity (Bonanomi et al., 2016; David et al., 2014; De Filippis et al., 2017b, 2017a; Garofalo et al., 2015; Stellato et al., 2015; Wang et al., 2015). To bring an overall perspective, most yeasts have been identified from sequence divergence in the D1/D2 domain of LSU rRNA (Kurtzman and Robnett, 1997). Despite the great resolution to recognized yeast species through 26S rRNA sequencing reactions, little is known about the potential uses and bias that can be introduced when using this target region in NGS. In this context, it is necessary for ecological studies to compare different targeting regions to describe the most accurate and reliable ecological populations in a food system. Given the nature of current challenges, the selection of a suitable genetic marker for the identification of fungi will help researchers to clear current issues insight into the selection of primer sets.

The main focus of this study is to address sequencing target regions and primer biases on one of the dominating taxonomic groups of fungi in the Dikarya, Ascomycota, which represents 53 % of the described species of true Fungi (Koljalg et al., 2013). This phylum is important in the food industry and serves as a source for biomass production, but also includes known human and plant pathogens (Bekatorou et al., 2006; Berbee, 2001). The present research focused on the assessment of two different targeting sites for amplicon-based Illumina NGS studies. We tested the 26S primer set, delivering high coverage and accurate taxonomic assignment of short (~ 400 bp) fungal amplicon *versus* the performance with the ITS2 region. This research intends to bring new insights in the field of taxonomic assignment, validation and resolution of uncertainties on using amplicon-sequencing approaches for fungi identification by using mock samples as well as fermented

samples. Attention was paid for monitoring fungi in mock communities and biological samples, where taxonomic assurance of the technique, and mapping and monitoring fungi dynamics are investigated for food applications.

2. Materials and methods

2.1. Primer selection and in silico analysis

Primer pairs targeting the ITS2 region (Tedersoo et al., 2015), and the D1 domain of 26S rRNA gene (Cocolin et al., 2000), were selected and reported in Table 1. For the amplification of the D1 domain of the 26S, we modified the LS2-MF primer sequence position from reverse to forward, corresponding to nucleotide position 266 of *Saccharomyces cerevisiae* 26S gene as described by Cocolin *et al.*, (2000) and a reverse primer NL4 (Jespersen et al., 2005). The Illumina overhang adapter sequences were added to locus-specific sequences. The D1 region from the 26S gene was amplified *in silico* to compare primer specificity and taxonomic coverage of both LS2-MF and NL4 by using Primer Prospector (Walters et al., 2011) against the constructed 26S databases and SILVA's database.

2.2. Mock community preparation, DNA extraction, and PCR amplification

Strains of yeast and filamentous fungi listed in Table 2 (DISAFA collection, Torino) were used and cultured on Malt Extract Agar (Oxoid, Milan, Italy) plus 25 mg l^{-1} streptomycin (Sigma, Milan, Italy) incubated at 28 °C for 72 h for yeast and 7 to 10 days for fungi. DNA extraction from yeast was carried out from a loopful of grown culture while 250 mg of mycelium was scraped from the plate for filamentous fungi. DNA extraction was carried out as described by Cocolin et al., (2000). DNA from each strain was quantified by using the Qubit dsDNA assay kit (Thermo Fisher Scientific, Milan, Italy) and standardized at 5 ng/µl. A pool (Mock-DNA) containing each of the standardized strain DNA was then obtained and subject to amplification of the ITS2 and the 26S regions. PCR was carried out for the two target regions using a PCR mixture prepared with 12.5 µl

of the 2X Kapa HiFi HotStart ReadyMix Taq (Roche, Milan, Italy), 1 μM each primer, 2.5 μl of DNA template, and PCR-grade water. Each PCR were subject to the following amplification conditions: thirty cycles of 30 s of denaturation (95 °C), 30 s of primer annealing (55 °C), and 30 s of primer elongation (72 °C), followed by a final elongation step (72 °C) of 10 min.

The amplification of each fungal strain was carried out by using the same couple of primers, each amplicon was then purified using the Agencourt AMPure XP beads (Beckman Coulter Genomics) and quantified using Qubit dsDNA assay kit. Based on the amplicon size of DNA assessed by using a Biorad experion workstation (Biorad, Milan, Italy), amplicons concentration was determined. Amplicons were diluted at 20 mM and aliquots of 10 μl were pooled together to construct a Mock-Amp. In total, two independent Mock-DNA and Mock-Amp were obtained by two independent DNA extraction, quantification and pooling procedure.

2.3. DNA extraction and PCR amplification of fermented cocoa beans

A total of twenty fermented cocoa beans samples were collected and DNA extracted as following original study (Mota-Gutierrez et al., 2018). Samples were collected during a fermentation period of 0, 48, 96 and 120 h. Detailed information of samples is reported in Mota-Gutierrez et al., 2018 and in supplementary table S1. Briefly, total DNA was extracted from the pellet of cocoa matrix by using the MasterPure Complete DNA & RNA Purification kit (Illumina Inc, San Diego, CA) following the manufacturer's instructions. DNA was quantified by using the Qubit dsDNA assay kit (Thermo Fisher Scientific), standardized at 5 ng/μl and subject to amplification of the two target regions using primers and procedure as described above.

2.4. Library preparation and sequencing

Sequencing was performed for the two target regions and for the three target samples (Mock-DNA, Mock-Amp and cocoa samples). After the first purification step following the Illumina sample preparation procedure, the library was combined with the sequencing adapters and

dual indices using the Nextera XT Index Kit (Illumina, San Diego, USA), obtaining the multiplexed paired-end libraries. Individual libraries concentration in nM were calculated based on the size of amplicons by using a Biorad Experion workstation (Biorad) and diluted to 4 nM, denaturated with 0.2 N NaOH and spiked with 20 % (v/v) of PhiX. The combination of pool library and PhiX were diluted to 12 pM and paired-end sequencing was performed on the MiSeq platform, using MiSeq Reagent Kit V3 (2 x 250bp) (Illumina, San Diego, USA), following the standard Illumina sequencing protocol.

2.5. Constructed 26S rRNA sequence database

The construction database of fungal rRNA gene sequence of the 26S gene was used to select primers, which amplify the D1 region of a broad fungal taxa. The sequences were downloaded from the Nucleotide database of the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/nucleotide/; accessed March 07, 2018). The database was constructed using the large subunit rRNA gene sequences, 23.381 sequences were downloaded using diverse taxonomic ID and the query word "26S rRNA". The final constructed database consisted of 4 phyla, 27 classes, 172 families, and 442 fungal strains. Incomplete sequences or sequences with absent taxonomies were removed. Duplicate sequences and sequences that clustered together at 99 % of similarity were discarded by using Prinseq and USEARCH respectively (Schmieder and Edwards, 2011). A taxonomy file, matching exactly seven taxonomic levels (root, subphylum, class, order, family, genus and species) was generated from the corresponding taxonomy strings to be compatible with implementation in the NGS analysis pipeline QIIME. Both files were manually curated for accuracy and consistency. Sequences obtained from the constructed 26S database from biological and mock samples were compared using SILVA database. All sequences identified by D1 domain of 26S rRNA sequence analysis from biological samples and mock communities were compared with our constructed database**.**

2.6. Bioinformatics

175 Paired-end reads (2x250 bp) were first merged using the FLASH software (Magoč and Salzberg,
176 2011), with default parameters. Joint reads were further quality filtered (Phred < Q20) using the
177 QIIME 1.9.0 software (Caporaso et al., 2011). Chimeras were then removed with the adopted
178 USEARCH version 8.1 software. Lastly, OTUs were picked at 99 % of similarity by means of
179 UCLUST clustering methods (Edgar, 2010) and representative sequences of each cluster were used
180 to assign taxonomy. For the 26S data, each cluster was used to assign taxonomy using the
181 Constructed 26S rRNA gene database and SILVA, while for the ITS dataset the UNITE rRNA ITS
182 database version 2012, by means of the RDP Classifier. Sequences were double-checked using the
183 BlastN search tool (http://www.ncbi.nlm.nih.gov/blast/) to confirm the taxonomy assignment.
184 Cocoa samples datasets (ITS and 26S) were rarefied at 10,018 reads after raw read quality filtering,
185 and OTU tables were filtered for OTUs occurring at 1 % of the relative abundance in at least 2
186 samples. While for mock community reads from the two target regions were rarefied at 17,313
187 reads.

188 *2.7. Statistical analyses*

189 Statistics and plotting were carried out in the R environment (www.r-project.org). Alpha
190 diversity indices were calculated using the diversity function of the vegan package (Dixon, 2007).
191 OTUs table were used to find differences between target regions by Anosim statistical test in R
192 environment. A two-sided permutation test with 999 permutations was performed to compare the
193 OTUs distribution and alpha diversity between the two datasets. Pairwise Kruskal-Wallis Wilcoxon
194 test or one way- ANOVA coupled with the Duncan honestly significant difference (HSD) test were
195 used as appropriate to determine significant differences in alpha diversity or OTU abundance from
196 mock communities and biological samples. Statistical analysis was acquired through the function
197 *aov* through the *stats* package and principal component analysis were plotted using the function
198 *dudi.pca* through the *made4* package using R version 3.3.2

199 *2.8. Accession numbers*

The ITS and 26S rRNA gene sequences are available at the Sequence Read Archive of the National Center for Biotechnology Information (NCBI), under the SRA accession number SRP126081 (fermented cocoa samples ITS) and SRP150401 (fermented cocoa samples 26S and mock sequences data).

3. Results

3.1. In silico performance of 26S primers

We performed an *in-silico* analysis of the 26S primer set against our constructed database and SILVA using Primer Prospector. LS2-MF primer showed the lowest weighted score (Fig. S1A) indicating higher coverage across the database sequences and lower number of mismatches if compared with NL4 (Fig. S1B). Comparing the taxonomic coverage of LS2-MF and NL4 against *Zygomycota*, *Glomeromycota*, *Ascomycota* and *Basidiomycota* sequence, LS2-MF showed the best performance with a coverage higher than 80 % for all the phyla except for *Glomeromycota* (Fig. S2A), while NL4 account for the 20 % of the coverage against our constructed database (Fig. S2B). Regarding the performance of the primer sets against the SILVA's database, the score of the primers was higher compared with our database (data not shown).

3.2. Performance of primers by mock community analysis

A mock community containing twenty fungal species (Table 2) was prepared to validate the performance of the two target regions. The possible effect of the bias introduced by PCR (Mock-DNA) and that of sequencing (Mock-Amp) was then evaluated. Amplicon length of the single species showed little variation when the 26S gene was amplified (461 ± 30 bp) while for ITS2 we observed greater dispersion in size (445 ± 55 bp) (Table 2). Significant difference in mycobiota composition (Anosim statistical test, $P < 0.05$) by using the two target regions or mock communities (DNA or AMP) was observed by Principal Component Analysis (Fig. 1).

In both samples (Mock-DNA and Mock-Amp), the target region ITS2 showed similar abundances with respect to the theoretical value for two fungal species, namely *Torulaspora delbrueckii* and *Plectosphaerella cucumerina* (Table 3). Similarly, with the 26S target gene, for six species, abundances retrieved from both Mock-DNA and Mock-Amp samples were comparable to the theoretical values (*Aspergillus fumigatus, Pichia membranifaciens, Pichia kudriavzevii, Penicillium glabrum, Penicillium brevicompactum* and *Starmerella bacillaris*). Furthermore, for the 26S region, the species *Alternaria alternata, Aspergillus flavus* and *Fusarium oxysporum* rendered different abundances in the Mock-DNA and the Mock-Amp but in both samples the values were comparable to the theoretical.

For 18 out of the 20 fungal species tested, the ITS2 region resulted in underestimation or overestimation with respect to the theoretical value in the Mock-DNA, Mock-Amp or both ($P < 0.05$). Four species were significantly overestimated (*A. fumigatus* (439bp), *F. verticillioides* (415bp), *K. marxianus* (521bp) and *P. brevicompactum* (428bp)) while other 4 were significantly underestimated (*Galactomyces geotrichum* (324bp), *Hanseniaspora opuntiae* (484bp), *Schizosaccharomyces pombe* (562bp) and *Starmerella bacillaris* (361bp)), in both samples. Interestingly, *G. geotrichum* and *S. pombe* were not detected in any of the two samples. Nine more species resulted to be significantly different from the theoretical value (either higher or lower) in the Mock-DNA or Mock-Amp sample only (Table 3).

When the 26S region was targeted, 11 out of the 20 species were either underestimated or overestimated in the Mock-DNA, Mock-Amp or both. More specifically, 4 species were underestimated in both types of samples (*Candida sake* (467bp), *H. opuntiae* (435 bp), *Kluyveromyces marxianus* (427 bp), *T. delbrueckii* (461 bp)) and only one (*G. geotrichum* (502bp)) was overestimated. Five more species were significantly different from the theoretical value (either higher or lower) in the Mock-DNA or Mock-Amp sample only (Table 2 and 3). It also should be

pointed out that we did not observe a clear correlation between amplicon size and over or underestimation.

In the Mock-Amp samples, correct relative quantification was obtained for 13 out of the 20 species targeting the 26S region and for 10 out of 20 species with the ITS2 region. In the Mock-DNA samples, correct relative quantification was obtained for 10 out of 20 species in the 26S region and for 4 out of 20 species targeting the ITS2 (Table 3).

Remarkably, *G. geotrichum* and *S. pombe* were only detected when the 26S region was targeted while *H. opuntiae* was the only species that was consistently underestimated, independently from the target region or sample. Overall, 26S sequencing data aligned better to theoretical abundance values for the fungal species tested than did the ITS sequencing data.

3.3. Mycobiota in biological samples

Sequencing of twenty fermented cocoa beans samples collected during a previous experiment, after amplification with the primers ITS2 and 26S showed a mean sequence length of 412 and 390 bp. respectively and an estimated sample coverage of 97.73 and 95.87 %, respectively (See Table S1). The 26S target region revealed greater OTU richness compared to the ITS2 region ($P < 0.05$) as shown in Fig. 2. Overall, 20 and 37 fungal OTUs were identified during the fermentations using the primer set ITS2 and 26S, respectively. In addition, we observed differences in length distribution across the two target genes. Histogram of reads length of 26S showed that the higher reads proportion were around 380 bp while for ITS2 we observed a varied distribution of the reads length around 370bp, 400bp, 420bp, and 450bp (See Fig. S3). Eleven OTUs, namely *Candida jaroonii, Candida tallmaniae, Fusarium, Hanseniaspora, H. opuntiae, Hanseniaspora uvarum, K. marxianus, S. cerevisiae, Saccharomycopsis crataegens, T. delbrueckii* and *Pichia pijperi* were detected by both targeting regions (Fig. 3). The relative abundance of several fungal species was significantly different according to the type of amplicon used ($P < 0.05$, Fig. 4), in which significantly higher relative abundance was found for *S. cerevisiae, P. pijperi,* and *H. uvarum* ($P <$

0.05) using ITS2 target gene, while *Hanseniaspora* showed higher abundance when using the 26S ($P < 0.05$, Fig. 4).

3.4. Performance of the new constructed 26S database against SILVA

To validate the new 26S database, biological samples and mock communities identified by D1 domain of 26S rRNA sequence analysis were compared with SILVA's database (Quast et al., 2013). Significant difference in mycobiota composition of the two databases was observed in mock communities (Anosim statistical test, $P < 0.05$). In detail, the constructed 26S database assigned successfully the twenty fungal species, while SILVA's database assigned only ten (*A. alternata, H. osmophila, K. marxianus, P. kudriavzevii, P. membranifaciens, S. ludwigii, S. pombe, S. bacillaris, T. delbrueckii,* Table 4). Interestingly, *S. cerevisiae* was not detected by using the SILVA's database from the mock communities. In contrast, no significant differences in the OTU distributions and the alpha diversity calculations were observed between the two datasets from biological samples. In depth, the constructed 26S database assigned 37 fungal species, described above, while SILVA's database assigned 35 (data not shown). However, we observed that *S. cerevisiae* and *H. uvarum* were not detected by using the SILVA's database.

4. Discussion

New tools and molecular techniques have been used to detect microbial ecology in the past decades. Recently, the interest in the use of amplicon sequencing to identify taxonomically relevant taxa in food has increased. However, this approach has potential biases as previously described (Bowers et al., 2015; Fouhy et al., 2016) where primer selection is considered one of the most important sources of biases (Bokulich et al., 2014; Bonanomi et al., 2016; David et al., 2014; De Filippis et al., 2017a; Ercolini, 2013; Garofalo et al., 2015; Stellato et al., 2015; Stielow et al., 2015; Wang et al., 2015). The 26S region (D1 domain) of the rRNA encoding gene and the ITS2 region have been proposed as good candidates for identifying fungal species when using NGS technologies due to the high taxonomic resolution (Tedersoo et al., 2015). In this study, we performed a

298 comparative evaluation of two regions as amplicon sequencing targets for the identification of fungi
299 and it also describes the mycobiota community in food matrices.

300 Recently the 26S region has been studied using the Roche 454 technology (De Filippis et al.,
301 2017b). However, this platform has been shown to result in high sequencing errors due to A and T
302 rich homopolymers (Luo et al., 2012), while Illumina does not present this sequencing error (Erlich
303 et al., 2008). Our results and a previous study (De Filippis et al., 2017b) reveal that 26S gene is a
304 reliable target site for both NGS technologies (Roche 454 and Illumina) for eukaryotic species. In
305 order to evaluate the effect of the target gene used we compared the sequencing results of both
306 DNA samples and mock communities. In our study, different relative abundances were obtained for
307 both mock communities and biological samples and these differences were based on the PCR target
308 used. We observed in such cases that ITS2 target region led to underestimations of species with
309 longer fragments (*S. pombe,* and *H. opuntiae*), while an overestimation of shorter fragments
310 occurred (*F. vercillioides, A. fumigatus* and *P. brevicompactum*). However, it should be pointed out
311 that this correlation between amplicon length and estimation is not valid for all the species
312 analysed. Apart from amplification length, other parameters that influence relative abundance
313 calculations of taxa within samples could be considered. Sampling errors, different primers
314 alignment efficiencies during PCR amplification, performance of degenerate primers used during
315 PCR amplification, result in wrong representation in terms of relative abundance of microbial
316 populations (Ihrmark et al., 2012; Polz and Cavanaugh, 1998).

317 In addition to underestimation of abundances, "identification bias" is also common to
318 amplicon-based analyses, where minor groups are poorly represented (Koljalg et al., 2013). The
319 lack of updated reliable public reference data set and the discrepancies to refer to fungal species
320 have been recently demonstrated for the ITS sequences (Koljalg et al., 2013). This is also in
321 accordance with our results, suggesting that our new database for the 26S, validated by the widely
322 used SILVA, proved to be a curated and rich database to be used. Differences between the two

databases regarding taxonomic classification of sequences were obtained. The newly constructed 26S database delivered a more precise taxonomic assignment of the sequences. This could be due to the fact that SILVA database comprised also non-microbial sequences, incomplete sequences and sequences with unassigned taxonomy. In contrast, each taxonomy in our database was double checked to get the higher taxonomic resolution, obtaining clearly more robustness results in terms of taxonomic assignment from the biological samples. Special attention must be paid on the miss-identification of fungal strain on the current available database. This current issue is pointed out in this study, in which *S. cerevisiae* and *H. uvarum* were misidentified from fermented samples, using SILVA's database.

Given the intricate nature of PCR, the amplification of biological samples has been problematic (Polz and Cavanaugh, 1998). Our results exhibited high proportion of fungal coverage (98 - 96 %) by both primer pair sets, which suggested that fungi account for roughly the complete eukaryote rRNA in the studied fermented cocoa beans on average. The results also highlight a lower biodiversity of fungal communities for ITS2 compared with the 26S region in fermented cocoa beans, which contradicts with previous studies where ITS region has been used in NGS studies as the universal primer set for fungi (Bokulich et al., 2014, 2012; Tedersoo et al., 2015). Such discrepancies between outcomes of different studies may arise on account of the biased quantification of relative abundance of taxa due to the uneven length of ITS fragments (Bellemain and Carlsen, 2010), the preferential amplification of rRNA genes for certain taxa by PCR, sequencing bias due to unequal amplification of the target gene or due to inaccurate taxonomic classification of reference databases (Simon and Daniel, 2011). Despite these challenges, the greater recovery trends in the community composition in the 26S target region observed here, have been supported from previous studies, where higher discrimination power of species identification in early diverging lineages of LSU compared with ITS was reported (Schoch et al., 2012).

Our study suggested that the 26S as a target showed greater biodiversity in biological samples compared with the universal primer ITS. However, it should be noted that the present study shows the performance of a new pair of primers targeting the 26S region for fungal strains. Therefore, the novelty of these primer sets is also our limitation, that can be successfully overcome through future research focusing on the use of small fragments of the LSU region to target fungal species, that could support our observations. Therefore, the combination of both target genes, where species identification can be performed applying ITS and phylogenetic analysis with 26S, is highly recommended and the use of both will depend on the purpose of taxa investigation (Klaubauf et al., 2010; Schoch et al., 2012). From a molecular microbial ecology perspective, in terms of classification of marker-gene sequences, there is evidently a need for more extensive testing of primers targeting different genes and *loci*, to support and identify all fungal species in NGS studies. Clearly, the benefits of characterizing fermented microbial diversity may bring important advancements to the food industry, such as discrimination of starter culture to improve food quality or to accelerate processes. This study provides new insight into the selection of better primers and taxonomical assignment to study fungal ecology, which should enable food research to gain better view of the microbial diversity present in a range of fermentations avoiding biases. One also notes, the limited availability of updated databases to assess ecological populations.

Acknowledgments

We would like to thank Dr. Sara Franco Ortega and Simona Prencipe for proving fungal strains.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Reference

Begerow, D., Nilsson, H., Unterseher, M., Maier, W., 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl. Microbiol. Biotechnol. 87, 99–108. https://doi.org/10.1007/s00253-010-2585-4

Bekatorou, A., Psarianos, C., Koutinas, A.A., 2006. Production of food grade yeasts. Food Technol. Biotechnol. 44, 407–415.

Bellemain, E., Carlsen, T., 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 10, 1–9.

Berbee, M.L., 2001. The phylogeny of plant and animal pathogens in the Ascomycota. Physiol. Mol. Plant Pathol. 59, 165–187. https://doi.org/10.1006/pmpp.2001.0355

Blackwell, M., 2011. The fungi: 1, 2, 3 ... 5.1 million species? Am. J. Bot. 98, 426–438. https://doi.org/10.3732/ajb.1000298

Bokulich, N.A., Bamforth, C.W., Mills, D.A., 2012. Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale. PLoS One 7. https://doi.org/10.1371/journal.pone.0035507

Bokulich, N.A., Mills, D.A., 2013. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. Appl. Environ. Microbiol. 79, 2519–2526. https://doi.org/10.1128/AEM.03870-12

Bokulich, N.A., Thorngate, J.H., Richardson, P.M., Mills, D.A., 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. Proc. Natl. Acad. Sci. 111, E139–E148. https://doi.org/10.1073/pnas.1317377110

Bonanomi, G., De Filippis, F., Cesarano, G., La Storia, A., Ercolini, D., Scala, F., 2016. Organic farming induces changes in soil microbiota that affect agro-ecosystem functions. Soil Biol. Biochem. 103, 327–336. https://doi.org/10.1016/j.soilbio.2016.09.005

Bowers, R.M., Clum, A., Tice, H., Lim, J., Singh, K., Ciobanu, D., Ngan, C.Y., Cheng, J.F., Tringe, S.G., Woyke, T., 2015. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. BMC Genomics 16, 1–12. https://doi.org/10.1186/s12864-015-2063-6

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107

Cocolin, L., Alessandria, V., Botta, C., Gorra, R., De Filippis, F., Ercolini, D., Rantsiou, K., 2013. NaOH-debittering induces changes in bacterial ecology during table olives fermentation. PLoS One 8. https://doi.org/10.1371/journal.pone.0069074

Cocolin, L., Bisson, L.F., Mills, D.A., 2000. Direct profiling of the yeast dynamics in wine fermentations. FEMS Microbiol. Lett. 189, 81–87. https://doi.org/10.1016/S0378-1097(00)00257-3

David, V., Terrat, S., Herzine, K., Claisse, O., Rousseaux, S., Tourdot-Maréchal, R., Masneuf-Pomarede, I., Ranjard, L., Alexandre, H., 2014. High-throughput sequencing of amplicons for monitoring yeast biodiversity in must and during alcoholic fermentation. J. Ind. Microbiol. Biotechnol. 41, 811–821. https://doi.org/10.1007/s10295-014-1427-2

De Filippis, F., La Storia, A., Blaiotta, G., 2017a. Monitoring the mycobiota during Greco di Tufo and Aglianico wine fermentation by 18S rRNA gene sequencing. Food Microbiol. 63, 117–122. https://doi.org/10.1016/j.fm.2016.11.010

De Filippis, F., Laiola, M., Blaiotta, G., Ercolini, D., 2017b. Different amplicon targets for sequencing-based studies of fungal. Appl. Environ. Microbiol. 83, 1–9.

Dixon, P., 2007. Vegan: Community ecology package for R. J. Veg. Sci. 14, 927–930. https://doi.org/10.1658/1100-9233(2003)014[0927:vaporf]2.0.co;2

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Ercolini, D., 2013. High-throughput sequencing and metagenomics: Moving forward in the culture-independent analysis of food microbial ecology. Appl. Environ. Microbiol. 79, 3148–3155. https://doi.org/10.1128/AEM.00256-13

Ercolini, D., De Filippis, F., La Storia, A., Iacono, M., 2012. "Remake" by high-throughput sequencing of the microbiota involved in the production of water Buffalo mozzarella cheese. Appl. Environ. Microbiol. 78, 8142–8145. https://doi.org/10.1128/AEM.02218-12

Erlich, Y., Mitra, P.P., delaBastide, M., McCombie, W.R., Hannon, G.J., 2008. Alta-Cyclic: A self-optimizing base caller for next-generation sequencing. Nat. Methods 5, 679–682. https://doi.org/10.1038/nmeth.1230

Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J., Cotter, P.D., 2016. 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. BMC Microbiol. 16, 1–13. https://doi.org/10.1186/s12866-016-0738-z

Garofalo, C., Osimani, A., Milanović, V., Aquilanti, L., De Filippis, F., Stellato, G., Di Mauro, S., Turchetti, B., Buzzini, P., Ercolini, D., Clementi, F., 2015. Bacteria and yeast microbiota in milk kefir grains from different Italian regions. Food Microbiol. 49, 123–133. https://doi.org/10.1016/j.fm.2015.01.017

Ihrmark, K., Bödeker, I.T.M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K.E., Lindahl, B.D., 2012. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. FEMS Microbiol. Ecol. 82, 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Jespersen, L., Nielsen, D.S., Hønholt, S., Jakobsen, M., 2005. Occurrence and diversity of yeasts involved in fermentation of West African cocoa beans. FEMS Yeast Res. 5, 441–453. https://doi.org/10.1016/j.femsyr.2004.11.002

Klaubauf, S., Inselsbacher, E., Zechmeister-Boltenstern, S., Wanek, W., Gottsberger, R., Strauss, J., Gorfer, M., 2010. Molecular diversity of fungal communities in agricultural soils from Lower Austria. Fungal Divers. 44, 65–75. https://doi.org/10.1007/s13225-010-0053-1

Koljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22, 5271–5277. https://doi.org/10.1111/mec.12481

Kurtzman, C.P., Robnett, C.J., 1997. Identification of clinically important ascomycetous yeasts based on nucleotide divergence in the 5 ' end of the large-subunit (26S) ribosomal DNA gene. J. Clin. Microbiol. 35, 1216. https://doi.org/0095-1137/97

Li, X.R., Ma, E.B., Yan, L.Z., Meng, H., Du, X.W., Zhang, S.W., Quan, Z.X., 2011. Bacterial and fungal diversity in the traditional Chinese liquor fermentation process. Int. J. Food Microbiol. 146, 31–37. https://doi.org/10.1016/j.ijfoodmicro.2011.01.030

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7, 30087. https://doi.org/10.1371/journal.pone.0030087

Magoč, T., Salzberg, S.L., 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mota-Gutierrez, J., Botta, C., Ferrocino, I., Giordano, M., Bertolino, M., Dolci, P., Cannoni, M., Cocolin, L., 2018. Dynamics and biodiversity of bacterial and yeast communities during the fermentation of cocoa beans. Appl. Environ. Microbiol. AEM.01164-18. https://doi.org/10.1128/AEM.01164-18

Polz, M.F., Cavanaugh, C.M., 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. 64, 3724–3730. https://doi.org/10.1007/s00253-012-4244-4

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, PacificBiosciences and Illumina MiSeq sequencers. BMC Genomics 13, 1–13. https://doi.org/10.1186/1471-2164-13-341

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596. https://doi.org/10.1093/nar/gks1219

Roh, S.W., Kim, K.H., Nam, Y. Do, Chang, H.W., Park, E.J., Bae, J.W., 2010. Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. ISME J. 4, 1–16. https://doi.org/10.1038/ismej.2009.83

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

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., et al., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. 109, 6241–6246. https://doi.org/10.1073/pnas.1117018109

Stellato, G., De Filippis, F., La Storia, A., Ercolini, D., 2015. Coexistence of lactic acid bacteria and potential spoilage microbiota in a dairy processing environment. Appl. Environ. Microbiol. 81, 7893–7904. https://doi.org/10.1128/AEM.02294-15

Stielow, J.B., Lévesque, C.A., Seifert, K.A., Meyer, W., et al., 2015. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. Persoonia - Mol. Phylogeny Evol. Fungi 35, 242–263. https://doi.org/10.3767/003158515X689135

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

Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N., Knight, R., 2011. Primer Prospector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27, 1159–1161. https://doi.org/10.1093/bioinformatics/btr087

Wang, C., García-Fernández, D., Mas, A., Esteve-Zarzoso, B., 2015. Fungal diversity in grape must and wine fermentation assessed by massive sequencing, quantitative PCR and DGGE. Front. Microbiol. 6, 1–8. https://doi.org/10.3389/fmicb.2015.01156

White TJ, Bruns T, Lee S, T.J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, In: PCR Pr. ed. Academic Press, Inc., New York.

Table legend

Table 1. Primers used for Illumina MiSeq sequencing

Table 2. Fungal strains used for sequencing analysis and respective amplicon length

Table 3. Relative abundance (%) of the fungal species identified in mock communities amplified using two different target regions. The expected concentration is referred to as theoretical

Figure legend

Fig 1. Principal component analysis (PCA) based on mock mycobiota composition.

Fig 2. Boxplots describe α-diversity measures (Chao1, Shannon index and number of observed species) of fermented cocoa bean samples. Individual points and brackets represent the richness estimate and the theoretical standard error range, respectively.

Fig 3. Distribution of OTUs in fermented cocoa bean samples in the amplicon datasets divided into 26S (upper figure) and ITS (lower figure). Only OTUs with an incidence above 1 % in at least 2 samples are shown.

Fig 4. Boxplots describe statistically different species detected in fermented cocoa bean samples analysed with two different target genes. Individual points and brackets represent the relative abundance and the theoretical standard error range, respectively.

Supplementary table

Table S1. Estimation of sample coverage of fungal community of fermented cocoa beans using 26S target gene (A). Estimation of sample coverage of fungal community of fermented cocoa beans using ITS2 target region (B)

Table S2. Relative abundance of mock communities identified by two different fungal databases

Supplementary figure

S1. Bar chart showing the evaluation of primer efficiency. Overall matches and weight score of 26S primer pair against our constructed database 26S. Assessment of LS2-MF forward primer and NL4 reversed primer

S2. Predictive taxonomic coverage of 26S primers. Numeric values above bins represent total sequence counts for each set, LS2-MF forward primer and NL4 reversed primer

S3. Histograms of read length of fungal communities using 26S target region and ITS2 target region

545 **Table 1**

Primer	Features	Primer sequence	Target region	Amplicon length	Reference
ITS3tagmix1	Fwd	5'-CTAGACTCGTCACCGATGAAGAACGCAG-3'	ITS2	385	Tedersoo *et al.*, 2015
ITS4ngs	Rev	5'- TTCCTSCGCTTATTGATATGC-3'	ITS2		Tedersoo *et al.*, 2015
LS2-MF	Fwd	5'-GAGTCGAGTTGTTTGGGAAT-3'	LSU D1	369	This study
NL-4	Rev	5'-GGTCCGTGTTTCAAGACGG-3'	LSU D1		Jespersen *et al.*, 2005

546

547

548

549 **Table 2**

Fungal species	Size bp (26S)	Size bp (ITS2)
Alternaria alternata	454	414
Aspergillus flavus	455	430
Aspergillus fumigatus	459	439
Candida sake	467	393
Fusarium oxysporum	462	405
Fusarium verticillioides	458	415
Galactomyces geotrichum	502	324
Hanseniaspora opuntiae	435	484
Hanseniaspora osmophila	462	520
Kluyveromyces marxianus	427	521
Penicillium brevicompactum	456	428
Penicillium glabrum	458	426
Pichia kudriavzevii	472	431
Pichia membranifaciens	466	398
Plectosphaerella cucumerina	457	441
Saccharomyces cerevisiae	460	496
Saccharomycodes ludwigii	452	470
Schizosaccharomyces pombe	488	562
Starmerella bacillaris	389	361
Torulaspora delbrueckii	461	518

550

551

552 **Table 3**

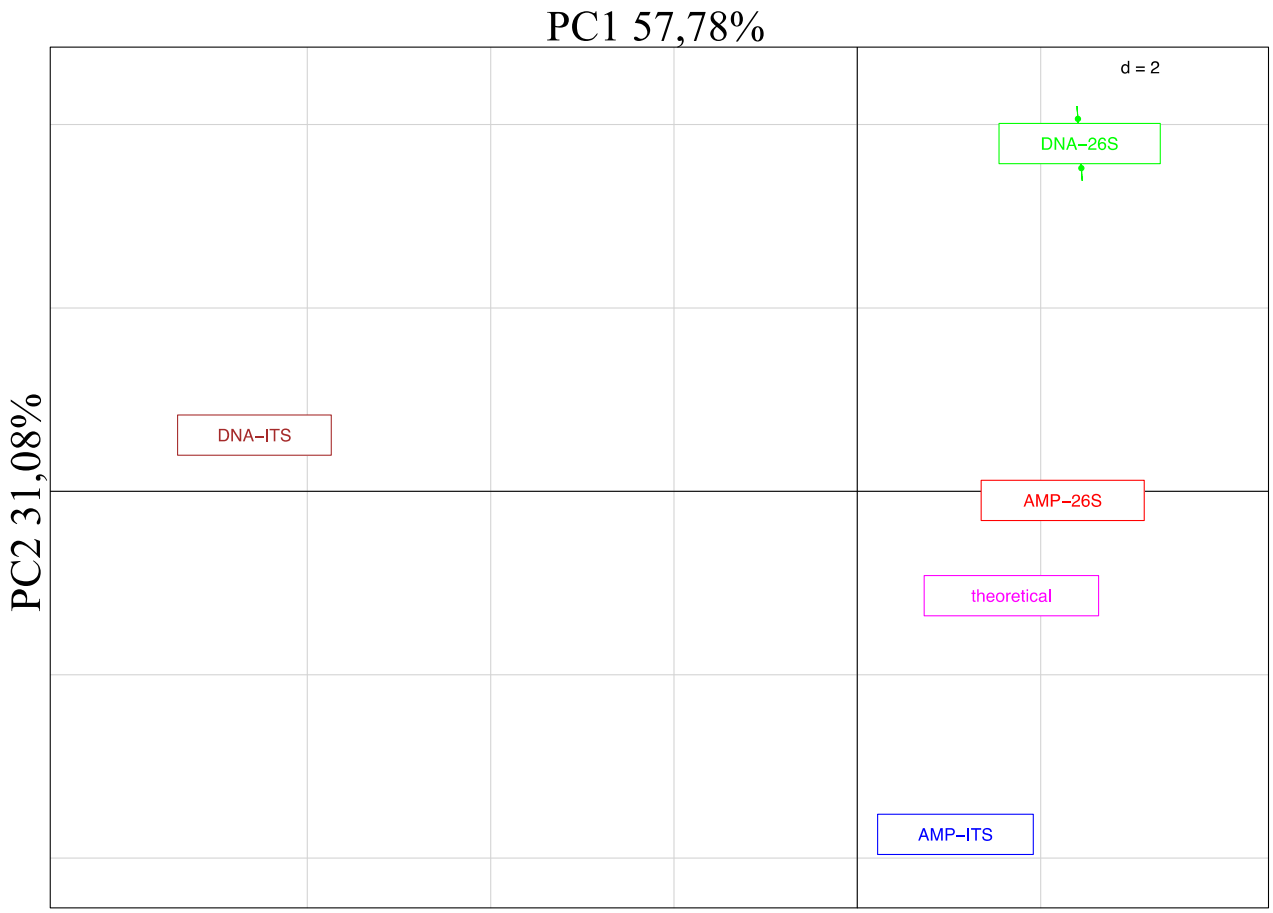
OTU	Theoretical	26S		Theoretical	ITS2	553
		DNA	AMP		DNA	AMP
Alternaria alternata	5 ^{ab}	3.62 ± 0.34 ^{b}	5.71 ± 0.47 ^{a}	5 ^{a}	7.67 ± 0.94 ^{b}	6.63 ± 0.09 ^{ab}
Aspergillus flavus	5 ^{ab}	5.78 ± 0.46 ^{a}	4.03 ± 0.07 ^{b}	5 ^{a}	9.39 ± 0.17 ^{b}	7.19 ± 0.26 ^{a}
Aspergillus fumigatus	5 ^{a}	4.82 ± 1.34 ^{a}	3.71 ± 0.09 ^{a}	5 ^{a}	8.41 ± 0.20 ^{b}	6.14 ± 0.01 ^{c}
Candida sake	5 ^{a}	0.24 ± 0.20 ^{b}	0.06 ± 0.02 ^{b}	5 ^{a}	4.57 ± 0.38 ^{a}	6.19 ± 0.03 ^{b}
Fusarium oxysporum	5 ^{ab}	3.99 ± 0.25 ^{b}	5.08 ± 0.22 ^{a}	5 ^{a}	8.11 ± 0.00 ^{b}	6.33 ± 0.04 ^{a}
Fusarium verticillioides	5 ^{a}	6.27 ± 0.04 ^{c}	4.63 ± 0.02 ^{b}	5 ^{a}	10.44 ± 0.12 ^{c}	6.60 ± 0.27 ^{b}
Galactomyces geotrichum	5 ^{a}	13.39 ± 0.61 ^{c}	6.94 ± 0.11 ^{b}	5 ^{b}	0.00 ± 0.00 ^{a}	0.00 ± 0.00 ^{a}
Hanseniaspora opuntiae	5 ^{b}	2.94 ± 0.16 ^{a}	3.46 ± 0.33 ^{a}	5 ^{c}	0.70 ± 0.07 ^{a}	3.67 ± 0.01 ^{b}
Hanseniaspora osmophila	5 ^{a}	10.63 ± 0.56 ^{b}	4.73 ± 0.13 ^{a}	5 ^{b}	1.63 ± 0.34 ^{a}	4.78 ± 0.08 ^{b}
Kluyveromyces marxianus	5 ^{b}	3.68 ± 0.08 ^{a}	3.33 ± 0.16 ^{a}	5 ^{a}	8.55 ± 0.33 ^{b}	6.71 ± 0.08 ^{b}
Penicillium brevicompactum	5 ^{a}	6.00 ± 0.92 ^{a}	3.55 ± 0.29 ^{a}	5 ^{a}	14.39 ± 0.09 ^{c}	9.04 ± 0.08 ^{b}
Penicillium glabrum	5 ^{a}	4.86 ± 0.19 ^{a}	4.88 ± 0.22 ^{a}	5 ^{b}	7.08 ± 0.11 ^{a}	5.69 ± 0.07 ^{b}
Pichia kudriavzevii	5 ^{a}	7.24 ± 1.89 ^{a}	10.76 ± 0.45 ^{a}	5 ^{b}	0.12 ± 0.14 ^{a}	2.11 ± 0.16 ^{b}
Pichia membranifaciens	5 ^{a}	6.63 ± 1.83 ^{a}	4.82 ± 0.28 ^{a}	5^{b}	2.58 ± 0.39 ^{a}	3.57 ± 0.13 ^{b}
Plectosphaerella cucumerina	5 ^{b}	1.11 ± 0.13 ^{a}	4.33 ± 0.02 ^{b}	5 ^{b}	3.51 ± 0.62 ^{a}	4.45 ± 0.26 ^{b}
Saccharomyces cerevisiae	5 ^{b}	1.86 ± 0.09 ^{a}	4.20 ± 0.49 ^{b}	5 ^{a}	4.74 ± 0.03 ^{a}	5.44 ± 0.10 ^{b}
Saccharomycodes ludwigii	5 ^{b}	3.51 ± 0.00 ^{a}	4.49 ± 0.00 ^{b}	5 ^{b}	0.68 ± 0.03 ^{a}	4.66 ± 0.03 ^{b}
Schizosaccharomyces pombe	5 ^{b}	5.00 ± 0.11 ^{b}	3.71 ± 0.19 ^{a}	5 ^{b}	0.01 ± 0.00 ^{a}	0.00 ± 0.00 ^{c}
Starmerella bacillaris	5 ^{a}	4.21 ± 1.68 ^{a}	5.91 ± 0.11 ^{a}	5 ^{b}	0.16 ± 0.07 ^{a}	0.19 ± 0.10 ^{b}
Torulaspora delbrueckii	5^{c}	2.46 ± 0.02 ^{a}	3.90 ± 0.17 ^{b}	5 ^{a}	5.38 ± 0.11 ^{a}	5.30 ± 0.52 ^{a}

554 Values are expressed as the mean from duplicate determinations (%). Different letters indicate statistical difference related to relative abundances of
555 mock communities using least significant difference test ($P < 0.05$). *P*-values were adjusted using Bonferroni's method. Different colour showed no
556 difference (grey), underestimation (light green) or overestimation (light blue) between mock samples and theoretical data.

557

558

559 **Figure 1.**



560

561

Figure 2.

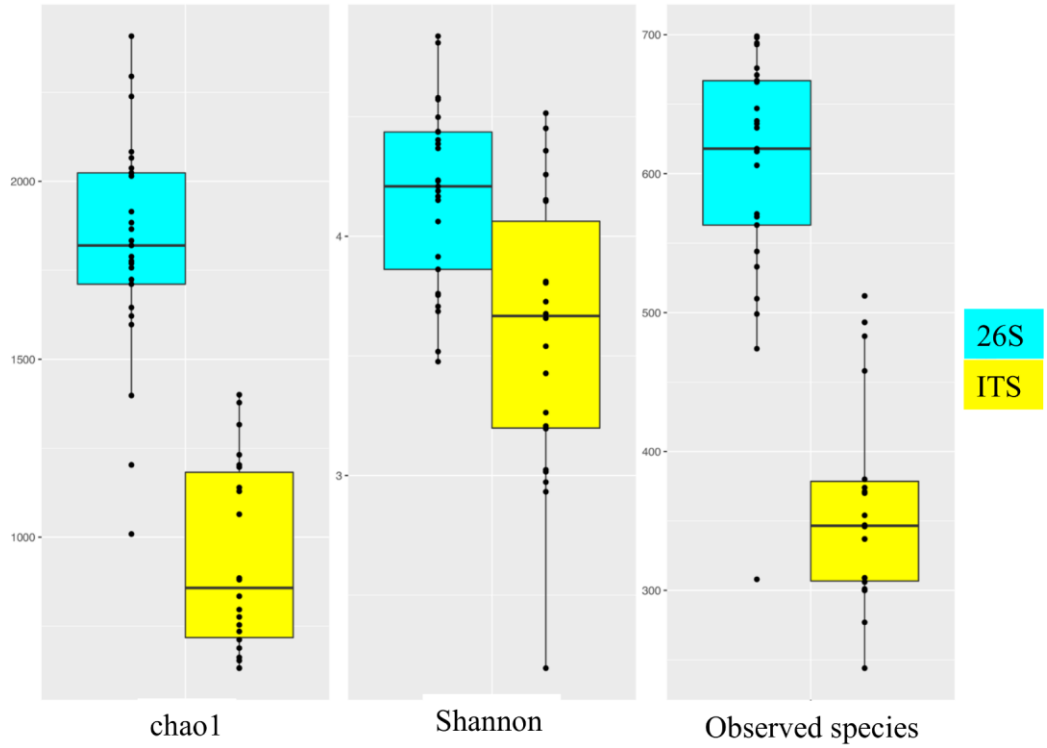

2000
1500
1000
4
3
700
600
500
400
300
26S
ITS
chao1
Shannon
Observed species

Figure 3.

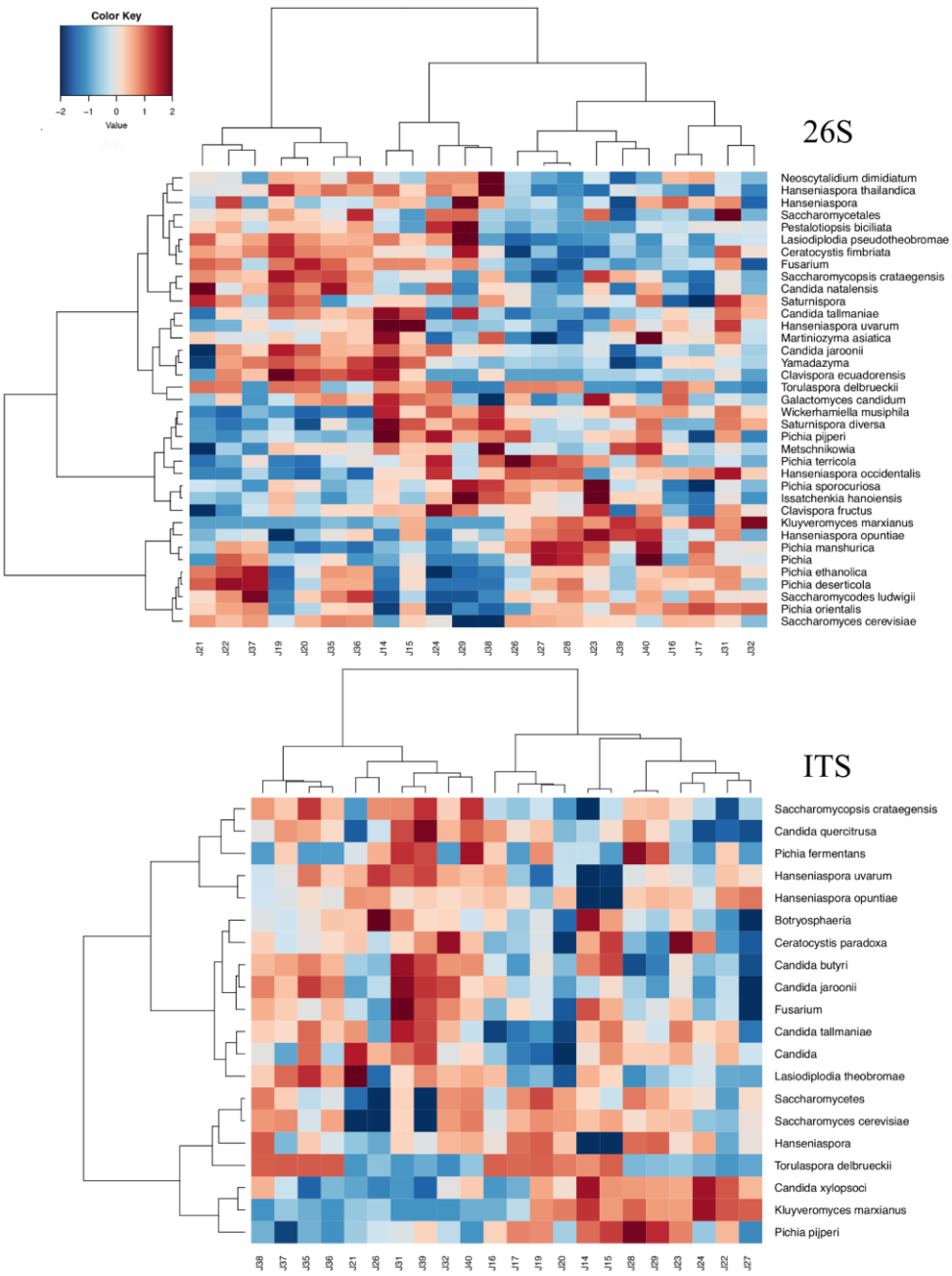
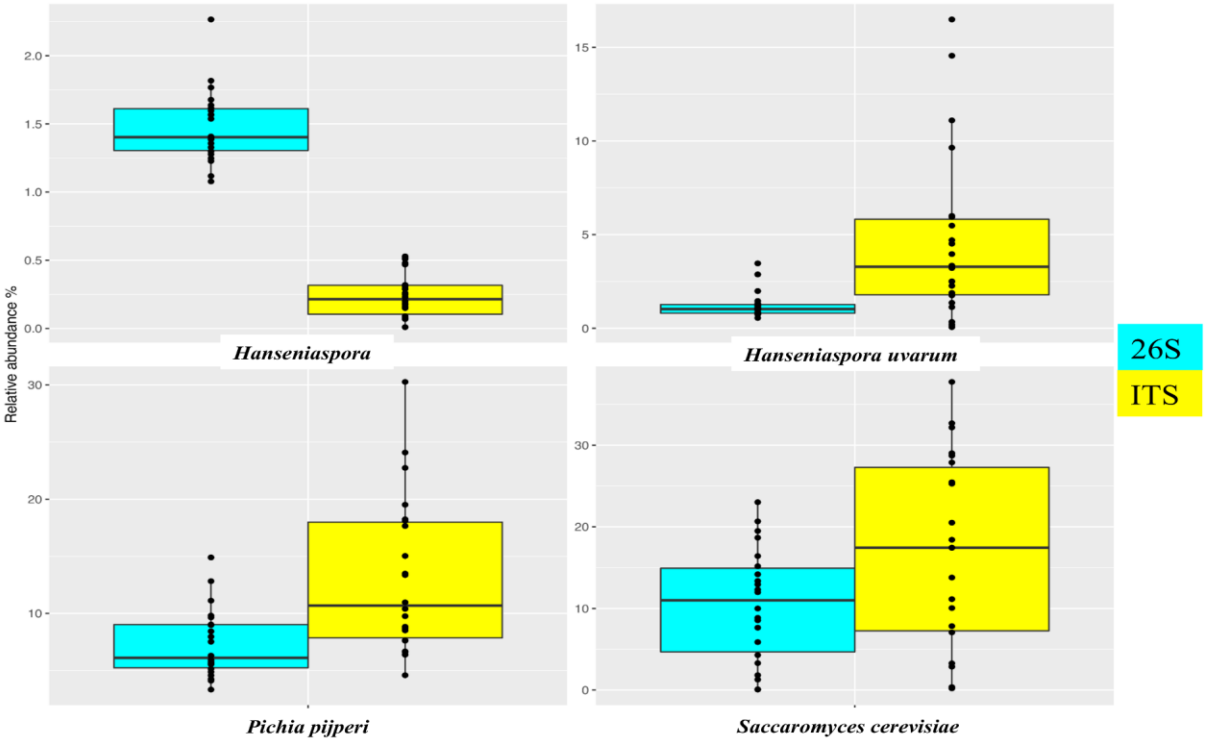

Color Key
−2 −1 0 1 2
Value
26S
Neoscytalidium dimidiatum
Hanseniaspora thailandica
Hanseniaspora
Saccharomycetales
Pestalotiopsis bicilliata
Lasiodiplodia pseudotheobromae
Ceratocystis fimbriata
Fusarium
Saccharomycopsis crataegensis
Candida natalensis
Saturnispora
Candida tallmaniae
Hanseniaspora uvarum
Martiniozyma asiatica
Candida jaroonii
Yamadazyma
Clavispora ecuadorensis
Torulaspora delbrueckii
Galactomyces candidum
Wickerhamiella musiphila
Saturnispora diversa
Pichia pijperi
Metschnikowia
Pichia terricola
Hanseniaspora occidentalis
Pichia sporocuriosa
Issatchenkia hanoiensis
Clavispora fructus
Kluyveromyces marxianus
Hanseniaspora opuntiae
Pichia manshurica
Pichia
Pichia ethanolica
Pichia deserticola
Saccharomycodes ludwigii
Pichia orientalis
Saccharomyces cerevisiae
J21 J22 J37 J19 J20 J35 J36 J14 J15 J24 J29 J38 J26 J27 J28 J23 J39 J40 J16 J17 J31 J32
ITS
Saccharomycopsis crataegensis
Candida quercitrusa
Pichia fermentans
Hanseniaspora uvarum
Hanseniaspora opuntiae
Botryosphaeria
Ceratocystis paradoxa
Candida butyri
Candida jaroonii
Fusarium
Candida tallmaniae
Candida
Lasiodiplodia theobromae
Saccharomycetes
Saccharomyces cerevisiae
Hanseniaspora
Torulaspora delbrueckii
Candida xylopsoci
Kluyveromyces marxianus
Pichia pijperi
J38 J37 J35 J36 J21 J26 J31 J39 J32 J40 J16 J17 J19 J20 J14 J15 J28 J29 J23 J24 J22 J27

Figure 4.


Relative abundance %
Hanseniaspora
Hanseniaspora uvarum
Pichia pijperi
Saccaromyces cerevisiae
26S
ITS