scientific reports

Check for updates

OPEN

A novel two-step metabarcoding approach improves soil microbiome biodiversity assessment

Marcin Musiałowski[1], Monika Mierzwa-Hersztek[2], Krzysztof Gondek[2] & Klaudia Dębiec-Andrzejewska[1]✉

The foundation of microbial ecology research is Next-Generation Sequencing (NGS), which allows for reconstruction of the soil microbiome taxonomical structure and the calculation of biodiversity metrics. However, obtaining reliable data on soil biodiversity poses several challenges, with accurate primer selection being one of the most critical. While 16S rDNA primers are widely used for their ability to broadly target bacterial communities, they can introduce biases. These primers may preferentially amplify certain bacterial groups, leading to a skewed representation of the microbial diversity in soil samples. To overcome the bias, we developed a novel, Two-Step Metabarcoding (TSM) approach to obtain more accurate and detailed data on soil microbiome structure and biodiversity. The first step involved sequencing of amplicons generated using universal 16S rDNA primers, provided an initial overview of the microbial community, and allowed the identification of key taxonomical groups. In the second step, we employed sequencing of amplicons generated with taxa-specific primers designed for the most abundant phyla in the community. We used the obtained data for a more reliable reconstruction of microbiome taxonomic structure and biodiversity. This two-step approach ensures a thorough exploration of the soil microbiome and promises to enhance our understanding of soil microbial dynamics and ecology.

Keywords Metabarcoding, Primers, Sequencing, Microbiome, Soil biodiversity

The development of DNA sequencing revolutionized soil microbiology by enabling researchers to study microorganisms beyond the limitations of culture-based methods[1]. The advent of Next-Generation Sequencing (NGS) further transformed the field by allowing for high-throughput, high-quality data generation, providing an unprecedented level of detail on microbial community structure and function, including unculturable microorganisms[2–4]. This capability enabled for the first comprehensive analysis of soil microbial populations, leading to deeper insights into ecosystem processes, microbial interactions, and the roles microorganisms play in global biogeochemical cycles[5,6].

Among the most important applications of NGS in the field of soil microbiology is the analysis of microbiome taxonomic structure[7]. Understanding the composition of microbial communities in soils is crucial, since microbiomes play a fundamental role in ecosystem functioning, including nutrient cycling, plant growth promotion, organic matter decomposition, and disease suppression[7–9]. A deeper comprehension of the microbiome's structure is also crucial for assessment of soil biodiversity, which is vital for the sustainability of the environment and the productivity of agricultural systems, since the variety of microorganisms within the soil creates a complex network that enhances soil structure, fertility, and health[10–13]. The importance of understanding microbiomes has been also demonstrated in large-scale soil microbiome surveys, which showed microbial biogeography patterns across continental gradients. Projects such as the French RMQS soil monitoring network and the pan-European LUCAS survey have revealed extensive insights into the environmental drivers of microbial diversity and structure at scale[14–16]. For example, compelling evidence of consistent ecological structuring across soils and geographic regions has been provided using such datasets. Due to the important role of microorganisms in ensuring ecosystems robustness, taxonomic structure analysis could be used for more precise management of soil health, enabling the development of targeted strategies for sustainable agriculture, ecosystem restoration, and climate change mitigation[17–19]. For this reason, constant improvement of research methodologies is important for delivery of more and more efficient and accurate tools.

[1]Department of Geomicrobiology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Warsaw, Poland. [2]Department of Agricultural and Environmental Chemistry, Faculty of Agriculture and Economics, University of Agriculture in Krakow, Krakow, Poland. ✉email: klaudia.debiec@uw.edu.pl

Studies of soil microbiomes are commonly performed using metabarcoding techniques, which involve the extraction of DNA, amplification of marker gene regions using PCR, and sequencing obtained amplicons to compare against reference databases[20,21]. In microbiology, the utilization of 16S rDNA universal primers, designed to target highly conserved regions flanking the more variable across a broad array of bacterial taxa, is considered the gold standard for elucidating the taxonomic structure and abundance within microbiomes[22–25]. Despite ensuring the maximum universality and applicability, enabling the detection of a diverse range of bacteria, this method encounters several limitations, particularly in environmental microbiology. The performance of universal primers is not uniformly effective across all bacterial groups, attributable to variable primers binding efficiencies, which introduces a selection bias during DNA amplification, which could lead either to under- or overrepresentation of particular taxa[26–28]. This could be especially notable on lower levels of taxonomic classification (e.g. genus), where universal primers resolution is often insufficient[29,30]. Several studies comparing PCR-based metabarcoding with shotgun metagenomic sequencing, an approach that avoids primer-related bias by directly sequencing environmental DNA, have demonstrated significant discrepancies in microbial community composition, particularly in the detection of rare taxa and in the relative abundance of dominant groups[13,31–33]. These findings highlight the limitations of PCR-dependent methods and reinforce the need for approaches that mitigate primer-induced biases. Differences in DNA template concentration could also affect the outcome of metabarcoding, since abundant species could outcompete rare taxa. Such biases can misrepresent microbial community structures, thereby compromising the accuracy of derived biodiversity metrics[34]. These inaccuracies might skew ecological assessments and interpretations, affecting our understanding of microbial diversity's roles and implications in environmental contexts.

One approach to overcome these limitations is to achieve higher sequencing resolution by complementing universal primers with use of group-specific primers[29,35–37]. Universal primers could provide a general picture of microbiome structure, which can be described in more detailed way with specific primers, such as those targeting phylum or class level. Investigating specific taxa rather offers a more comprehensive understanding of individual taxon's ecology, as they frequently exhibit diverse responses to environmental factors. In recent years, the development of group-specific primers has become more feasible, thanks to the increasing number of 16S rDNA sequences available in public databases, e.g. Silva or Greengenes[38–40]. Several studies have not only identified efficient 16S rDNA primers for important soil bacterial groups, such as Actinobacteria, Acidobacteria, Firmicutes, or Proteobacteria, but have also demonstrated their ability to provide a significantly richer and more diverse representation of the taxonomic composition within these groups compared to the use of universal primers[29,36,37,41,42]. Extended depth of taxonomic structure was obtained due to the higher resolution of specific primers on lower taxonomic levels (e.g. genus), where universal primers often underperform. This is particularly important in biodiversity and functional estimations, which rely on accurate identification of microbial taxa on fine level.

Despite the demonstration of the usefulness of specific primers in current studies, their role in investigation of entire microbiomes has been underestimated, since they were rather used in the context of single taxa[37,41,42]. We believe that this approach limits the potential of specific primers, which due to their improved resolution, could supplement and clarify the results obtained with universal primers on microbiome level. Therefore, the core aim of our study was to develop a straightforward and efficient approach, which allows to combine advantages of universal and specific primers to obtain a more detailed picture of soil microbiome taxonomic structure. In this pursuit, we developed a novel Two-Step Metabarcoding (TSM) methodology, which we tested on agricultural soil. In this study, we focused on controlled microcosm experiments, allowing to reduce environmental variability and better isolate the methodological effects of primer choice on soil microbiome assessment. The first step of TSM involved traditional/classic metabarcoding with 16S rDNA, followed by second step of metabarcoding with phylum/class specific 16S rDNA primers for the most abundant taxa in the sample. The use of universal primers in the first step enabled to outline the microbiome structure and pinpoint predominant bacterial groups enabling design of the scaffold of the taxonomic structure. Subsequently in the second step, we employed specific 16S rDNA primers targeting key taxa to acquire a more nuanced understanding of the soil microbiome composition. Upon completing these analyses, we reconstructed taxonomic structure for both universal and specific primers, computed alpha diversity metrics and performed in silico functional profiling to compare the efficacy of taxa-specific primers against universal primers in revealing microbiome structure and diversity. The knowledge gained from this study could offer a fresh perspective and methodology for soil microbial biodiversity analysis, delivering a novel tool for more precise and in-depth depiction of the intricacies of subterranean life. Such enhanced understanding is pivotal for elucidating soil biological processes and among others developing sustainable agricultural practices.

Material & methods

Soil microcosm experimental setup

The experiment was performed using agricultural topsoil (30 cm) collected in Otwock County in Mazovian Voivodeship, Poland (GPS coordinates: 52°03′20″ N, 21°13′12″ E). Soil was air-dried and sieved through a 2 mm mesh to remove rocks and bigger particles and mixed thoroughly. Five PVC containers were filled with 600 g of soil and sealed with permeable membrane. The containers were transferred inside a climatic chamber at a constant temperature of 20 °C ± 2 °C for 120 days to stabilize the chemical and biological conditions between samples. The humidity of soil was kept at the level of 40% maximum water capacity. At the end of the experiment, after 120 days of incubation, soil samples were collected. Soil samples for chemical analysis were air-dried and sieved through a 2 mm mesh while fresh samples for DNA extraction were collected and kept in −80 °C for further analysis.

Soil chemical and physical properties

Total carbon, nitrogen and sulfur content were measured with CHNS analyzer Vario MAX Cube (Elementar Analysensysteme GmbH, Germany). Organic carbon was measured with loss-on-ignition method by heating soil to 550 °C for 8 hours and calculation of sample weight loss attributed to organic matter. For determination of total content of P, K, Mg, Ca, Fe samples were burned in a muffle furnace (at 450 °C for 12 h) and dissolved in a mixture of HNO_3 and $HClO_4$ acids (3:1 v/v). The total concentration of the elements in acid-digested soil solutions was determined using an inductively coupled plasma optical emission spectrometer ICP-OES Optima 7300DV (Perkin Elmer, United States) and was calculated on a soil dry mass basis. The pH was measured in a soil-to-water suspension (1:5). Soil granulometric composition was analyzed using the dry sieving method to determine the proportion of sand, silt, and clay fractions. Air-dried soil samples were gently crushed to break up aggregates and passed through a series of sieves with progressively smaller mesh sizes. The sieves, arranged in descending order of aperture size, were shaken mechanically for a fixed duration to ensure separation of particles based on size. The retained material on each sieve was weighed to calculate the percentage of sand (particles 2–0.05 mm), silt (0.002–0.05 mm), and clay (< 0.002 mm) fractions. The summary of soil properties was presented in Table 1.

Isolation of soil DNA

Isolation of DNA from soil was performed in five biological replicates. The extraction was carried out using samples weighing 500 mg, with the FastDNA®SPIN Kit for Feces DNA Extraction Kit and a bead beater (MP Biomedica, United States), following the manufacturer's instructions. The concentration and quality of the extracted DNA were evaluated using Qubit Fluorometer (Invitrogen, United States) and gel electrophoresis.

16S rDNA sequencing

Universal primers targeting V3-V4 regions of 16S rDNA gene were used for the PCR amplification on the level of the bacterial domain. Phylum/Class specific primers were selected according to the taxonomic classification from sequencing with universal primers, with a goal to represent in summary at least 80% of soil microbiome. Literature data were searched to find primers targeting these selected groups. The selected primers had to meet the following criteria: (i) a length suitable for Illumina sequencing with 2 × 300 bp chemistry (approximately 350–550 bp) and (ii) group specificity for the selected taxa. In the case of Planctomycetes, the product of primers described by Mühling et al. (2008) was 587 bp, which could potentially result in low-quality assembly due to excessive length[43]. For this reason, trials using both the original pair (Planctomycetes-long) and a modified version, where the universal primer was used as the forward primer (Planctomycetes-short) were conducted. Selected primers were presented in Table 2. All target sequences of primers used in this study were preceded by adapter sequence (F: 5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCT –3', R: 5' - GACTGGAG TTCAGACGTGTGCTCTTCCGATCT –3'). Total DNA extracted from soil microcosms was subjected to shotgun metagenomic sequencing to serve as an amplification-independent reference for microbial community composition. All PCR amplification products and total DNA were sequenced with Illumna MiSeq using 2 × 300 chemistry by Eurofins Genomics GmbH (Germany).

Chemical composition	
Element	**Total content [g/kg dry soil]**
Total C	16.540
Organic C	14.652
N	1.900
P	0.993
K	2.141
S	0.160
Mg	3.380
Ca	5.240
Fe	8.481
Chemical properties	
Parameter	**Value**
pH	7.64
EC	207.5 µS/cm
Granulometric structure	
Fraction	**Percentage [%]**
Sand (2.0–0.05 mm)	24
Silt (0.05–0.002 mm)	67
Clay (Ił < 0.002 mm)	9

Table 1. Chemical and physical properties of analyzed soil.

Targeted Taxon	Forward primer sequence (5'–3')	Reverse primer sequence (5'–3')	Product length (bp)	PCR conditions
Universal	CCTACGGGNGGCWGCAG[44]	GACTACHVGGGTATCTAATCC[44]	450	95 °C for 3 min 24 cycles x (98 °C for 20 s, 62 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min
Acidobacteria	GCTCAGAATSAACGCTGG[35]	TTACCGCGGCKGCTG[45]	500	95 °C for 3 min 24 cycles x (98 °C for 20 s, 62.5 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min
Actinobacteria	AAACTCAAAGGAATTGACGG[36]	CTTCCTCCGAGTTGACCC[36]	385	95 °C for 3 min 24 cycles x (98 °C for 20 s, 65 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min
Alphaprotebacteria	CGGTAATACGRAGGGRGYT[29]	GGTAAGGTTCTGCGCGTT[29]	421	95 °C for 3 min 26 cycles x (98 °C for 20 s, 63 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min
Firmicutes	CCGCGGTAATACGTAGGT[29]	ACCATGCACCACCTGTC[29]	500	95 °C for 3 min 28 cycles x (98 °C for 20 s, 62.5 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min
Planctomycetes long	GGCTGCAGTCGAGRATCT[43]	TGTGTGAGCCCCCGTCAA[43]	587	95 °C for 3 min 25 cycles x (98 °C for 20 s, 64 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min
Planctomycetes short	CAGCMGCCGCGGTAA[47]	TGTGTGAGCCCCCGTCAA[43]	410	95 °C for 3 min 24 cycles x (98 °C for 20 s, 60 °C for 20 s, 72 °C for 15 s) 72 °C for 1 min

Table 2. List of used primers and PCR reaction conditions used to generate amplicons.

Bioinformatic and statistical analysis

Reads obtained from sequencing were analyzed with use of QIIME2 framework version 2023.9[48]. Reads were processed with DADA2 plugin in independent pooling denoising mode and with the application of consensus chimera detection method. Taxonomic classification of obtained amplicon sequence variants (ASVs) was performed with the use of feature-classifier plugin. Classifiers for each pair of used primers were trained individually on SILVA SSU NR99 reference database version 138. Normalization of microbiome data was performed individually for every primer pair with the use of scaling with ranked subsampling (SRS) algorithm implemented in the Qiime2 framework as a q2-srs plugin. Analysis of alpha-diversity in samples was conducted with the Qiime2 plugin diversity alpha. The differences in taxonomic structure were analyzed with the use of the PERMANOVA method from vegan R package[49]. The results were presented with ggplot2 v3.3.5 and ggbiplot v0.55 packages[50]. Statistical analysis of alpha diversity Shannon index was performed with ANOVA using R.

To extract and classify 16S rRNA gene sequences from metagenomic data, the PhyloFlash pipeline (v3.4), was used[51]. Raw reads were first mapped to the SILVA SSU rRNA database (release 138.1) using BBMap. PhyloFlash then performed assembly of 16S rRNA gene fragments using SPAdes and assigned taxonomy via a Lowest Common Ancestor (LCA) approach. Beta-diversity patterns were analyzed using Non-metric Multidimensional Scaling (NMDS) based on Bray–Curtis dissimilarities, implemented in the *vegan* package in R. NMDS was performed using the *metaMDS* function on genus-level relative abundance data for selected taxonomic groups and visualized with ggplot2. Euclidean distances between group centroids were calculated using the *betadisper* function, which was also used to assess multivariate dispersion within groups. Mean distances of samples to their respective centroids were extracted to quantify the degree of dispersion for each sample type.

Results

Overall taxonomic structure and specific primers selection

The taxonomic structure of the samples reconstructed using universal primers (UPT – universal primers taxonomy) allowed to identify the most abundant phyla, which had more than 3% relative abundance in the samples. These phyla included Actinobacteria (25.56–38.39%), Acidobacteria (7.38–16.65%), Chloroflexi (10.71–21.42%), Firmicutes (3.94–4.81%), Planctomycetes (6.46–9.03%), and Proteobacteria (12.31–18.30%) (Fig. 1). Similarly, in the taxonomy obtained from shotgun sequencing (SMT – shotgun metagenome taxonomy), these phyla were identified as the most abundant, with a more pronounced presence of Actinobacteria (37.33–42.22%), lower abundance of Chloroflexi (6.45–8.58%), Planctomycetes (3.21–3.95%), and Acidobacteria (9.26–10.41%). The abundances of Firmicutes (3.19–4.31%) and Proteobacteria (15.25–18.67%) remained within a similar range.

Given the complexity of Proteobacteria and their well-established taxonomic divisions, specific primers are typically designed at the class level rather than the phylum level. For this reason, we chose primer pairs targeted at Alphaproteobacteria, the most abundant class (6.73–9.62%) in the investigated samples, as a marker for Proteobacteria. In the case of Chloroflexi, we were unable to identify primers specific to the entire group or significant classes, and for this reason, it was omitted from further analysis. For Acidobacteria and Actinobacteria, the selected primer pairs consisted of one specific primer and one universal primer. The length of the products did not exactly match the optimal value for 2 × 300 bp Illumina chemistry, thus in both cases, universal primers were substituted to obtain products of the appropriate length.

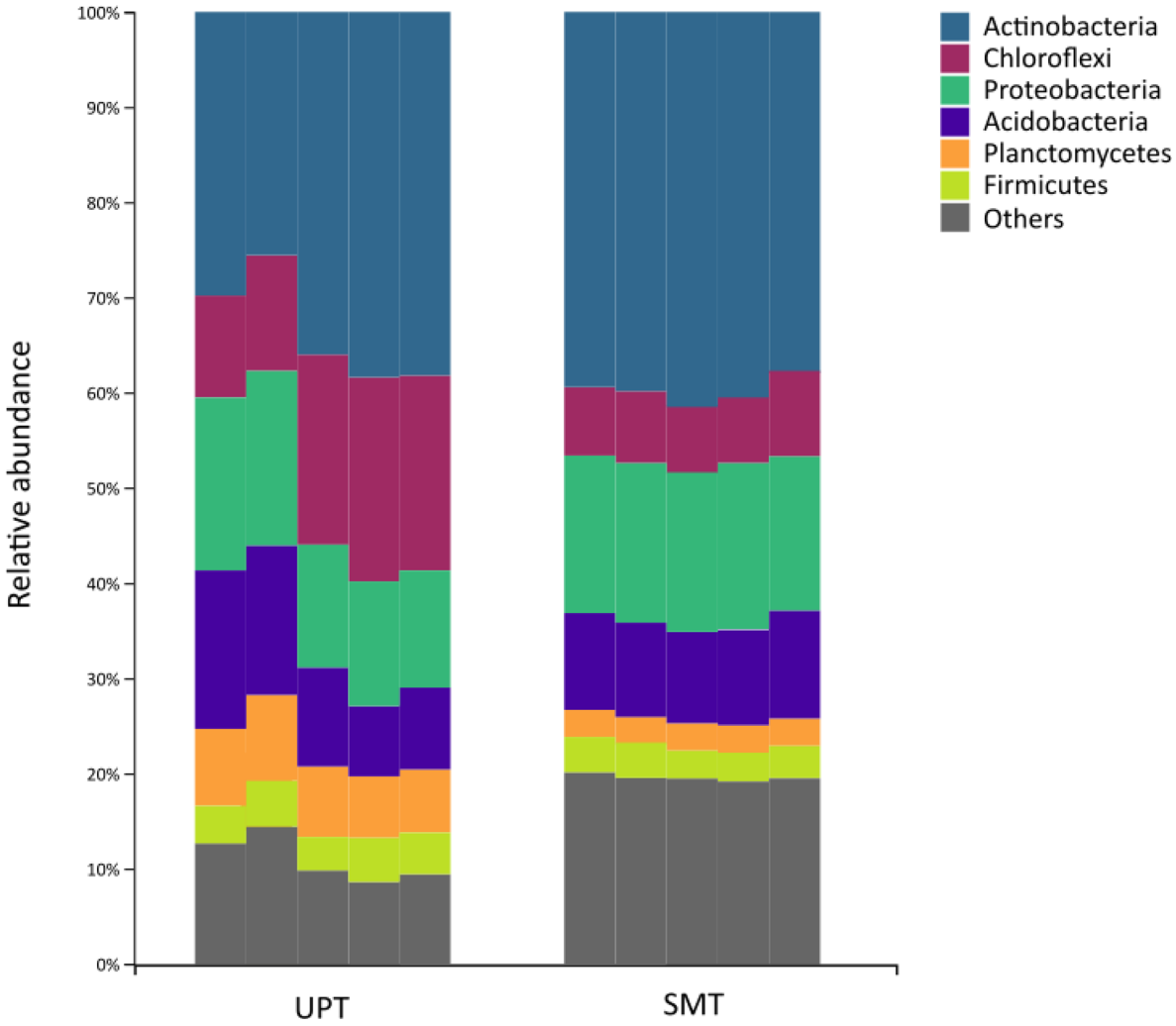


Fig. 1. Phyla relative abundance (%) in amplicons generated with 16S rDNA universal primers (UPT) and shotgun metagenome sequencing (SMT). Phyla under 3% relative abundance are presented as "Others". Five bars represent five biological samples collected from investigated soil.

Taxon	Reads	Non-chimeric (%)	Target sequences (%)
Acidobacteria	32 393	42.20	69.34
Actinobacteria	51 490	71.76	100
Alphaproteobacteria	67 424	85.11	93.49
Firmicutes	42 124	71.04	48.79
Planctomycetes-Long	59 412	17.85	100
Planctomycetes-Short	73 094	47.12	100

Table 3. Amplicons sequencing statistics for specific primers.

Microbial community composition across specific primer sets

Selected primers were used to obtain amplicons specific for the targeted taxa. For Planctomycetes, the results obtained with the Planctomycetes-Long primers were unsatisfactory due to a low level of non-chimeric reads (approximately 20%). Consequently, the Planctomycetes-Short primer pair was selected for further analysis. Sequencing of the other amplicons yielded a sufficient number of non-chimeric reads after filtering (Table 3). However, the performance of the primers varied in terms of precision (number of target sequences), with highest values of this parameter for Actinobacteria (100% target sequences), Planctomycetes, and Alphaproteobacteria (93.49% target sequences). In Acidobacteria amplicons, which contained 69.34% target sequences, there was a significant population of Proteobacteria in the sample (approximately 26.08%). The lowest number of target

sequences was observed for Firmicutes (48.79%), where approximately 50% of the sample was composed of Actinobacteria.

Taxonomic structure from universal and specific primers

For further analysis, we curated the obtained data to provide comparable conditions between the taxonomic structures obtained with universal primers (UPT) and specific primers (SPT – Specific Primers Taxonomy). First, we filtered the specific primers' amplicons to eliminate non-specific taxa and filtered the universal primers' results to obtain subsamples representing only selected taxa. Subsequently, we performed a rarefaction analysis to ensure that samples, after filtration, captured the full diversity. After data curation we reconstructed the microbiome structure on order level for both UPT and SPT to compare obtained results (Fig. 2). PERMANOVA statistical analysis showed that differences in taxonomic structure obtained using universal and specific primers were significant ($p = 0.0014–0.0084$) for every analyzed taxon. At the order level, the most noticeable differences were observed within the most abundant taxon in the microbiome – Actinobacteria. In SPT the two biggest groups were Propionibacteriales and Micrococales. Noticeable differences were also observed for MB-A2-108 and Solirubrobacteriales (abundant in UPT) or Streptomycetales, Pseudonocardiales, and Micromonosporales (abundant in SPT). Contrasting results were also observed for Planctomycetes. In all samples, the three most abundant taxa were the same (Tepidispherales, Pirellulales, Gemmatales); however, in UPT, Tepidispherales was dominant, whereas in SPT, their distribution was more even. In Acidobacteria, Alphaproteobacteria, and Firmicutes, the overall structure at the order level was more equal, with similar dominant taxa. In SMT, the most abundant taxa within each examined phylum/class were similar to those in UPT and SPT, whereas the most notable difference was the higher proportion of taxa below 3%.

Comparison of alpha-diversity metrics

To investigate biodiversity and alpha-diversity in the samples we have analyzed their richness by estimating the number of identified taxa on class, order, family, and genus level (Table 4). We compared it also with a taxonomy obtained from shotgun sequencing (SMT- shotgun metagenome taxonomy). Use of specific primers resulted in higher numbers for every tested taxa, both for family and genus. Significant differences were observed on genus level for Actinobacteria, Alphaproteobacteria and Firmicutes where number of identified taxa increased more than twofold. On class and order level use of the majority of specific primers also resulted in higher numbers of taxa, with only exception of Actinobacteria. Moreover the results showed the number of taxa observed in SMT was generally closer to SPT (being slightly higher) and noticeable surpassing numbers form UPT.

Next, the distribution of unique genera obtained using both types of primers was evaluated (Fig. 3). In Acidobacteria and Planctomycetes, the highest percentages of genera were shared between UPT and SPT (46% and 44%, respectively). In both taxa, the proportion of genera shared between specific primers (SPT) and metagenomic sequencing (SMT) was higher (17% and 12%, respectively) than between universal primers (UPT) and metagenomic sequencing (2% and 3%, respectively). The highest number of unique genera was observed in MT (22% and 20%, respectively), followed by SPT (7% and 10%) and UPT (0% and 5%).

In Alphaproteobacteria, the highest proportion of genera was shared among all three methods (25%), followed by those shared between SPT and SMT (22%), and those unique to either SPT (25%) or SMT (23%). In Actinobacteria and Firmicutes, the highest numbers of unique genera were found in SMT (36% and 35%, respectively), followed by genera shared between SMT and SPT (18% and 24%), and those unique to SPT (10% and 26%).

As a part of biodiversity analysis we also calculated alpha diversity metric Shannon index (SI) for universal/specific primers and shotgun metagenomics. Results were significantly different in almost all samples, with only exception of Acidobacteria (ANOVA $p = 0.0674$), where it spanned between 3.57 and 4.15 (Fig. 4). The most significant difference was observed for Alphaproteobacteria (ANOVA $p = 2e\text{-}16$), where SI was almost twofold higher for SPT (5.31–5.44) and SMT (5.26–5.39) than for UPT (3.04–3.28). Also for Firmicutes there was a significant (ANOVA $p = 3.1e\text{-}14$) increase in SI for SPT samples (4.59–4.88) and SMT (4.44–4.64) in comparison to UPT (2.54–2.96). In Actinobacteria SI for SMT was significantly higher than UPT and SPT (ANOVA $p = 2.22e\text{-}8$). Planctomycetes was the only group in which the statistically significant difference (ANOVA $p = 6.22e\text{-}13$) was in favor of higher SI in UPT samples (3.32–3.58) than in SPT (2.79–2.98) or SMT (2.92–3.14).

Comparison with shotgun metagenomic and beta-diversity analysis

To investigate beta-diversity patterns across the obtained taxonomic groups, Non-metric Multidimensional Scaling (NMDS) based on Bray–Curtis dissimilarities was performed (Fig. 5). The analysis revealed an overall higher level of similarity between SMT and SPT. This was particularly evident for Planctomycetes, where all SPT samples were entirely confined within the SMT confidence ellipse, and the Euclidean distance between the SMT and SPT centroids (0.253) was notably shorter than between SMT and UPT (0.593). In the case of Acidobacteria, overlapping confidence ellipses were observed between SMT and SPT, with centroid distances of 0.163 (SMT–SPT) and 0.254 (SMT–UPT). For Actinobacteria and Alphaproteobacteria, no overlap in confidence ellipses was observed; however, the SPT centroids were closer to SMT (0.166 and 0.387, respectively) than those of UPT (0.807 and 0.519, respectively). The only exception was observed for Firmicutes, where the UPT centroid was closer to SMT (0.162) than the SPT centroid (0.428).

The NMDS analysis also revealed a consistent pattern across all examined taxonomic groups regarding the dispersion of samples obtained with different methods. For each phylum/class, significant differences in dispersion were observed among SMT, SPT, and UPT, statistically tested with pairwise PERMANOVA (Table 5). The significantly most tightly clustered samples were observed in SPT, with the lowest mean Euclidean distances to centroids (Acidobacteria: 0.039; Alphaproteobacteria: 0.046; Firmicutes: 0.069; Planctomycetes: 0.033). Only in case of Actinobacteria the dispersion in SPT (0.074) and UPT (0.084) were not significantly different, followed

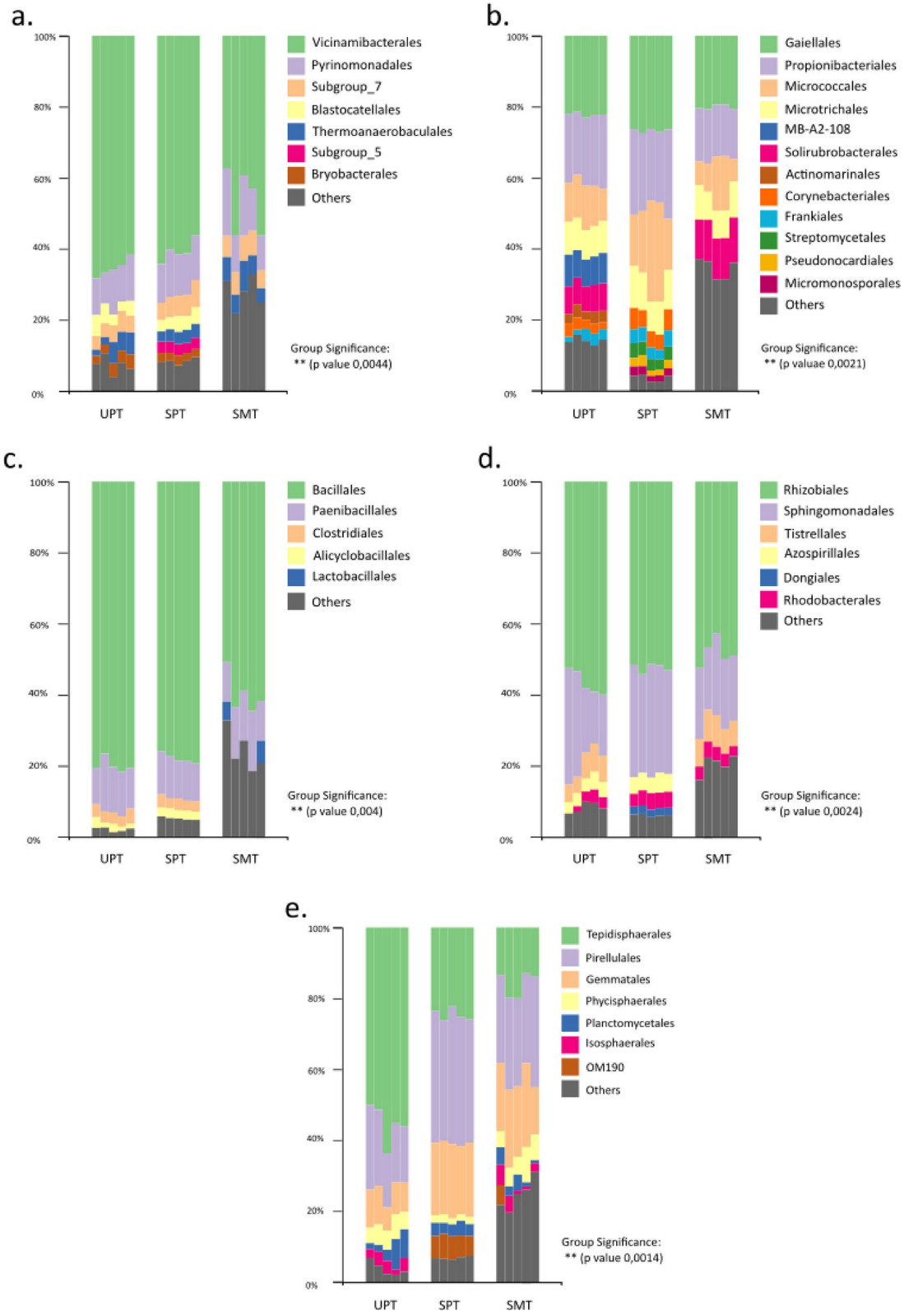


Fig. 2. Comparison of order relative abundance (%) in amplicons generated with 16S rDNA universal (UPT) specific (SPT) 16S rDNA primers and shotgun metagenome sequencing (SMT) for (**a**) Acidobacteria, (**b**) Actinobacteria, (**c**) Alphaproteobacteria, (**d**) Firmicutes and (**e**) Planctomycetes. In each panel "Universal" bars represent amplicons generated with 16S rDNA universal primers and "Specific" bars represent amplicons generated with specific primers for given phylum/class. Orders under 3% relative abundance are presented as "Others". Five bars represent five biological samples collected from investigated soil.

	Class			Order			Family			Genus		
	UPT	SPT	SMT	UPT	SPT	SMT	UPT	SPT	SMT	UPT	SPT	SMT
Acidobacteria	11	11	11	19	24	28	20	27	28	25	35	37
Actinobacteria	6	4	8	24	21	31	38	42	58	67	107	162
Alphaproteobacteria	-	-		12	19	23	21	42	37	44	111	92
Firmicutes	5	7	13	10	31	37	14	50	60	29	124	121
Planctomycetes	5	8	8	12	18	23	17	23	30	33	40	39

Table 4. Number of taxons identified on class, family and genus level for universal primers (UPT), specific primers (SPT) and shotgun metagenome taxonomy (SMT).

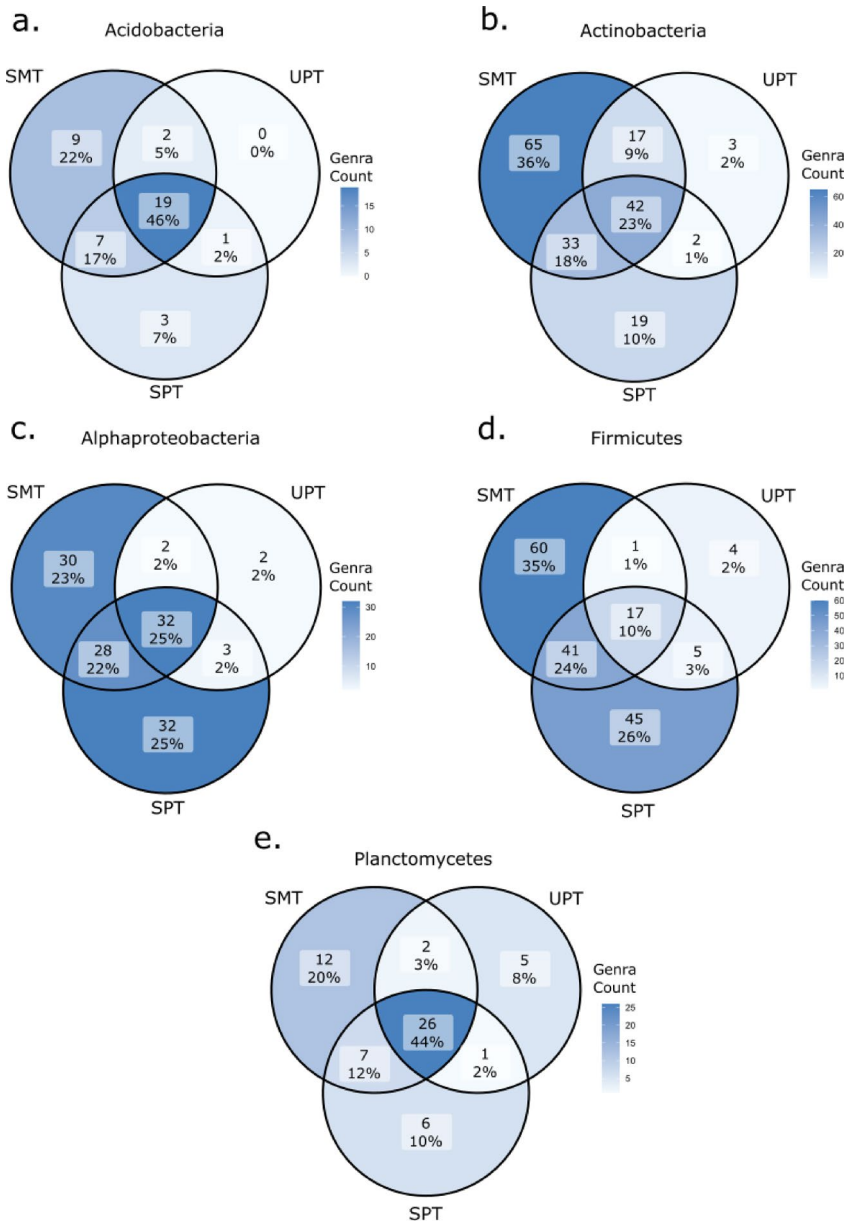


Fig. 3. The distribution of unique genera in soil samples for (**a**) Acidobacteria, (**b**) Actinobacteria, (**c**) Alphaproteobacteria, (**d**) Firmicutes and (**e**) Planctomycetes. Each panel represents the comparison between amplicons generated with 16S rDNA universal primers (UPT), specific primers for given phylum/class (SPT) and shotgun metagenome taxonomy (SMT). The number in the "UPT" field indicates the number of genera unique for universal primers, in "SPT" number of genera unique for specific primers. The number in the "SMT" field indicates the number of genera unique for metagenome taxonomy; intersection indicates the number of shared genera.

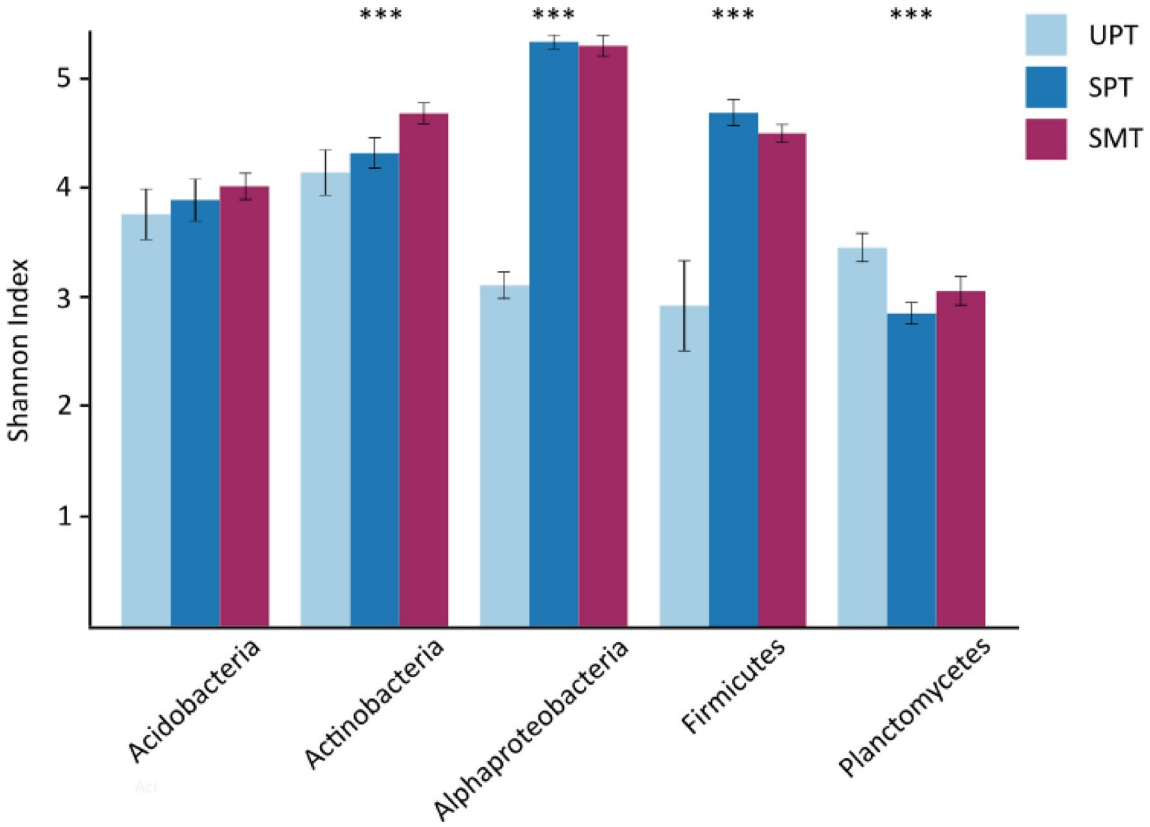


Fig. 4. Comparison of alpha diversity Shannon index between universal (UPT), specific primers (SPT) and shotgun metagenome sequencing (SMT). Each group of three bars represents results for amplicons generated with 16S rDNA universal primers, specific primers and shotgun metagenome for given phylum/class. Asterisk above groups indicate results of statistical analysis with ANOVA (*** - $p < 0.001$, ** - $p \leq 0.01$, * $p \leq 0.05$).

by highest dispersion in SMT (0.127). In Alphaproteobacteria, Firmicutes, and Planctomycetes, the highest dispersion was recorded in SMT (0.191; 0.285; 0.179, respectively), with intermediate levels in UPT (0.118; 0.147; 0.125). In Acidobacteria, no significant differences in dispersion were observed between UPT (0.089) and SMT (0.091).

Discussion

Phylum- or class-specific primers have been successfully used in many studies in the field of microbial ecology across various environments. For example, primers for Actinobacteria have been used to elucidate the microbiome complexity of this taxon in samples from the Atacama Desert, West Istria Sea marine sediments, and even fish eggs[52–54]. Acidobacteria-specific primers have been used to analyze the microbiome associated with black soil in northeast China and to study the soil Acidobacteria community of soybean in Amazon Forest[42,55]. However, in previous research using specific primers, the goal was typically to investigate a single taxon, separate from the context of the entire microbiome and without comparing the efficiency of specific primers to universal primers. In this study, the methodology that allows the use of specific primers alongside universal primers to create a more detailed picture of the soil microbiome was developed. With this approach, specific primers can be used to perform more comprehensive studies.

The microcosm-based experimental design was applied in this study to eliminate environmental heterogeneity, which is inherent in field samples, and thus better isolate the methodological effects of primer selection on microbial diversity assessments. By maintaining controlled and consistent conditions across replicates, confounding variables such as soil structure, moisture, plant-microbe interactions, and nutrient availability were minimized. This ensured that observed differences in microbial profiles could be attributed more confidently to the primer sets used rather than environmental heterogeneity.

We observed that the taxonomic structures of phyla/classes obtained using specific primers are significantly different from those obtained with universal primers. Observed differences in taxonomic structure could be explained by the better resolution of specific primers, as demonstrated in numerous studies[37,52,55]. According to the literature, the highest level of bias and underperformance in case of universal primers is observed on lower levels of taxonomic structure, especially genus, which could lead to recovery of limited numbers of taxon, favoring the most abundant ones[27,34,37]. This is supported by results of our study, since use of specific primer pairs was correlated with a significantly higher number of identified genera in comparison to universal primers.

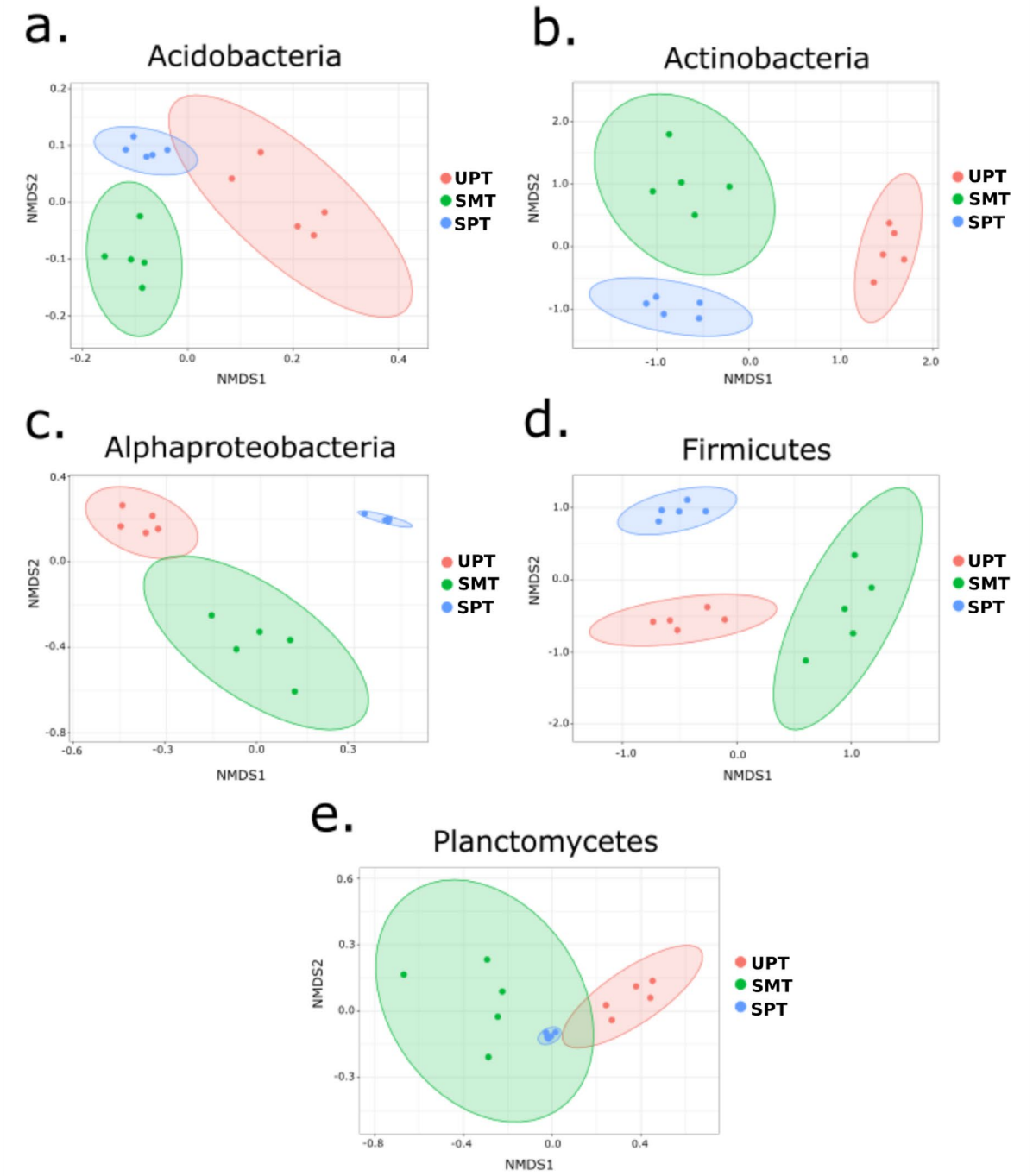


Fig. 5. The NMDS plots generated for (**a**) Acidobacteria, (**b**) Actinobacteria, (**c**) Alphaproteobacteria, (**d**) Firmicutes and (**e**) Planctomycetes. Each panel represents the comparison between amplicons generated with 16S rDNA universal primers (UPT), specific primers for given phylum/class (SPT) and shotgun metagenome sequencing (SMT). The ellipsis around points show the 95% confidence interval.

The most spectacular results were observed in the case of Firmicutes (124 genera in SPT vs. 29 in UPT) and Alphaproteobacteria (111 genera in SPT vs. 44 in UPT).

In our study, shotgun metagenomic sequencing was performed alongside amplicon-based profiling, allowing a direct comparison of universal and phylum-specific primer results against an amplification-free benchmark. The Two-Step Metabarcoding approach, combining universal and group-specific primers, recovered a significantly higher number of ASVs and more diverse taxa compared to the universal primer alone, and more importantly, the community structure derived from this combined approach showed closer similarity to the metagenomic profiles. In contrast, the universal primer set al.one failed to recover many genera that were

	SMT vs. UPT	SMT vs. SPT	SPT vs. UPT
Acidobacteria	0.0783	0.0151	0.00124
Actinobacteria	0.0321	0.0212	0.0856
Alphaproteobacteria	0.0122	1.31e-5	2.41e-4
Firmicutes	0.0322	4.31e-3	3.22e-3
Planctomycetes	1.56e-3	6.76e-4	3.42e-4

Table 5. The results of pairwise PERMANOVA statistical analysis of sample dispersion in NDMS analysis. In columns were presented results for every pairwise comparison between universal (UPT), specific primers (SPT) and shotgun metagenome sequencing (SMT), number represent p value for given pair.

detected in the metagenome dataset, particularly from phyla such as Actinobacteria, Alphaproteobacteria, and Firmicutes known to be underrepresented due to primer mismatches[56–58]. These findings are consistent with previous studies reporting that group-specific primers provide deeper coverage for their target lineages and align more accurately with shotgun-based taxonomic distributions[13,59,60].

The metagenomic dataset thus served as an internal reference for evaluating amplicon bias, and our results clearly indicate that the Two-Step Metabarcoding (TSM) approach captures a taxonomic composition more reflective of the true microbial community. This is particularly significant in soil, where diversity is immense and many taxa are rare or uncultivated. The reduced within-treatment dispersion observed in the Two-Step datasets further reinforces the idea that this method yields a more stable and less stochastic view of community composition. Dispersion metrics such as Bray-Curtis and Aitchison distances are frequently used as indicators of technical consistency and methodological reliability[61,62]. Therefore, both in terms of richness and compositional fidelity, our findings suggest that the TSM approach not only expands the detectable diversity but also improves accuracy relative to metagenomic standards.

Among the most important results of this study is the taxa uniqueness analysis between universal primer-based (UPT) and specific primer-based (SPT) datasets. A critical condition for the usefulness of the presented TSM methodology is its ability to expand taxonomic coverage without sacrificing the information already captured by universal primers. Our results confirmed that this goal was achieved: a substantial portion of the taxa detected by specific primers were either unique to SPT or shared between UPT and SPT, while only a minority were exclusive to UPT. Importantly, this broader taxonomic coverage was validated against the shotgun metagenomic (metagenome) dataset, which served as a primer-free reference. The taxa recovered using the combined UPT and SPT approach showed higher overlap with the metagenome than UPT alone, demonstrating that specific primers helped recover phyla and genera that universal primers tend to under-detect.

This improved taxonomic resolution translated into significantly higher alpha diversity, as reflected in elevated Shannon Index values for microbiomes reconstructed using the combined amplicon data. Notably, these values were closer to those obtained from shotgun metagenomic sequencing, suggesting that the TSM approach provides a more faithful representation of true community diversity. This added value represents a key advantage of the method, as microbial biodiversity is one of the most critical indicators of soil health, fertility, and ecological resilience[63,64]. Tools that improve the accuracy and resolution of biodiversity estimation are therefore indispensable for applications in sustainable agriculture, ecosystem restoration, and environmental monitoring[65–68]. By enabling a more complete and reproducible microbial inventory, the TSM approach facilitates better-informed decisions regarding soil management, conservation strategies, and the assessment of ecosystem vulnerability in the face of climate change.

From a practical standpoint, the findings of this study offer valuable guidance for researchers in the field of soil microbiome analysis who must often balance methodological depth with cost-efficiency and scalability. While shotgun metagenomic sequencing offers the most comprehensive and unbiased view of microbial communities, it remains significantly more expensive and computationally demanding than targeted amplicon sequencing, particularly when high replication is needed for robust ecological studies[69–71]. The TSM approach introduced here represents a cost-effective middle ground. By combining universal primers with a small panel of phylum- or class-specific primers, researchers can substantially enhance taxonomic resolution and community coverage without the need to sequence entire genomes. This is particularly useful in soil environments, where microbial diversity is extremely high and many taxa are present at low abundance. Moreover, the increased reproducibility and lower dispersion observed with the TSM method suggest that it may reduce the number of replicates or sequencing depth required to detect biologically meaningful patterns, which translates into additional cost savings. The flexibility of the method also allows researchers to tailor primer combinations based on the ecological or functional focus of their study, for instance, targeting Actinobacteria in studies on antibiotic-producing taxa or Planctomycetes in nitrogen cycling research. Finally, because the method is compatible with standard Illumina workflows and does not require major changes in library preparation protocols, it is easy to adopt in both well-resourced and resource-limited settings.

In literature critical methodological considerations are being raised concerning the use of full-length adapter-tagged primers in the initial PCR step, which could potentially introduce amplification bias. Including adapters and barcodes during the first round of PCR has been shown to influence community composition due to altered primer binding efficiency or secondary structure effects[72]. This bias has led some studies to adopt a two-step PCR strategy, where an initial amplification is conducted with untagged primers, followed by a second, short-cycle PCR to add sequencing adapters and barcodes[16,73]. In our case, the employment of one-step PCR with primers containing Illumina overhangs, a method widely applied in protocols validated by the Earth Microbiome Project

and other large-scale soil microbiome initiatives[74]. Although minor amplification bias due to this approach cannot be entirely ruled out, low dispersion among replicates and high consistency with shotgun metagenome data were observed, indicating that any potential adapter-related bias was not dominant or did not obscure the community structure. Furthermore, all primer sets were treated identically in terms of amplification protocol, ensuring that any bias introduced by the use of tagged primers was systematically distributed across samples. Nonetheless, we agree this is an important consideration, and future work could implement a two-step PCR strategy to further minimize possible primer-induced distortion and strengthen methodological robustness.

Challenges for specific primers

Despite the advantages of specific primers, they are still not widely used in microbial ecology. This could be due to several challenges that need to be overcome. One major issue is the quality and size of databases on environmental microbiomes[74]. Despite constant development in the field of (meta)genomics, there is still a need for gathering novel data, especially from diverse environments[75,76]. The complexity and variability of conditions in various habitats cause challenges in primers design and data interpretation.

The expansion of databases through the deposition of new sequences also has significant implications for phylogenetic research, as it provides a more comprehensive reference for phylogenetic studies, leading to better taxonomic classification and identification of novel lineages[77,78]. Improved phylogenetic resolution can lead to the reclassification of microbial taxa, enhancing the accuracy of microbial taxonomy. Results obtained in our studies demonstrated the importance of accurate phylogeny for the quality of specific primer design. In cases where primers were designed for highly supported monophyletic groups (e.g., Actinobacteria, Alphaproteobacteria, Planctomycetes), the targeted sequences comprised approximately 100% of the sample. However, the highest number of non-target sequences (approximately 50%) was observed in Firmicutes, which is a polyphyletic group due to the unstable positions of Tenericutes and Fusobacteria within this clade[78]. This poses a challenge for accurate primer design and could result in non-specific amplifications, especially in environmental samples.

This result also indicates that the use of specific primers for quantitative analysis should be approached with caution. For instance, the aforementioned primers have been used to conduct qPCR analysis of Firmicutes communities in various environments[79]. While it has been demonstrated that this primer pair can provide enhanced information about Firmicutes biodiversity, it also amplifies a significant number of Actinobacteria sequences. Although this issue can be mitigated by filtering the results in community analysis, it may lead to unreliable results in quantitative approaches.

The development of new sequencing technologies could also help broaden the possibilities for primer design and selection. One of the main limitations of Illumina sequencing platforms regarding amplicons is the restriction on target sequence length. Even when using the 2×300 bp chemistry, the maximum length of the amplicon should be approximately 500–550 bp to ensure proper assembly of the sequences[80]. In this study, this problem was encountered with primers for Planctomycetes. This limitation allows for the use of primers spanning only 1–2 variable regions out of the total 9 present in the complete gene sequence[34]. Given that taxon-specific fragments could occur in various gene regions, this excludes a vital part of possible combinations from practical use. In the near future the progress in novel sequencing technologies, such as Oxford Nanopore Technology (ONT), could allow us to overcome this limitation by enabling the sequencing of fragments as long as 50,000 bp. With this approach, a much broader selection of specific 16S rDNA primers could become feasible[81]. This capability could be beneficial for the methodology presented in this study, as it provides a potential for enhancing second step efficiency and accuracy.

Data availability

The raw sequence data reported in this paper have been deposited in the NCBI Sequence Read Archive (SRA), BioProject ID: PRJNA1191376.

Received: 12 December 2024; Accepted: 4 September 2025
Published online: 29 September 2025

References

1. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & WIllerslev, E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050 (2012).
2. Rizk, S. M. et al. Culturable and unculturable potential heterotrophic Microbiological threats to the oldest pyramids of the Memphis necropolis, Egypt. *Frontiers Microbiology* **14**, 1167083 (2023).
3. Rosa, L. H. et al. DNA metabarcoding uncovers fungal diversity in soils of protected and non-protected areas on deception island, Antarctica. *Sci. Rep.* **10**, 21986 (2020).
4. Mohd Salleh, M. H., Esa, Y., Ngalimat, M. S. & Chen, P. N. Faecal DNA metabarcoding reveals novel bacterial community patterns of critically endangered Southern river terrapin, Batagur affinis. *PeerJ* **10**, e12970 (2022).
5. Joos, L. et al. Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genom.* **21**, 733 (2020).
6. Orgiazzi, A., Dunbar, M. B., Panagos, P., de Groot, G. A. & Lemanceau, P. Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biol. Biochem.* **80**, 244–250 (2015).
7. Hartmann, M. & Six, J. Soil structure and Microbiome functions in agroecosystems. *Nat. Reviews Earth Environ.* **4**, 4–18 (2023).
8. Burns, J. H., Anacker, B. L., Strauss, S. Y. & Burke, D. J. Soil microbial community variation correlates most strongly with plant species identity, followed by soil chemistry, Spatial location and plant genus. *AoB PLANTS.* **7**, plv030 (2015).
9. Philippot, L., Chenu, C., Kappler, A. & Rillig, M. C. Fierer, N. The interplay between microbial communities and soil properties. *Nat. Rev. Microbiol.* **22**, 226–239 (2024).
10. Paes da Costa. Soil fertility impact on recruitment and diversity of the soil Microbiome in sub-humid tropical pastures in Northeastern Brazil. *Sci. Rep.* **14**, 3919 (2024).
11. Liu, S. et al. Composition and diversity of soil microbial community associated with land use types in the Agro–Pastoral area in the upper yellow river basin. *Frontiers Plant. Science* **13**, 819661 (2022).

12. Pratibha, G. et al. Soil bacterial community structure and functioning in a long-term conservation agriculture experiment under semi-arid rainfed production system. *Frontiers Microbiology* **14**, 1102682 (2023).
13. Hong, X. et al. Metagenomic sequencing reveals the relationship between microbiota composition and quality of Chinese rice wine. *Sci. Rep.* **6**, 26621 (2016).
14. Bahram, M. et al. Structure and function of the global topsoil Microbiome. *Nature* **560**, 233–237 (2018).
15. Köninger, J. et al. Ecosystem type drives soil eukaryotic diversity and composition in Europe. *Glob. Change Biol.* **29**, 5706–5719 (2023).
16. Karimi, B. et al. Biogeography of soil bacterial networks along a gradient of cropping intensity. *Sci. Rep.* **9**, 3812 (2019).
17. Ortiz, A. & Sansinenea, E. Recent advancements for microorganisms and their natural compounds useful in agriculture. *Appl. Microbiol. Biotechnol.* **105**, 891–897 (2021).
18. Vejan, P., Abdullah, R., Khadiran, T. & Ismail, S. & Nasrulhaq boyce, A. Role of plant growth promoting rhizobacteria in agricultural Sustainability-A review. *Molecules* **21**, 573 (2016).
19. Mendes, R., Garbeva, P. & Raaijmakers, J. M. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* **37**, 634–663 (2013).
20. Hassan, S. et al. Environmental DNA metabarcoding: A novel contrivance for documenting terrestrial biodiversity. *Biology (Basel)* **11**, 1297 (2022).
21. Ficetola, G. F. & Taberlet, P. Towards exhaustive community ecology via DNA metabarcoding. *Mol. Ecol.* **32**, 6320–6329 (2023).
22. Baker, G. C., Smith, J. J. & Cowan, D. A. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Methods*. **55**, 541–555 (2003).
23. Lu, Y. Z., Ding, Z. W., Ding, J., Fu, L. & Zeng, R. J. Design and evaluation of universal 16S rRNA gene primers for high-throughput sequencing to simultaneously detect DAMO microbes and anammox bacteria. *Water Res.* **87**, 385–394 (2015).
24. Wang, Y., Tian, R. M., Gao, Z. M., Bougouffa, S. & Qian, P. Y. Optimal eukaryotic 18S and universal 16S/18S ribosomal RNA primers and their application in a study of symbiosis. *PLoS One*. **9**, e90053 (2014).
25. Yang, N. et al. Novel primers for 16S rRNA gene-based archaeal and bacterial community analysis in oceanic trench sediments. *Appl. Microbiol. Biotechnol.* **106**, 2795–2809 (2022).
26. Hongoh, Y., Yuzawa, H., Ohkuma, M. & Kudo, T. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. *FEMS Microbiol. Lett.* **221**, 299–304 (2003).
27. Hansen, M. C., Tolker-Nielsen, T., Givskov, M. & Molin, S. Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol. Ecol.* **26**, 141–149 (1998).
28. Brooks, J. P. et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* **15**, 66 (2015).
29. Pfeiffer, S. et al. Improved group-specific primers based on the full SILVA 16S rRNA gene reference database. *Environ. Microbiol.* **16**, 2389–2407 (2014).
30. Farris, M. H. & Olson, J. B. Detection of actinobacteria cultivated from environmental samples reveals bias in universal primers. *Lett. Appl. Microbiol.* **45**, 376–381 (2007).
31. Schmidt, A. et al. Shotgun metagenomics of soil invertebrate communities reflects taxonomy, biomass, and reference genome properties. *Ecol. Evol.* **12**, e8991 (2022).
32. Alteio, L. V. et al. A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biol. Biochem.* **160**, 108357 (2021).
33. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35**, 833–844 (2017).
34. Cai, L., Ye, L., Tong, A. H. Y., Lok, S. & Zhang, T. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLOS ONE*. **8**, 1–11 (2013).
35. Lee, S. H. & Cho, J. C. Group-specific PCR primers for the phylum Acidobacteria designed based on the comparative analysis of 16S rRNA gene sequences. *J. Microbiol. Methods*. **86**, 195–203 (2011).
36. Schäfer, J., Jäckel, U. & Kämpfer, P. Development of a new PCR primer system for selective amplification of actinobacteria. *FEMS Microbiol. Lett.* **311**, 103–112 (2010).
37. Zhao, K. et al. Actinobacteria associated with Chinaberry tree are diverse and show antimicrobial activity. *Sci. Rep.* **8**, 11103 (2018).
38. Yoon, S. H. et al. Introducing ezbiocloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* **67**, 1613–1617 (2017).
39. DeSantis, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
40. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–596 (2013).
41. Liu, J. et al. Diversity and distribution patterns of acidobacterial communities in the black soil zone of Northeast China. *Soil Biol. Biochem.* **95**, 212–222 (2016).
42. Sui, X. et al. Soil acidobacterial community composition changes sensitively with wetland degradation in Northeastern of China. *Frontiers Microbiology* **13**, 1052161 (2022).
43. Mühling, M., Woolven-Allen, J., Murrell, J. C. & Joint, I. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J.* **2**, 379–392 (2008).
44. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
45. Zhou, M. et al. Community competition is the microorganism feedback to sedimentary carbon degradation process in aquaculture tidal flats. *Frontiers Mar. Science* **9**, 880120 (2022).
46. Ashigar, M. A. & Majid, A. H. A. 16S rRNA metagenomic data of microbial diversity of pheidole decarinata Santschi (Hymenoptera: Formicidae) workers. *Data Brief.* **31**, 106037 (2020).
47. Wang, Y. & Qian, P. Y. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One*. **4**, e7401 (2009).
48. Bolyen, E. et al. Reproducible, interactive, scalable and extensible Microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
49. Oksanen, J. et al. Vegan: Community Ecology Package. (2024).
50. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* (Springer International Publishing, 2016).
51. Gruber-Vodicka Harald, R. & Seah Brandon, K. B. & Pruesse Elmar. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems* **5**, (2020). 10.1128/msystems.00920 – 20
52. Duran, R. et al. Exploring actinobacteria assemblages in coastal marine sediments under contrasted human influences in the West Istria sea, Croatia. *Environ. Sci. Pollut. Res.* **22**, 15215–15229 (2015).
53. Liu, Y. et al. Deciphering microbial landscapes of fish eggs to mitigate emerging diseases. *ISME J.* **8**, 2002–2014 (2014).
54. Idris, H., Goodfellow, M., Sanderson, R., Asenjo, J. A. & Bull, A. T. Actinobacterial rare biospheres and dark matter revealed in habitats of the Chilean Atacama desert. *Sci. Rep.* **7**, 8373 (2017).
55. Navarrete, A. A. et al. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. *FEMS Microbiol. Ecol.* **83**, 607–621 (2013).

56. Soergel, D. A. W., Dey, N., Knight, R. & Brenner, S. E. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* **6**, 1440–1444 (2012).
57. Tremblay, J. et al. Primer and platform effects on 16S rRNA Tag sequencing. *Frontiers Microbiology* **6**, 00771 (2015).
58. Francioli, D., Lentendu, G., Lewin, S. & Kolb, S. DNA Metabarcoding for the Characterization of Terrestrial Microbiota-Pitfalls and Solutions. *Microorganisms* 9, (2021).
59. Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* **469**, 967–977 (2016).
60. Abellan-Schneyder, I. et al. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere* 6, (2021).
61. Weiss, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
62. Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-Based studies. *PLOS ONE*. **6**, e27310 (2011).
63. Girvan Martina, S. et al. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Appl. Environ. Microbiol.* **69**, 1800–1809 (2003).
64. Van Der Heijden, M. G. A., Bardgett, R. D. & Van Straalen, N. M. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).
65. Delgado-Baquerizo, M. et al. Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* **7**, 10541 (2016).
66. Mendes, R. et al. Deciphering the rhizosphere Microbiome for Disease-Suppressive bacteria. *Science* **332**, 1097–1100 (2011).
67. Fierer, N. Embracing the unknown: disentangling the complexities of the soil Microbiome. *Nat. Rev. Microbiol.* **15**, 579–590 (2017).
68. Maron Pierre-Alain et al. High microbial diversity promotes soil ecosystem functioning. *Appl. Environ. Microbiol.* **84**, e02738–e02717 (2018).
69. Edwin Niranjana et al. Consistent microbial insights across sequencing methods in soil studies: the role of reference taxonomies. *mSystems* 0, e01059-24 (2025).
70. Liu, Y. X. et al. A practical guide to amplicon and metagenomic analysis of Microbiome data. *Protein Cell.* **12**, 315–330 (2021).
71. Usyk, M. et al. Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. *Cell. Rep. Methods*. **3**, 100391 (2023).
72. Berry David, B. M. & Karim, W. Loy alexander. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* **77**, 7846–7849 (2011).
73. Terrat, S. et al. Mapping and predictive variations of soil bacterial richness across France. *PLOS ONE*. **12**, e0186766 (2017).
74. Thompson, L. R. et al. A communal catalogue reveals earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
75. Anthony, W. E. et al. From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology. *Environ. Microbiome*. **19**, 56 (2024).
76. Lobanov, V., Gobet, A. & Joyce, A. Ecosystem-specific microbiota and Microbiome databases in the era of big data. *Environ. Microbiome*. **17**, 37 (2022).
77. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
78. Zhu, Q. et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains bacteria and archaea. *Nat. Commun.* **10**, 5477 (2019).
79. Abdelmoneim, T. K., Mohamed, M. S. M., Abdelhamid, I. A., Wahdan, S. F. M. & Atia, M. A. M. Development of rapid and precise approach for quantification of bacterial taxa correlated with soil health. *Front. Microbiol.* **13**, 1095045 (2022).
80. Hu, T., Chitnis, N., Monos, D. & Dinh, A. Next-generation sequencing technologies: an overview. *Hum. Immunol.* **82**, 801–811 (2021).
81. Zhang, T. et al. The newest Oxford nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl. Environ. Microbiol.* **89**, e0060523 (2023).

Acknowledgements

This work was funded by the National Centre for Research and Development (Poland) in the frame of the LIDER program, grant no. LIDER/13/0051/L-11/NCBR/2020.

Author contributions

Conceptualization: MM, KD-A; Data curation: MM; Funding acquisition: KD-A; Investigation: MM, KDA; Methodology: MM, KG, MM-H, KD-A; Project administration: KD-A; Resources: KD-A; Supervision: KD-A; Validation: MM; Visualization: MM; Writing—original draft: MM, KD-A; Writing—review & editing: MM, KG, MM-H, KD-A. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.D.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025