



Review

# DNA Metabarcoding for the Characterization of Terrestrial Microbiota—Pitfalls and Solutions

Davide Francioli <sup>1,\*</sup> , Guillaume Lentendu <sup>2</sup> , Simon Lewin <sup>1</sup> and Steffen Kolb <sup>1</sup>

<sup>1</sup> Microbial Biogeochemistry, Research Area Landscape Functioning, Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany; Simon.Lewin@zalf.de (S.L.); steffen.kolb@zalf.de (S.K.)

<sup>2</sup> Laboratory of Soil Biodiversity, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland; guillaume.lentendu@unine.ch

\* Correspondence: davide.francioli@zalf.de; Tel.: +49-334-328-2488

**Abstract:** Soil-borne microbes are major ecological players in terrestrial environments since they cycle organic matter, channel nutrients across trophic levels and influence plant growth and health. Therefore, the identification, taxonomic characterization and determination of the ecological role of members of soil microbial communities have become major topics of interest. The development and continuous improvement of high-throughput sequencing platforms have further stimulated the study of complex microbiota in soils and plants. The most frequently used approach to study microbiota composition, diversity and dynamics is polymerase chain reaction (PCR), amplifying specific taxonomically informative gene markers with the subsequent sequencing of the amplicons. This methodological approach is called DNA metabarcoding. Over the last decade, DNA metabarcoding has rapidly emerged as a powerful and cost-effective method for the description of microbiota in environmental samples. However, this approach involves several processing steps, each of which might introduce significant biases that can considerably compromise the reliability of the metabarcoding output. The aim of this review is to provide state-of-the-art background knowledge needed to make appropriate decisions at each step of a DNA metabarcoding workflow, highlighting crucial steps that, if considered, ensures an accurate and standardized characterization of microbiota in environmental studies.

**Keywords:** DNA metabarcoding workflow; high-throughput sequencing; terrestrial ecosystem; bacteria; archaea; fungi; protists; soil and plant-associated microorganisms



**Citation:** Francioli, D.; Lentendu, G.; Lewin, S.; Kolb, S. DNA Metabarcoding for the Characterization of Terrestrial Microbiota—Pitfalls and Solutions. *Microorganisms* **2021**, *9*, 361. <https://doi.org/10.3390/microorganisms9020361>

Academic Editor: Anna Reale

Received: 23 November 2020

Accepted: 9 February 2021

Published: 12 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Soil microorganisms have been recognized as an integral part of terrestrial ecosystems because they play a central role in nutrient transformation and in plant community productivity, composition and diversity [1]. However, our knowledge of soil microbiota is limited by the huge microbial diversity that characterizes terrestrial ecosystems and by the complexity of soil–plant–microbe interactions [2]. Indeed, soil has often been dubbed a “black box” because of the high abundance of soil microbial populations ( $10^8$ – $10^{11}$  cells per gram) and the methodological challenges to characterize them [3,4]. Currently, this black box is beginning to be pried open, largely due to advances in molecular tools that have paved the way forward for soil microbial ecologists to unravel the composition and function of the soil microbiota in terrestrial ecosystems [5]. Novel molecular approaches, which employ polymerase chain reaction (PCR) and high-throughput sequencing (HTS), have revolutionized the way to study the soil microbiota. Application of these methods has demonstrated that a large fraction of terrestrial microbes can be detected solely using molecular approaches, thus discouraging the need for laboratory isolation and culturing of specimens. Furthermore, with the decrease of sequencing price and high-throughput

samples analysis by various bioinformatics tools, the use of massively parallel sequencing (MPS) in soil microbial ecology has become a standard approach.

Prokaryotes (Archaea and Bacteria) and fungi are the most studied microbes in soils and plants. The “other” microbes in soils are grouped under the term protists [6], and despite their relative lower abundance compared to their prokaryotic and fungal counterparts, they carry significant functional roles at all trophic levels [7]. The characterization of the soil microbial community is commonly carried out via PCR amplification of taxonomic marker genes (called “DNA barcodes”). These markers are typically 100 to 600 bp long, and they need to be sufficiently variable to provide deep taxonomic resolution and are simultaneously flanked by conserved regions to cover a broad range of taxa. The combination of HTS with barcoding has been named “metabarcoding” [8]. The relative short length of these markers does not always allow a resolution to species level, so alternative approaches like single-cell metagenomics or isolation via cultivation are needed to fully discriminate microbial species. Despite this limitation, DNA metabarcoding has rapidly emerged as a powerful, repeatable and cost-effective method for characterizing microbial communities in small and large-scale studies. This comprehensive approach has enabled soil microbiologists to explore important ecological aspects related to soil–plant–microbe systems, such as the identification of microbial taxa that are (i) dominant or low in abundance across different terrestrial ecosystems; (ii) involved in specific processes (e.g., litter decomposition, nitrogen cycling, degradation of toxic compounds and many more); (iii) more sensitive to abiotic and biotic factors. DNA metabarcoding further allows assessing soil microbial biodiversity (also in terms of phylogenetic relatedness), and to compare soil communities subjected to experimental conditions or geographical distance. It is also a cost-effective method for biomonitoring as DNA metabarcoding is more frequently used for monitoring agricultural practices, restoration efforts or forensics [9–12]. Presently, it represents the most used molecular approach to characterize microbiota in environmental samples.

In this review, we focus on all the steps in the identification of soil and plant-associated microbes using DNA metabarcoding (Figure 1). This approach consists of multiple laboratory procedures and requires bioinformatics and computational statistics. Therefore, sufficient technical knowledge and informed choice at each step are essential for successful microbial detection and taxonomic identification [13]. In addition, the use of DNA metabarcoding for microbial identification has some important limitations, including the variable number of copies of the selected gene markers in microbial genomes, the low taxonomic resolution at the species level for some microbial groups and biases in the taxonomic annotation of sequences depending on the variable region chosen for the analysis [14,15]. Hence, the choice of a proper modus operandi for all the steps in metabarcoding workflows is crucially important. Inappropriate methods in microbiota studies may generate insufficient and fallacious biological inferences [16,17]. Indeed, significant biases can occur from the cumulative effect of both systematic and random errors throughout the whole workflow, including sampling, DNA extraction, amplicon library preparation, sequencing and bioinformatics [18,19].

Based on literature review and experience, we provide a comprehensive overview of the positive and negative aspects related to each step of the metabarcoding workflow for microbiota studies on samples associated with terrestrial ecosystems (Figure 1). Since sampling procedures for soil- and plant-associated microbiome were already covered in other reviews [18,20,21], we here concentrate mainly on the molecular aspects of the metabarcoding workflow. Therefore, in the next sections, we first discuss practical sample handling procedures and molecular approaches fundamental in the preparation of the sequencing library. This will provide guidance on important methodological issues that might be overlooked. Second, we describe useful software tools that are typically employed in the bioinformatics data processing and in the taxonomic characterization of the detected microbial taxa. Finally, we discuss potential future applications of next-generation sequencing (NGS) platforms and technologies in unraveling the relationships between microbial biodiversity and ecosystem functions.

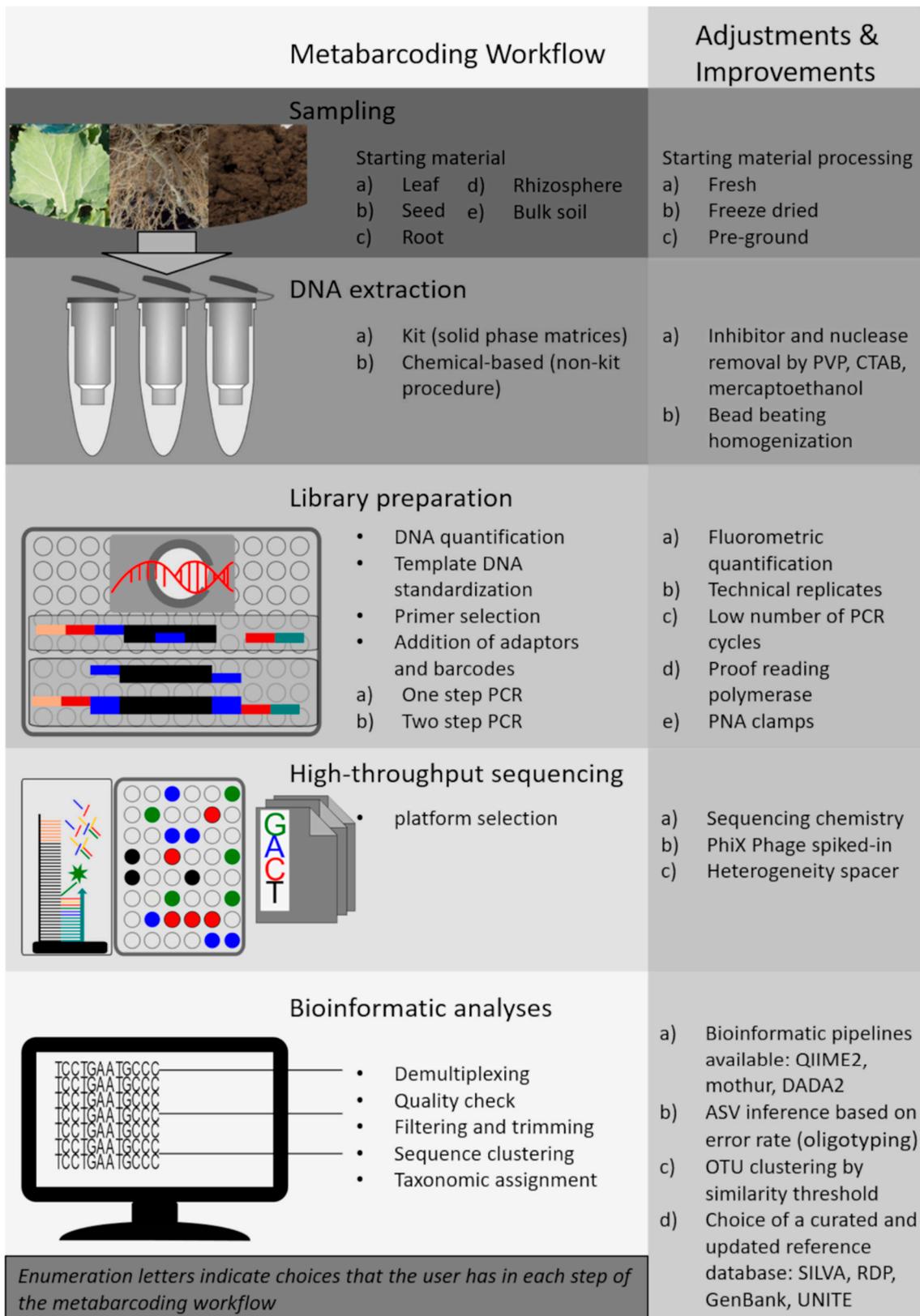


Figure 1. DNA metabarcoding workflow with suggested adjustments and improvements.

## 2. DNA Extraction Procedure

Extraction of the genetic material from environmental samples is the first step in the metabarcoding workflow. Total genomic DNA extraction represents a crucial stage in which the potential biases have to be minimized using appropriate laboratory protocols. The analytical success of molecular techniques is significantly affected by a successful DNA extraction, which involves the effective sample homogenization and disruption of cells, denaturation of proteins and nucleoprotein complexes, inactivation of nucleases, removal of humic acids and other PCR inhibitors and recovery of the DNA. Presently, these steps are performed using commercial kits that employ both chemicals and solid-phase matrices. Such DNA extraction kits are simple to use and rapid, and most of them do not include harmful solutions. However, chemical-based DNA extraction protocols that do not involve the use of commercial kits, such as phenol-chloroform-based extraction method, are still in use [22]. Such DNA extraction procedures are usually cheaper per extraction compared to commercial kits, in addition to their good quality and quantity of the extracted DNA. Moreover, the different steps and solutions of such a procedure can be optimized to the sample material. However, solution-based DNA extraction protocols can be quite laborious, since (i) all the steps are manual, (ii) they often need fresh-made solutions and (iii) they use toxic chemicals.

For the isolation of total genomic DNA from terrestrial environments (soil and plant material), many commercial kits and protocols for soil, seeds and plant tissue are available (Table 1). However, the choice of the suitable DNA extraction procedure can be more complicated when dealing with multiple sample types such as bulk soil, rhizosphere, stem and leaf. In the case that the experimental aim is the comparison of microbiota across compartments, then the DNA associated with such compartments must be extracted with the same method. This is necessary to avoid protocol-specific biases when comparing, for example, rhizospheric soil to root or either of those to the leaf or stem tissue. However, each compartment can be extracted with the method that works best for it when the comparison across compartments is not the aim. This will provide a better snapshot of the community associated with each compartment, but with the loss of the capability to compare between them. Based on our experience, the soil DNA extraction kits listed in Table 1 can be employed for the extraction of genomic material from different types of samples (soil, sediments and plant material) with satisfactory results in terms of quantity and quality of the DNA.

Another important aspect to consider concerning the DNA extraction procedure is that Gram-positive and -negative Bacteria, Archaea, fungi and protists are differentially sensitive to cell disruption. Thus, sample homogenization and disruption of cells can represent a major cause of bias in the microbiota composition. Thereby, bead-beating in combination with chemical lysis agents was shown to be most efficient for soil and plant material [23]. Thus, the downstream analyses will not be confounded to less or highly resistant microorganisms. Furthermore, when there is the need to process a large number of samples, bead beating-based kits can represent a much better choice than tedious and time-consuming “home-made” protocols (e.g., phenol-chloroform-based methods), although the kits tend to be more costly. It is worth mentioning that the bead beating procedure requires a dedicated bead beater homogenizer, which can be prohibitive due to its cost (from a few thousand up to 10,000 dollars depending on the homogenizer features).

Additionally, other extraction methodologies can be employed when the objective is to extract not the soil total genomic DNA (also known as environmental DNA or eDNA [24]) but specific fractions of it. For instance, to collect the extracellular DNA fraction, which can be released from dead prokaryotic and eukaryotic cells and can be protected against nuclease degradation by its adsorption on soil colloids and sand particles, protocols that avoid the lysis of the cells by using only low centrifugation speeds and mild chemical concentrations are generally used [25,26]. Another approach, named “indirect DNA extraction”, is employed when the aim is to individually collect different microbial DNA fractions. This method involves the initial separation of prokaryotic and eukaryotic cells from the

soil matrix by density gradient centrifugation prior to their lysis [27,28]. Such isolated cell communities could then be further sorted at the single-cell level using flow cytometry or microfluidic devices before DNA extraction and subsequent metabarcoding [29,30].

Therefore, the choice of a particular DNA extraction protocol depends on the type and number of samples, study purpose, equipment availability and financial constraints. Finally, the extracted DNA can be stored at  $-20\text{ }^{\circ}\text{C}$  or  $-80\text{ }^{\circ}\text{C}$  for further processing. It is also worth noting that RNA could be extracted in parallel to DNA using dedicated kits, but this aspect was covered elsewhere [31–33] and we will not go into detail here.

**Table 1.** Commonly used DNA extraction kits and methods for soil and plant-associated microbiota.

Kit Manufacturer or Method	Sample Type	Homogenization and Cell Lysis	DNA Purification and Concentration	Relative Cost Per Sample [Low (\$) to High (\$\$\$)]
<i>DNeasy PowerSoil</i> Qiagen, USA	Soil, compost, manure, plant material	Bead beating + chemical lysis	Silica membrane binding	\$\$\$
<i>FastDNA Kit for Soil</i> MP Biomedicals, USA	Soil, compost, manure	Bead beating + chemical lysis	Silica membrane binding	\$\$\$
<i>Plant DNeasy Mini kit</i> Qiagen, USA	Plant and fungal tissue.	Mortar/pestel or TissueLyzer + chemical lysis	Silica membrane binding	\$\$
<i>Quick-DNA Fecal/Soil Microbe Miniprep Kit</i> Zymo Research, Germany	Soil, biofilm, animal and human samples	Bead beating + chemical lysis	Silica membrane binding	\$\$
Phenol-chloroform-isoamyl alcohol-Extraction [22]	Soil	Bead beating + CTAB <sup>a</sup>	PEG <sup>b</sup> 6000 + ethanol precipitation	\$
Phenol-chloroform-isoamyl alcohol-Extraction modified [31]	Soil	Bead beating + CTAB <sup>a</sup> + PVP <sup>c</sup>	PEG <sup>b</sup> 6000 + ethanol precipitation	\$
Phenol-chloroform-isoamyl alcohol-Extraction modified [32]	Soil	Bead beating + CTAB <sup>a</sup> + PVPP <sup>d</sup>	Isopropanol precipitation	\$\$
Sodiumphosphate extraction [34]	Sediments	Bead beating + Sodiumphosphate buffer + PVP <sup>c</sup>	Silica membrane binding + GuaHCL <sup>e</sup> precipitation	\$\$

<sup>a</sup> hexadecyltrimethylammonium bromide; <sup>b</sup> polyethylene glycol; <sup>c</sup> polyvinylpyrrolidone; <sup>d</sup> polyvinylpolypyrrolidone; <sup>e</sup> guanidium-hydrochlorid.

### 3. Amplicon Library Preparation

#### 3.1. DNA Quality and Quantity

The next step in the metabarcoding workflow is the preparation of the sequencing library. In this stage, several key points deserve careful consideration regardless of the sequencing platform that will be employed once the amplicon library is complete. First, the DNA template that will be used for the subsequent PCRs should be checked for its quality and quantity. The easiest way to assess DNA quality is by a spectrophotometer. Nucleic acids (DNA and RNA) absorb maximally at a wavelength of 260 nm. Protein absorbs best at 280 nm and organic compounds and chaotropic salts at 230 nm. In general, the A260/A280 ratio is used as an indicator of DNA purity, and its value should range between 1.8 and 2.0. The A260/A230 ratio is also a metric for DNA quality, and it is best if it is greater than 1.5. If these ratios are appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that may be introduced by extraction procedures and can act as PCR inhibitors. To overcome PCR inhibition due to a

low purity of the extracted DNA, preliminary PCR tests using a serial dilution of template DNA, additional purification procedures (commercial kit or manual ethanol, isopropanol, polyethylene glycol precipitation) and/or addition of PCR-enhancing or -stabilizing agents (dimethyl sulfoxide, betaine, bovine serum albumin) can be performed. As alternatives, PVP (polyvinylpyrrolidone) and 2-mercaptoethanol can be added during the cell lysis step in the DNA extraction procedure to remove negatively charged polysaccharides and polyphenols.

For accurate quantification of the extracted DNA, the use of fluorimetric determination is recommended, which utilizes fluorescent dyes that bind to double-strand DNA. This quantification method is more sensitive than using a spectrophotometer, especially when samples with low, e.g., nanomolar, DNA concentration are measured. After DNA quantification, it is recommended to standardize DNA concentrations prior to PCR, because DNA concentration might be highly variable among samples. The importance of this latter step is to have approximately the same amount of template DNA for the subsequent PCR amplification. Typically, 10–20 ng of template DNA is sufficient for amplification of ribosomal marker genes (see below for further details), but higher amounts might be required if rare gene markers are targeted [35].

### 3.2. Amplification of a Target Marker Gene

The success of DNA metabarcoding mainly depends on the selection of the appropriate DNA marker gene, which requires careful consideration. Ideally, such gene markers should have sufficiently conserved flanking primer-binding sites to minimize taxonomic bias during PCR amplification, while the intervening sequence is sufficiently variable for taxonomic identification [36]. *In silico* PCR is thus a critical step in the development of a primer in order to control for appropriate coverage of the target group (i.e., taxonomic coverage and breadth), the efficient exclusion of outgroups (i.e., taxonomic specificity) and the ability to discriminate taxa based on nucleotide variability of the amplified marker (i.e., taxonomic resolution). Integrated tools, such as TestPrime [37], are available to perform *in silico* PCR directly on a specific database (e.g., SILVA rDNA database). More generic tools that search for primers can be used on any set of reference sequences and allow for the computation of standard coverage and specificity indices, like *ecoPCR* or *cutadapt* [38,39].

Moreover, amplicon length is a critical aspect, as longer sequences will substantially increase annotation accuracy and phylogenetic resolution [40]. Amplicon libraries created for being sequenced using Illumina paired-end technology will produce amplicon sizes up to  $2 \times 300$  bp. For longer amplicons, third-generation NGS technology, such as those of Pacific Biosciences [41] and Oxford Nanopore Technologies [42], can be employed. The major advantage of third-generation NGS technology over broadly established technologies is the capability to produce ultra-long reads spanning genomic fragments measured in tens of thousands of bases [43]. At present, the benefits of the third-generation sequencing come at cost of sequencing accuracy [44]. However, Illumina technology is, so far, the most accurate technology that has been used in nearly all metabarcoding studies. It provides reads of 100 to 500 bp, which in most cases is sufficient for the analysis of typical gene markers, such as the informative regions of 16S/18S rRNA gene of prokaryotes/eukaryotes or the ITS region of fungi. Hence, we will focus only on amplicons library preparation conceived for Illumina sequencing in the next sections.

#### 3.2.1. Identification of Prokaryotes from Environmental Samples

Characterization of prokaryotic communities (Bacteria and Archaea) in environmental samples targeting regions of the 16S rRNA gene has been widely employed, unless primers have been designed to detect individual species and/or genera. 16S rRNA gene primer pairs usually target a single stretch of the hypervariable regions of the ~1500 bp prokaryote 16S rRNA gene [45,46]. Thus, the choice of the hypervariable region (V-region) targeted and the corresponding primer set should be done meticulously in order to provide coverage and accurate representation of the prokaryotic profiles in microbiota analyses [47,48]. In

this line, using suboptimal primer pairs can lead to under-representation of certain or selection against single taxa, which can lead to incorrect results and conclusions [49]. The various evaluated primer sets commonly employed to identify bacteria are listed in Table 2.

**Table 2.** Primer pairs targeting the 16S rRNA gene that have been frequently used to characterize Bacteria biodiversity in studies based on Illumina sequencing.

Primer Pair	Sequence 5'-3'	T <sub>m</sub> (°C) *	Amplified Region	Amplicon Length	Reference
515fB 806rB	GTGYCAGCMGCCGCGGTAA GGACTACNVGGGTWTCTAAT	63.6 51.2	V4	253	[50] [51]
515fB 926r	GTGYCAGCMGCCGCGGTAA CCGYCAATTYMTTTRAGTTT	63.6 48.9	V4-V5	394	[50] [52]
341f B805r	CCTACGGGAGGCAGCAG GACTACHVGGGTATCTAATCC	58.2 51.3	V3-V4	418	[37]
799f 1115r	AACMGGATTAGATACCKG AGGTTGCGCTCGTTG	50.9 56.1	V5-V6	301	[53] [54]
799f 1193r	AACMGGATTAGATACCKG ACGTCATCCCCACCTTCC	50.9 57.1	V5-V7	377	[53] [55]
967f 1391r	CAACGCGAAGAACCTTACC GACGGGCGGTGWGTRCA	53.8 59.5	V6-V8	405	[56] [57]
68f 518r	TNANACATGCAAGTCGRRCG WTTACCGCGGCTGCTG	55.5 56	V1-V3	438	[58] [59]

\* Average melting temperature as calculated with OligoAnalyzer using default parameter ([www.idtdna.com/calc/analyser](http://www.idtdna.com/calc/analyser), accessed on 13 January 2021).

Two of the most used sets of primers for soil samples are 515fB [50] and 806rB [51]. This primer pair, which was designed for use with the Illumina platform [60], is recommended for the identification of Bacteria and Archaea from soil samples by the international scientific consortium Earth Microbiome Project (EMP) [61]. However, a recent study on the performance of different Archaea-specific primers reported that the 515fB/806rB primer set performed worst for analysis of Archaea by producing only 2.1% of Archaea reads (on average) and covering only the phyla Euryarchaeota and Thaumarchaeota [62]. This suggests that the diversity of Archaea can be largely underestimated when utilizing the primers 515fB and 806rB, while the primer sets SSU1ArF/SSU520R and 340f/806rB yielded a higher sequencing coverage of the archaeal diversity using Illumina platform [62]. A list of specific primer sets to identify Archaea from soil samples is reported in Table 3.

**Table 3.** Primer pairs targeting the 16S rRNA gene that have been frequently used to characterize Archaea biodiversity in studies based on Illumina sequencing.

Primer Pair	Sequence 5'-3'	T <sub>m</sub> (°C) *	Amplified Region	Amplicon Length	Reference
515fB 806rB	GTGYCAGCMGCCGCGGTAA GGACTACNVGGGTWTCTAAT	63.6 51.2	V4	253	[50] [51]
340f 806rB	CCCTAYGGGGYGCASCAG GGACTACNVGGGTWTCTAAT	61.3 51.2	V3-V4	388	[63] [51]
SSU1ArF SSU520R	TCCGTTGATCCYGCBRG GCTACGRRYGYTTTARRC	59.2 51	V1-V4	491	[62]
349f 519r	GYGCASCAGKCGMGAAW TTACCGCGGCKGCTG	57.7 57.6	V3-V4	111	[64] [37]
Parch519f Arch915r	CAGCCGCCGCGGTAA GTGCTCCCCGCCAATTCCT	59.4 62.9	V4-V5	386	[65] [66]
1106F 1378R	TTWAGTCAGGCAACGAGC TGTGCAAGGAGCAGGGAC	52.5 57.9	V7-V8	280	[67]

\* Average melting temperature as calculated with OligoAnalyzer using default parameter ([www.idtdna.com/calc/analyser](http://www.idtdna.com/calc/analyser), accessed on 13 January 2021).

Several other primer sets have been tested and proposed as suitable candidates for the characterization of Bacteria diversity of soil samples (Table 2). For instance, a recent study [46] reported that the primer pair 341f/B805r [37], which targets the V3 to V4 region, outperformed the other three primer sets in terms of operational taxonomic unit (OTU) numbers, phylogenetic richness and Shannon diversity. The 341f/B805r primers are also recommended in the official protocol for the amplification of 16S rRNA genes released by Illumina [68].

The choice of prokaryotic primer pairs becomes more difficult when amplifying regions of the 16S rRNA gene from plant-associated samples. In this type of sample, it is crucial to reduce the amplification of non-target DNA-sequences, such as those co-extracted from plastid (mostly chloroplast) and mitochondria. Hence, the homology between bacterial 16S rRNA gene, mitochondrial and chloroplast 16S rRNA genes complicates the selection of the appropriate primers to study plant–bacteria interactions [48]. The preferred method to reduce the impact of these contaminant sequences is the use of specific mismatching primers, which amplify bacterial 16S rRNA genes while discriminating against chloroplast 16S rRNA genes. The chloroplast mismatch primer 799f [53] has been widely used in combination with the reverse primer 1193r [55] to characterize the bacterial community of plant samples, especially of roots. This primer combination has also revealed the lowest co-amplification levels of chloroplast and mitochondrial 16S rRNA gene reads among the other three bacterial primers tested [46]. It generates ~380 bp amplicons from the hypervariable region V5 to V7 of the bacterial 16S rRNA gene. Mitochondrial 16S rRNA gene amplicons with length of 800 bp are also produced, but they can be easily removed via agarose gel purification. For stem and leaf material, the primer set 799f/1115r [53] can be selected, as recommended in previous works [69,70]. These chloroplast 16S rRNA gene-discriminating primers are commonly utilized for the identification of phyllosphere associated Bacteria [71–73] because these primers do not amplify host-plant nor cyanobacterial DNA; cyanobacteria are known to be rare in the phyllosphere [74,75].

Alternative techniques, such as the use of peptide-nucleic acid (PNA) PCR-clamps [45] can be employed to reduce the co-amplification of non-target DNA sequences. PNA clamps are synthetic oligomers that bind tightly and specifically to a unique signature in the contaminant sequence and physically block its amplification [76,77]. In brief, they are designed to suppress plant host plastid and mitochondrial 16S rRNA gene contamination in the PCR reaction. For instance, the widely used primer set 515fB/806rB showed a high affinity for chloroplast 16S rRNA gene (up to 97% of the total number of reads) when used to characterize the plant-associated Bacteria from leaves and roots [78]. However, very low chloroplast co-amplification levels have been reported when this primer set is used in combination with PNA clamps [79–81], although their employment might also lead to the exclusion of certain microbial taxa [82]. It is worth mentioning that the efficacy of these approaches in reducing host-organelle 16S rRNA gene amplification significantly varies across plant species [83].

### 3.2.2. Identification of Fungi from Environmental Samples

The common marker DNA sequence used to identify fungi from soil and plant material is the internal transcribed spacer (ITS) region, which has an average length of 500 and 600 base pairs (bp) [84,85]. The ITS region includes the ITS1 and ITS2 sublocus, separated by the 5.8S rRNA gene, and it is situated between the 18S (SSU) and 28S (LSU) rRNA genes in the eukaryotic rRNA cistron [86]. The entire ITS region was described as the genetic marker with the highest probability of successful identification for a very broad range of fungi [87]. Further studies have supported the use of the ITS region as a suitable universal fungal barcode [88,89]. Consequently, most of the environmental and ecological research studies have used and are using the ITS region in combination with NGS for the identification of fungal taxa in environmental samples. Thus, large numbers of ITS sequences have been collected from terrestrial environments that are available in different reference databases, such as UNITE and GenBank (see below in Section 4.2 for more details),

making the ITS region the most ubiquitous gene marker for taxonomic characterization of fungal biodiversity.

However, with the rapid establishment of Illumina technology as the most popular sequencing platform, only short fragments can be sequenced, which constrains the choice to one of the subloci that compose the ITS region, ITS1 or ITS2 (Table 4). Therefore, the primer set selection for the characterization of fungal diversity has created a crucial and critical issue. There is some controversy on the selection of ITS markers for metabarcoding, and yet there is no consensus about which ITS sublocus is the best. Comparisons between ITS1 and ITS2 for fungal profiles have been assessed in many studies, which yielded contrasting conclusions. For example, ITS1 was thought to be more variable and hence should allow for better distinction among fungal species than ITS2 [90,91]. However, the opposite has been shown [92,93]. Nonetheless, both of these ITS regions have meaningful drawbacks and limitations in assessing fungal diversity, such as a taxonomic bias relative to the length of the amplified region, unsuitability for phylogenetic studies, co-amplification of plant DNA and exclusion of specific fungal taxonomic groups [94]. More and detailed information on the differences between the primer sets targeting the ITS1 and ITS2 regions can be found elsewhere (i.e., [95–98]).

**Table 4.** Primer pairs targeting the ITS region that have been frequently used to characterize fungal biodiversity in studies based on Illumina sequencing.

Primer Pair	Sequence 5'-3'	T <sub>m</sub> (°C) *	Amplified Region	Amplicon Length	Reference
ITS1f ITS2r	CTGGTCATTAGAGGAAGTAA GCTGCGTTCTTCATCGATGC	49.7 57	ITS1	357	[99]
ITS1F_KYO2 ITS2_KYO2	TAGAGGAAGTAAAAGTCGTAA TTYRCTRCGTTCTTCATC	48 48.4	ITS1	358	[100]
ITS3 ITS4	GCATCGATGAAGAACGCAGC TCCTCCGCTTATTGATATGC	57 52.1	ITS2	306	[99]
gITS7 ITS4 <sub>ngs</sub>	GTGARTCATCGARTCTTTG TCCTSCGCTTATTGATATGC	48.3 52.9	ITS2	288	[101] [102]
fITS7 ITS4	GTGARTCATCGAATCTTTG TCCTCCGCTTATTGATATGC	47.3 52.1	ITS2	292	[101] [99]
ITS86f ITS4	GTGAATCATCGAATCTTTGAA TCCTCCGCTTATTGATATGC	48.6 52.1	ITS2	290	[103] [99]

\* Average melting temperature as calculated with OligoAnalyzer using default parameter ([www.idtdna.com/calc/analyzer](http://www.idtdna.com/calc/analyzer), accessed on 13 January 2021).

Although the ITS region has been described, and frequently utilized, as the universal barcode for fungi [87], it has consistently demonstrated poor resolution for the arbuscular mycorrhizal fungi (AMF; phylum Glomeromycota) compared with the 18S rRNA gene (SSU markers) [104]. In Glomeromycota, species are multinucleate with extreme intraspecies divergence in nuclear ribosomal sequences, which creates additional challenges for the use of ITS for species discrimination [105]. Specifically, primer sets targeting the ITS1 sublocus have limited coverage for AMF [106], whereas recent research has highlighted that ITS2 primers can be successfully employed to characterize the most abundant AMF taxa from soil samples [107,108]. However, AMF-specific 18S rRNA gene primers might be able to amplify more families and provide a broader view of the AMF community than fungal ITS2 primers [107]. In this regard, the primer pair AMV4.5NF/AMDGR [109] is widely used to characterize fungal members affiliated with the Glomeromycota using Illumina platforms [110–112]. These primers amplify a ~258 bp fragment internal to the 18S rRNA gene. A direct comparison with other AMF-specific primers revealed that the AMV4.5NF/AMDGR outperformed the other tested primer pairs in terms of number of Glomeromycota reads (AMF specificity and coverage) [113,114]. However, these primers tended to preferentially amplify Glomeraceae at the expense of other major families (i.e., Ambisporaceae, Claroideoglomeraceae, Paraglomeraceae) of Glomeromycota [113].

Another disadvantage of the ITS region is its poor resolution for phylogenetic analysis. Diverging levels of genetic variation, due to different rates of evolution, have been observed for the three separate regions (18S rRNA gene, ITS and 28S rRNA gene) that compose the fungal nuclear ribosomal operon. The 18S rRNA gene possesses a low amount of variation among fungal taxa because it evolves slowly compared to the ITS region, which evolves the fastest and exhibits the highest variation among the three rRNA gene regions [115,116]. For phylogenetic analysis at higher taxonomic levels, such as family, order, class and phyla, former studies recommended targeting the 18S regions V1 to V5 with the primer set NS1 and NS4 [99,117]. However, these primers produce sequences of incompatible length for high-throughput sequencing, so new primers targeting the V7-V8 regions of the 18S have been proposed to target fungi in environmental samples when using Illumina sequencing [118]. These primers also have the advantage to cover well the basal fungal groups (i.e., Blastocladiomycota, Chytridiomycota, Entomophthoromycotina, Glomeromycota, Kickxellomycotina, Mucoromycota and Zoopagomycotina) when ITS primers are biased toward Dikarya. Fungal diversity could also be assessed jointly with protists using general eukaryotic primers, particularly the one targeting the V4 18S rRNA gene (see next section, e.g., [119]). The last alternative is to target the 28S rRNA gene with the primer combination LROR and LR3, with the 100 nucleotide (nt) region before the reverse primer being the best discriminant region for fungi [120]. However, this primer pair also amplifies a too-long fragment for Illumina sequencing (~600 nt), so that different strategies to shorten the reads (e.g., nested PCR, sequence fragmentation) have to be carefully investigated before routine high-throughput sequencing. Consequently, the 18S and 28S rRNA genes are more suitable for investigating the phylogenetic relationship among higher rank fungal taxa, while the ITS region can be used alone or in combination with other protein-coding genes for genus- to species-level taxonomic identification [76]. Hence, it is important to recognize and account for biases and limitations inherent to universal barcodes, especially in fungal studies, where the primer selection might have a significant impact on the taxonomic identification.

### 3.2.3. Identification of Protists from Environmental Samples

The major issue when selecting a primer pair for protists is the paraphyletic nature of this group. Protists are composed of all eukaryotic clades except Fungi, Metazoa and Embryophyta (i.e., higher plants). Except for a few protist clades that are found almost exclusively in marine environments (e.g., Diplonemea, Picozoa, Radiolaria, Telonemia, see [121]), all other clades were detected in soil, and thus only general eukaryotic primer can cover the complete biodiversity of terrestrial protists. Analogous to prokaryotes, the 18S rRNA gene has established as the standard gene for protist metabarcoding. The hypervariable regions V4 and V9 are the most commonly used, but multiple other hypervariable regions have been identified as suitable to cover the diversity of protists [122]. The EMP selected the primer pairs 1391F and EukBr targeting the V9 region for their standard protocol [123,124] while multiple other studies use slight variants with the primer 1380F/1389F and 1510R [125,126] (Table 5).

**Table 5.** Primer pairs targeting the 18S rRNA gene that have been frequently used to characterize protists biodiversity in studies based on Illumina sequencing.

Primer Pair	Sequence 5'-3'	T <sub>m</sub> (°C) *	Amplified Region	Amplicon Length	Reference
NS1/Euk20f Euk516r	GTAGTCATATGCTTGTCTC ACCAGACTTGCCCTCC	47.2 54.3	V1-V3	507	[99,127] [128]
18S_0067a_deg NSR 399	AAGCCATGCATGYCTAAGTATMA TCTCAGGCTCCYTCTCCGG	54.4 59.7	V1-V3	310	[129]
fw_366 rv_586	ATTAGGGTTCGATTCCGGAGAGG CTGGAATTACCGCGGCTGCTG	58.2 61	V3	180	[130]
TAREuk454FWD1/ V4_1f TAREukREV3	CCAGCASCYCGGTAATTCC/ CCAGCASCYCGGTAATWCC ACTTTCGTTCTTGATYRA	60.1/59.9 45.9	V4	391	[131] [132]
616*f 1132r	TTAAARVGYTCGTAGTYG CCGTCAATTHCTTYAART	47.1 45.4	V4-V5	504	[133]
18S_allshorts-f 18S_allshort-r	TTTGCTGSITTAATTSCG TCACAGACCTGTTATTGC	47.7 49.4	V7	109	[134]
V8f 1510R	ATAACAGGTCTGTGATGCCCT CCTTCYGCAGGTTACCTAC	55.9 56.6	V8-V9	339	[135] [125]
1380F/1389F 1510R	CCCTGCCHTTTGTACACAC/ TTGTACACACCGCCC CCTTCYGCAGGTTACCTAC	54.6/51.9 56.6	V9	141/136	[125]
1391F EukBr	GTACACACCGCCCGTC TGATCCTTCTGCAGGTTACCTAC	56.1 58.4	V9	127	[123] [124]

\* Average melting temperature as calculated with OligoAnalyzer using default parameter ([www.idtdna.com/calc/analyzer](http://www.idtdna.com/calc/analyzer), accessed on 13 January 2021).

In parallel, the V4 region has also been established as an equally powerful region to resolve protist diversity when amplified with the TAREuk primer pair [132,136]. Other primer pairs have been designed to target V1-V3, V4-V5 and V7 regions, and they cover the biodiversity of protist clades well (see Table 5). However, no comparison has been thoroughly conducted of the performances of these primer pairs on terrestrial samples, and only in silico studies are available comparing them with the bias of database completeness for each region [18,121,122]. Moreover, considering that the Illumina sequencing of 2 × 300 bp now delivers almost identical quality to the 2 × 150 bp variant, a promising combination of primer amplifying 400 to 500 nt spanning regions can be tested like, for example, the V7 to V9 regions. The same primers have been used to study plant-associated protists. This is, for example, the case of *Sphagnum* and peatland-mosses-associated protists for which both V4 (TAREuk) and V9 (1380F/1510R) primers have been used [137,138]. Both V4 (V4\_1f/TAREukREV3) and V9 (1380F/1510R) primers have also been employed to study rhizospheric protists [139,140]. Although plant sequences could represent the majority of reads in such plant-associated protist metabarcoding datasets, strategies to reduce the co-amplification of the associated plant(s), for example, the utilization of blocking oligos, have not yet been implemented. Furthermore, the use of general eukaryotic primers can come at the cost of reduced taxonomic coverage, which is not limited anymore by the primers and sequencing depth but by the competition between all target DNA during the PCR amplification. Indeed, specific primers have been shown to cover two to three times more diversity than general eukaryotic primers [141]. Likewise, clades often under-represented in general eukaryotic datasets, like Myxomycetes, can be recovered with clade-specific primers [142]. Lists of clade-specific primer pairs targeting either the same gene (18S) or other genes (e.g., 28S, ITS, COI, *rbcL*) are provided elsewhere [143,144].

### 3.3. Further Recommendations for Library Preparation

Once a proper primer pair has been selected, the library preparation workflow should be checked and evaluated for its compatibility with the chosen sequencing platform. In

the case of Illumina sequencing technology, adaptor sequences and short barcodes must be added to the target gene primer sequence to enable the sequencing of many samples in parallel. This can be achieved with three different approaches. The Illumina standard workflow recommends a two-steps procedure in which the template is first amplified with the target gene primers that include the Illumina's adaptors, while barcodes are added in a second PCR [68]. The second procedure involves only a single PCR step, in which the primers already incorporate the barcodes and adaptors [60]. This latter approach is used and recommended by the Earth Microbiome Project [61]. The third alternative is to perform the first PCR as for the Illumina standard workflow, and then to use a ligation-based kit, originally developed for shotgun sequencing, in order to reduce cost and avoid potential cross-contamination during the second PCR [145]. For this third approach, it is important to note that different steps in the ligation protocol (e.g., blunt ending, post-ligation PCR) can considerably increase the amount of tag-jump (sequencing outputs with false forward and reverse combinations of used tags) when pooling multiple tagged amplicons in the same library, and that adaptation of original kit protocol is necessary [146].

Several other factors related to library preparation and sequencing technology can significantly influence the accuracy of the metabarcoding procedure. For example, it is advisable to perform technical replicates for each sample during the PCR step and subsequently pool them before sequencing. This procedure allows one to minimize PCR-introduced biases on relative abundance and to efficiently saturate the diversity estimates of soil microbes [147]. To further reduce primer bias in the amplification process, it is important to determine the optimal annealing temperature for the primer pair chosen to avoid the formation of unspecific products. The optimal annealing temperature was found to be a function of the melting temperatures of the primers [148], and it should be determined empirically usually using the gradient PCR method. The use of proofreading DNA polymerases is strongly recommended to reduce chimera formation during PCR amplification, which may result in an overestimation of community richness [149].

Another important argument to consider is that Illumina sequencing platforms are known to cause biases when sequencing DNA libraries with low gene diversity, such as samples containing exclusively 16S rRNA gene or ITS amplicons [49,150]. To artificially increase sequence diversity, especially in the primer region, the addition of genomic DNA from the phage PhiX to the amplicon library is a common procedure. On the other hand, this results in a loss of sequence recovery because between 5 and 50% of the capacity of an Illumina sequencing run may have to be allocated to PhiX DNA sequencing. However, the amount of PhiX DNA to be used varies between Illumina platforms [151]. Alternatively, the design of heterogeneity spacers, short sequences of 1–7 bp linked to index adaptors or the gene-specific primers, can be utilized to reduce the amount of PhiX DNA added to amplicon library pools to create the base diversity needed [152,153]. However, designing index adaptors or primers comprising different variable-length sequences can be a complicated and challenging approach with additional technical limitations [154]. However, this approach has been tested for multiple targets and allowed for an increased reads recovery and increased base quality at the 3' end [155–158]. Another possibility to increase the base diversity is to sequence multiple targets in the same sequencing run (i.e., 16S, 18S and ITS gene libraries of the same samples), which is pertinent in research projects interested in multiple target taxa but should be restricted to marker gene of comparable length.

The addition of negative controls is needed in order to estimate potential contamination during the DNA extraction and PCR preparation. It is thus recommended to use negative controls during each DNA extraction and each PCR preparation [8]. For DNA extraction, soil or plant material can be replaced with sterile water to create the negative control. This extracted material will then be used in PCR as a template to control for contamination during the DNA extraction. PCR negative controls use sterile water to replace DNA template in PCR in order to check for contamination during the PCR preparation. Even if no bands are visible on agarose gels for these negative controls, it is necessary to in-

clude them in the sequencing pool in order to detect potential low abundant contaminants. Sequences assigned to a PCR negative control need to be removed from any other sample from which the DNA was PCR-amplified together with this control. A particular case may arise when using double-tagging, as tag-jump could potentially produce sequences with an unused combination of tags by recombination of sequences from different samples in the sequencing pool. In such a situation, sequences assigned to negative controls by their tags could originate from other original samples and would thus contain a set of sequences mainly composed of the most abundant sequences found in the other samples sharing the same forward or reverse tag. Consequently, double tagging has to be used with caution, and multiple approaches have been developed to mitigate this issue [155,159].

The addition of mock communities (DNA pools of multiple known species) or positive controls (single-species DNA) into run libraries is also a common practice that can be helpful to (i) assess the primer bias and error rate of the sequenced run, (ii) benchmark bioinformatic tools, (iii) control for false positive in the case of tag-jumping, (iv) determine a relative abundance threshold to remove putative artifact out and (v) correct for compositional bias in case of differential abundance analyses. Initial Illumina MiSeq metabarcoding studies combining error rate estimates and bioinformatic tool benchmarking were based on sequencing bacterial, fungal and protists mock communities [135,160,161]. In general, mock communities are needed to validate new molecular (e.g., primer evaluation) and bioinformatic (e.g., sequence grouping algorithm) methods but are not crucial to analyze samples with established methodologies. Mock or positive controls can also be used to determine a threshold below which an OTU can be considered as an artifact. This threshold can be either a fixed number of reads [142] or a per-sample relative abundance when multiple positive controls were sequenced [162]. Most recent studies advocate for the use of separate or spike-in mock communities in order to use the recovered relative abundance of the known mixed species to apply a correction factor to a sample's relative abundances [163]. This approach appears to be particularly crucial in differential abundance analyses when taking into account the compositional bias of amplicon sequencing data [164,165].

## 4. Bioinformatic Processing

### 4.1. Pre-Processing of the Metabarcoding Dataset

The typical metabarcoding bioinformatics pipeline consists of several steps, including (i) the demultiplexing of barcoded samples, (ii) pair-end assembly, (iii) removal of chimeric reads, (iv) quality filtering, (v) sequence grouping and (vi) comparison of the representative sequences to a reference database (Figure 1). QIIME and MOTHUR are the most-used platforms to perform bioinformatic analyses of metabarcoding data [166,167]. These software pipelines provide the capability to customize the analysis of high-throughput metabarcoding data using a wide choice of tools. However, many other pipelines and bioinformatics tools have been developed for the processing of amplicon sequencing data, such as PEMA [168], PipeCraft [169], SLIM [170], BioMas/Galaxy [171], PIPITS [172] USEARCH [173], VSEARCH [174], OBITools [175] and DADA2 [176]. Most of the above-mentioned platforms and pipelines are particularly well-suited for beginners in the field because they provide smooth wrappers around commonly used command-line tools as well as well-documented tutorials and examples [177]. It is important to note that some equivalent tools have been preferred in the analyses of certain target genes due to preference among the scientific communities, but most of them can be used for any metabarcoding target.

After the filtering and quality procedures, a key step in the bioinformatics analysis workflow is the clustering of reads based on their homology. Traditionally, during clustering, reads sharing a predefined level of similarity (generally between 95% and 99%) are assembled into Operational Taxonomic Units (OTUs) [178]. This step is intended to eliminate erroneous sequences produced by PCR and sequencing errors [18] as well as to merge intraspecific variance on diverging alleles or gene copies. However, such a global OTU clustering approach has several limitations [179]. For example, the 97% similarity cut-off used for V4 16S is to a large degree arbitrary, since different taxa might differ by

a small percentage in their nucleotide sequence but still represent ecologically distinct clades [180,181]. In other words, there might be the risk that multiple similar species can be grouped into one single OTU with their true individual identifications being lost, while on the other hand, reads of a unique species may end up in different OTUs when the intra-specific variability is high. Other disadvantages of this method are associated with (i) the addition of data outputs, such as OTUs, that exclusively consist of PCR amplification or sequencing errors and (ii) the biologically meaningful interpretations/annotations of the inferred OTUs [181].

Recently, novel methods that use either single-linkage local clustering or error model correction algorithms have been developed to produce high-resolution representative sequences independently from a determined similarity threshold. The first approach was developed in the tool Swarm [182,183]. It allowed tackling the main issue of arbitrary similarity threshold of the global clustering approach. Swarm has allowed better discrimination of reads from closely related species, which is acknowledged by its wide adoption in the analysis of 18S rRNA metabarcoding datasets [136,184]. The second approach is called oligotyping [185] and is now mainly computed using the algorithm DADA2 [176]. DADA2 has been developed to control errors sufficiently to produce amplicon sequence variants (ASVs) that can be resolved exactly, down to the level of single-nucleotide differences over the sequenced region. This approach avoids clustering sequences at an arbitrarily defined similarity threshold (e.g., 97%) and instead uses only unique, identical sequences for downstream community analyses. Furthermore, because ASVs are exact sequences generated without clustering or reference databases, ASVs output can be readily compared between studies using the same target region and the same primers [186]. Several studies have reported that ASV-level pipelines allow for easier inter-study integration of biological features, as ASVs have intrinsic biological meaning, independent of reference database or study context [187–189]. The ASVs approach has also been described as being more effective than OTU clustering for recovering richness and composition of fungal [190] and bacterial [191] communities from environmental samples. Indeed, the DADA2 algorithm has shown to find more ASVs than other denoising pipelines when analyzing sequencing data from soil datasets, suggesting that it could be better at finding rare organisms, but at the expense of possible false positives [192]. For the aforementioned reasons, most of the recent metabarcoding studies on bacterial and fungal microbiota associated with soil and plant material have chosen ASVs over OTUs [193–198]. However, fungal and bacterial diversity patterns appear to be equally well described by both OTU and ASV, which does not appear to change the conclusion on alpha and beta diversity analyses over contrasted samples along elevation gradients [199].

For a meaningful interpretation and reliable analysis of amplicon sequencing data, after the OTU/ASVs generation stage, additional steps should be considered. Primarily, post-clustering algorithm should be used when a high amount of artefactual sequence variants are suspected [200,201]. Then, an adequate coverage, in terms of sequencing depth, is crucial to generate reliable information on the composition and taxonomic structure of the microbial community investigated. Rarefaction and accumulation curves can provide useful information to assess whether the sequencing depth yielded sufficient reads to describe most of the diversity in the samples. For example, if the coverage per sample is too low, the diversity of the microbiota being studied is likely to be underrepresented, as rarer members of the microbiota are less likely to be detected [19]. In general, a satisfactory coverage can be achieved with 10,000 to 100,000 sequences per sample, but it largely depends on the complexity of the microbiota, type of starting material (soil or plant), the targeted gene and the desired resolution [35].

Additional filtering steps will increase the quality and resolution of the output dataset. For example, the exclusion of rare OTUs or ASVs, which may be sequencing artifacts, is commonly recommended [202]. However, there is no consensus on the threshold number of sequences below which an OTU/ASV can be considered rare [190]. The suggested thresholds might range from 1 to 10 sequences [203] or depend on the relative abundance

of OTUs/ASVs [160]. Another filtering option is to remove OTUs/ASVs that have been detected solely in one or a few samples from a single sequencing run, but such an approach strictly depends on the number of samples that constitute the entire dataset and if multiple sequencing runs were used.

#### 4.2. Taxonomic Profiling

The taxonomic annotation of the OTUs/ASVs identified is the last step of the metabarcoding workflow. It provides valuable information on the OTUs/ASVs in light of what is known about these taxa from previous works, and, more broadly, it allows comparison across microbiota studies [18]. Essentially, the taxonomical identification of microbes relies on sequence similarity searches in reference databases. It is noteworthy that taxonomy assignment based on different reference databases might lead to different results [204]. So far, there is no consensus on which reference database to use for taxonomic assignment of the detected OTUs/ASVs. In this section, we report the most common options utilized by bioinformaticians and microbial ecologists.

Reference databases for 16S rRNA gene taxonomy assignment include SILVA [205], the Ribosomal Database Project (RDP) [206], Greengenes [207] and the National Center for Biotechnology Information (NCBI) [208]. Since all these databases are widely used for taxonomical identification of prokaryotic sequences, we provide here a quick overview of each of them (Table 6).

**Table 6.** List of the main reference databases used for the taxonomic annotation of the representative sequences in metabarcoding studies of terrestrial microbial communities.

Database/Release	Marker/Taxa	URL *	Reference
SILVA/138.1	16S, 18S SSU, 23S, 28S, LSU rRNA sequences/Archaea, Prokaryotes, Eukaryotes	<a href="http://www.arb-silva.de">www.arb-silva.de</a>	[205]
Ribosomal Database Project (RDP)/11	16S, 28S rRNA sequences/Prokaryotes, Archaea and Fungi	<a href="http://rdp.cme.msu.edu">rdp.cme.msu.edu</a>	[207]
Greengenes/12_10	16S rRNA sequences/Archaea and Bacteria	<a href="http://greengenes.secondgenome.com">greengenes.secondgenome.com</a>	[207]
National Center for Biotechnology Information (NCBI) GenBank/241.0	raw sequences/Archaea, Prokaryotes, Eukaryotes	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a>	[208]
UNITE/8.2	nuclear ribosomal ITS region sequences/Eukaryotes	<a href="http://unite.ut.ee">unite.ut.ee</a>	[209],
Protist Reference Database (PR2)/4.12.0	18S rRNA sequences/Eukaryotes	<a href="http://pr2-database.org">pr2-database.org</a>	[210]

\*, accessed on 13 January 2021.

The SILVA database provides a phylogenetic classification for the small and large rRNA subunits for Bacteria, Archaea and Eukarya in the European Nucleotide Archive (ENA) [211]. It is based primarily on phylogenies for small subunit rRNAs (16S rRNA gene for prokaryotes and 18S rRNA gene for eukaryotes), and its taxonomic rank assignment is manually curated. To date, the last SILVA database update was on 27.08.2020 with the 138.1 release. Interestingly, the QIIME2 platform makes available pre-formatted SILVA reference databases to QIIME2 users in order to provide a fast and standardized workflow in the taxonomy assignment step. The RDP database also contains rRNA sequences from the three domains, but it provides primarily phylogenetic classification for prokaryotic organisms. It contains sequences available from the International Nucleotide Sequence Database Collaboration (INSDC) [212]. The RDP classifier was updated to version 2.13, which was released on 30 July 2020. Greengenes is a database that provides a phylo-

genetic classification of prokaryotic organisms, and most of the sequences are retrieved from the NCBI GenBank [213]. The last update of the Greengenes database occurred on 5 January 2019. The NCBI taxonomy database contains the names of all organisms associated with submissions to the NCBI sequence databases. Specifically, the NCBI Taxonomy database is the standard nomenclature and classification repository for the International Nucleotide Sequence Database Collaboration (INSDC), comprising the GenBank, European Nucleotide Archive (ENA) and DNA Data Bank of Japan (DDBJ) databases [208].

For the taxonomic identification of fungi, three main reference ITS databases spanning the fungal kingdom are available: UNITE [209], Warcup ITS [214] and RDP. Among them, UNITE is considered as the main reference ITS database for the identification of fungi. It represents a middle ground between including the very latest sequences and offering detailed taxonomic annotation [95]. Indeed, UNITE clusters the ITS sequences at different sequence similarity thresholds to obtain approximate species-level OTUs referred to as species hypotheses (SHs) [215]. These SHs (458,797 as of August 2018) have a unique digital object identifier (DOI) to allow stable, unambiguous reference across studies [216]. Its last update was on 20 February 2020 with the release version 8.2. It is worth noting the existence of two ITS reference databases sequences associated with a specific ITS sublocus. This is the case of ITSoneDB [217], which is a curated collection of eukaryotic ITS1 sequences, and the ITS2 Database [218], which is a eukaryotic ITS2 database.

Other reference databases for fungal annotation are used if the target marker gene amplified via PCR differs from the ITS region, such as LSU or SSU regions of the fungal rRNA gene. In this case, SILVA, RDP and NCBI databases are ubiquitously employed. Interestingly, for the specific taxonomic classification of fungal taxa affiliated to the phylum Glomeromycota, the MaarjAM database [219] was created in 2010. This database associates information about geography, habitat and climate to Glomeromycota sequences, which cluster in “Virtual Taxa”, a proxy for fungal species [220]. The MaarjAM database is manually curated, and its last update occurred on 5 June 2019.

The main reference database for the eukaryotic 18S rRNA gene is the Protist Reference Database (PR2; [210], now accessible at <https://github.com/pr2database/pr2database>, accessed on 13 January 2021). It is a curated reference 18S sequence collection that follows the most up-to-date higher ranks taxonomic classification of eukaryotes [143]. The classification is provided in a fixed eight-rank taxonomy, which eases the statistical analyses. The last version is 4.12.0 from 8 August 2019. Alternatively, the SSU Ref NR 99 SILVA reference database can also be used, which can be particularly interesting when using the aligned version of the database.

Overall, the selection and availability of curated reference databases are crucial to characterize on a large scale the taxonomic complexity of microbiota from various environments through metabarcoding.

## 5. Importance of Metadata Standards and Archiving Practices

As DNA metabarcoding has become a routine approach for the characterization of microbial communities across different environments, in recent years a surge in the volume of the sequences archived in public genetic repositories has been recorded [221]. Presently, the deposition of sequencing data in genetic databases has become standard practice, mainly because it is a more frequent requirement for the publication of studies in peer-reviewed journals. The electronic archiving of sequencing data is primarily centralized in three public genetic databases that are routinely synchronized and members of the INSDC: NCBI’s Sequence Read Archive (SRA), the EBI’s European Nucleotide Archive (ENA) and DDJ’s Sequence Read Archives (DRA) [212]. These archives represent an invaluable resource as they create a window of opportunity for data reuse and synthesis in microbiome research. Therefore, it is crucial that the sequencing data are correctly uploaded and made available in public genetic repositories with appropriate formatting and metadata to allow others to reuse them. The standardization of protocols and metadata collection, alongside a simple and straightforward process of data storage, accessibility and

sharing, is vital for ensuring that microbiome data are findable, accessible, interoperable and reusable (FAIR) [222].

Several research groups and consortiums have pioneered and coordinated the generation of community-driven standards for collecting and managing relevant contextual information associated with genomic data. So far, the minimum information standards (MIxS: minimum information about any (x) sequence) established by the Genomic Standards Consortium (GSC) [223] is the most accepted and adopted initiative by the public genetic databases in order to provide rich information on the uploaded sequences [224]. The MIxS standards consist of checklists for describing minimum information about marker genes (MIMARKS), genomes (MIGS) and metagenomes (MIMS), and of 15 different environmental packages that can be used to specify the environmental context of a sequenced microbial community, particularly for soil and plant-associated samples [225]. In parallel, MIMARKS standards have been developed by GSC for reporting information about metabarcoding studies [226], and the MIMARKS checklist is provided on the GSC website (<https://gensc.org/mixs/>, accessed on 13 January 2021). The implementation of this checklist alongside the sequencing data is fundamental to facilitate the ability to retrieve appropriate contextual information for marker genes, frequently referred to as “metadata”, enabling the reusability and sharing of the sequencing data to allow for reproducibility, meta-analyses and cross-comparison among studies.

Although many efforts have been made to demonstrate and promote the importance of having systematic reporting conventions and standards to accurately describe any chosen workflow, a recent study on the deposited sequencing data of 26,927 microbial studies published between January 2015 and March 2019 showed gaps in the availability and reusability of these data [227]. The authors of this study identified the lack of metadata, improper file formatting and data deposition to inappropriate repositories as the main causes of data loss. In particular, the lack or the incorrect information reported in the metadata, which includes all information concerning the description of the sample, sample processing, experimental design, library creation and sequencing platform configuration, represents a common issue that hinders the reusability of the sequencing data available in genetic databases. In light of these findings, we would like to emphasize the importance of improving data archiving practices to enhance the value of the sequencing data in repurposing and better sharing of microbial datasets.

## 6. Future Perspective and Challenges

Within the past decade, metabarcoding has become the gold standard for the characterization of complex microbial communities associated with environmental samples. Although this approach may not successfully identify all the taxa in a sample, the output generated by a proper metabarcoding workflow provides reliable information for adequate biological inferences. However, generating accurate and verifiable data, such as biodiversity estimates and taxonomic assignment, requires robust methods and generally accepted standards [228]. So far, metabarcoding workflows have relied primarily on Illumina sequencing technology, which constrains the length of the amplicons to a maximum of 600 bp. This represents a considerable limitation in terms of taxonomic resolution for many bacterial and fungal taxa, as the taxonomic assignment of short-reads at the species or even genus level is often elusive. Third-generation sequencing technologies, such as the MinION and PromethION platform from Oxford Nanopore Technologies (ONT) or PacBio from Pacific Biosciences, are emerging as promising sequencing systems to overcome many of the limitations of short-read sequencing. Considering that ONT technology allows for the design of primers covering the whole length of the 16S rRNA gene or ITS region, it is then plausible to conceive a better phylogenetic inference and higher taxonomic resolution in microbial ecology studies. However, despite the apparent potential advantages of the application of ONT technology in metabarcoding, there are still several factors limiting its implementation in microbial ecology research. For instance, there is only a limited number of bioinformatic tools and protocols designed for the specific analysis of long reads. Thus,

it is challenging to carry out a specialized taxonomic analysis compared with previous sequencing technologies [229]. Another major drawback of this technology is the high read error rates, which hampers accurate read classification [230]. Furthermore, it is a relatively novel technology for which standards are still largely absent, thus complicating the standardization and reproducibility of results [231].

Other methodological approaches can also be employed in the characterization of complex microbiota from environmental samples. Metagenomics, or the shotgun sequencing technique, which refers to the recovery and sequencing of the collective genomic material in environmental samples, are largely used to investigate the functional complement of the microbiota as a whole. Nonetheless, the data output generated by this approach can also be utilized for taxonomic profiling. A significant advantage of metagenomics over metabarcoding is that metagenomic approaches do not rely on the amplification of specific genomic sequences, avoiding all the bias introduced by PCR procedures. However, important drawbacks are associated with shotgun sequencing in biodiversity studies. The efficiency of shotgun metagenomics is mainly constrained to adequate read depths in order to obtain accurate results, which can be difficult to achieve from complex samples like soil. Hence, huge increases in sequencing power to acquire adequate sequencing depth often result in prohibitive costs. Another main disadvantage associated with shotgun metagenomics is the lack of curated reference databases of bacterial and fungal genomes. Specifically, fungal and protist genome databases are rare at present and, in particular, compared with bacterial genome databases [95,232]. As a result, the proportion of sequences identified as fungal is low even in metagenomes with high fungal abundance, such as topsoil metagenomes [233]. Lastly, challenges and difficulties frequently occur in analyzing metagenomics datasets because of the extensive filtering that is required as a result of the sequencing of all sampled DNA. This leads to datasets of significantly larger orders of magnitude compared to the ones produced by metabarcoding approaches. Consequently, analyses of shotgun metagenomics data take much longer to perform and require far more computational power and expertise.

Capture by hybridization also represents a promising approach for the enrichment of a target gene as an alternative to PCR amplification [234]. It has the advantage of allowing the use of multiple probes annealing to the target gene and allows the conservation of long DNA fragments, which is suitable for third-generation high-throughput sequencing. This novel technique also has the potential to unravel new hidden diversity missed by the traditional PCR approach [235].

In conclusion, DNA metabarcoding represents a powerful approach to explore the microbial biodiversity of environmental samples. With further technological advances, procedure optimization and refinement, metabarcoding will likely emerge as a fundamental tool for several scientific tasks not only in biodiversity monitoring in terrestrial environments but also in other research and application areas such as diet analysis, air, water and food quality testing and monitoring [15]. Moreover, the future of DNA metabarcoding deeply relies on the quality and completeness of reference sequence databases, which should be also designed and further curated to allow efficient data mining and report generation. Finally, we believe that the combination of different sequencing methodologies, such as DNA metabarcoding and metagenomics, together with gene expression, including metatranscriptomics, stable isotope labeling and canonical cultivation and enrichment techniques, represents the best approach to open the soil black box in order to unravel the complex dynamics of the soil–plant–microbe system and to get further insight into soil microbial functions on the level of complex terrestrial microbiota.

**Author Contributions:** Conceptualization, D.F. and S.K.; methodology, D.F., G.L. and S.L.; validation, D.F., G.L. and S.K.; resources, D.F., G.L. and S.L.; writing—original draft preparation, D.F.; writing—review and editing, D.F., G.L. and S.K.; visualization, D.F. and S.L.; supervision, D.F. and S.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** The study was funded by the Leibniz Competition Program project “Volcorn—Volatilome of a Cereal Crop Microbiota Complex under drought and Flooding” (K102/2018) (Leibniz Association). The work of G.L. was supported by a grant from the Swiss National Science Foundation (project number: 182531).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Nannipieri, P.; Ascher-Jenull, J.; Ceccherini, M.T.; Pietramellara, G.; Renella, G.; Schloter, M. Beyond microbial diversity for predicting soil functions: A mini review. *Pedosphere* **2020**, *30*, 5–17. [\[CrossRef\]](#)
2. Francioli, D.; Schulz, E.; Buscot, F.; Reitz, T. Dynamics of Soil Bacterial Communities Over a Vegetation Season Relate to Both Soil Nutrient Status and Plant Growth Phenology. *Microb. Ecol.* **2018**, *75*, 216–227. [\[CrossRef\]](#)
3. Francioli, D.; Schulz, E.; Purahong, W.; Buscot, F.; Reitz, T. Reinoculation elucidates mechanisms of bacterial community assembly in soil and reveals undetected microbes. *Biol. Fertil. Soils* **2016**, *52*, 1073–1083. [\[CrossRef\]](#)
4. Cortois, R.; De Deyn, G.B. The curse of the black box. *Plant Soil* **2012**, *350*, 27–33. [\[CrossRef\]](#)
5. Delmont, T.O.; Francioli, D.; Jacquesson, S.; Laoudi, S.; Mathieu, A.; Nesme, J.; Ceccherini, M.T.; Nannipieri, P.; Simonet, P.; Vogel, T.M. Microbial community development and unseen diversity recovery in inoculated sterile soil. *Biol. Fertil. Soils* **2014**, *50*, 1069–1076. [\[CrossRef\]](#)
6. Caron, D.A.; Worden, A.Z.; Countway, P.D.; Demir, E.; Heidelberg, K.B. Protists are microbes too: A perspective. *ISME J.* **2009**, *3*, 4–12. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Geisen, S.; Mitchell, E.A.D.; Adl, S.; Bonkowski, M.; Dunthorn, M.; Ekelund, F.; Fernández, L.D.; Jousset, A.; Krashevskaya, V.; Singer, D.; et al. Soil protists: A fertile frontier in soil biology research. *FEMS Microbiol. Rev.* **2018**, *42*, 293–323. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Taberlet, P.; Coissac, E.; Pompanon, F.; Brochmann, C.; Willerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **2012**, *21*, 2045–2050. [\[CrossRef\]](#)
9. Giampaoli, S.; Berti, A.; Di Maggio, R.M.; Pili, E.; Valentini, A.; Valeriani, F.; Gianfranceschi, G.; Barni, F.; Ripani, L.; Romano Spica, V. The environmental biological signature: NGS profiling for forensic comparison of soils. *Forensic Sci. Int.* **2014**, *240*, 41–47. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Szelez, I.; Lösch, S.; Seppely, C.V.W.; Lara, E.; Singer, D.; Sorge, F.; Tschui, J.; Perotti, M.A.; Mitchell, E.A.D. Comparative analysis of bones, mites, soil chemistry, nematodes and soil micro-eukaryotes from a suspected homicide to estimate the post-mortem interval. *Sci. Rep.* **2018**, *8*, 25. [\[CrossRef\]](#) [\[PubMed\]](#)
11. van der Heyde, M.; Bunce, M.; Dixon, K.; Wardell-Johnson, G.; White, N.E.; Nevill, P. Changes in soil microbial communities in post mine ecological restoration: Implications for monitoring using high throughput DNA sequencing. *Sci. Total Environ.* **2020**, *749*, 142262. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Vischetti, C.; Casucci, C.; De Bernardi, A.; Monaci, E.; Tiano, L.; Marcheggiani, F.; Ciani, M.; Comitini, F.; Marini, E.; Taskin, E.; et al. Sub-Lethal Effects of Pesticides on the DNA of Soil Organisms as Early Ecotoxicological Biomarkers. *Front. Microbiol.* **2020**, *11*. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Inderbitzin, P.; Robbertse, B.; Schoch, C.L. Species Identification in Plant-Associated Prokaryotes and Fungi Using DNA. *Phytobiomes J.* **2020**, *4*, 103–114. [\[CrossRef\]](#)
14. Poretsky, R.; Rodriguez-R, L.M.; Luo, C.; Tsementzi, D.; Konstantinidis, K.T. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE* **2014**, *9*, e93827. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Ruppert, K.M.; Kline, R.J.; Rahman, M.S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **2019**, *17*, e00547. [\[CrossRef\]](#)
16. van Ruijven, J.; Ampt, E.; Francioli, D.; Mommer, L. Do soil-borne fungal pathogens mediate plant diversity–productivity relationships? Evidence and future opportunities. *J. Ecol.* **2020**, *108*, 1810–1821. [\[CrossRef\]](#)
17. Zinger, L.; Bonin, A.; Alsos, I.G.; Bálint, M.; Bik, H.; Boyer, F.; Chariton, A.A.; Creer, S.; Coissac, E.; Deagle, B.E.; et al. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Mol. Ecol.* **2019**, *28*, 1857–1862. [\[CrossRef\]](#)
18. Hugerth, L.W.; Andersson, A.F. Analysing Microbial Community Composition through Amplicon Sequencing: From Sampling to Hypothesis Testing. *Front. Microbiol.* **2017**, *8*, 1561. [\[CrossRef\]](#)
19. Pollock, J.; Glendinning, L.; Wisedchanwet, T.; Watson, M. The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies. *Appl. Environ. Microbiol.* **2018**, *84*, e02627. [\[CrossRef\]](#)
20. Černohlávková, J.; Jarkovský, J.; Nešporová, M.; Hofman, J. Variability of soil microbial properties: Effects of sampling, handling and storage. *Ecotoxicol. Environ. Saf.* **2009**, *72*, 2102–2108. [\[CrossRef\]](#)
21. Öhlinger, R. Soil Sampling and Sample Preparation. In *Methods in Soil Biology*; Schinner, F., Öhlinger, R., Kandeler, E., Margesin, R., Eds.; Springer: Berlin/Heidelberg, Germany, 1996; pp. 7–11. [\[CrossRef\]](#)
22. Griffiths, R.I.; Whiteley, A.S.; O’Donnell, A.G.; Bailey, M.J. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **2000**, *66*, 5488–5491. [\[CrossRef\]](#)

23. Lakay, F.M.; Botha, A.; Prior, B.A. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J. Appl. Microbiol.* **2007**, *102*, 265–273. [[CrossRef](#)] [[PubMed](#)]
24. Pawlowski, J.; Apothéoz-Perret-Gentil, L.; Altermatt, F. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Mol. Ecol.* **2020**, *29*, 4258–4264. [[CrossRef](#)] [[PubMed](#)]
25. Ceccherini, M.T.; Ascher, J.; Agnelli, A.; Borgogni, F.; Pantani, O.L.; Pietramellara, G. Experimental discrimination and molecular characterization of the extracellular soil DNA fraction. *Antonie Van Leeuwenhoek* **2009**, *96*, 653–657. [[CrossRef](#)] [[PubMed](#)]
26. Taberlet, P.; Prud'Homme, S.M.; Campione, E.; Roy, J.; Miquel, C.; Shehzad, W.; Gielly, L.; Rioux, D.; Choler, P.; Clément, J.C.; et al. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.* **2012**, *21*, 1816–1820. [[CrossRef](#)] [[PubMed](#)]
27. Courtois, S.; Frostegård, Å.; Göransson, P.; Depret, G.; Jeannin, P.; Simonet, P. Quantification of bacterial subgroups in soil: Comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ. Microbiol.* **2001**, *3*, 431–439. [[CrossRef](#)] [[PubMed](#)]
28. Holmsgaard, P.N.; Norman, A.; Hede, S.C.; Poulsen, P.H.B.; Al-Soud, W.A.; Hansen, L.H.; Sørensen, S.J. Bias in bacterial diversity as a result of Nycodenz extraction from bulk soil. *Soil Biol. Biochem.* **2011**, *43*, 2152–2159. [[CrossRef](#)]
29. Eichorst, S.A.; Strasser, F.; Woyke, T.; Schintlmeister, A.; Wagner, M.; Wobken, D. Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. *Fems Microbiol. Ecol.* **2015**, *91*. [[CrossRef](#)]
30. Lentendu, G.; Hübschmann, T.; Müller, S.; Dunker, S.; Buscot, F.; Wilhelm, C. Recovery of soil unicellular eukaryotes: An efficiency and activity analysis on the single cell level. *J. Microbiol. Methods* **2013**, *95*, 463–469. [[CrossRef](#)]
31. Sharma, S.; Mehta, R.; Gupta, R.; Schloter, M. Improved protocol for the extraction of bacterial mRNA from soils. *J. Microbiol. Methods* **2012**, *91*, 62–64. [[CrossRef](#)]
32. Lim, N.Y.N.; Roco, C.A.; Frostegård, Å. Transparent DNA/RNA Co-extraction Workflow Protocol Suitable for Inhibitor-Rich Environmental Samples That Focuses on Complete DNA Removal for Transcriptomic Analyses. *Front. Microbiol.* **2016**, *7*. [[CrossRef](#)]
33. Lever, M.A.; Torti, A.; Eickenbusch, P.; Michaud, A.B.; Šantl-Temkiv, T.; Jørgensen, B.B. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front. Microbiol.* **2015**, *6*. [[CrossRef](#)]
34. Alawi, M.; Schneider, B.; Kallmeyer, J. A procedure for separate recovery of extra- and intracellular DNA from a single marine sediment sample. *J. Microbiol. Methods* **2014**, *104*, 36–42. [[CrossRef](#)] [[PubMed](#)]
35. Schöler, A.; Jacquiod, S.; Vestergaard, G.; Schulz, S.; Schloter, M. Analysis of soil microbial communities based on amplicon sequencing of marker genes. *Biol. Fertil. Soils* **2017**, *53*, 485–489. [[CrossRef](#)]
36. Liu, M.; Clarke, L.J.; Baker, S.C.; Jordan, G.J.; Burridge, C.P. A practical guide to DNA metabarcoding for entomological ecologists. *Ecol. Entomol.* **2020**, *45*, 373–385. [[CrossRef](#)]
37. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F.O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **2013**, *41*, e1. [[CrossRef](#)]
38. Ficitola, G.F.; Coissac, E.; Zundel, S.; Riaz, T.; Shehzad, W.; Bessière, J.; Taberlet, P.; Pompanon, F. An In silico approach for the evaluation of DNA barcodes. *BMC Genom.* **2010**, *11*, 434. [[CrossRef](#)] [[PubMed](#)]
39. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet. J.* **2011**, *17*, 3. [[CrossRef](#)]
40. Singer, E.; Bushnell, B.; Coleman-Derr, D.; Bowman, B.; Bowers, R.M.; Levy, A.; Gies, E.A.; Cheng, J.-F.; Copeland, A.; Klenk, H.-P.; et al. High-resolution phylogenetic microbial community profiling. *ISME J.* **2016**, *10*, 2020–2032. [[CrossRef](#)]
41. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)]
42. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **2016**, *17*, 239. [[CrossRef](#)] [[PubMed](#)]
43. Mahmoud, M.; Zywicki, M.; Twardowski, T.; Karlowski, W.M. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics* **2019**, *111*, 43–49. [[CrossRef](#)] [[PubMed](#)]
44. Overholt, W.A.; Hölzer, M.; Geesink, P.; Diezel, C.; Marz, M.; Küsel, K. Inclusion of Oxford Nanopore long reads improves all microbial and viral metagenome-assembled genomes from a complex aquifer system. *Environ. Microbiol.* **2020**, *22*, 4000–4013. [[CrossRef](#)]
45. Lundberg, D.S.; Yourstone, S.; Mieczkowski, P.; Jones, C.D.; Dangl, J.L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* **2013**, *10*, 999–1002. [[CrossRef](#)] [[PubMed](#)]
46. Thijs, S.; Op De Beeck, M.; Beckers, B.; Truyens, S.; Stevens, V.; Van Hamme, J.D.; Weyens, N.; Vangronsveld, J. Comparative Evaluation of Four Bacteria-Specific Primer Pairs for 16S rRNA Gene Surveys. *Front. Microbiol.* **2017**, *8*, 494. [[CrossRef](#)]
47. Tremblay, J.; Singh, K.; Fern, A.; Kirton, E.; He, S.; Woyke, T.; Lee, J.; Chen, F.; Dangl, J.; Tringe, S. Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* **2015**, *6*, 771. [[CrossRef](#)] [[PubMed](#)]
48. Ghyselincx, J.; Pfeiffer, S.; Heylen, K.; Sessitsch, A.; De Vos, P. The Effect of Primer Choice and Short Read Sequences on the Outcome of 16S rRNA Gene Based Diversity Studies. *PLoS ONE* **2013**, *8*, e71360. [[CrossRef](#)]
49. Lear, G.; Dickie, I.; Banks, J.C.; Boyer, S.; Buckley, H.L.; Buckley, T.R.; Cruickshank, R.; Dopheide, A.; Handley, K.M.; Hermans, S.; et al. Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. *N. Z. J. Ecol.* **2018**, *42*, 10A–50A. [[CrossRef](#)]
50. Parada, A.E.; Needham, D.M.; Fuhrman, J.A. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **2016**, *18*, 1403–1414. [[CrossRef](#)]

51. Apprill, A.; McNally, S.; Parsons, R.; Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **2015**, *75*, 129–137. [[CrossRef](#)]
52. Quince, C.; Lanzen, A.; Davenport, R.J.; Turnbaugh, P.J. Removing Noise From Pyrosequenced Amplicons. *BMC Bioinform.* **2011**, *12*, 38. [[CrossRef](#)] [[PubMed](#)]
53. Chelius, M.K.; Triplett, E.W. The Diversity of Archaea and Bacteria in Association with the Roots of *Zea mays* L. *Microb. Ecol.* **2001**, *41*, 252–263. [[CrossRef](#)]
54. Redford, A.J.; Bowers, R.M.; Knight, R.; Linhart, Y.; Fierer, N. The ecology of the phyllosphere: Geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ. Microbiol.* **2010**, *12*, 2885–2893. [[CrossRef](#)] [[PubMed](#)]
55. Bodenhausen, N.; Horton, M.W.; Bergelson, J. Bacterial Communities Associated with the Leaves and the Roots of *Arabidopsis thaliana*. *PLoS ONE* **2013**, *8*, e56329. [[CrossRef](#)] [[PubMed](#)]
56. Sogin, M.L.; Morrison, H.G.; Huber, J.A.; Welch, D.M.; Huse, S.M.; Neal, P.R.; Arrieta, J.M.; Herndl, G.J. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12115–12120. [[CrossRef](#)]
57. Walker, J.J.; Pace, N.R. Phylogenetic Composition of Rocky Mountain Endolithic Microbial Ecosystems. *Appl. Environ. Microbiol.* **2007**, *73*, 3497–3504. [[CrossRef](#)]
58. McAllister, S.M.; Davis, R.E.; McBeth, J.M.; Tebo, B.M.; Emerson, D.; Moyer, C.L. Biodiversity and Emerging Biogeography of the Neutrophilic Iron-Oxidizing Zetaproteobacteria. *Appl. Environ. Microbiol.* **2011**, *77*, 5445–5457. [[CrossRef](#)] [[PubMed](#)]
59. Lee, T.K.; Van Doan, T.; Yoo, K.; Choi, S.; Kim, C.; Park, J. Discovery of commonly existing anode biofilm microbes in two different wastewater treatment MFCs using FLX Titanium pyrosequencing. *Appl. Microbiol. Biotechnol.* **2010**, *87*, 2335–2343. [[CrossRef](#)] [[PubMed](#)]
60. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4516–4522. [[CrossRef](#)]
61. Gilbert, J.A.; Jansson, J.K.; Knight, R. The Earth Microbiome project: Successes and aspirations. *BMC Biol.* **2014**, *12*, 69. [[CrossRef](#)] [[PubMed](#)]
62. Bahram, M.; Anslan, S.; Hildebrand, F.; Bork, P.; Tedersoo, L. Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environ. Microbiol. Rep.* **2019**, *11*, 487–494. [[CrossRef](#)]
63. Gantner, S.; Andersson, A.F.; Alonso-Sáez, L.; Bertilsson, S. Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. *J. Microbiol. Methods* **2011**, *84*, 12–18. [[CrossRef](#)] [[PubMed](#)]
64. Takai, K.; Horikoshi, K. Rapid Detection and Quantification of Members of the Archaeal Community by Quantitative PCR Using Fluorogenic Probes. *Appl. Environ. Microbiol.* **2000**, *66*, 5066–5072. [[CrossRef](#)]
65. Ovreås, L.; Forney, L.; Daae, F.L.; Torsvik, V. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol.* **1997**, *63*, 3367–3373. [[CrossRef](#)] [[PubMed](#)]
66. Raskin, L.; Stromley, J.M.; Rittmann, B.E.; Stahl, D.A. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol.* **1994**, *60*, 1232–1240. [[CrossRef](#)] [[PubMed](#)]
67. Watanabe, T.; Kimura, M.; Asakawa, S. Dynamics of methanogenic archaeal communities based on rRNA analysis and their relation to methanogenic activity in Japanese paddy field soils. *Soil Biol. Biochem.* **2007**, *39*, 2877–2887. [[CrossRef](#)]
68. Illumina. 16S metagenomic sequencing library preparation - Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System (Illumina Technical Note 15044223). Available online: [http://support.illumina.com/documents/documentation/chemistry\\_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf](http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf) (accessed on 8 September 2020).
69. Laforest-Lapointe, I.; Messier, C.; Kembel, S.W. Tree Leaf Bacterial Community Structure and Diversity Differ along a Gradient of Urban Intensity. *mSystems* **2017**, *2*, e00087-17. [[CrossRef](#)]
70. Kembel, S.W.; O’Connor, T.K.; Arnold, H.K.; Hubbell, S.P.; Wright, S.J.; Green, J.L. Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13715–13720. [[CrossRef](#)]
71. Laforest-Lapointe, I.; Messier, C.; Kembel, S.W. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome* **2016**, *4*, 27. [[CrossRef](#)]
72. Miura, T.; Sánchez, R.; Castañeda, L.E.; Godoy, K.; Barbosa, O. Shared and unique features of bacterial communities in native forest and vineyard phyllosphere. *Ecol. Evol.* **2019**, *9*, 3295–3305. [[CrossRef](#)]
73. Ulrich, K.; Becker, R.; Behrendt, U.; Kube, M.; Ulrich, A. A Comparative Analysis of Ash Leaf-Colonizing Bacterial Communities Identifies Putative Antagonists of *Hymenoscyphus fraxineus*. *Front. Microbiol.* **2020**, *11*, 966. [[CrossRef](#)]
74. Gdanetz, K.; Trail, F. The Wheat Microbiome Under Four Management Strategies, and Potential for Endophytes in Disease Protection. *Phytobiomes J.* **2017**, *1*, 158–168. [[CrossRef](#)]
75. Vorholt, J.A. Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* **2012**, *10*, 828–840. [[CrossRef](#)] [[PubMed](#)]
76. Sakai, M.; Ikenaga, M. Application of peptide nucleic acid (PNA)-PCR clamping technique to investigate the community structures of rhizobacteria associated with plant roots. *J. Microbiol. Methods* **2013**, *92*, 281–288. [[CrossRef](#)] [[PubMed](#)]
77. Ray, A.; Nordén, B. Peptide nucleic acid (PNA): Its medical and biotechnical applications and promise for the future. *FASEB J.* **2000**, *14*, 1041–1060. [[CrossRef](#)] [[PubMed](#)]
78. Santhanam, R.; Groten, K.; Meldau, D.G.; Baldwin, I.T. Analysis of Plant-Bacteria Interactions in Their Native Habitat: Bacterial Communities Associated with Wild Tobacco Are Independent of Endogenous Jasmonic Acid Levels and Developmental Stages. *PLoS ONE* **2014**, *9*, e94710. [[CrossRef](#)] [[PubMed](#)]

79. Toju, H.; Kurokawa, H.; Kenta, T. Factors Influencing Leaf- and Root-Associated Communities of Bacteria and Fungi Across 33 Plant Orders in a Grassland. *Front. Microbiol.* **2019**, *10*, 241. [[CrossRef](#)] [[PubMed](#)]
80. Wagner, M.R.; Lundberg, D.S.; del Rio, T.G.; Tringe, S.G.; Dangl, J.L.; Mitchell-Olds, T. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat. Commun.* **2016**, *7*, 12151. [[CrossRef](#)]
81. Wagner, M.R.; Busby, P.E.; Balint-Kurti, P. Analysis of leaf microbiome composition of near-isogenic maize lines differing in broad-spectrum disease resistance. *New Phytol.* **2020**, *225*, 2152–2165. [[CrossRef](#)]
82. Jackrel, S.L.; Owens, S.M.; Gilbert, J.A.; Pfister, C.A. Identifying the plant-associated microbiome across aquatic and terrestrial environments: The effects of amplification method on taxa discovery. *Mol. Ecol. Resour.* **2017**, *17*, 931–942. [[CrossRef](#)]
83. Fitzpatrick, C.R.; Lu-Irving, P.; Copeland, J.; Guttman, D.S.; Wang, P.W.; Baltrus, D.A.; Dlugosch, K.M.; Johnson, M.T.J. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome* **2018**, *6*, 144. [[CrossRef](#)]
84. Begerow, D.; Nilsson, H.; Unterseher, M.; Maier, W. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Appl. Microbiol. Biotechnol.* **2010**, *87*, 99–108. [[CrossRef](#)] [[PubMed](#)]
85. Nilsson, R.H.; Tedersoo, L.; Ryberg, M.; Kristiansson, E.; Hartmann, M.; Unterseher, M.; Porter, T.M.; Bengtsson-Palme, J.; Walker, D.M.; de Sousa, F.; et al. A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts. *Microbes Environ.* **2015**, *30*, 145–150. [[CrossRef](#)]
86. Bellemain, E.; Carlsen, T.; Brochmann, C.; Coissac, E.; Taberlet, P.; Kausserud, H. ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. *BMC Microbiol.* **2010**, *10*, 189. [[CrossRef](#)]
87. Schoch, C.L.; Seifert, K.A.; Huhndorf, S.; Robert, V.; Spouge, J.L.; Levesque, C.A.; Chen, W. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6241–6246. [[CrossRef](#)]
88. Li, S.; Deng, Y.; Wang, Z.; Zhang, Z.; Kong, X.; Zhou, W.; Yi, Y.; Qu, Y. Exploring the accuracy of amplicon-based internal transcribed spacer markers for a fungal community. *Mol. Ecol. Resour.* **2020**, *20*, 170–184. [[CrossRef](#)] [[PubMed](#)]
89. Xu, J. Fungal DNA barcoding. *Genome* **2016**, *59*, 913–932. [[CrossRef](#)]
90. Nilsson, R.H.; Kristiansson, E.; Ryberg, M.; Hallenberg, N.; Larsson, K.-H. Intraspecific ITS Variability in the Kingdom Fungi as Expressed in the International Sequence Databases and Its Implications for Molecular Species Identification. *Evol. Bioinform.* **2008**, *4*, EBO-S653. [[CrossRef](#)] [[PubMed](#)]
91. Wang, X.-C.; Liu, C.; Huang, L.; Bengtsson-Palme, J.; Chen, H.; Zhang, J.-H.; Cai, D.; Li, J.-Q. ITS1: A DNA barcode better than ITS2 in eukaryotes? *Mol. Ecol. Resour.* **2015**, *15*, 573–586. [[CrossRef](#)] [[PubMed](#)]
92. Bazzicalupo, A.L.; Bálint, M.; Schmitt, I. Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecol.* **2013**, *6*, 102–109. [[CrossRef](#)]
93. Yang, R.-H.; Su, J.-H.; Shang, J.-J.; Wu, Y.-Y.; Li, Y.; Bao, D.-P.; Yao, Y.-J. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PLoS ONE* **2018**, *13*, e0206428. [[CrossRef](#)] [[PubMed](#)]
94. Yahr, R.; Schoch, C.L.; Dentinger, B.T.M. Scaling up discovery of hidden diversity in fungi: Impacts of barcoding approaches. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*, 20150336. [[CrossRef](#)]
95. Nilsson, R.H.; Anslan, S.; Bahram, M.; Wurzbacher, C.; Baldrian, P.; Tedersoo, L. Mycobiome diversity: High-throughput sequencing and identification of fungi. *Nat. Rev. Microbiol.* **2019**, *17*, 95–109. [[CrossRef](#)] [[PubMed](#)]
96. Blaaliid, R.; Kumar, S.; Nilsson, R.H.; Abarenkov, K.; Kirk, P.M.; Kausserud, H. ITS1 versus ITS2 as DNA metabarcodes for fungi. *Mol. Ecol. Resour.* **2013**, *13*, 218–224. [[CrossRef](#)] [[PubMed](#)]
97. Monard, C.; Gantner, S.; Stenlid, J. Utilizing ITS1 and ITS2 to study environmental fungal diversity using pyrosequencing. *Fems Microbiol. Ecol.* **2013**, *84*, 165–175. [[CrossRef](#)]
98. Tedersoo, L.; Lindahl, B. Fungal identification biases in microbiome projects. *Environ. Microbiol. Rep.* **2016**, *8*, 774–779. [[CrossRef](#)]
99. White, T.J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols*; Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J., Eds.; Academic Press: San Diego, CA, USA, 1990; pp. 315–322. [[CrossRef](#)]
100. Toju, H.; Tanabe, A.S.; Yamamoto, S.; Sato, H. High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. *PLoS ONE* **2012**, *7*, e40863. [[CrossRef](#)]
101. 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.; et al. New primers to amplify the fungal ITS2 region – evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* **2012**, *82*, 666–677. [[CrossRef](#)]
102. Tedersoo, L.; Bahram, M.; Pölme, S.; Kõljalg, U.; Yorou, N.S.; Wijesundera, R.; Ruiz, L.V.; Vasco-Palacios, A.M.; Thu, P.Q.; Suija, A.; et al. Global diversity and geography of soil fungi. *Science* **2014**, *346*, 1256688. [[CrossRef](#)]
103. Turenne, C.Y.; Sanche, S.E.; Hoban, D.J.; Karlowsky, J.A.; Kabani, A.M. Rapid Identification of Fungi by Using the ITS2 Genetic Region and an Automated Fluorescent Capillary Electrophoresis System. *J. Clin. Microbiol.* **1999**, *37*, 1846–1851. [[CrossRef](#)]
104. Öpik, M.; Davison, J.; Moora, M.; Zobel, M. DNA-based detection and identification of Glomeromycota: The virtual taxonomy of environmental sequences. *Botany* **2013**, *92*, 135–147. [[CrossRef](#)]
105. Krüger, M.; Krüger, C.; Walker, C.; Stockinger, H.; Schüßler, A. Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytol.* **2012**, *193*, 970–984. [[CrossRef](#)] [[PubMed](#)]

106. Francioli, D.; Schulz, E.; Lentendu, G.; Wubet, T.; Buscot, F.; Reitz, T. Mineral vs. organic amendments: Microbial community structure, activity and abundance of agriculturally relevant microbes are driven by long-term fertilization strategies. *Front. Microbiol.* **2016**, *7*. [[CrossRef](#)]
107. Lekberg, Y.; Vasar, M.; Bullington, L.S.; Sepp, S.-K.; Antunes, P.M.; Bunn, R.; Larkin, B.G.; Öpik, M. More bang for the buck? Can arbuscular mycorrhizal fungal communities be characterized adequately alongside other fungi using general fungal primers? *New Phytol.* **2018**, *220*, 971–976. [[CrossRef](#)]
108. Berruti, A.; Desirò, A.; Visentin, S.; Zecca, O.; Bonfante, P. ITS fungal barcoding primers versus 18S AMF-specific primers reveal similar AMF-based diversity patterns in roots and soils of three mountain vineyards. *Environ. Microbiol. Rep.* **2017**, *9*, 658–667. [[CrossRef](#)] [[PubMed](#)]
109. Sato, K.; Suyama, Y.; Saito, M.; Sugawara, K. A new primer for discrimination of arbuscular mycorrhizal fungi with polymerase chain reaction-denature gradient gel electrophoresis. *Grassl. Sci.* **2005**, *51*, 179–181. [[CrossRef](#)]
110. Cui, X.; Hu, J.; Wang, J.; Yang, J.; Lin, X. Reclamation negatively influences arbuscular mycorrhizal fungal community structure and diversity in coastal saline-alkaline land in Eastern China as revealed by Illumina sequencing. *Appl. Soil Ecol.* **2016**, *98*, 140–149. [[CrossRef](#)]
111. Higo, M.; Tatewaki, Y.; Iida, K.; Yokota, K.; Isobe, K. Amplicon sequencing analysis of arbuscular mycorrhizal fungal communities colonizing maize roots in different cover cropping and tillage systems. *Sci. Rep.* **2020**, *10*, 6039. [[CrossRef](#)] [[PubMed](#)]
112. Faggioli, V.; Menoyo, E.; Geml, J.; Kempainen, M.; Pardo, A.; Salazar, M.J.; Becerra, A.G. Soil lead pollution modifies the structure of arbuscular mycorrhizal fungal communities. *Mycorrhiza* **2019**, *29*, 363–373. [[CrossRef](#)]
113. Van Geel, M.; Busschaert, P.; Honnay, O.; Lievens, B. Evaluation of six primer pairs targeting the nuclear rRNA operon for characterization of arbuscular mycorrhizal fungal (AMF) communities using 454 pyrosequencing. *J. Microbiol. Methods* **2014**, *106*, 93–100. [[CrossRef](#)] [[PubMed](#)]
114. Suzuki, K.; Takahashi, K.; Harada, N. Evaluation of primer pairs for studying arbuscular mycorrhizal fungal community compositions using a MiSeq platform. *Biol. Fertil. Soils* **2020**, *56*, 853–858. [[CrossRef](#)]
115. Mitchell, J.I.; Zuccaro, A. Sequences, the environment and fungi. *Mycologist* **2006**, *20*, 62–74. [[CrossRef](#)]
116. Misra, J.K.; Tewari, J.P.; Deshmukh, S.K. *Systematics and Evolution of Fungi*; Misra, J., Tewari, J., Deshmukh, S., Eds.; CRC Press: Boca Raton, FL, USA, 2011. [[CrossRef](#)]
117. Raja, H.A.; Miller, A.N.; Pearce, C.J.; Oberlies, N.H. Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. *J. Nat. Prod.* **2017**, *80*, 756–770. [[CrossRef](#)]
118. Banos, S.; Lentendu, G.; Kopf, A.; Wubet, T.; Glöckner, F.O.; Reich, M. A comprehensive fungi-specific 18S rRNA gene sequence primer toolkit suited for diverse research issues and sequencing platforms. *BMC Microbiol.* **2018**, *18*, 190. [[CrossRef](#)] [[PubMed](#)]
119. De Gruyter, J.; Weedon, J.T.; Bazot, S.; Dauwe, S.; Fernandez-Garberí, P.-R.; Geisen, S.; De La Motte, L.G.; Heinesch, B.; Janssens, I.A.; Leblans, N.; et al. Patterns of local, intercontinental and interseasonal variation of soil bacterial and eukaryotic microbial communities. *FEMS Microbiol. Ecol.* **2020**, *96*, fiae018. [[CrossRef](#)]
120. Liu, K.-L.; Porras-Alfaro, A.; Kuske, C.R.; Eichorst, S.A.; Xie, G. Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Appl. Environ. Microbiol.* **2012**, *78*, 1523–1533. [[CrossRef](#)] [[PubMed](#)]
121. Singer, D.; Seppely, C.V.W.; Lentendu, G.; Dunthorn, M.; Bass, D.; Belbahri, L.; Blandenier, Q.; Debroas, D.; de Groot, G.A.; de Vargas, C.; et al. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ. Int.* **2021**, *146*, 106262. [[CrossRef](#)] [[PubMed](#)]
122. Hadziavdic, K.; Lekang, K.; Lanzen, A.; Jonassen, I.; Thompson, E.M.; Troedsson, C. Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLoS ONE* **2014**, *9*, e87624. [[CrossRef](#)]
123. Lane, D.J. 6S/23S rRNA Sequencing. In *Nucleic Acid Techniques in Bacterial Systematic*; Stackebrandt, E., Goodfellow, M., Eds.; John Wiley and Sons: New York, NY, USA, 1991; pp. 115–175.
124. Medlin, L.; Elwood, H.J.; Stickel, S.; Sogin, M.L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **1988**, *71*, 491–499. [[CrossRef](#)]
125. Amaral-Zettler, L.A.; McCliment, E.A.; Ducklow, H.W.; Huse, S.M. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE* **2009**, *4*, e6372. [[CrossRef](#)]
126. Seppely, C.V.W.; Singer, D.; Dumack, K.; Fournier, B.; Belbahri, L.; Mitchell, E.A.D.; Lara, E. Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. *Soil Biol. Biochem.* **2017**, *112*, 68–76. [[CrossRef](#)]
127. Euringer, K.; Lueders, T. An optimised PCR/T-RFLP fingerprinting approach for the investigation of protistan communities in groundwater environments. *J. Microbiol. Methods* **2008**, *75*, 262–268. [[CrossRef](#)] [[PubMed](#)]
128. Amann, R.I.; Binder, B.J.; Olson, R.J.; Chisholm, S.W.; Devereux, R.; Stahl, D.A. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **1990**, *56*, 1919–1925. [[CrossRef](#)] [[PubMed](#)]
129. Dollive, S.; Peterfreund, G.L.; Sherrill-Mix, S.; Bitteringer, K.; Sinha, R.; Hoffmann, C.; Nabel, C.S.; Hill, D.A.; Artis, D.; Bachman, M.A.; et al. A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol.* **2012**, *13*, R60. [[CrossRef](#)] [[PubMed](#)]
130. Nolte, V.; Pandey, R.V.; Jost, S.; Medinger, R.; Ottenwalder, B.; Boenigk, J.; Schlotterer, C. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.* **2010**, *19*, 2908–2915. [[CrossRef](#)]

131. Bass, D.; Silberman, J.D.; Brown, M.W.; Pearce, R.A.; Tice, A.K.; Jousset, A.; Geisen, S.; Hartikainen, H. Coprophilic amoebae and flagellates, including *Guttulinopsis*, *Rosculus* and *Helkesimastix*, characterise a divergent and diverse rhizarian radiation and contribute to a large diversity of faecal-associated protists. *Environ. Microbiol.* **2016**, *18*, 1604–1619. [[CrossRef](#)]
132. Stoeck, T.; Bass, D.; Nebel, M.; Christen, R.; Jones, M.D.; Breiner, H.W.; Richards, T.A. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **2010**, *19* (Suppl. 1), 21–31. [[CrossRef](#)]
133. Hugerth, L.W.; Muller, E.E.L.; Hu, Y.O.O.; Lebrun, L.A.M.; Roume, H.; Lundin, D.; Wilmes, P.; Andersson, A.F. Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia. *PLoS ONE* **2014**, *9*, e95567. [[CrossRef](#)]
134. Guardiola, M.; Uriz, M.J.; Taberlet, P.; Coissac, E.; Wangenstein, O.S.; Turon, X. Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. *PLoS ONE* **2015**, *10*, e0139633. [[CrossRef](#)]
135. Bradley, I.M.; Pinto, A.J.; Guest, J.S. Design and Evaluation of Illumina MiSeq-Compatible, 18S rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. *Appl. Environ. Microbiol.* **2016**, *82*, 5878–5891. [[CrossRef](#)]
136. Mahé, F.; de Vargas, C.; Bass, D.; Czech, L.; Stamatakis, A.; Lara, E.; Singer, D.; Mayor, J.; Bunge, J.; Sernaker, S.; et al. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* **2017**, *1*, 0091. [[CrossRef](#)]
137. Heger, T.J.; Giesbrecht, I.J.W.; Gustavsen, J.; Del Campo, J.; Kellogg, C.T.E.; Hoffman, K.M.; Lertzman, K.; Mohn, W.W.; Keeling, P.J. High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environ. Microbiol.* **2018**, *20*, 1185–1203. [[CrossRef](#)]
138. Singer, D.; Metz, S.; Unrein, F.; Shimano, S.; Mazei, Y.; Mitchell, E.A.D.; Lara, E. Contrasted Micro-Eukaryotic Diversity Associated with Sphagnum Mosses in Tropical, Subtropical and Temperate Climatic Zones. *Microb. Ecol.* **2019**, *78*, 714–724. [[CrossRef](#)]
139. Guo, S.; Xiong, W.; Xu, H.; Hang, X.; Liu, H.; Xun, W.; Li, R.; Shen, Q. Continuous application of different fertilizers induces distinct bulk and rhizosphere soil protist communities. *Eur. J. Soil Biol.* **2018**, *88*, 8–14. [[CrossRef](#)]
140. Xiong, W.; Song, Y.; Yang, K.; Gu, Y.; Wei, Z.; Kowalchuk, G.A.; Xu, Y.; Jousset, A.; Shen, Q.; Geisen, S. Rhizosphere protists are key determinants of plant health. *Microbiome* **2020**, *8*, 27. [[CrossRef](#)]
141. Lentendu, G.; Wubet, T.; Chatzinotas, A.; Wilhelm, C.; Buscot, F.; Schlegel, M. Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: A multiple barcoding approach. *Mol. Ecol.* **2014**, *23*, 3341–3355. [[CrossRef](#)]
142. Fiore-Donno, A.M.; Rixen, C.; Rippin, M.; Glaser, K.; Samolov, E.; Karsten, U.; Becker, B.; Bonkowski, M. New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Mol. Ecol. Resour.* **2018**, *18*, 229–239. [[CrossRef](#)]
143. Adl, S.M.; Bass, D.; Lane, C.E.; Lukeš, J.; Schoch, C.L.; Smirnov, A.; Agatha, S.; Berney, C.; Brown, M.W.; Burki, F.; et al. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **2019**, *66*, 4–119. [[CrossRef](#)]
144. Pawlowski, J.; Audic, S.; Adl, S.; Bass, D.; Belbahri, L.; Berney, C.; Bowser, S.S.; Cepicka, I.; Decelle, J.; Dunthorn, M.; et al. CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biol.* **2012**, *10*, e1001419. [[CrossRef](#)]
145. Zizka, V.M.A.; Elbrecht, V.; Macher, J.-N.; Leese, F. Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Mol. Ecol. Resour.* **2019**, *19*, 893–899. [[CrossRef](#)]
146. Carøe, C.; Bohmann, K. Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *Mol. Ecol. Resour.* **2020**, *20*, 1620–1631. [[CrossRef](#)]
147. Dopheide, A.; Xie, D.; Buckley, T.R.; Drummond, A.J.; Newcomb, R.D. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods Ecol. Evol.* **2019**, *10*, 120–133. [[CrossRef](#)]
148. Rychlik, W.; Spencer, W.J.; Rhoads, R.E. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **1990**, *18*, 6409–6412. [[CrossRef](#)] [[PubMed](#)]
149. Oliver, A.K.; Brown, S.P.; Callahan, M.A.; Jumpponen, A. Polymerase matters: Non-proofreading enzymes inflate fungal community richness estimates by up to 15%. *Fungal Ecol.* **2015**, *15*, 86–89. [[CrossRef](#)]
150. Krueger, F.; Andrews, S.R.; Osborne, C.S. Large Scale Loss of Data in Low-Diversity Illumina Sequencing Libraries Can Be Recovered by Deferred Cluster Calling. *PLoS ONE* **2011**, *6*, e16607. [[CrossRef](#)]
151. Illumina. How much PhiX spike-in is recommended when sequencing low diversity libraries on Illumina platforms? Available online: <https://support.illumina.com/bulletins/2017/02/how-much-phix-spike-in-is-recommended-when-sequencing-low-divers.html> (accessed on 11 November 2020).
152. De Muinck, E.J.; Trosvik, P.; Gilfillan, G.D.; Hov, J.R.; Sundaram, A.Y.M. A novel ultra high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform. *Microbiome* **2017**, *5*, 68. [[CrossRef](#)]
153. Holm, J.B.; Humphrys, M.S.; Robinson, C.K.; Settles, M.L.; Ott, S.; Fu, L.; Yang, H.; Gajer, P.; He, X.; McComb, E.; et al. Ultrahigh-Throughput Multiplexing and Sequencing of >500-Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform. *MSystems* **2019**, *4*, e00029-19. [[CrossRef](#)]
154. Glenn, T.C.; Pierson, T.W.; Bayona-Vásquez, N.J.; Kieran, T.J.; Hoffberg, S.L.; Thomas Iv, J.C.; Lefever, D.E.; Finger, J.W.; Gao, B.; Bian, X.; et al. Adapterama II: Universal amplicon sequencing on Illumina platforms (TaggiMatrix). *PeerJ* **2019**, *7*, e7786. [[CrossRef](#)]
155. Esling, P.; Lejzerowicz, F.; Pawlowski, J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* **2015**, *43*, 2513–2524. [[CrossRef](#)]

156. Fadrosch, D.W.; Ma, B.; Gajer, P.; Sengamalay, N.; Ott, S.; Brotman, R.M.; Ravel, J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2014**, *2*, 6. [[CrossRef](#)] [[PubMed](#)]
157. Jensen, E.A.; Berryman, D.E.; Murphy, E.R.; Carroll, R.K.; Busken, J.; List, E.O.; Broach, W.H. Heterogeneity spacers in 16S rDNA primers improve analysis of mouse gut microbiomes via greater nucleotide diversity. *BioTechniques* **2019**, *67*, 55–62. [[CrossRef](#)]
158. Taberlet, P.; Bonin, A.; Zinger, L.; Coissac, E. Environmental DNA: For Biodiversity Research and Monitoring. *Environ. Dna: Biodivers. Res. Monit.* **2018**, 1–253. [[CrossRef](#)]
159. Schnell, I.B.; Bohmann, K.; Gilbert, M.T.P. Tag jumps illuminated—Reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Resour.* **2015**, *15*, 1289–1303. [[CrossRef](#)]
160. Bokulich, N.A.; Subramanian, S.; Faith, J.J.; Gevers, D.; Gordon, J.I.; Knight, R.; Mills, D.A.; Caporaso, J.G. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **2013**, *10*, 57–59. [[CrossRef](#)] [[PubMed](#)]
161. Kozich, J.J.; Westcott, S.L.; Baxter, N.T.; Highlander, S.K.; Schloss, P.D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **2013**, *79*, 5112–5120. [[CrossRef](#)] [[PubMed](#)]
162. De Barba, M.; Miquel, C.; Boyer, F.; Mercier, C.; Rioux, D.; Coissac, E.; Taberlet, P. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Mol. Ecol. Resour.* **2014**, *14*, 306–323. [[CrossRef](#)] [[PubMed](#)]
163. McLaren, M.R.; Willis, A.D.; Callahan, B.J. Consistent and correctable bias in metagenomic sequencing experiments. *eLife* **2019**, *8*, e46923. [[CrossRef](#)] [[PubMed](#)]
164. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **2017**, *8*. [[CrossRef](#)]
165. Harrison, J.G.; John Calder, W.; Shuman, B.; Alex Buerkle, C. The quest for absolute abundance: The use of internal standards for DNA-based community ecology. *Mol. Ecol. Resour.* **2021**. [[CrossRef](#)]
166. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [[CrossRef](#)]
167. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [[CrossRef](#)]
168. Zafeiropoulos, H.; Viet, H.Q.; Vasileiadou, K.; Potirakis, A.; Arvanitidis, C.; Topalis, P.; Pavloudi, C.; Pafilis, E. PEMA: A flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience* **2020**, *9*. [[CrossRef](#)] [[PubMed](#)]
169. Anslan, S.; Bahram, M.; Hiiesalu, I.; Tedersoo, L. PipeCraft: Flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Mol. Ecol. Resour.* **2017**, *17*, e234–e240. [[CrossRef](#)] [[PubMed](#)]
170. Dufresne, Y.; Lejzerowicz, F.; Perret-Gentil, L.A.; Pawlowski, J.; Cordier, T. SLIM: A flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinform.* **2019**, *20*, 88. [[CrossRef](#)] [[PubMed](#)]
171. Fosso, B.; Santamaria, M.; Marzano, M.; Alonso-Aleman, D.; Valiente, G.; Donvito, G.; Monaco, A.; Notarangelo, P.; Pesole, G. BioMaS: A modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS. *BMC Bioinform.* **2015**, *16*, 203. [[CrossRef](#)] [[PubMed](#)]
172. Gweon, H.S.; Oliver, A.; Taylor, J.; Booth, T.; Gibbs, M.; Read, D.S.; Griffiths, R.I.; Schonrogge, K. PIPITS: An automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods Ecol. Evol.* **2015**, *6*, 973–980. [[CrossRef](#)]
173. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)]
174. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [[CrossRef](#)] [[PubMed](#)]
175. Boyer, F.; Mercier, C.; Bonin, A.; Le Bras, Y.; Taberlet, P.; Coissac, E. obitools: A unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **2016**, *16*, 176–182. [[CrossRef](#)] [[PubMed](#)]
176. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)]
177. Porter, T.M.; Hajibabaei, M. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Mol. Ecol.* **2018**, *27*, 313–338. [[CrossRef](#)] [[PubMed](#)]
178. Rideout, J.R.; He, Y.; Navas-Molina, J.A.; Walters, W.A.; Ursell, L.K.; Gibbons, S.M.; Chase, J.; McDonald, D.; Gonzalez, A.; Robbins-Pianka, A.; et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2014**, *2*, e545. [[CrossRef](#)]
179. Nguyen, N.-P.; Warnow, T.; Pop, M.; White, B. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *Npj Biofilms Microbiomes* **2016**, *2*, 16004. [[CrossRef](#)]
180. Gevers, D.; Cohan, F.M.; Lawrence, J.G.; Spratt, B.G.; Coenye, T.; Feil, E.J.; Stackebrandt, E.; de Peer, Y.V.; Vandamme, P.; Thompson, F.L.; et al. Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* **2005**, *3*, 733–739. [[CrossRef](#)]
181. Schmidt, T.S.B.; Matias Rodrigues, J.F.; von Mering, C. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environ. Microbiol.* **2015**, *17*, 1689–1706. [[CrossRef](#)]

182. Mahé, F.; Rognes, T.; Quince, C.; de Vargas, C.; Dunthorn, M. Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ* **2015**, *3*, e1420. [[CrossRef](#)]
183. Mahé, F.; Rognes, T.; Quince, C.; de Vargas, C.; Dunthorn, M. Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ* **2014**, *2*, e593. [[CrossRef](#)] [[PubMed](#)]
184. De Vargas, C.; Audic, S.; Henry, N.; Decelle, J.; Mahé, F.; Logares, R.; Lara, E.; Berney, C.; Le Bescot, N.; Probert, I.; et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **2015**, *348*, 1261605. [[CrossRef](#)]
185. Eren, A.M.; Morrison, H.G.; Lescault, P.J.; Reveillaud, J.; Vineis, J.H.; Sogin, M.L. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **2015**, *9*, 968–979. [[CrossRef](#)]
186. Callahan, B.J.; Wong, J.; Heiner, C.; Oh, S.; Theriot, C.M.; Gulati, A.S.; McGill, S.K.; Dougherty, M.K. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* **2019**, *47*, e103. [[CrossRef](#)]
187. Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017**, *11*, 2639–2643. [[CrossRef](#)]
188. Prodan, A.; Tremaroli, V.; Brolin, H.; Zwinderman, A.H.; Nieuwdorp, M.; Levin, E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS ONE* **2020**, *15*, e0227434. [[CrossRef](#)]
189. Estaki, M.; Jiang, L.; Bokulich, N.A.; McDonald, D.; González, A.; Kosciulek, T.; Martino, C.; Zhu, Q.; Birmingham, A.; Vázquez-Baeza, Y.; et al. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. *Curr. Protoc. Bioinform.* **2020**, *70*, e100. [[CrossRef](#)]
190. Pauvert, C.; Buée, M.; Laval, V.; Edel-Hermann, V.; Fauchery, L.; Gautier, A.; Lesur, I.; Vallance, J.; Vacher, C. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecol.* **2019**, *41*, 23–33. [[CrossRef](#)]
191. Caruso, V.; Song, X.; Asquith, M.; Karstens, L. Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. *MSystems* **2019**, *4*, e00163-18. [[CrossRef](#)]
192. Nearing, J.T.; Douglas, G.M.; Comeau, A.M.; Langille, M.G.I. Denoising the Denoisers: An independent evaluation of microbiome sequence error-correction approaches. *PeerJ* **2018**, *6*, e5364. [[CrossRef](#)]
193. Semchenko, M.; Leff, J.W.; Lozano, Y.M.; Saar, S.; Davison, J.; Wilkinson, A.; Jackson, B.G.; Pritchard, W.J.; De Long, J.R.; Oakley, S.; et al. Fungal diversity regulates plant-soil feedbacks in temperate grassland. *Sci. Adv.* **2018**, *4*, eaau4578. [[CrossRef](#)]
194. Beirinckx, S.; Viaene, T.; Haegeman, A.; Debode, J.; Amery, F.; Vandenabeele, S.; Nelissen, H.; Inzé, D.; Tito, R.; Raes, J.; et al. Tapping into the maize root microbiome to identify bacteria that promote growth under chilling conditions. *Microbiome* **2020**, *8*, 54. [[CrossRef](#)] [[PubMed](#)]
195. Yergeau, É.; Quiza, L.; Tremblay, J. Microbial indicators are better predictors of wheat yield and quality than N fertilization. *FEMS Microbiol. Ecol.* **2019**, *96*, fiz205. [[CrossRef](#)]
196. Fitzpatrick, C.R.; Copeland, J.; Wang, P.W.; Guttman, D.S.; Kotanen, P.M.; Johnson, M.T.J. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1157–E1165. [[CrossRef](#)]
197. Rocca, J.D.; Simonin, M.; Blaszczyk, J.R.; Ernakovich, J.G.; Gibbons, S.M.; Midani, F.S.; Washburne, A.D. The Microbiome Stress Project: Toward a Global Meta-Analysis of Environmental Stressors and Their Effects on Microbial Communities. *Front. Microbiol.* **2019**, *9*, 3272. [[CrossRef](#)]
198. Francioli, D.; van Ruijven, J.; Bakker, L.; Mommer, L. Drivers of total and pathogenic soil-borne fungal communities in grassland plant species. *Fungal Ecol.* **2020**, *48*, 100987. [[CrossRef](#)]
199. Glassman, S.I.; Martiny, J.B.H. Broad-scale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units. *MSphere* **2018**, *3*, e00148-18. [[CrossRef](#)] [[PubMed](#)]
200. Forster, D.; Lentendu, G.; Filker, S.; Dubois, E.; Wilding, T.A.; Stoeck, T. Improving eDNA-based protist diversity assessments using networks of amplicon sequence variants. *Environ. Microbiol.* **2019**, *21*, 4109–4124. [[CrossRef](#)]
201. Frøslev, T.G.; Kjeller, R.; Bruun, H.H.; Ejrnæs, R.; Brunbjerg, A.K.; Pietroni, C.; Hansen, A.J. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* **2017**, *8*, 1188. [[CrossRef](#)] [[PubMed](#)]
202. Bálint, M.; Bahram, M.; Eren, A.M.; Faust, K.; Fuhrman, J.A.; Lindahl, B.; O'Hara, R.B.; Öpik, M.; Sogin, M.L.; Unterseher, M.; et al. Millions of reads, thousands of taxa: Microbial community structure and associations analyzed via marker genes. *FEMS Microbiol. Rev.* **2016**, *40*, 686–700. [[CrossRef](#)]
203. Brown, S.P.; Veach, A.M.; Rigdon-Huss, A.R.; Grond, K.; Lickteig, S.K.; Lothamer, K.; Oliver, A.K.; Jumpponen, A. Scraping the bottom of the barrel: Are rare high throughput sequences artifacts? *Fungal Ecol.* **2015**, *13*, 221–225. [[CrossRef](#)]
204. Balvočiūtė, M.; Huson, D.H. SILVA, RDP, Greengenes, NCBI and OTT—How do these taxonomies compare? *BMC Genom.* **2017**, *18*, 114. [[CrossRef](#)] [[PubMed](#)]
205. Pruesse, E.; Quast, C.; Knittel, K.; Fuchs, B.M.; Ludwig, W.; Peplies, J.; Glöckner, F.O. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **2007**, *35*, 7188–7196. [[CrossRef](#)]
206. Cole, J.R.; Wang, Q.; Cardenas, E.; Fish, J.; Chai, B.; Farris, R.J.; Kulam-Syed-Mohideen, A.S.; McGarrell, D.M.; Marsh, T.; Garrity, G.M.; et al. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **2008**, *37*, D141–D145. [[CrossRef](#)]

207. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072. [[CrossRef](#)]
208. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **2011**, *40*, D136–D143. [[CrossRef](#)]
209. Abarenkov, K.; Henrik Nilsson, R.; Larsson, K.-H.; Alexander, I.J.; Eberhardt, U.; Erland, S.; Høiland, K.; Kjølter, R.; Larsson, E.; Pennanen, T.; et al. The UNITE database for molecular identification of fungi—Recent updates and future perspectives. *New Phytol.* **2010**, *186*, 281–285. [[CrossRef](#)]
210. Guillou, L.; Bachar, D.; Audic, S.; Bass, D.; Berney, C.; Bittner, L.; Boutte, C.; Burgaud, G.; de Vargas, C.; Decelle, J.; et al. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **2012**, *41*, D597–D604. [[CrossRef](#)]
211. Leinonen, R.; Akhtar, R.; Birney, E.; Bower, L.; Cerdano-Tárraga, A.; Cheng, Y.; Cleland, I.; Faruque, N.; Goodgame, N.; Gibson, R.; et al. The European Nucleotide Archive. *Nucleic Acids Res.* **2010**, *39*, D28–D31. [[CrossRef](#)]
212. Nakamura, Y.; Cochrane, G.; Karsch-Mizrachi, I. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* **2012**, *41*, D21–D24. [[CrossRef](#)]
213. Benson, D.A.; Karsch-Mizrachi, I.; Clark, K.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2011**, *40*, D48–D53. [[CrossRef](#)] [[PubMed](#)]
214. Deshpande, V.; Wang, Q.; Greenfield, P.; Charleston, M.; Porras-Alfaro, A.; Kuske, C.R.; Cole, J.R.; Midgley, D.J.; Tran-Dinh, N. Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* **2016**, *108*, 1–5. [[CrossRef](#)]
215. Kõljalg, U.; Nilsson, R.H.; Abarenkov, K.; Tedersoo, L.; Taylor, A.F.S.; Bahram, M.; Bates, S.T.; Bruns, T.D.; Bengtsson-Palme, J.; Callaghan, T.M.; et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **2013**, *22*, 5271–5277. [[CrossRef](#)]
216. Nilsson, R.H.; Larsson, K.-H.; Taylor, A.F.S.; Bengtsson-Palme, J.; Jeppesen, T.S.; Schigel, D.; Kennedy, P.; Picard, K.; Glöckner, F.O.; Tedersoo, L.; et al. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **2018**, *47*, D259–D264. [[CrossRef](#)]
217. Santamaria, M.; Fosso, B.; Licciulli, F.; Balech, B.; Larini, I.; Grillo, G.; De Caro, G.; Liuni, S.; Pesole, G. ITSoneDB: A comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences. *Nucleic Acids Res.* **2018**, *46*, D127–D132. [[CrossRef](#)]
218. Ankenbrand, M.J.; Keller, A.; Wolf, M.; Schultz, J.; Förster, F. ITS2 Database V: Twice as Much. *Mol. Biol. Evol.* **2015**, *32*, 3030–3032. [[CrossRef](#)] [[PubMed](#)]
219. Öpik, M.; Vanatoa, A.; Vanatoa, E.; Moora, M.; Davison, J.; Kalwij, J.M.; Reier, Ü.; Zobel, M. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **2010**, *188*, 223–241. [[CrossRef](#)] [[PubMed](#)]
220. Martorelli, I.; Helwerda, L.S.; Kerkvliet, J.; Gomes, S.I.F.; Nuytinck, J.; Werff, C.R.A.v.d.; Ramackers, G.J.; Gulyaev, A.P.; Merckx, V.S.F.T.; Verbeek, F.J. Fungal metabarcoding data integration framework for the MycoDiversity DataBase (MDDb). *J. Integr. Bioinform.* **2020**, *17*, 20190046. [[CrossRef](#)]
221. Kodama, Y.; Shumway, M.; Leinonen, R.; on behalf of the International Nucleotide Sequence Database, C. The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Res.* **2012**, *40*, D54–D56. [[CrossRef](#)]
222. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]
223. Yilmaz, P.; Gilbert, J.A.; Knight, R.; Amaral-Zettler, L.; Karsch-Mizrachi, I.; Cochrane, G.; Nakamura, Y.; Sansone, S.-A.; Glöckner, F.O.; Field, D. The genomic standards consortium: Bringing standards to life for microbial ecology. *ISME J.* **2011**, *5*, 1565–1567. [[CrossRef](#)] [[PubMed](#)]
224. ten Hoopen, P.; Finn, R.D.; Bongo, L.A.; Corre, E.; Fosso, B.; Meyer, F.; Mitchell, A.; Pelletier, E.; Pesole, G.; Santamaria, M.; et al. The metagenomic data life-cycle: Standards and best practices. *GigaScience* **2017**, *6*, gix047. [[CrossRef](#)] [[PubMed](#)]
225. Glass, E.M.; Dribinsky, Y.; Yilmaz, P.; Levin, H.; Van Pelt, R.; Wendel, D.; Wilke, A.; Eisen, J.A.; Huse, S.; Shipanova, A.; et al. MIXS-BE: A MIXS extension defining a minimum information standard for sequence data from the built environment. *ISME J.* **2014**, *8*, 1–3. [[CrossRef](#)]
226. Yilmaz, P.; Kottmann, R.; Field, D.; Knight, R.; Cole, J.R.; Amaral-Zettler, L.; Gilbert, J.A.; Karsch-Mizrachi, I.; Johnston, A.; Cochrane, G.; et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat. Biotechnol.* **2011**, *29*, 415–420. [[CrossRef](#)]
227. Jurburg, S.D.; Konzack, M.; Eisenhauer, N.; Heintz-Buschart, A. The archives are half-empty: An assessment of the availability of microbial community sequencing data. *Commun. Biol.* **2020**, *3*, 474. [[CrossRef](#)]
228. Cristescu, M.E. From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* **2014**, *29*, 566–571. [[CrossRef](#)]
229. Santos, A.; van Aerle, R.; Barrientos, L.; Martinez-Urtaza, J. Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 296–305. [[CrossRef](#)]

- 
230. Nicholls, S.M.; Quick, J.C.; Tang, S.; Loman, N.J. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience* **2019**, *8*, giz043. [[CrossRef](#)]
  231. Winand, R.; Bogaerts, B.; Hoffman, S.; Lefevre, L.; Delvoye, M.; Braekel, J.V.; Fu, Q.; Roosens, N.H.; Keersmaecker, S.C.D.; Vanneste, K. Targeting the 16S rRNA Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. *Int. J. Mol. Sci.* **2019**, *21*, 298. [[CrossRef](#)]
  232. Sokol, H.; Leducq, V.; Aschard, H.; Pham, H.-P.; Jegou, S.; Landman, C.; Cohen, D.; Liguori, G.; Bourrier, A.; Nion-Larmurier, I.; et al. Fungal microbiota dysbiosis in IBD. *Gut* **2017**, *66*, 1039–1048. [[CrossRef](#)]
  233. Bahram, M.; Hildebrand, F.; Forslund, S.K.; Anderson, J.L.; Soudzilovskaia, N.A.; Bodegom, P.M.; Bengtsson-Palme, J.; Anslan, S.; Coelho, L.P.; Harend, H.; et al. Structure and function of the global topsoil microbiome. *Nature* **2018**, *560*, 233–237. [[CrossRef](#)]
  234. Ribière, C.; Beugnot, R.; Parisot, N.; Gasc, C.; Defois, C.; Denonfoux, J.; Boucher, D.; Peyretailade, E.; Peyret, P. Targeted Gene Capture by Hybridization to Illuminate Ecosystem Functioning. In *Microbial Environmental Genomics (MEG)*; Martin, F., Uroz, S., Eds.; Springer: New York, NY, USA, 2016; pp. 167–182. [[CrossRef](#)]
  235. Gasc, C.; Peyret, P. Hybridization capture reveals microbial diversity missed using current profiling methods. *Microbiome* **2018**, *6*, 61. [[CrossRef](#)]