DARWINIANA
NUEVA SERIE
www.ojs.darwin.edu.ar
ISSN 0011-6793 impresa - ISSN 1850-1699 en línea

DARWINIANA, nueva serie 11(2): 402-419. 2023
Versión de registro, efectivamente publicada el 23 de agosto de 2023
DOI: 10.14522/darwiniana.2023.112.1127

ASSESSING FUNGAL ENDOPHYTE DIVERSITY: A COMPARATIVE STUDY OF THREE AUTOMATED METABARCODING PIPELINES

Lucía Molina[1,2,*], Mario Rajchenberg[1,2], M. Catherine Aime[3] & M. Belén Pildain[1,2,4]

[1] *Área de Fitopatología y Microbiología Aplicada, Centro de Investigación y Extensión Forestal Andino Patagónico (CIEFAP), Ruta 259 Km 16.24, 9200 Esquel, Chubut, Argentina; * lmolina@ciefap.org.ar* (author for correspondence).
[2] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.*
[3] *Department of Botany and Plant Pathology, Purdue University, 915 W State St, West Lafayette, IN 47907, USA.*
[4] *Facultad de Ciencias Naturales y Ciencias de la Salud, Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), Ruta 259 Km 16.4, 9200 Esquel, Chubut, Argentina.*

Abstract. Molina, L.; M. Rajchenberg, M. C. Aime & M. B. Pildain. 2023. Assessing fungal endophyte diversity: a comparative study of three automated metabarcoding pipelines. *Darwiniana*, nueva serie 11(2): 402-419.

High-throughput sequencing approaches have become frequent in the study of endophyte communities allowing the cumulative description of fungal diversity in the last decade. However, they brought new challenges to researchers in terms of programming and developing of informatics tools. Currently, there is no consensus concerning the appropriate bioinformatics to process such sequence data. The aim of this study was to compare the performance of three pipelines of two cost-free toolkits designed to be friendly to non-programmer users, and specifically developed for fungal data: AMPtk and PIPITS. The sapwood-inhabiting fungal assemblages of two *Nothofagus* species from the Patagonian Forests were assessed through metabarcoding of the internal transcribed spacer (ITS) and compared with an extant sequence dataset obtained from culture prospection in the same study sites and trees. The AMPtk toolkit has performed better concerning community description in terms of precision of taxa clustering, mainly due to the DADA2 algorithm; PIPITS evidenced a higher sensitivity in detecting taxa known to be present, hence it is potentially useful for future specific taxa detection surveys. Because of a current lack of information of the reference databases, both bioinformatic toolkits performed poorly as to taxonomy assignment. It is imperative to continue studying these ecosystems to, concomitantly, improve databases and the explanatory potential of the new technologies.

Keywords. AMPtk; environmental DNA; fungal endophytes; *Nothofagus* forests; PIPITS.

Resumen. Molina, L.; M. Rajchenberg, M. C. Aime & M. B. Pildain. 2023. Evaluación de la diversidad de endófitos fúngicos: un estudio comparativo de tres flujos de trabajo automatizados de metabarcoding. *Darwiniana*, nueva serie 11(2): 402-419.

La utilización de la secuenciación de alto rendimiento se ha vuelto frecuente en el estudio de comunidades endófitas. Estas tecnologías han permitido la descripción acumulativa de diversidad fúngica a lo largo de la última década. No obstante, también han implicado nuevos desafíos para los investigadores de las áreas involucradas en términos de la necesidad de contar con herramientas de programación y habilidades desarrolladoras. Hoy en día no existe un consenso sobre las herramientas bioinformáticas más adecuadas para procesar los datos crudos de secuencias que estas tecnologías arrojan. El objetivo de este trabajo fue comparar el rendimiento de tres flujos de trabajo realizados en dos plataformas gratuitas diseñadas para ser amigables con usuarios que no son programadores y desarrolladas específicamente para estudios de hongos: AMPtk y PIPITS. Evaluamos los ensambles de hongos que habitan en la albura de dos especies de *Nothofagus* de los bosques patagónicos y comparamos el conjunto de datos de *metabarcoding* del espaciador transcrito interno (ITS) con un conjunto de datos de secuencias existente, obtenido de la prospección de cultivos de los mismos árboles y sitios de estudio.

Original recibido el 3 de marzo de 2023, aceptado el 9 de julio de 2023
Editor Asociado: Mario Carlos Saparrat

La plataforma AMPtk se desempeñó mejor con respecto a la descripción de la comunidad, en términos de precisión del agrupamiento de taxones, principalmente debido al algoritmo DADA2. El flujo de trabajo PIPITS evidenció una mayor sensibilidad en la detección de taxones conocidos presentes, por lo que es potencialmente útil para futuros estudios que persigan la detección de taxones específicos. Debido a la falta de información que exhiben las bases de datos de referencia sobre el ecosistema en estudio, ambas plataformas tuvieron un desempeño deficiente en cuanto a la asignación de taxonomías. Es imperativo seguir estudiando estos ecosistemas y mejorar las bases de datos para aumentar el potencial explicativo de las nuevas tecnologías.

Palabras clave. ADN ambiental; AMPtk; bosques de *Nothofagus*; endófitos fúngicos; PIPITS.

INTRODUCTION

Temperate forests have been demonstrated to be reservoirs of an outsize fungal endophyte diversity living in standing trees (Unterseher, 2011). It is known that this mycobiota plays a key role in the fitness and functioning of the trees through complex dynamics (Baldrian, 2016), and whose roles fell along a continuum from mutualism, commensalism, and parasitism that can elapse even through the same fungal organism lifetime (Saikkonen et al., 1998, Stone et al., 2004). Plant-associated mycobiota also contribute to large-scale patterns of plant diversity in forest ecosystems (Wang et al., 2019). However, there is a huge gap in the understanding of fungal endophyte diversity, the drivers that modulate such communities, and the nature of the interactions they establish with plants (Suryanarayanan, 2020).

In the last decades, novel high-throughput sequencing technologies (HTS) become frequent and accessible, leading to progress in plant mycobiomes investigations, especially through metabarcoding approaches. They are cost and time-efficient and have allowed increasing sensitivity and rate at which biomes can be assessed (Terhonen et al., 2019). Nonetheless, they brought great challenges in data processing, in terms of sequence quality filtering and curation, assemblage, clustering, and taxonomic assignment of such a huge volume of output sequences. This led to the development of the many different bioinformatic tools assessing each of the steps of the workflow, whose performances are still under assessment. HTS metabarcoding approaches also required from researchers certain bioinformatics and programming skills. In this sense, numerous developments came to light, aiming to provide accessible and integrated tools for the whole data processing (Gweon et al., 2015; Rognes et al., 2016; Palmer et al. 2018; Jalili et al., 2020). There have been various efforts to compare the performance of different individual tools aimed at different steps of the workflow (Schloss & Westcott, 2011; Edgar & Flyvbjerg, 2015), but there are limited efforts to assess integrated pipelines useful to microbiologists and mycologists that are not programmers or developers (Mysara et al., 2017). Many of these tools have been developed for bacterial 16S amplicon analysis and have been subsequently adapted to fungal data (Anslan et al., 2018). Among the available tools, PIPITS (Gweon et al., 2015) and the Amplicon toolkit (AMPtk) (Palmer et al., 2018) have been created, specifically, to process fungal ITS-sequencing data and have demonstrated to perform better for fungal ITS amplicon analysis among other available pipelines (Anslan et al., 2018, Nilsson et al., 2019). Both PIPITS and AMPtk are command-line cost-free toolkits, designed with easy-to-use straightforward pipelines that allow performing all data analysis from raw sequences to final operational taxonomic unit (OTU) tables and taxonomy. They take advantage of extant methods from other toolkits and, also, develop new tools for certain steps of the workflow.

Recently, we first described the endophytic mycobiota of the Andean Patagonian Forest through culture prospection, in a study that assessed wood endophyte communities of *Nothofagus* trees (Molina et al., 2020). Even though we described a rich and heterogeneous diversity, we concluded that our results underestimated such diversity, and that fungal endophyte diversity harboured by *Nothofagus* trees was unable to be fully assessed through culture prospection; to further describe fungal endophyte diversity and elucidate beta diversity patterns, a culture-independent approach was proposed (Molina et al., 2020, 2022).

In this study, we aim to compare the performances of three pipelines from two automated toolkits (AMPtk and PIPITS) in the assessment of wood fungal endophytes assemblages of *Nothofagus* trees from the Patagonian Forests; this is accomplished by comparing the ITS metabarcoding dataset with another sequence dataset derived from culture prospection isolates, from the same sites and trees.

MATERIALS AND METHODS

Study area and sampling procedure

The study was conducted in Los Alerces National Park in Argentinian Patagonia, from May 2016 to April 2018. The sampling collection was described in Molina et al. (2020). Briefly, at each site, roots and stems were sampled seasonally, from ten trees of similar diameter at breast height. The sampling was performed in seven sites: three stands of *Nothofagus pumilio* (Poepp. & Endl.) Krasser and four stands of *Nothofagus dombeyi* (Poepp. & Endl.) Krasser. Sapwood cores of 5 mm diameter and 15 mm length were extracted by using an increment borer sterilized with 70% ethanol (v/v) and flaming between samples. A total of 280 trees were sampled, taking different samples for the culture-dependent and culture-independent approaches, but from the same trees.

Culture-dependent database construction

Sampling collection, fungal isolation, and molecular identification methods were reported in Molina et al. (2020). Briefly, the sapwood tissue of the core samples were cut into 5 mm pieces, surface sterilized, put into Ascomycota and Basidiomycota selective media and incubated at 20-24 °C for up to 4 months. Pure axenic cultures were used for DNA extraction, and ITS sequencing.

Culture-independent database construction

The sampling processing, library preparation and amplicon sequencing were described in Molina & Pildain (2022). Briefly, sapwood samples were recovered using a sterilized increment borer, about 50 mg of wood was ground to powder, for each wood sample, according to Dumolin et al. (1995). Total DNA was extracted using DNeasy Power Plant Pro Kit (QIAGEN, Hilden, Germany). Internal Transcribed Spacer 1 (ITS1) library was prepared using the TrueSeq dual indexing strategy. ITS1 amplification was performed by using the primers pair TS-ITS1-F and TS-ITS2-R (White et al., 1990; Gardes & Bruns, 1993) and MyTaqTM Mix (Bioline, USA, Inc., Memphis) in a total volume of 25 μL per reaction, with the following cycling conditions: 94 °C for 5 minutes, 32 cycles of 94 °C for 45 seconds, 50 °C for 45 seconds, and 72 °C for 1 minute, and a final extension at 72 °C for 7 minutes. PCR products were purified using ExoSap-IT (USB Corporation, Cleveland, OH). The purified PCR products were indexed by using sample-specific barcodes combinations of the TruSeq primers pairs i5-TS-DI-5xx and i7-TS-DI-7xx with the following cycling conditions: 95 °C for 3 minutes, 8 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 30 seconds, and a final extension at 72 °C for 5 minutes. PCR products were purified as already mentioned, and quantified by using a NanoDrop spectrophotometer (ThermoFisher, Waltham, MA). Negative controls (from both the DNA extraction and PCR runs) and non-biological synthetic mock communities (*SynMock*; Palmer et al., 2018) were simultaneously processed, and sequenced. Synthetic mock communities are non-biological constructs specifically designed to mimic the composition and complexity of real-world fungal communities, serving as essential references for validating and benchmarking experimental procedures in mycological research. All samples were randomly separated in two groups and each one received pooled purified ITS amplicons in equimolar ratios (multiplexed). Libraries were sequenced at the Purdue Genomics Core Facilities (Purdue University, West Lafayette, IN) with a MiSeq version 2 Reagent kit of 500 cycles in the Illumina MiSeq platform (2 x 250 bp).

Bioinformatic analysis

Data processing was carried out through the different toolkits: PIPITS (v2.7.12) and the Amplicon toolkit (AMPtk) (v1.2.4). Also, AMPtk was performed by using two different clustering methods. The detailed workflow of the three different pipelines (hereafter PIPITS, AMPtk-UPARSE, and AMPtk-DADA2) is illustrated in Fig. 1. Pre-clustering steps were conducted under default conditions.

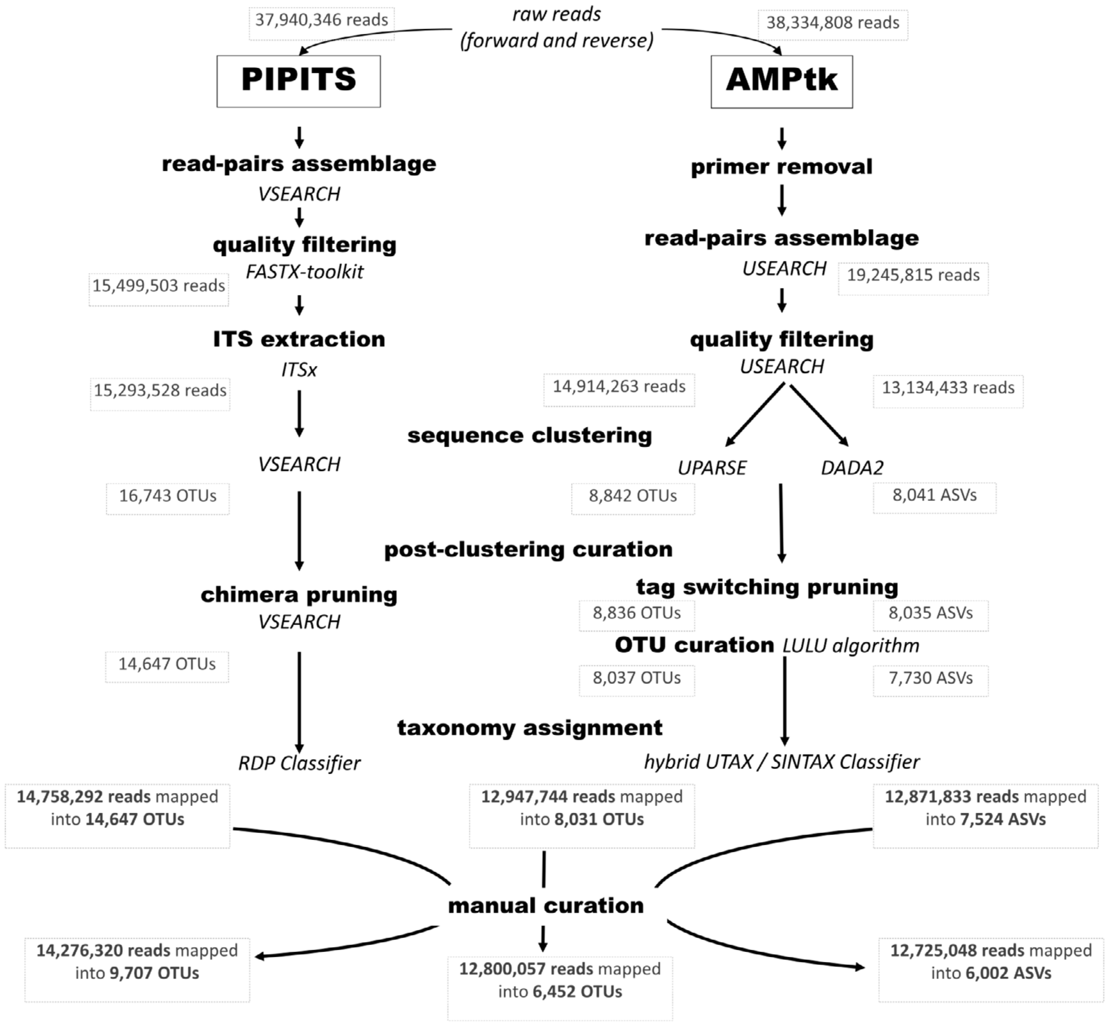

Fig. 1. Workflow of the three pipelines performed for the data processing of the ITS-sequences from MiSeq Illumina (2 x 250) platform: PIPITS and AMPtk toolkit using UPARSE and DADA2 clustering tools. Sequence and OUT/ASV input and output are informed for each step.

Both toolkits differ in the order and the tools used for quality filtering, read-pair assemblage, and sequence trimming to get ITS amplicon. AMPtk uses USEARCH tools (v9.2.64; Edgar, 2010) while PIPITS takes advantage of the open-source alternative: VSEARCH (v2.7.0; Rognes et al., 2016).

The Figure 1 shows that the first step in the PIPITS toolkit is the read-pairs assembling by using VSEARCH; then, it filters the assemblage reads by quality using FASTX-toolkit (v0.0.13; Gordon & Hannon, 2010). Sequences are dereplicated by using VSEARCH, this eliminates the redundant sequences from the large dataset, streamlining downstream data processing and enhancing overall computational efficiency. Next, the pipeline calls the ITS extractor algorithm (ITSx, v 1.1b1; Bengtsson-Palme et al., 2013) to identify the ITS1 region and to extract it from the reads, deleting any conserved region or primer sequences. After data processing, the pipeline re-inflates the replicated sequences in order to keep the reads abundances information. Conversely, the AMPtk toolkit first trims the short reads, next trims the primer sequences from the reads; then, merges the paired-end reads by using USEARCH and performs the quality filtering (Edgar & Flyvbjerg, 2015).

For OTU clustering, the threshold was set at 97% identity. PIPITS uses VSEARCH to cluster OTUs, and for chimera detection and deletion by confronting the UNITE UCHIME reference data set (http://unite.ut.ee/repository.php). In the AMPtk toolkit, the algorithms used for the taxonomic units definition were: OTU clustering with UPARSE (v9.2.6.4; Edgar, 2013) and the DADA2 pipeline (v1.6.0; Callahan et al., 2016) which also performs chimeras' detection and removal. The DADA2 pipeline does not cluster OTUs but defines amplicon sequence variants (ASVs). Conversely to OTU, ASV represents unique sequences without clustering (Callahan et al., 2017). The AMPtk toolkit provides an additional functionality to address cross-contamination errors by leveraging the sequences of the SynMock community. It accomplishes this by identifying the SynMock sequences, estimating their frequencies, and calculating the tag-switching index. Also, allows running the LULU algorithm (v0.1.0; Frøslev et al., 2017), which is a post-clustering curation pipeline that combines co-occurrence patterns and sequence identity analysis to detect and delete (or merge) erroneous OTUs from the set.

Finally, to assign taxonomic classifications to the defined OTUs, PIPITS employs the RDP Classifier (v2.10.2; Wang et al., 2007) which is a machine learning approach. This approach utilizes computational algorithms to automatically analyze and classify the obtained sequences based on patterns and characteristics present in the data. The RDP Classifier compares the obtained sequences against the carefully curated reference dataset of fungal ITS regions UNITE (https://sourceforge.net/projects/rdp-classifier/files/RDP_Classifier_TrainingData). Conversely, a "hybrid" approach was used to assign the taxonomy with the AMPtk toolkit. This approach combines classification from a global alignment, with classification from the UTAX (RC Edgar, http://drive5.com/usearch/manual9.2/cmd_utax.html) and SINTAX (Edgar, 2016) approaches. This hybrid method chooses the best taxonomy from the three approaches, by prioritizing the global alignment result, if the threshold is higher than 97%, or selecting the higher confidence score from the other approaches. If there is a conflict between the taxonomies, the algorithm chooses the last common ancestor taxonomy (Palmer et al., 2018), meaning the last taxonomic rank in which there would be no conflict.

Manual curation of the three pipeline outputs was performed by following Brown et al. (2015) recommendations, thus three additional databases were evaluated (hereafter, curated data). Non-fungal and kingdom undefined OTUs/ASVs, as well as OTUs/ASVs represented by less than 10 reads, were removed from the set.

Database comparisons

Sequences obtained from fungal strains isolated from cultures, were used as reference to assess the performance of the HTS methodologies on finding and describing accurately the fungal taxa present in the sapwood of *Nothofagus* species under study. To achieve this, the ITS sequences obtained from cultures by Molina et al. (2020) were compared, by using the BLAST algorithm (Altschul et al., 1997), to the datasets obtained from the PIPITS, the AMPtk-UPARSE, and the AMPtk-DADA2 pipelines. The comparisons were performed in Geneious Prime (v2020.1.1; Biomatters Ltd, https://www.geneious.com/).

The databases used were the output FASTA files from each pipeline, so that all the OTUs with a similarity above 97% were reported. The culture-based dataset that was blasted to them, consisted of 72 sequences of the full ITS regions of rDNA (ITS1, the intervening 5.8 RNA gene, and ITS2) obtained through Sanger sequencing from Basidiomycota, Ascomycota and Mucoromycota phyla isolated from the same study sites following the same sampling methods. These taxa are known to be present in the studied system as a result of the previous culture prospection study (Molina et al., 2020). The sensitivity of each pipeline to detect the taxa registered was evaluated, as well as the precision in reaching the taxonomy assignment. Five parameters were defined to assess pipeline performances : a) the percentage of cultured taxa that matched to at least one OTU/ASV defined bioinformatically from HTS (as an estimator of sensitivity), b) the percentage of cultured taxa that were detected as the same OTU/ASV by pipeline algorithms (i.e., different cultured taxa that were merged by the culture-independent approach), c) the percentage of cultured taxa that were assigned a wrong taxonomy by a pipeline, d) average number of OTUs/ASVs that matched to a cultured taxon and e) maximum number of OTUs/ASVs that matched with a single cultured taxon in the HTS pipeline (both parameters assessing the redundancy in OTUs/ASVs clustering and post-clustering curation).

Statistical analyses

Differences in alpha diversity between the three pipelines were assessed by using the Friedman test and the Bonferroni test. Statistical analyses and graphics were performed in R (R Core Team, 2022) with packages biomformat (v1.8.0; McMurdie & Paulson, 2016), phyloseq (v1.24.2; McMurdie & Holmes, 2013), ggplot2 (v3.3.3; Wickham, 2016), vegan (v2.5.7; Oksanen et al., 2020).

RESULTS

Bioinformatic pipelines comparisons

The sequencing experiment yielded a mean depth of 133 855 reads per sample (paired-end raw reads) and a total depth of 38 million raw reads.

With identical computing power, the AMPtk toolkit was much more time-efficient than PIPITS. The AMPtk-UPARSE pipeline took 79 minutes of total run time, and AMPtk-DADA2 took 205 minutes. The difference was originated by the clustering algorithm of either pipeline. The PIPITS pipeline used 11 418 minutes mainly used to complete the "pipits-funits" step: reads dereplicate, ITS extraction, and back to reads replicate. Also, the clustering method from PIPITS (119 minutes) was also more time-consuming than the AMPtk-UPARSE method (38 minutes).

The OTU/ASV richness strongly differed between pipelines; PIPITS generated almost twice as many OTUs (14 647 OTUs) as AMPtk pipelines [UPARSE (8 031 OTUs), and DADA2 (7 524 ASVs)] and the differentiation occurred mainly at the clustering step (Fig. 1). Also, differences in OTU/ASV richness were observed between toolkits at the sample level (Fig. 2, above). Furthermore, both AMPtk pipelines evidenced significant variations in their richness per sample (Wilcoxon test, $p<0.001$) evidencing that the clustering algorithm performed affects richness results. However, the rarefaction curves approximated asymptotes for the three datasets when manually curated data was considered (Fig. 2, below). Deleting OTUs/ASVs with low reads abundance and bad taxonomic resolution improved the representation of the fungal community for the three pipelines tested. Manual curation reduced the differences in OTUs/ASVs richness between pipelines, although those were still significant (Fig. 1; Fig. 2, above). This is because 34% of PIPITS final OTUs lacked taxonomic assignment at the Kingdom level (against 20% from the AMPtk pipelines), which might indicate a redundant OTU clustering and a lower performance of post-clustering curation methods in this pipeline. After these taxa were removed from the set, the taxonomic resolution did not differ significantly between pipelines, although PIPITS showed a slightly higher proportion of OTUs assigned to Class or lower taxonomic levels (48% against 37% and 36% in AMPtk-UPARSE and AMPtk-DADA2, respectively).

Comparison with culture prospection dataset

The sequence matches between cultured and uncultured datasets, and their estimated identity percentage, are listed in Table 1. There were certain cultured taxa that the pipelines did not detect in the HTS experiment. In that sense, PIPITS was the most sensitive pipeline, with 19% of the cultured taxa undetected, followed by AMPtk-DADA2 (24%) and USEARCH (28%) (Table 2). Cultured taxa were molecularly identified at genus or species ranks in Molina et al. (2020), whereas the same uncultured OTUs/ASVs were identified at higher taxonomic ranks: 40% and 47% of the matched OTUs/ASVs were assigned at Family level or higher in AMPtk and PIPITS pipelines, respectively. Despite low taxonomic resolution, the AMPtk pipelines did not mistake the taxonomic identifications (Table 2) whereas PIPITS exhibited 1.39% of misassignment.

Clustering redundancy was low for AMPtk pipelines, only 5.6% of the cultured taxa matched with multiple OTUs/ASVs (maximum 3) resulting in a mean value of 1.10 OTUs and 1.12 ASVs per taxon for UPARSE and DADA2, respectively (Table 2). In contrast, PIPITS evidenced high redundancy in OTU clustering: 33.33% of the cultured taxa matched with multiple OTUs, the maximum number of matched OTUs for the same cultured taxon was 35, giving an average number of hits of 1.85 OTUs per taxon.

Table 1. BLASTn results of the comparison between Sanger sequences set of Molina et al. (2020), against the OUT/ASV sequence data set of each pipeline.

Sanger	**AMPtk - UPARSE**			**AMPtk - DADA2**			**PIPITS**		
Molina et al. (2020)		%identity	UNITE		%identity	UNITE		%identity	UNITE
Aleurodiscus patagonicus	**OTU2205**	98.639	*Stereaceae*	**OTU2257**	98.639	*Stereaceae*	-		
Ambrosiozyma sp.	**OTU4296**	100.000	*Ascomycota*	**OTU4807**	99.558	*Ascomycota*	**OTU591**	100	*Saccharomycetales*
Anthostomella sp.	**OTU5800**	99.111	*Xylariaceae*	**OTU5133**	99.111	*Xylariaceae*	**OTU4781**	98.895	Fungi
Anthracobia muelleri	-			-			-		
Anthracobia sp.	-			-			**OTU3073**	100	*Tricharina praecox*
Arambarria destruens	-			-			-		
Armillaria umbrinobrunnea	**OTU2654**	98.485	*Armillaria*	**OTU2335**	98.485	*Armillaria*	**OTU1098**	98.039	*Armillaria*
	OTU4824	98.413	*Armillaria*	**OTU3483**	100	*Armillaria*	**OTU1181**	98.039	*Armillaria*
Arthrinium sacchari	**OTU364**	100.000	*Arthrinium sacchari*	**OTU282**	100.000	*Arthrinium sacchari*	**OTU4874**	100	*Arthrinium sacchari*
Arthrinium sp.	**OTU76**	99.160	Fungi	**OTU75**	99.160	Fungi	**OTU4800**	98.343	Fungi
							OTU4943	98.333	Fungi
							OTU4979	98.333	Fungi
							OTU4985	98.87	Fungi
							OTU5006	98.333	Fungi
							OTU5010	98.333	Fungi
							OTU5034	98.333	Fungi
							OTU5056	98.324	Fungi
							OTU5120	97.778	Fungi
Ascocoryne cylichnium	-			-			**OTU2201**	100	*Ascocoryne cylichnium*
Ascocoryne sp.	**OTU12**	99.608	*Ascocoryne*	**OTU10**	99.608	*Ascocoryne*	**OTU1938**	99.519	*Ascocoryne cylichnium*
Ascocoryne sarcoides							**OTU2141**	99.476	*Ascocoryne cylichnium*
Atheliaceae	**OTU1621**	100.000	Fungi	**OTU1923**	100.000	Fungi	**OTU7975**	100	Agaricomycetes
Aurantiporus albidus	**OTU110**	100.000	*Aurantiporus albidus*	-			**OTU2325**	98.125	Agaricomycetes
							OTU2363	99.367	Agaricomycetes
							OTU2425	99.367	Agaricomycetes
							OTU2426	98	Agaricomycetes
							OTU2431	98.734	Agaricomycetes
							OTU2444	99.355	Agaricomycetes
							OTU2445	98.101	Agaricomycetes
							OTU2446	98	Agaricomycetes
							OTU2447	98.101	Agaricomycetes
							OTU2448	98.101	Agaricomycetes
							OTU2449	98.101	Agaricomycetes
							OTU2450	98	Agaricomycetes
							OTU2452	98.101	Agaricomycetes
Beauveria sp.	**OTU862**	100.000	*Beauveria bassiana*	**OTU1010**	100.000	*Beauveria bassiana*	**OTU9070**	100	*Beauveria pseudobassiana*
Cadophora sp.	**OTU403**	100.000	Helotiales	**OTU426**	100.000	Helotiales	**OTU8490**	100	Helotiales
							OTU8521	98.78	Helotiales

Table 1. (Continuation). BLASTn results of the comparison between Sanger sequences set of Molina et al. (2020), against the OUT/ASV sequence data set of each pipeline.

Sanger	**AMPtk - UPARSE**			**AMPtk - DADA2**			**PIPITS**		
Molina et al. (2020)		%identity	UNITE		%identity	UNITE		%identity	UNITE
							OTU8306	98.182	*Cadophora*
							OTU8605	98.16	Helotiales
							OTU8753	98.16	Helotiales
							OTU8742	98.137	Helotiales
							OTU8608	97.546	*Cadophora*
							OTU8741	97.546	Helotiales
Capronia sp.	**OTU505**	100.000	*Capronia kleinmondensis*	**OTU472**	100.000	*Capronia kleinmondensis*	**OTU2689**	99.49	Herpotrichiellaceae
							OTU2860	97.436	Herpotrichiellaceae
Cladosporium sp.	**OTU19**	99.502	*Cladosporium*	**OTU23**	99.502	*Cladosporium*	**OTU8584**	99.359	Fungi
							OTU9435	98.052	Fungi
							OTU10332	97.436	*Cladosporium ramotenellum*
							OTU10565	97.419	Fungi
Coniochaeta sp1	-			**OTU721**	99.465	*Coniochaeta*	**OTU6998**	98.817	*Coniochaeta*
							OTU6696	98.256	*Coniochaeta*
							OTU7198	98.246	*Coniochaeta*
							OTU6982	97.647	*Coniochaeta*
							OTU6981	97.076	*Coniochaeta*
Coniochaeta sp2	**OTU700**	99.465	*Coniochaeta*	**OTU1847**	97.382	*Coniochaeta ligniaria*	**OTU6285**	97.701	*Coniochaeta ligniaria*
							OTU6262	97.11	*Coniochaeta*
							OTU6455	97.093	Coniochaetales
Coprinellus sp.	**OTU4015**	100.000	*Coprinellus*	**OTU3770**	100.000	*Coprinellus*	**OTU1276**	100	*Coprinellus*
Ophiostoma novae-zelandiae	-								
Cordycipitaceae	-			-			-		
Cosmospora sp.	-			**OTU1299**	100.000	Nectriaceae	**OTU10892**	100	Hypocreales
Curreya sp.	**OTU104**	100.000	Pleosporales	**OTU64**	100.000	Pleosporales	**OTU6708**	99.419	Teichosporaceae
							OTU6857	98.361	Teichosporaceae
							OTU6950	98.4	Teichosporaceae
							OTU7002	100	Pleosporales
							OTU7027	98.246	Teichosporaceae
							OTU7059	98.235	Pleosporales
							OTU7073	98.235	Teichosporaceae
							OTU7087	98.235	Teichosporaceae
							OTU7094	99.194	Teichosporaceae
							OTU7109	98.387	Pleosporales
							OTU7134	100	Teichosporaceae
							OTU7135	98.387	Pleosporales

Table 1. (Continuation). BLASTn results of the comparison between Sanger sequences set of Molina et al. (2020), against the OUT/ASV sequence data set of each pipeline.

Sanger		AMPtk - UPARSE			AMPtk - DADA2			PIPITS	
Molina et al. (2020)		%identity	UNITE		%identity	UNITE		%identity	UNITE
							OTU7137	98.235	Pleosporales
							OTU7149	98.387	Teichosporaceae
							OTU7155	98.387	Pleosporales
							OTU7157	99.194	Pleosporales
							OTU7181	97.619	Teichosporaceae
							OTU7217	98.235	Pleosporales
							OTU7218	98.235	Pleosporales
							OTU7220	98.235	Teichosporaceae
							OTU7221	98.387	Pleosporales
							OTU7222	98.235	Teichosporaceae
							OTU7223	98.235	Pleosporales
							OTU7224	98.333	Teichosporaceae
							OTU7225	98.374	Teichosporaceae
							OTU7226	98.387	Teichosporaceae
							OTU7228	98.817	Teichosporaceae
							OTU7230	98.235	Pleosporales
							OTU7234	98.387	Pleosporales
							OTU7235	98.347	Teichosporaceae
							OTU7236	98.81	Teichosporaceae
							OTU7432	97.581	Teichosporaceae
							OTU7681	97.633	Teichosporaceae
							OTU8067	100	Teichosporaceae
Cytospora sp.	**OTU2105**	100.000	*Valsa cypri*	**OTU2153**	99.597	*Valsa cypri*	**OTU2462**	100	*Valsa cypri*
Fistulina antarctica	**OTU91**	99.571	*Fistulina*	**OTU2962**	100.000	*Fistulina*	**OTU839**	100	*Fistulina*
				OTU92	98.589	*Fistulina*	**OTU861**	98.326	*Fistulina*
				OTU7105	98.958	Fungi	**OTU888**	97.881	*Fistulina*
							OTU889	97.458	*Fistulina*
							OTU893	97.458	*Fistulina*
Ganoderma australe	-			-			-		
Gyromitra sp.	**OTU3049**	99.425	*Gyromitra esculenta*	**OTU3733**	99.425	*Gyromitra esculenta*	-	-	
Helotiales	**OTU966**	100.000	Helotiales	**OTU1001**	100.000	Helotiales	**OTU7438**	100	Helotiales
							OTU6643	100	Helotiales
Hyaloscypha sp1	-			-			**OTU7848**	97.024	Helotiales
Hyaloscypha sp2	**OTU1534**	99.537	*Hyaloscypha*	**OTU1524**	99.537	*Hyaloscypha*	**OTU10358**	99.351	*Hyaloscypha*
Hypholoma frowardii	**OTU338**	99.609	*Hypholoma australe*	**OTU263**	99.609	*Hypholoma australe*	**OTU1192**	99.163	*Hypholoma*
							OTU1243	97.479	*Hypholoma*
							OTU1296	97.468	*Hypholoma*
							OTU1301	97.468	*Hypholoma*

Table 1. (Continuation). BLASTn results of the comparison between Sanger sequences set of Molina et al. (2020), against the OUT/ASV sequence data set of each pipeline.

Sanger	AMPtk - UPARSE			AMPtk - DADA2			PIPITS		
Molina et al. (2020)		%identity	UNITE		%identity	UNITE		%identity	UNITE
							OTU1313	97.468	*Hypholoma*
							OTU1314	97.468	*Hypholoma*
							OTU1315	97.468	*Hypholoma*
							OTU1393	97.468	*Hypholoma*
							OTU1378	97.436	*Hypholoma*
							OTU1644	99.539	*Hypholoma*
Lachnum sp	**OTU1985**	98.726	*Lachnum*	**OTU1917**	98.734	*Lachnum*	**OTU8630**	100	Helotiales
							OTU8696	100	Helotiales
Laetiporus portentosus	**OTU1680**	100.000	Fomitopsidaceae	**OTU1593**	100.000	Fomitopsidaceae	**OTU4147**	99.459	*Laetiporus*
Leptodontidium sp.	**OTU1344**	97.807	*Leptodontidium*	**OTU1331**	97.807	*Leptodontidium*	**OTU7446**	99.405	Helotiales
	OTU1651	97.807	*Leptodontidium*	**OTU1737**	97.368	*Leptodontidium*	**OTU7496**	97.633	Helotiales
	OTU2899	97.368	*Leptodontidium trabinellum*	**OTU3097**	97.368	*Leptodontidium trabinellum*			
Metapochonia sp.	**OTU768**	100.000	*Metapochonia*	**OTU835**	100.000	*Metapochonia*	**OTU6814**	98.837	Fungi
							OTU6542	97.688	Fungi
Meyerozyma caribbica	**OTU282**	100.000	*Debaryomycetaceae*	**OTU246**	100.000	*Debaryomycetaceae*	**OTU7478**	100	*Meyerozyma*
Meyerozyma guilliermondii							**OTU7342**	99.342	*Meyerozyma*
Microcera sp.	**OTU83**	100.000	*Microcera*	**OTU88**	100.000	*Microcera*	**OTU10146**	99.363	*Microcera*
							OTU10402	100	Nectriaceae
							OTU10533	98.052	*Microcera*
Nemania sp.	-			-			-		
Obba valdiviana	-			-			-		
Oidiodendron sp1	**OTU1373**	99.548	*Oidiodendron*	**OTU1465**	99.548	*Oidiodendron*	**OTU7439**	99.405	*Oidiodendron*
Ophiostoma nothofagi	**OTU1135**	99.621	Sordariomycetes	**OTU1225**	99.621	Sordariomycetes	**OTU3658**	99.465	Sordariomycetes
	OTU1695	97.753	Sordariomycetes				**OTU3221**	97.895	Sordariomycetes
Ophiostoma valdivianum	-			-			-		
Paecilomyces / Isaria	**OTU924**	99.167	*Isaria*	**OTU1003**	99.167	*Isaria*	**OTU4876**	100	Hypocreales
							OTU2663	100	Hypocreales
Paraphoma sp.	-			-			**OTU1999**	99.528	Pleosporales
Pezicula sp.	**OTU21**	100.000	Leotiomycetes	**OTU27**	100.000	Leotiomycetes	**OTU6720**	100	Dermateaceae
							OTU7515	100	Dermateaceae
							OTU6622	99.371	Dermateaceae
							OTU7121	99.363	Dermateaceae
Phacidium sp.	**OTU54**	100.000	Phaciciadeae	**OTU55**	100.000	Phacidiaceae	**OTU7096**	100	Fungi
Phanerochaete velutina / sordida	**OTU1086**	100.000	*Phanerochaete velutina*	**OTU1100**	100.000	*Phanerochaete velutina*	**OTU2329**	99.505	*Phanerochaete*
							OTU2366	97.015	*Phanerochaete velutina*
							OTU2389	97	*Phanerochaete velutina*

Table 1. (Continuation). BLASTn results of the comparison between Sanger sequences set of Molina et al. (2020), against the OUT/ASV sequence data set of each pipeline.

Sanger	AMPtk - UPARSE			AMPtk - DADA2			PIPITS		
Molina et al. (2020)		%identity	UNITE		%identity	UNITE		%identity	UNITE
Phialocephala sp1	-			-			-		
Phialocephala sp2	**OTU4346**	100.000	*Helotiales*	**OTU3925**	100.000	*Helotiales*	**OTU10524**	100	Helotiales
	OTU8649	98.013	*Helotiales*				**OTU10734**	100	Helotiales
							OTU10695	97.521	Helotiales
Phlebia sp.	**OTU2762**	100.000	*Phlebia*	**OTU3016**	100.000	*Phlebia*	**OTU2543**	100	*Phlebia*
Pholiota baeosperma	**OTU2626**	100.000	*Pholiota*	**OTU2521**	99.606	*Pholiota*	**OTU1121**	100	Agaricales
Pholiota brunnescens	-			-			-		
Pholiota multicingulata	**OTU2521**	100.000	*Pholiota*	**OTU2405**	100.000	*Pholiota*	**OTU1434**	100	*Pholiota*
Pleurostoma sp.	**OTU4819**	99.539	Ascomycota	**OTU484**	97.235	Sordariomycetes	**OTU6249**	98.851	Ascomycota
Postia dissecta	**OTU1288**	100.000	*Postia dissecta*	**OTU1344**	100.000	*Postia dissecta*	**OTU3052**	100	*Postia dissecta*
Postia pelliculosa	**OTU1645**	100.000	*Postia pelliculosa*	**OTU1730**	100.000	*Postia pelliculosa*	**OTU3118**	98.958	*Postia pelliculosa*
	OTU850	99.119	*Postia*	**OTU796**	99.119	*Postia*	**OTU3483**	98.413	*Postia*
Pseudoinonotus crustosus	-			-			-		
Pseudovalsaria sp.	**OTU2323**	100.000	*Pseudovalsaria*	**OTU2253**	100.000	*Pseudovalsaria*	**OTU6930**	99.415	Boliniaceae
							OTU7177	97.015	Fungi
Rasamsonia sp.	-			-			-		
Sarocladium sp.	**OTU2330**	97.585	*Sarocladium*	**OTU2702**	97.858	*Sarocladium*	**OTU8624**	99.387	Sordariomycetes
Sistotrema brinkmannii	**OTU7907**	99.038	*Sistotrema brinkmannii*	**OTU7408**	99.038	*Sistotrema brinkmannii*	**OTU6717**	100	Cantharellales
Sporothrix cabralii	**OTU1261**	98.605	*Sporothrix*	**OTU1326**	98.605	*Sporothrix*	**OTU3542**	97.386	Ophiostomataceae
	OTU6747	97.120	*Sporothrix variecibatus*						
Stereum sp.	**OTU963**	99.545	*Stereum*	**OTU940**	99.099	*Stereum hirsutum*	**OTU3591**	98.936	*Stereum hirsutum*
							OTU3463	97.884	*Stereum hirsutum*
							OTU3810	97.861	*Stereum*
							OTU3842	97.861	*Stereum hirsutum*
							OTU3872	97.861	*Stereum hirsutum*
Tolypocladium album	**OTU212**	100.000	*Tolypocladium*	**OTU203**	100.000	*Tolypocladium*	**OTU8327**	98.788	*Tolypocladium*
Umbelopsis vinacea	**OTU339**	100.000	*Umbelopsis*	**OTU266**	100.000	*Umbelopsis*	**OTU2241**	98.864	*Umbelopsis*
				OTU706	98.936	*Umbelopsis*	**OTU2314**	98.857	*Umbelopsis*
				OTU71	100.000	Umbelopsidales	**OTU2230**	97.159	*Umbelopsis*
Umbelopsis changbaiensis	**OTU94**	100.000	Umbelopsidales	**OTU71**	100.000	Umbelopsidales	**OTU2272**	97.059	*Umbelopsis*
							OTU2192	99.512	*Umbelopsis changbaiensis*
							OTU2288	97.688	*Umbelopsis changbaiensis*
Umbelopsis nana-dimorpha	-			**OTU299**	99.554	*Umbelopsis versiformis*	**OTU6856**	98.844	*Umbelopsis dimorpha*
Umbelopsis ramanianna	-			-			-		
Xylariales	**OTU50**	98.421	Fungi	**OTU60**	98.421	Fungi	**OTU9902**	98.11	Ascomycota
							OTU10432	98.065	Fungi
							OTU10430	97.351	Xylariales

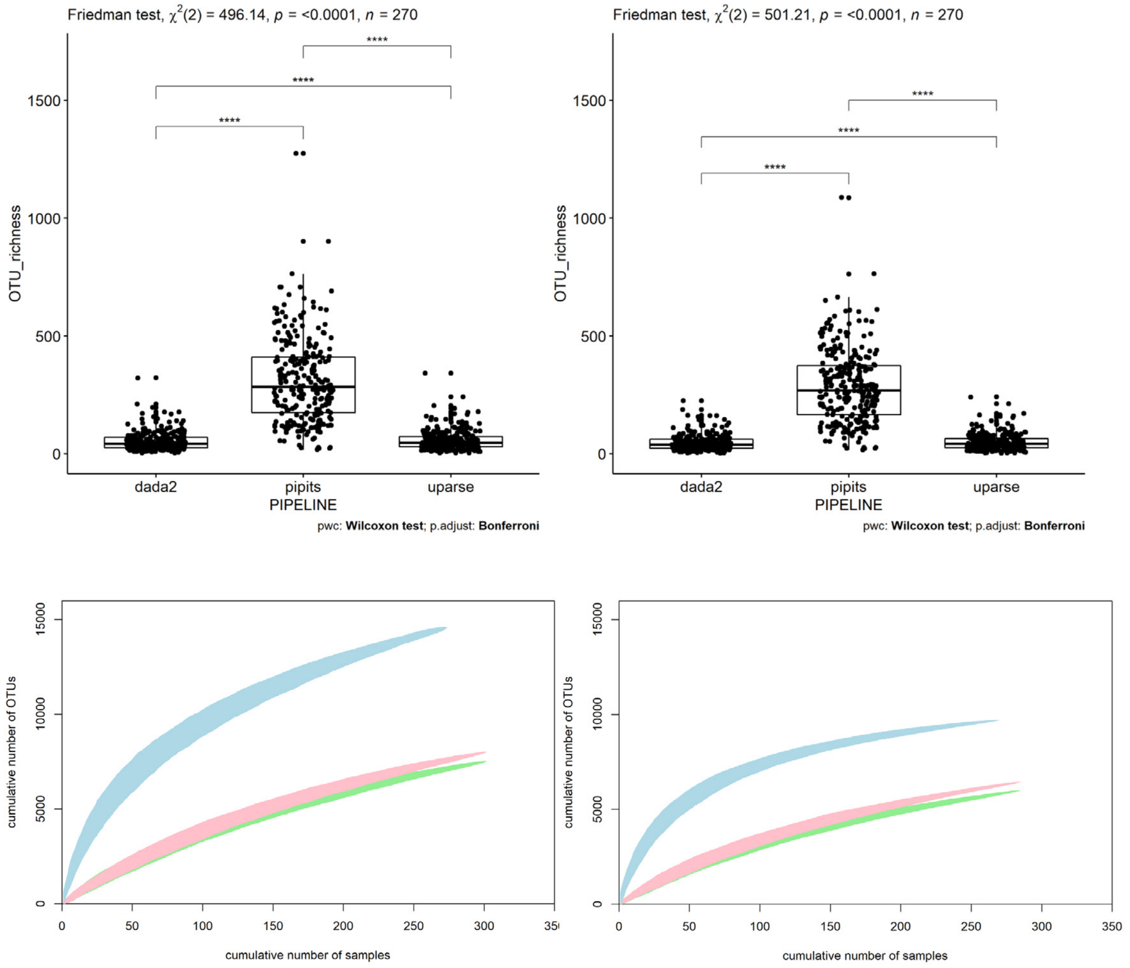

Fig. 2. Richness per sample comparisons (above): significant differences were detected between the three pipelines for curated (right) and non-curated data (left). Rarefaction curves (below): cumulative number of OTUs as a function of cumulative number of samples for the manually curated (right) and non-curated data (left) of the three pipelines tested: PIPITS (blue), AMPtk-UPARSE (pink) and AMPtk-DADA2 (green). Color version at https://www.ojs.darwin.edu.ar/index.php/darwiniana/article/view/1127/1309

Table 2. Parameters describing the comparison between culture-dependent and culture-independent approaches for each pipeline.

	AMPtk		PIPITS
	UPARSE	DADA2	
Matched taxa (%)	72.22	76.39	80.56
Merged taxa (%)	2.78	4.17	2.78
Misassigned taxa (%)	0.00	0.00	1.39
OTUs per taxa (mean)	1.10	1.12	1.85
Maximum number of OTU per taxa	3	3	35

DISCUSSION

In recent years, HTS metabarcoding approaches have revolutionized fungal ecology, increasing our ability to assess biodiversity in a wide range of habitats (Alberdi et al., 2018). As an emerging technology, it implied new methodological and theoretical challenges in terms of data processing and integration of knowledge production with existing backgrounds. Despite the number of studies that compare the performance of the different tools developed for data processing, there is still no consensus on the most appropriate bioinformatics approach (Anslan et al., 2018; Pauvert et al., 2019). These studies use comparison criteria with mock communities to evaluate the results of the pipelines. This is the first study that assesses the performance of different automated bioinformatic toolkits in the characterization of natural fungal communities, and that uses diversity data -obtained from culture methods- from the same sites and samples, as comparison criteria. This is a key step to further characterize endophyte communities and their beta diversity patterns through metabarcoding methodologies.

Bioinformatic pipelines comparisons

When evaluating the performance of a bioinformatic pipeline for analyzing community data, important factors to consider include the runtime, sensitivity, and precision. These parameters provide valuable insights into the efficiency and accuracy of the pipeline, serving as important tools for making informed decisions on which pipeline to apply based on the study goals and capabilities. The runtime of a pipeline determines the computational efficiency and speed of processing the large volumes of sequencing data. The sensitivity of the pipeline is directly impacted by the reads filtering, clustering, and chimera detection steps. These steps influence the ability to accurately capture the true fungal taxa present in the samples because the more reads are removed from the dataset during the filtering steps, the greater the risk of inadvertently eliminating existing taxa in the biological community. On the other hand, precision, which measures the pipeline's accuracy in identifying true fungal taxa without introducing false positives, relies on stringent filtering and error correction methods, such as chimera removal, clustering, and taxonomy assignment. The choice of clustering algorithm affects precision, with conservative algorithms creating distinct clusters, reducing the risk of merging sequences from different taxa. Conversely, sensitive algorithms may capture more taxa but have a higher chance of including false positives. In this context, it becomes evident the trade-off between sensitivity and precision (Weiss et al., 2016). Increasing sensitivity by relaxing filtering criteria may lead to a higher chance of false positives or including artifactual taxa, whereas prioritizing precision through more stringent filtering may result in the loss of rare or low-abundance taxa (Baldrian et al., 2021).

In the context of this study, it is worth noting that the filtering steps result in differences in reading recruitment with the fastx-toolkit applied in the PIPITS pipeline being less strict compared to the error trimming pipeline used in AMPtk. However, it is the clustering method that explains most of the differences between pipeline performances in this study. The VSEARCH and USEARCH algorithms did not perform similarly. VSEARCH clustering resulted in almost double OTU richness than UPARSE, even though it removed singletons before clustering. On the one hand, the PIPITS pipeline overestimates taxa, as evidenced in the high OTU redundancy (number of OTUs per taxa), that is, it showed less precision. Moreover, the highest richness obtained from this approach is partially reflecting actual taxa in the fungal community that the other pipelines could not recover, as is evidenced by the higher coverage over the cultured dataset that was reported for PIPITS (the ratio of cultured sequences that matches). These results agree with previous studies that have pointed out that the VSEARCH clustering method is a more sensitive approach (Rognes et al., 2016). However, others have found similar results in richness and sensitivity between VSEARCH and USEARCH when applied in other pipelines/toolkits (Anslan et al., 2018; Pauvert et al., 2019). Here, we found that applied in AMPtk toolkit, VSEARCH clustering method results in lower sensitivity. In fact, the AMPtk-UPARSE pipeline was the one that recovered the lower ratio of cultured taxa (that is, showed less sensitivity) and the one that evidenced the best OTU redundancy parameters (that is, more precision).

Regarding the latter, it might be an effect of the LULU algorithm that filtered 800 OTUs from the AMPtk-UPARSE pipeline. This tool was developed to detect taxa splitted in more than one OTU/ASV during the clustering method and merge them; plus, it is not available in the PIPITS toolkit.

The AMPtk-DADA2 pipeline offered the best balance concerning both precision and sensitivity. It yields low ASV redundancy aligning with sequences from culture and more than 76% coverage for the cultured dataset and, also, achieving the lowest richness among the three pipelines. These results are consistent with other studies that found that the DADA2 pipeline achieves the most approximate characterization of mock community alfa diversity (Pauvert et al., 2019). DADA2 is a clustering-free method developed to enable result comparisons between studies and improve taxonomic resolution (Callahan et al., 2017). It evaluates reads at the sample level and combines the identification of ASVs with chimera detection and removal. This approach assumes that highly similar ASVs within the same sample represent errors when they occur in very low abundances. It allows the detection of single-nucleotide polymorphisms that may indicate different fungal species while reducing OTU redundancy (Callahan et al., 2016).

The taxonomy assignment lacks mycological accuracy for the system under study in both PIPITS and AMPtk toolkits: up to 35% of total manually curated OTUs could not go further than the Kingdom Fungi determination. This is a common issue in metabarcoding approach studies from diverse environments (Kirker et al., 2017; Purahong et al., 2019), especially when plant endophyte mycobiota is being assessed (James et al., 2020). In general, sequence-based identification depends on informative sequence databases (Costello et al., 2013). In particular, bioinformatic methods for taxonomy assignment, and especially the learning machine approaches, are sensitive to the incompleteness of the reference databases because the algorithms perform better when there are multiple representatives for each group (Gdanetz et al., 2017). There is a current lack of knowledge about fungal diversity in certain environments and about entire fungal lineages that keep the public databases incomplete (Halwachs et al., 2017). Apart from the general limitations regarding the available databases that both toolkits faced, PIPITS and AMPtk pipelines performed differently concerning taxonomy assignment. On the one hand, PIPITS resulted in a larger proportion of unassigned OTUs at the Kingdom rank, which might be due to the more relaxed algorithms for reads filtering and post-clustering curation. On the other hand, this pipeline resulted in slightly better taxonomic resolution, meaning a larger proportion of assignments at Class rank or less. However, it also evidenced a proportion of misassignment when validating with the cultured dataset. AMPtk toolkit has achieved lower resolution but with no such errors. Unlike PIPITS, which implements the RDP Classifier method, AMPtk uses an approach in which taxonomy is assigned through the consensus of three different methods. Evidently, the hybrid is a more conservative approach, which loses resolution by assigning taxonomy with a "last common ancestor" criterion but allowing reducing errors.

Cultured taxa from our reference dataset are known to be present in the studied ecosystem, however, there is a fraction of their sequences that the HTS experiment did not recover. Culture-dependent and culture-independent experiments were not carried out from the same samples (although they were taken from the same individuals in a year of culture prospection), therefore some taxa could be absent in one of the approaches, especially if those taxa are rare. Nevertheless, some of the sequences absent in the pipeline's outputs were from taxa reported in high frequency in the culture-prospection study (Molina et al., 2020). Furthermore, certain of these genera were not informed at all in the culture-independent approaches, such as *Ophiostoma* or *Arambarria*. Here, we might be witnessing the bias and limitations of the HTS metabarcoding approach and of the ITS amplicon itself. During total DNA extraction from an environmental sample, the cell wall properties of the different fungal taxa or types (Vesty et al., 2017) and the variable number of nuclei per cell across taxa (Roper et al., 2011) will affect the DNA recruitment. Besides, it is well documented that markers differ in their capacity to recover OTUs/ASVs across fungal lineages (Tedersoo et al., 2015).

The Internal Transcribed Spacer region has been used as a universal fungal marker because of the optimal multi-copy characteristics and its variation rate across lineages (Schoch et al., 2012). However, there are some drawbacks for HTS metabarcoding studies. For instance, the ITS region is highly variable in length (Schoch et al., 2014) and GC content (Wang et al., 2015); longer and higher GC content barcodes are reported as difficult templates to amplify in NGS because of the unequal competition for primers (Aird et al., 2011). Plus, longer barcodes are less likely to be recovered in short-read-based approaches, because the quality falls in tails and low-quality sequences are problematic to pair (Baldrian et al., 2021).

It is noticeable, however, that compared with this study the few studies that have compared the alpha diversity achieved by HTS approaches versus morphological studies have got a very small overlapping of taxa between approaches (Porter et al., 2008, Heine et al., 2021). Furthermore, the findings of this study align with previous assessments of the sensitivity of HTS workflows that employ mock community approaches (Pauvert et al., 2019), demonstrating consistent results in terms of the percentage of recovered sequences. All this suggests that sampling and sequencing efforts in this study were satisfactory.

Sequencing depth achieved here was higher than that of similar wood fungal endophytes studies (Küngas et al., 2020; Migliorini et al., 2021). Sequence depth is the more important variable in HTS experimental design that aims to assess beta diversity (Smith & Peay, 2014). However, high sequencing depth increases the potential for cross-contamination and errors during sequencing (Baldrian et al., 2021). In this study, we combine a high sequence depth with an approach to correct cross-contamination errors by using a synthetic mock community.

In summary, the AMPtk toolkit showed to be more precise in terms of false positives and taxonomy assignment than PIPITS. Both AMPtk pipelines had similar performances but the pipeline that uses the DADA2 clustering algorithm showed lower redundancy and higher sensitivity. The AMPtk-DADA2 would be chosen to perform community patterns analyses, however, PIPITS showed itself as a more sensitive pipeline and would be considered in studies aiming for species detection.

DATA AVAILABILITY

Raw sequence reads are deposited in the Short Read Archive of the National Center for Biotechnology Information (BioProject ID: PRJNA785007).

Sanger DNA sequences from culture are available on GenBank (accessions MT076081-MT07685).

ACKNOWLEDGEMENTS

We thank Verónica El Mujtar for her advice in laboratory work and Jorge Ronny Díaz-Valderrama for his assistance during bioinformatic analysis. This work was supported by the ANPCyT 2018/3234 (to MR), CONICET 11220200101167CO (to MBP) and UNPSJB 80020220100003UP (to MBP and LM).

BIBLIOGRAPHY

Aird, D.; M. G. Ross, W.-S. Chen, M. Danielsson, T. Fennell, C. Russ, D. B. Jaffe, C. Nusbaum & A. Gnirke. 2011. Analyzing and Minimizing PCR Amplification Bias in Illumina Sequencing Libraries. *Genome Biology* 12 (2): R18. DOI: https://doi.org/10.1186/gb-2011-12-2-r18

Altschul, S.; T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 25 (17): 3389-3402. DOI: https://doi.org/10.1093/nar/25.17.3389

Anslan, S.; R. H. Nilsson, C. Wurzbacher, P. Baldrian, L. Tedersoo & M. Bahram. 2018. Great Differences in Performance and Outcome of High-Throughput Sequencing Data Analysis Platforms for Fungal Metabarcoding. *MycoKeys* 39: 29-40. DOI: https://doi.org/10.3897/mycokeys.39.28109

Baldrian, P. 2016. Forest Microbiome: Diversity, Complexity and Dynamics. *FEMS Microbiology Reviews* 41 (2): 109-30. DOI: https://doi.org/10.1093/femsre/fuw040

Baldrian, P.; T. Větrovský, C. Lepinay & P. Kohout. 2021. High-Throughput Sequencing View on the Magnitude of Global Fungal Diversity. *Fungal Diversity* 114: 539-547. DOI: https://doi.org/10.1007/s13225-021-00472-y

Bengtsson-Palme, J.; M. Ryberg, M. Hartmann, S. Branco, Z. Wang, A. Godhe, P. De Wit 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 in Ecology and Evolution* 4 (10): 914-919. DOI: https://doi.org/10.1111/2041-210X.12073

Brown, S. P.; A. M. Veach, A. R. Rigdon-Huss, K. Grond, S. K. Lickteig, K. Lothamer, A. K. Oliver & A. Jumpponen. 2015. Scraping the Bottom of the Barrel: Are Rare High Throughput Sequences Artifacts? *Fungal Ecology* 13: 221-225. DOI: https://doi.org/10.1016/j.funeco.2014.08.006

Callahan, B. J.; P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson & S. P. Holmes. 2016. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nature Methods* 13 (7): 581-83. DOI: https://doi.org/10.1038/nmeth.3869

Callahan, B. J.; P. J. McMurdie & S. P Holmes. 2017. Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis. *The ISME Journal* 11 (12): 2639-2643. DOI: https://doi.org/10.1038/ismej.2017.119

Costello, M. J.; R. M. May & N. E. Stork. 2013. Can We Name Earth's Species Before They Go Extinct? *Science* 339 (6118): 413-416. DOI: https://doi.org/10.1126/science.1230318

Dumolin, S.; B. Demesure & R. J. Petit. 1995. Inheritance of Chloroplast and Mitochondrial Genomes in Pedunculate Oak Investigated with an Efficient PCR Method. *Theoretical and Applied Genetics* 91 (8): 1253-1256. DOI: https://doi.org/10.1007/BF00220937

Edgar, R. C. 2010. Search and Clustering Orders of Magnitude Faster than BLAST. *Bioinformatics* 26 (19): 2460-2461. DOI: https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C. 2013. UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads. *Nature Methods* 10 (10): 996-998. DOI: https://doi.org/10.1038/nmeth.2604

Edgar, R. C. 2016. SINTAX: A Simple Non-Bayesian Taxonomy Classifier for 16S and ITS Sequences. *BioRxiv*, 074161. DOI: https://doi.org/https://doi.org/10.1101/074161

Edgar, R. C. & H. Flyvbjerg. 2015. Error Filtering, Pair Assembly and Error Correction for next-Generation Sequencing Reads. *Bioinformatics* 31 (21): 3476-3482. DOI: https://doi.org/10.1093/bioinformatics/btv401

Frøslev, T. G.; R. Kjøller, H. H. Bruun, R. Ejrnæs, A. K. Brunbjerg, C. Pietroni & A. J. Hansen. 2017. Algorithm for Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates. *Nature Communications* 8 (1): 1188. DOI: https://doi.org/10.1038/s41467-017-01312-x

Gardes, M. & T. D. Bruns. 1993. ITS Primers with Enhanced Specificity for Basidiomycetes - Application to the Identification of Mycorrhizae and Rusts. *Molecular Ecology* 2 (2): 113-118. DOI: https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Gdanetz, K.; G. M. N. Benucci, N. Vande Pol & G. Bonito. 2017. CONSTAX: A Tool for Improved Taxonomic Resolution of Environmental Fungal ITS Sequences. *BMC Bioinformatics* 18 (1): 538. DOI: https://doi.org/10.1186/s12859-017-1952-x

Gordon, A. & G. J. Hannon. 2010. Fastx-Toolkit. *FASTQ/A Short-Reads Preprocessing Tools (Unpublished) Http://Hannonlab. Cshl. Edu/Fastx_toolkit*.

Gweon, H. S.; A. Oliver, J. Taylor, T. Booth, M. Gibbs, D. S. Read, R. I. Griffiths & K. Schonrogge. 2015. PIPITS: An Automated Pipeline for Analyses of Fungal Internal Transcribed Spacer Sequences from the Llumina Sequencing Platform. *Methods in Ecology and Evolution* 6 (8): 973-980. DOI: https://doi.org/10.1111/2041-210X.12399

Halwachs, B.; N. Madhusudhan, R. Krause, R. H. Nilsson, C. Moissl-Eichinger, C. Högenauer, G. G. Thallinger & G. Gorkiewicz. 2017. Critical Issues in Mycobiota Analysis. *Frontiers in Microbiology* 8: 1-12. DOI: https://doi.org/10.3389/fmicb.2017.00180

Heine, P.; J. Hausen, R. Ottermanns & M. Roß-Nickoll. 2021. Comparing EDNA Metabarcoding with Morphological Analyses: Fungal Species Richness and Community Composition of Differently Managed Stages along a Forest Conversion of Norway Spruce towards European Beech in Germany. *Forest Ecology and Management* 496: 119429. DOI: https://doi.org/10.1016/j.foreco.2021.119429

Jalili, V.; E. Afgan, Q. Gu, D. Clements, D. Blankenberg, J. Goecks, J. Taylor & A. Nekrutenko. 2020. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2020 Update. *Nucleic Acids Research* 48 (W1): W395-402. DOI: https://doi.org/10.1093/nar/gkaa434

James, T. Y.; J. E. Stajich, C. T. Hittinger & A. Rokas. 2020. Toward a Fully Resolved Fungal Tree of Life. *Annual Review of Microbiology* 74 (1): 291-313. DOI: https://doi.org/10.1146/annurev-micro-022020-051835

Kirker, G. T.; A. B. Bishell, M. A. Jusino, J. M. Palmer, W. J. Hickey & D. L. Lindner. 2017. Amplicon-Based Sequencing of Soil Fungi from Wood Preservative Test Sites. *Frontiers in Microbiology* 8: 1997. DOI: https://doi.org/10.3389/fmicb.2017.01997

Küngas, K.; M. Bahram & K. Põldmaa. 2020. Host Tree Organ Is the Primary Driver of Endophytic Fungal Community Structure in a Hemiboreal Forest. *FEMS Microbiology Ecology* 96 (2): 1-10. DOI: https://doi.org/10.1093/femsec/fiz199

McMurdie, P. J. & S. Holmes. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Edited by Michael Watson. *PLoS ONE* 8 (4): e61217. DOI: https://doi.org/10.1371/journal.pone.0061217

McMurdie, P. J. & J. N. Paulson. 2016. Biomformat: An Interface Package for the BIOM File Format. *R/Bioconductor Package*.

Migliorini, D.; M. Messal, A. Santini, A. P. Ramos, P. Talhinhas, M. J. Wingfield & T. Burgess. 2021. Metabarcoding Reveals Southern Hemisphere Fungal Endophytes within Wood of Cultivated Proteaceae in Portugal. *European Journal of Plant Pathology* 160 (1): 173-184. DOI: https://doi.org/10.1007/s10658-021-02233-8

Molina, L.; M. Rajchenberg, A. de Errasti, M. C. Aime & M. B. Pildain. 2020. Sapwood-Inhabiting Mycobiota and Nothofagus Tree Mortality in Patagonia: Diversity Patterns According to Tree Species, Plant Compartment and Health Condition. *Forest Ecology and Management* 462: 117997. DOI: https://doi.org/10.1016/j.foreco.2020.117997

Molina, L.; M. Rajchenberg, A. de Errasti, B. Vogel, M. P. A. Coetzee, M. C. Aime & M. B. Pildain. 2022. Sapwood mycobiome varies across host, plant compartment and environments in *Nothofagus* forests from Northern Patagonia. *Molecular Ecology* 00: 1-20. DOI: https://doi.org/10.1111/mec.16771

Molina, L. & M. B. Pildain. 2022. Uso de la secuenciación de segunda generación (NGS) para descubrir la diversidad de hongos degradadores de la madera en los bosques Andino Patagónicos. *Lilloa* 59 (Suplemento): 155-172. DOI: https://doi.org/10.30550/j.lil/2022.59.S/2022.09.22

Mysara, M.; M. Njima, N. Leys, J. Raes & P. Monsieurs. 2017. From Reads to Operational Taxonomic Units: An Ensemble Processing Pipeline for MiSeq Amplicon Sequencing Data. *GigaScience* 6 (2): 1-10. DOI: https://doi.org/10.1093/gigascience/giw017

Nilsson, R. H.; S. Anslan, M. Bahram, C. Wurzbacher, P. Baldrian & L. Tedersoo. 2019. Mycobiome Diversity: High-Throughput Sequencing and Identification of Fungi. *Nature Reviews Microbiology* 17 (2): 95-109. DOI: https://doi.org/10.1038/s41579-018-0116-y

Oksanen, J.; F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson & P. Solymos. 2020. Vegan: Community Ecology Package. R Package Version 2.5.7. 2020.

Palmer, J. M.; M. A. Jusino, M. T. Banik & D. L. Lindner. 2018. Non-Biological Synthetic Spike-in Controls and the AMPtk Software Pipeline Improve Mycobiome Data. *PeerJ* 6 (5): e4925. DOI: https://doi.org/10.7717/peerj.4925

Pauvert, C.; M. Buée, V. Laval, V. Edel-Hermann, L. Fauchery, A. Gautier, I. Lesur, J. Vallance & C. Vacher. 2019. Bioinformatics Matters: The Accuracy of Plant and Soil Fungal Community Data Is Highly Dependent on the Metabarcoding Pipeline. *Fungal Ecology* 41: 23-33. DOI: https://doi.org/10.1016/j.funeco.2019.03.005

Porter, T. M.; J. E. Skillman & J.-M. Moncalvo. 2008. Fruiting Body and Soil RDNA Sampling Detects Complementary Assemblage of Agaricomycotina (Basidiomycota, Fungi) in a Hemlock-Dominated Forest Plot in Southern Ontario. *Molecular Ecology* 17 (13): 3037-3050. DOI: https://doi.org/10.1111/j.1365-294X.2008.03813.x

Purahong, W.; A. Mapook, Y.-T. Wu & C.-T. Chen. 2019. Characterization of the *Castanopsis carlesii* Deadwood Mycobiome by Pacbio Sequencing of the Full-Length Fungal Nuclear Ribosomal Internal Transcribed Spacer (ITS). *Frontiers in Microbiology* 10: 1-14. DOI: https://doi.org/10.3389/fmicb.2019.00983

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/

Rognes, T.; T. Flouri, B. Nichols, C. Quince & F. Mahé. 2016. VSEARCH: A Versatile Open Source Tool for Metagenomics. *PeerJ* 4 (10): e2584. DOI: https://doi.org/10.7717/peerj.2584

Roper, M.; C. Ellison, J. W. Taylor & N. L. Glass. 2011. Nuclear and Genome Dynamics in Multinucleate Ascomycete Fungi. *Current Biology* 21 (18): R786-93. DOI: https://doi.org/10.1016/j.cub.2011.06.042

Saikkonen, K.; S. H. Faeth, M. Helander & T. J. Sullivan. 1998. Fungal Endophytes: A Continuum of Interactions with Host Plants. *Annual Review of Ecology and Systematics* 29 (1): 319-343. DOI: https://doi.org/10.1146/annurev.ecolsys.29.1.319

Schloss, P. D. & S. L. Westcott. 2011. Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S RRNA Gene Sequence Analysis. *Applied and Environmental Microbiology* 77 (10): 3219-3226. DOI: https://doi.org/10.1128/AEM.02810-10

Schoch, C. L.; K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen et al. 2012. Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* 109 (16): 6241-6246. DOI: https://doi.org/10.1073/pnas.1117018109

Schoch, C. L, B. Robbertse, V. Robert, D. Vu, G. Cardinali, L. Irinyi, W. Meyer et al. 2014. Finding Needles in Haystacks: Linking Scientific Names, Reference Specimens and Molecular Data for Fungi. *Database* 2014: bau061-bau061. DOI: https://doi.org/10.1093/database/bau061

Smith, D. P. & K. G. Peay. 2014. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. *PLoS ONE* 9 (2): e90234. DOI: https://doi.org/10.1371/journal.pone.0090234

Stone, J. K.; J. D. Polishook & J. F. J. White. 2004. Endophytic Fungi. In *Biodiversity of Fungi: Inventory and Monitoring Methods*, edited by G. M. Mueller; G. F. Billis & M. S. Foster, 241-270. Burlinghton: Elsevier. DOI: https://doi.org/10.13140/RG.2.1.2497.0726

Suryanarayanan, T. S. 2020. The Need to Study the Holobiome for Gainful Uses of Endophytes. *Fungal Biology Reviews* 34 (3): 144-150. DOI: https://doi.org/10.1016/j.fbr.2020.07.004

Tedersoo, L.; S. Anslan, M. Bahram, S. Põlme, T. Riit, I. Liiv, U. Kõljalg et al. 2015. Shotgun Metagenomes and Multiple Primer Pair-Barcode Combinations of Amplicons Reveal Biases in Metabarcoding Analyses of Fungi. *MycoKeys* 10: 1-43. DOI: https://doi.org/10.3897/mycokeys.10.4852

Terhonen, E.; K. Blumenstein, A. Kovalchuk & F. O. Asiegbu. 2019. Forest Tree Microbiomes and Associated Fungal Endophytes: Functional Roles and Impact on Forest Health. *Forests* 10 (1): 42. DOI: https://doi.org/10.3390/f10010042

Unterseher, M. 2011. Diversity of Fungal Endophytes in Temperate Forest Trees. In *Endophytes of Forest Trees: Biology and Applications*, edited by A. M. Pirttilä & A. C. Frank, 80: 31-46. Forestry Sciences. Dordrecht: Springer Netherlands. DOI: https://doi.org/10.1007/978-94-007-1599-8_2

Vesty, A.; K. Biswas, M. W. Taylor, K. Gear & R. G. Douglas. 2017. Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities. *PLOS ONE* 12 (1): e0169877. DOI: https://doi.org/10.1371/journal.pone.0169877

Wang, P.; Y. Chen, Y. Sun, S. Tan, S. Zhang, Z. Wang, J. Zhou et al. 2019. Distinct Biogeography of Different Fungal Guilds and Their Associations With Plant Species Richness in Forest Ecosystems. *Frontiers in Ecology and Evolution* 7: 216. DOI: https://doi.org/10.3389/fevo.2019.00216

Wang, Q.; G. M. Garrity, J. M. Tiedje & J. R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of RRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology* 73 (16): 5261-5267. DOI: https://doi.org/10.1128/AEM.00062-07

Wang, X.-C.; C. Liu, L. Huang, J. Bengtsson-Palme, H. Chen, J.-H. Zhang, D. Cai & J.-Q. Li. 2015. ITS1: A DNA Barcode Better than ITS2 in Eukaryotes? *Molecular Ecology Resources* 15 (3): 573-586. DOI: https://doi.org/10.1111/1755-0998.12325

Weiss, S.; W. Van Treuren, C. Lozupone, K. Faust, J. Friedman, Y. Deng, L. C. Xia et al. 2016. Correlation Detection Strategies in Microbial Data Sets Vary Widely in Sensitivity and Precision. *The ISME Journal* 10 (7): 1669-1681. DOI: https://doi.org/10.1038/ismej.2015.235

White, T. J.; T. Bruns, S. J. W. T. Lee & J. Taylor. 1990. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*, edited by T. J. White; T. Bruns, S. J. W. T. Lee, J. Taylor, M. A. Innis, D. H. Gelfand & J. J. Sninsky, 315-322. New York: Academic Press.

Wickham, H. 2016. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. DOI: https://ggplot2.tidyverse.org