



OPEN ACCESS

EDITED BY

Yongxin Lin,
Fujian Normal University,
China

REVIEWED BY

Rui Tao,
Shaoxing University,
China
Chris Sedlacek,
University of Vienna,
Austria

*CORRESPONDENCE

Willm Martens-Habbena
✉ w.martenshabbena@ufl.edu

SPECIALTY SECTION

This article was submitted to
Microbiological Chemistry and
Geomicrobiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 09 January 2023

ACCEPTED 30 January 2023

PUBLISHED 23 February 2023

CITATION

Zhao J, Rodriguez J and
Martens-Habbena W (2023) Fine-scale
evaluation of two standard 16S rRNA gene
amplicon primer pairs for analysis of total
prokaryotes and archaeal nitrifiers in differently
managed soils.
Front. Microbiol. 14:1140487.
doi: 10.3389/fmicb.2023.1140487

COPYRIGHT

© 2023 Zhao, Rodriguez and Martens-
Habbena. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Fine-scale evaluation of two standard 16S rRNA gene amplicon primer pairs for analysis of total prokaryotes and archaeal nitrifiers in differently managed soils

Jun Zhao, Jonathan Rodriguez and Willm Martens-Habbena*

Fort Lauderdale Research and Education Center, University of Florida, Davie, FL, United States

The advance of high-throughput molecular biology tools allows in-depth profiling of microbial communities in soils, which possess a high diversity of prokaryotic microorganisms. Amplicon-based sequencing of 16S rRNA genes is the most common approach to studying the richness and composition of soil prokaryotes. To reliably detect different taxonomic lineages of microorganisms in a single soil sample, an adequate pipeline including DNA isolation, primer selection, PCR amplification, library preparation, DNA sequencing, and bioinformatic post-processing is required. Besides DNA sequencing quality and depth, the selection of PCR primers and PCR amplification reactions arguably have the largest influence on the results. This study tested the performance and potential bias of two primer pairs, i.e., 515F (Parada)-806R (Apprill) and 515F (Parada)-926R (Quince) in the standard pipelines of 16S rRNA gene Illumina amplicon sequencing protocol developed by the Earth Microbiome Project (EMP), against shotgun metagenome-based 16S rRNA gene reads. The evaluation was conducted using five differently managed soils. We observed a higher richness of soil total prokaryotes by using reverse primer 806R compared to 926R, contradicting to *in silico* evaluation results. Both primer pairs revealed various degrees of taxon-specific bias compared to metagenome-derived 16S rRNA gene reads. Nonetheless, we found consistent patterns of microbial community variation associated with different land uses, irrespective of primers used. Total microbial communities, as well as ammonia oxidizing archaea (AOA), the predominant ammonia oxidizers in these soils, shifted along with increased soil pH due to agricultural management. In the unmanaged low pH plot abundance of AOA was dominated by the acid-tolerant NS-Gamma clade, whereas limed agricultural plots were dominated by neutral-alkaliphilic NS-Delta/NS-Alpha clades. This study stresses how primer selection influences community composition and highlights the importance of primer selection for comparative and integrative studies, and that conclusions must be drawn with caution if data from different sequencing pipelines are to be compared.

KEYWORDS

16S rRNA gene amplicon Illumina sequencing, metagenomics, soil, ammonia-oxidizing archaea, primer evaluation, microbial communities, land use change

1. Introduction

Soil harbors high abundance and diversity of prokaryotic microorganisms, with 1 g of soil containing up to 10 billion microbial cells (Torsvik and Øvreås, 2002) and 10^3 – 10^6 phylotypes (Bickel and Or, 2020), including many microorganisms that play critical biogeochemical roles (Crowther et al., 2019). With continuous expansion of prokaryotic diversity and the development of next-generation sequencing tools (Hug et al., 2016), it is important to select a proper pipeline for in-depth analysis of microbial communities. Amplicon-based high throughput sequencing targeting 16S rRNA genes is the most frequently used approach to generating overall profiles of prokaryotic diversity in various environments (Yarza et al., 2014), but the key aspects, e.g., primer design, have been regularly modified and optimized (Fischer et al., 2016; Walters et al., 2016; Fadeev et al., 2021). To facilitate systematic analysis of global microbial taxonomic diversity on the planet and to allow data comparison between different environments and studies, the Earth Microbiome Project (EMP) developed several standard protocols for environmental microbiome analysis, including one for prokaryotic microorganisms using paired-end 16S rRNA gene sequencing on the Illumina platform (Thompson et al., 2017). Currently, two sets of primers for 16S rRNA genes are recommended in the EMP standard pipeline,¹ i.e., primers 515F–806R for the V4 region (Apprill et al., 2015; Parada et al., 2016), and primers 515F–926R, targeting V4–V5 region (Quince et al., 2011; Parada et al., 2016) of 16S rRNA gene. Both primers have been widely used in previous studies, although *in silico* evaluation of these two primers indicated 515F–926R covers higher prokaryotic diversity than 515F–806R (Parada et al., 2016). However, *in vitro* test using mock community (Fouhy et al., 2016; Parada et al., 2016; Wear et al., 2018) or field samples (Fischer et al., 2016; Thijs et al., 2017; Wear et al., 2018; Fadeev et al., 2021; Bai et al., 2022) is also necessary for comprehensive evaluation of the primer performance. Indeed, it has been demonstrated that *in silico* evaluation of primers does not always reflect *in vitro* performance obtained from experimental tests (Fischer et al., 2016). Especially, evaluating primers using field samples is important in addition to using mock communities, as the complexity is expected to be higher in soils (Rappé and Giovannoni, 2003; Pedrós-Alió and Manrubia, 2016; Lloyd et al., 2018), which could magnify the primer bias toward specific taxa (Eloe-Fadrosh et al., 2016; Parada et al., 2016). The increasing availability of shotgun metagenomes, deemed to be less prone to compositional bias than PCR-based methods, now allows for further ground-truthing of amplicon sequencing approaches (Eloe-Fadrosh et al., 2016).

Cultivation-independent tools have rapidly expanded the understanding of the diversity, evolution and ecology of functional microbial guilds, particularly within the domain of *Archaea*, which also play an important role in sustaining the terrestrial microbial biodiversity and biogeochemistry (Offre et al., 2013; Baker et al., 2020). Importantly, the ammonia oxidizing archaea (AOA) within the archaeal class of *Nitrososphaeria* (previously known as phylum *Thaumarchaeota*) are widespread and often account for the highest proportion of archaea in soil (Bates et al., 2011; Wang et al., 2015; Zhao et al., 2015; Liu et al., 2019), and are critical for terrestrial nitrogen cycling (Gubry-Rangin et al., 2010; Zhang et al., 2010, 2012; Hink et al., 2018; Zhao et al., 2020). Oxygen availability, pH and temperature have been recognized to drive

the radiative evolutionary diversification of terrestrial AOA lineages and their differential niche adaptation (Gubry-Rangin et al., 2015; Ren et al., 2019; Yang et al., 2021), leading to a consortium of phylogenetically distinct phylotypes co-existing and competing in the same soil (Gubry-Rangin et al., 2011; Zhao et al., 2015; Huang et al., 2021). Due to the low rate of horizontal gene transfer within *Nitrososphaeria* (Sheridan et al., 2020), the congruence is well supported between genome-resolved phylogeny and some single-copy marker gene phylogenies, including the 16S rRNA gene (Oton et al., 2016; Sheridan et al., 2020; Wang et al., 2021). So far, 12 putative family-level AOA clades can be confidently identified from established 16S rRNA gene database (Wang et al., 2021), and continuous expansion of metagenome assembled genomes (MAGs) of AOA allow characterization of novel lineages using 16S rRNA genes as biomarker. However, there is little information about the potential bias toward specific AOA clade by using different 16S rRNA gene primers.

Environmental disturbances due to anthropogenic activity can drastically alter microbial diversity and ecosystem functions in terrestrial ecosystems (Rillig et al., 2019, 2021). Particularly, the expansion of cropping systems by the conversion of native land will change soil physiochemical properties, leading to shift in microbial diversity and nutrient cycling (Zhou et al., 2020), including nitrogen transformations (Zhao et al., 2020, 2021; Klimasmith and Kent, 2022; Sun et al., 2022). For instance, it is a common practice to modify soil pH to neutral or near-neutral condition for maximum crop productivity, e.g., by liming of acidic forest soil or reclamation of saline-sodic soil (Zhao et al., 2020; Sun et al., 2021). Soil pH is recognized as the main driver for microbial assembly and distribution (Rousk et al., 2010; Tripathi et al., 2018), and the major determinant for niche differentiation of soil AOA (Gubry-Rangin et al., 2011). Other important environmental parameters, such as soil moisture and organic matter, can also be affected by agricultural managements and affect soil microbial distribution and activity (Frindte et al., 2019; Domeignoz-Horta et al., 2021), including ammonia oxidizers (Thion and Prosser, 2014). While correlation analysis is often used to predict the main influencing factors for soil microbial community change after land use conversion, the potential effect of selected primers on identifying these environmental factors remains less clear.

In this study, we examined the performance of the two standard primer pairs targeting the 16S rRNA gene for analyzing total soil microbiome composition and the specific functional guild of nitrifying archaea in soils under different land uses. The 16S rRNA genes retrieved from shotgun metagenome data, which should have no primer-induced bias, were analyzed as a reference point to compare with the amplicon sequencing results.

2. Materials and methods

2.1. Soil sample collection

A total of 90 soil samples were used for this research. Soils were collected in five different plots under different land uses in the Everglades Agricultural Area (EAA) including an unmanaged native plot (plot 1), a sugarcane-spinach-fallow-sugarcane rotation plot (plot 2), a sugarcane-sweetcorn-rice-sugarcane rotation plot (plot 3), a sugarcane-sweetcorn-fallow-sugarcane rotation plot (plot 4), and a year-round sugar cane plantation plot (plot 5). Plots 2–5 had been historically cultivated with sugar cane for 2 (plot 5) or 3 (plot 2–4) years. Starting

¹ <https://earthmicrobiome.org/protocols-and-standards/16s/>

from January 2017, plot 2 was planted with spinach till May 2017, followed by a 28-week fallow period before sugar cane plantation in December 2017. Plot 3 was planted with sweet corn from January to May 2017, followed by rice cultivation from May to October 2017 and an 8-week fallow period before sugar cane plantation in December 2017. Plot 4 was planted with sweet corn from January to May 2017, also followed by a 28-week fallow before sugar cane plantation in December 2017. Plot 5 was cultivated with sugar cane for a third consecutive year. Soils were collected during May 2017 to April 2018 at 2–3-month intervals. The area of each plot is $\sim 20,000\text{ m}^2$ ($\sim 100 \times 200\text{ m}$ square). Biological triplicates were sampled in each plot, each containing 9 random cores of 10-cm topsoil within a subplot of $\sim 10 \times 10\text{ m}$ square, and the samples were immediately transported to the laboratory on ice. Part of the soil samples were stored in -20°C before DNA extraction.

2.2. Soil physiochemistry measurements

Soil pH was measured in slurry with a soil-to-water ratio of 1:2 (10 g of soil in 20 ml deionized water) after shaking for 1 h at room temperature. Soil moisture was determined by calculating weight loss after drying 20 g fresh soil at 105°C for 24 h. Soil ammonium and nitrate were extracted with 2 M KCl by mixing 5 g fresh soil with 45 ml of KCl solution, followed by homogenization on a shaker at 50 rpm for 1 h and centrifuge at $3,000\text{ g}$ for 10 min. The supernatant was filtered through a $0.45\ \mu\text{m}$ Nylon filter and the ammonium and nitrate concentrations were then determined colorimetrically following salicylate method (Bower et al., 1980) and VCl_3 -Griess method (García-Robledo et al., 2014), respectively. Total phosphorus (TP) and available phosphorus (AP) concentrations were measured using a continuous segmented flow analyzer (Auto Analyzer 3, Seal Analytical Inc., Mequon, WI, United States) following the EPA method 365.1 (Determination of Phosphorus by Semi-Automated Colorimetry, Revision 2.0). Soil organic matter (OM) content was measured by calculating loss on ignition and dissolved organic carbon (DOC) content was measured using a TOC analyzer (Shimadzu, Kyoto, Japan).

2.3. DNA extraction and amplicon sequencing

Soil DNA was extracted using 0.25 g of soil by DNeasy PowerSoil Kit (QIAGEN) following manufacturer's instruction, except that bead-beating was performed twice using a FastPrep-24 (MP Biomedicals, Santa Ana, CA, United States) at 4.5 m s^{-1} for 15 s. Soil DNA extract was diluted to a concentration of $10\text{ ng }\mu\text{L}^{-1}$ to minimize variability during PCR and stored at -20°C before PCR amplification.

Wet-lab work for amplicon sequencing including PCR amplification, library preparation and DNA sequencing was performed at the Environmental Sample Preparation and Sequencing Facility at Argonne National Laboratory. Amplification of 16S rRNA gene fragments by PCR followed EMP standard pipeline, using either primers 515F–806R (Apprill et al., 2015; Parada et al., 2016) or 515F–926R (Quince et al., 2011; Parada et al., 2016). The forward primer 515F (GTGYCAGCMGCCGCGGTAA) was the same for both primer pairs while the reverse primers were different (806R: GGACTACNVGGGTWTCTAAT and 926R: CCGYCAATYMTTTRAGTTT), resulting in amplicon size of ~ 253 and $\sim 374\text{ bp}$ length, respectively. Adapter sequences were fused to the

forward primer and reverse primer contained 12 base barcode sequence. PCR reactions consisted of $9.5\ \mu\text{l}$ of DNase-free PCR Water (MoBio, Carlsbad, CA, United States), $12.5\ \mu\text{l}$ of 2x AccuStart II PCR ToughMix (QuantaBio, Beverly, MA, United States), $1\ \mu\text{l}$ of 200 pM forward primer and reverse primer, and $1\ \mu\text{l}$ of template DNA. The PCR thermocycling conditions were as follows: denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 45 s, 50°C for 60 s, and 72°C for 90 s, and final extension at 72°C for 10 min. PCR products were quantified using PicoGreen (Invitrogen, Carlsbad, CA, United States) in a 96-well microplate reader (Infinite 200 PRO, Tecan, Grödig, Austria). PCR products were pooled in equimolar ratios, purified using AMPure XP Beads (Beckman Coulter, Brea, CA, United States), and quantified using Qubit DNA quantification kit (Invitrogen, Carlsbad, CA, United States). PCR products were diluted to 2 nM for denaturation, and then diluted to 6.75 pM final concentration with 10% PhiX spike. Sequencing was performed on an Illumina MiSeq instrument using 2×150 (for 515F–806R PCR products) or 2×250 (for 515F–926R PCR products) reagent kit (Illumina, San Diego, CA, United States). The raw 16S rRNA gene amplicon sequencing data were deposited to NCBI under BioProject accession number PRJNA831877.

Paired-end reads were processed by QIIME2 version 2020.11 (Caporaso et al., 2010). The DADA2 plug-in with standard settings was used for quality filtering, denoising, paired-end read merging, chimera removal, dereplication and generation of amplicon sequence variants (ASVs; Callahan et al., 2016). This resulted in $89,103 \pm 3,615$ and $21,752 \pm 529$ reads per sample from primers of 515F–806R and 515F–926R, respectively, which were then taxonomically classified against SILVA database release 138.1 (Quast et al., 2013) using the scikit-learn classifier (Pedregosa et al., 2011) embedded in QIIME2. For more detailed taxonomic classification, sequences of ASVs belonging to class *Nitrososphaeria* (excluding non-AOA group 1.1c) were picked out and re-classified into putative family-level clades using BLASTN against a custom AOA database (Wang et al., 2021). A total of $4,302 \pm 217$ and $1,130 \pm 50$ reads AOA per sample were obtained from primers of 515F–806R and 515F–926R, respectively.

2.4. Metagenome shotgun sequencing

It is expected that the results generated from shotgun metagenome data should not be affected by primer-induced bias and thus can be deemed as proxy of the actual community composition. Therefore, to assess the accuracy of the different primer pairs for analyzing microbial composition in soils, the results of amplicon sequences were compared with 16S rRNA genes obtained from shotgun metagenomes. Shotgun metagenome sequencing was conducted in three of the five soil plots (plot 1, 3 and 5) in July and December of 2017 at the Department of Energy Joint Genome Institute using standard ($2 \times 151\text{ bp}$) library preparation and sequencing workflow for Illumina NovaSeq S4 instrument platform and yielded between $80,108,594 \pm 4,478,157$ per sample high quality metagenome reads (read length $> 150\text{ bp}$, Q-score > 30). A total of $12,420 \pm 671$ 16S rRNA gene reads per sample were retrieved from metagenomes by SortMeRNA v4.0.086 (Kopylova et al., 2012). These sequences were then dereplicated using “qiime vsearch dereplicate-sequences” plug-in in QIIME2 and the representative sequences were taxonomically classified against SILVA database using the scikit-learn classifier as described above. AOA sequences were extracted and reclassified by BLASTN against the same custom AOA database as used for analyzing amplicon sequencing data (see above).

Metagenomic shotgun sequences are deposited at NCBI under SRA study IDs SRP258623, SRP258632, SRP258633, SRP258644, SRP258647, SRP258661, SRP258662, SRP258690, SRP258692, SRP258713, SRP258714, SRP258717, SRP258719, SRP258720, SRP258807, SRP258809, SRP258810, SRP270154.

2.5. Statistical analyses

Alpha diversity (number of ASVs and Pielou's evenness) was calculated after normalization of the sequences to the depth of 12,000 reads per sample. Principal coordinates analysis (PCoA) of weighted UniFrac distance matrices was performed using the vegan package v2.5–7 (Dixon, 2003) in R version 4.1.1, and permutational multivariate analysis of variance (PERMANOVA) was used to assess the compositional difference between different land uses and primer pairs. To characterize the correlation of environmental variables and microbial community compositions, the 'envfit' function of the vegan package was used to test the significance of eight soil physiochemical properties (soil pH, OM, moisture, DOC, TP, AP, nitrate and ammonium concentrations) for the PCoA ordinations. Differences at $p < 0.05$ were considered statistically significant.

The proportion of each taxon in a soil was calculated. We selected the major phyla (with average proportion $> 0.5\%$ in all soils by amplicon sequencing) to evaluate the performance of the primers. For each major phylum identified by both primer pairs, the proportions were compared against each other using a paired sample t -test across all samples from each soil plot, as similarly used previously (Toole et al., 2021). We further calculated the proportion of each AOA clade relative to total AOA sequences, and performed paired sample t -test to compare the proportions by using different primer pairs. Differences at $p < 0.05$ were considered statistically significant.

To evaluate the accuracy of amplicon sequencing for estimating the proportion of a specific taxon in a plot, we calculated \log_2 -transformed ratios by dividing the proportion of the taxon from amplicon sequencing data (from 515F-806R or 515F-926R) by the proportion within each sample's respective metagenome-16S rRNA, as similarly used previously (Wear et al., 2018). The absolute values of \log_2 ratios between 515F-806R and 515F-926R were compared with each other using paired sample t -test across all samples in each soil plot. Differences at $p < 0.05$ were considered statistically significant and a significantly smaller absolute value indicated a more accurate estimation of proportion compared to metagenome result.

The R package "ggplot2" (Wickham, 2016) was used to plot the results of diversity indices and relative abundances of different taxa. The \log_2 ratio results were plotted using OriginPro 2016.² And Adobe Illustrator 2019 was used to merge different figure plots.

2.6. *In silico* evaluation of the primer coverage

The theoretical coverage of the two primer pairs for analyzing total prokaryotes was tested *in silico* by using the TestProbe function of the

SILVA database 138.1 server.³ Additionally, the primers were also evaluated for the coverage of specific AOA clades by Probe Match function of ARB (Ludwig et al., 2004) against the AOA 16S rRNA gene database after removing reads that do not contain the primer locus region (Wang et al., 2021).

3. Results

3.1. *In silico* evaluation of the two primer pairs

The *in silico* evaluation was in agreement with previous results (Parada et al., 2016). Primers 515F-926R resulted in higher coverage of the total database sequences, as well as higher coverage of the sequences from all domains and the common phyla (Table 1). Only *Crenarchaeota* and *Firmicutes* were consistently less represented by 515F-926R compared to 515F-806R. Primers 515F-926R also resulted in less coverage of *Entotheonellaeota* than using 515F-806R when perfect match was required, but this trend was reversed when one mismatch was allowed. We also tested the primer coverage to a well classified AOA 16S rRNA database (Wang et al., 2021) and found both primer pairs matched perfectly to all sequences (102 sequence) in the database. The degeneracy of the primers did not affect the 515F or 806R matching, as a single primer variant was responsible for the perfect match in both primer binding sites. However, 926R had 101 perfect matches with one specific primer sequence and one perfect match with a different primer sequence, showing that the degeneracy was necessary.

3.2. Comparison of total prokaryotic communities from 16S rRNA amplicon sequences using different primer pairs

To our surprise, much higher microbial richness (indicated by number of ASVs) was observed in all soil plots by using 515F-806R ($1,926 \pm 52$ ASVs, MEAN \pm SE) compared to that from 515F-926R (772 ± 17 ; Figure 1), after sequencing 90 soil samples from the five differently managed plots in EAA. In contrast, higher evenness of microbial communities (indicated by Pielou's evenness index) was detected by 515F-926R (0.931 ± 0.001) than using 515F-806R (0.912 ± 0.002). Despite the difference in alpha diversity of microbial communities detected by these two primer pairs, both captured significantly lower richness of microbes in the unmanaged plot (plot 1) compared to those in the agriculturally managed plots (plot 2–5) ($p < 0.001$; Figure 1).

Different primer pairs can affect the characterization of microbial composition in the soils (Supplementary Figure S1). However, PCoA on the unweighted UniFrac distance matrix revealed the same pattern of variation in microbial composition between different plots (Pairwise PERMANOVA, $p < 0.001$ between any two plots), irrespective of primer pair used (Figure 2). All eight environmental variables tested in this study (data in Supplementary Table S1) showed significant correlation with the variation in microbial composition among different soils ($p < 0.002$; Figure 2). The results from both primer pairs indicated that

² <https://www.originlab.com/origin>

³ <https://www.arb-silva.de/search/testprobe/>

TABLE 1 *In silico* assessment of the coverage (percentage of matched sequences) against SILVA SSU 138.1 database of the two different primer pairs 515F-806R or 515F-926R, allowing zero or one mismatch, respectively.

	0 mismatch		1 mismatch	
	515F-806R	515F-926R	515F-806R	515F-926R
Total taxa in SILVA database	75.6%	85.1%	83.2%	93.6%
Domain-level taxa				
<i>Archaea</i>	83.5%	81.0%	94.6%	90.8%
<i>Bacteria</i>	83.6%	84.5%	92.8%	93.4%
<i>Eukaryota</i>	0.1%	80.8%	2.0%	91.4%
Phylum-level taxa				
<i>Proteobacteria</i>	87.9%	89.4%	93.5%	94.3%
<i>Acidobacteriota</i>	90.7%	92.7%	95.5%	97.1%
<i>Actinobacteriota</i>	78.6%	82.1%	91.8%	92.8%
<i>Chloroflexi</i>	56.6%	79.4%	87.7%	94.5%
<i>Planctomycetota</i>	81.9%	85.0%	94.1%	95.7%
<i>Bacteroidota</i>	86.9%	88.3%	94.7%	94.8%
<i>Crenarchaeota</i>	85.4%	83.4%	95.8%	90.3%
<i>Myxococcota</i>	90.6%	91.9%	96.0%	97.5%
<i>Verrucomicrobiota</i>	79.2%	82.4%	94.4%	95.0%
<i>Gemmatimonadota</i>	88.2%	92.3%	93.5%	96.5%
<i>Methylomirabilota</i>	88.9%	93.0%	93.8%	98.4%
<i>Firmicutes</i>	84.4%	83.6%	92.0%	91.9%
<i>Nitrospirota</i>	89.6%	91.5%	95.3%	96.5%
<i>Entotheonellaeota</i>	88.2%	75.7%	95.4%	99.3%
<i>Desulfobacterota</i>	87.4%	89.8%	94.9%	95.5%
<i>Cyanobacteria</i>	75.7%	85.2%	89.0%	92.0%
<i>Latescibacterota</i>	89.7%	90.1%	95.4%	95.8%

The assessment was conducted within three domains and 17 major phyla characterized in the soils tested in this study.

soil pH had the strongest correlation ($r^2 > 0.92$) with the microbial composition variation, followed by available phosphorus and organic matter content. RDA on species distribution consistently characterized the pH as the dominant environmental variables explaining majority of the microbial variations between different plots (806R: 63.7%, pseudo- $F = 124$, $p = 0.002$; 926R: 55.9%, pseudo- $F = 90.2$, $p = 0.002$), while other environmental variables explained much less (all $\leq 3.7\%$, see [Supplementary Table S2](#)).

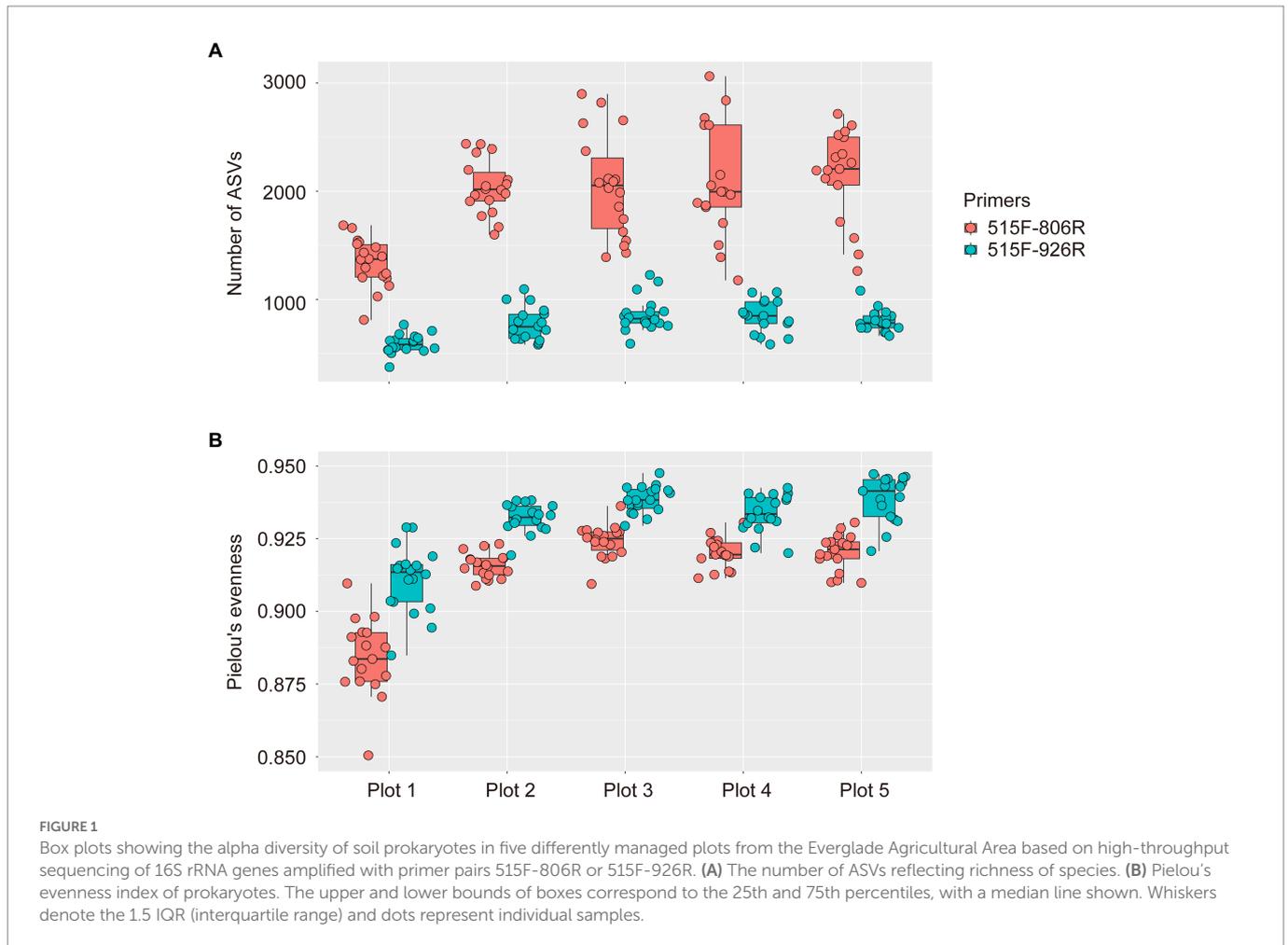
To provide detailed information of the effect of different primer pairs on assessing the relative abundance of the major taxa, we selected the commonly detected phyla in soils (with average relative abundance $> 0.5\%$ in amplicon sequencing data) and compared their proportions by using different primer pairs. A total of 17 phyla were selected ([Figure 3A](#); [Supplementary Figure S2](#)), among which four (*Actinobacteriota*, *Myxococcota*, *Verrucomicrobiota*, *Gemmatimonadota*) showed consistently higher proportions in the 515F-806R data, than in the 515F-926R data, and six phyla (*Acidobacteriota*, *Chloroflexi*, *Planctomycetota*, *Bacteroidota*, *Methylomirabilota*, *Entotheonellaeota*) showed the opposite trend ([Figure 3B](#)). The other seven phyla (*Proteobacteria*, *Crenarchaeota*, *Firmicutes*, *Nitrospirota*, *Desulfobacterota*, *Cyanobacteria*, *Latescibacterota*) showed inconsistent patterns of proportions using 515F-806R versus 515F-926R primer pairs, depending on soil management. For instance, 515F-806R resulted in significantly higher proportion of *Proteobacteria* in the agricultural soil plots 2–5, but the proportion of *Proteobacteria* did not

change in the unmanaged plot 1, compared to the results obtained from 515F-926R.

3.3. Comparison of AOA communities from 16S rRNA gene amplicon sequences using different primer pairs

Ammonia oxidizing archaea represented the dominant archaea in all our soils, irrespective of primers used. The proportion of AOA relative to total archaeal reads ranged between $92.3\% \pm 1.3$ to $99.2\% \pm 0.2\%$ and between $90.4\% \pm 1.4$ to $99.7\% \pm 0.2\%$ in different plots by using primer pairs of 515F-806R and 515F-926R, respectively. Both primer pairs also resulted in similar proportions of AOA relative to total prokaryotes, ranging between $4.0\% \pm 0.1$ to $6.0\% \pm 0.3\%$ and between $4.1\% \pm 0.1$ to $6.2\% \pm 0.3\%$ in different plots by using 515F-806R and 515F-926R, respectively. As expected, ammonia oxidizing bacteria (AOB) were rarely detected in these soils (proportion $\leq 0.08\%$ and $\leq 0.05\%$ relative to total prokaryotes using 515F-806R and 515F-926R, respectively), since no chemical nitrogen fertilizer was applied to stimulate AOB growth and activity ([Verhamme et al., 2011](#); [Tao et al., 2021b,c](#)).

All AOA sequences were classified into putative family-level clades based on a highly resolved phylogeny ([Alves et al., 2018](#); [Wang et al.,](#)



2021). A total of nine AOA clades were identified in our soils by 16S rRNA gene amplicon sequences (Figure 4A). Four of the AOA clades (NS-Alpha, NS-Delta, NS-Gamma and NS-Zeta) belonging to the order *Nitrososphaerales* constituted the majority of AOA in the tested soils. We also observed four *Nitrosopumilales*-related AOA clades accounting for smaller portions, and very rarely a *Ca. Nitrosotaleales* clade (NT-Alpha).

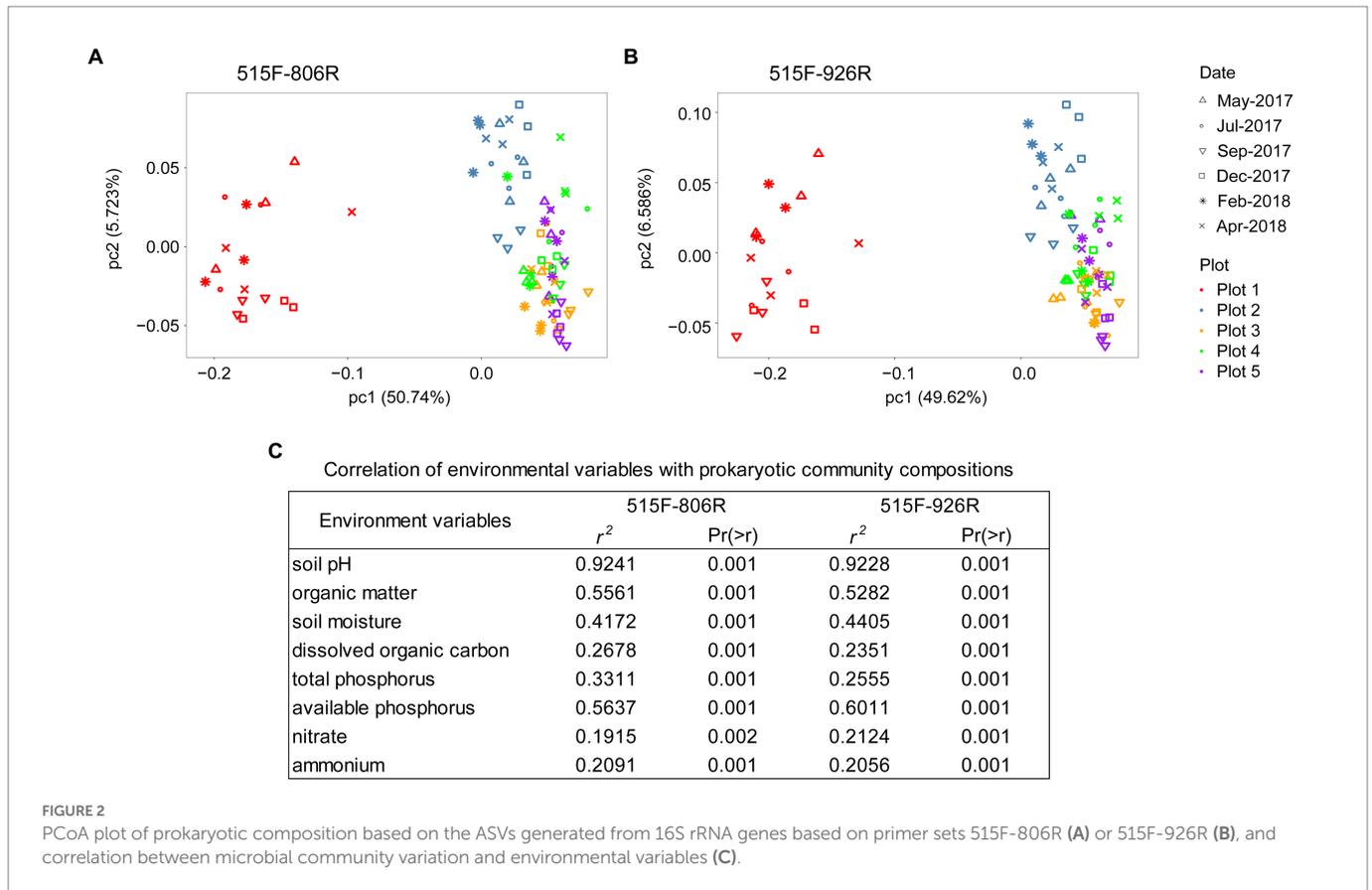
Different primer pairs led to significant change in relative proportion of some major AOA clades relative to total AOA abundance (Figure 4B). For example, primers 515F-926R resulted in higher proportions of NS-Alpha and NP-Delta clades, compared to results from 515F-806R. In contrast, 515F-806R generated consistently higher proportion of NS-Zeta. For characterizing two of the clades, i.e., NS-Delta and NP-Eta, the effect of the different primer pairs was not consistent when used in different soils. The *Ca. Nitrosotaleales* clade (NT-Alpha) was rarely detected by both primer sets.

3.4. Comparison of microbial communities from 16S rRNA gene amplicon sequences and shotgun metagenomes

Shotgun metagenomes (total 18 metagenomic datasets, including replicates) were performed in 3 of the plots (plot 1, 3, and 5) in two time points (July and December of 2017). Metagenomic results showed that some taxonomic patterns were comparable to the amplicon-based

sequencing results (Figure 5). However, there was variation in relative abundance by amplicon-based sequencing compared to that from metagenome data. Among all 17 major phyla in the soils, the relative abundances of the metagenome-16S rRNA gene results of eight phyla, i.e., *Acidobacteriota*, *Actinobacteriota*, *Planctomycetota*, *Bacteroidota*, *Crenarchaeota*, *Gemmatimonadota*, *Firmicutes*, and *Entotheonellaeota*, were better approximated by 515F-806R, than by 515F-926R primers. In contrast, the results from 515F-926R primers were more similar to metagenomic results for *Proteobacteria*, *Chloroflexi*, *Myxococcota*, *Verrucomicrobiota*, *Nitrospirota*, and *Desulfobacterota*. No significant differences were observed in the relative abundances of *Methylomirabilota*, *Cyanobacteria*, and *Latescibacterota* by using different primer pairs and shotgun sequencing approaches. We further evaluated the accuracy of different primers for characterizing the major genera (average proportion > 0.5%). Most of the phylum-level patterns also applied to the affiliated genera (Supplementary Figure S3). However, genera from the phyla *Actinobacteriota* and *Chloroflexi*, as well as some genera from *Proteobacteria* and *Acidobacteriota* showed opposite patterns compared to the respective phylum-level results. For instance, the proportions of the four major genera from *Chloroflexi* by 515F-806R were more similar to metagenomic results.

Similarly, we assessed the accuracy of amplicon-based sequencing results of AOA, in comparison to metagenome-16S rRNA gene results. Due to lack of reads of some AOA clades in the metagenome-16S rRNA gene data, we can only evaluate the results of six AOA clades (Figure 6). We found 515F-806R results of NS-Delta, NS-Zeta and NP-Eta had



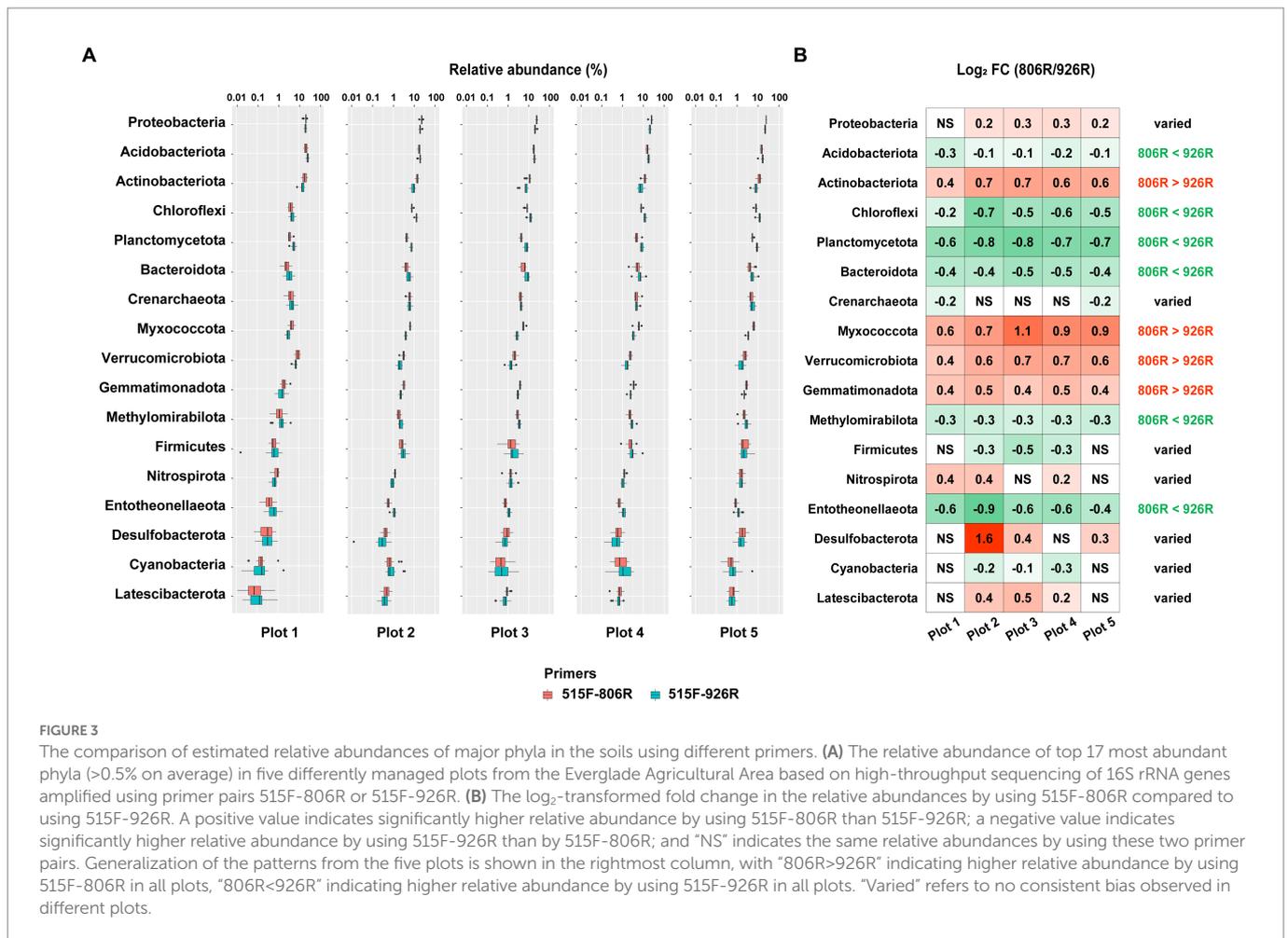
higher similarity to metagenome results, while 515F-926R showed more accurate results of NS-Alpha. The effect of these two primer pairs on estimation of NP-Delta proportion was the same, but the effect on NS-Gamma proportion was not consistent in different soils.

4. Discussion

One of the major criteria for selecting primers is their coverage of target microbes. In this study, the two primer pairs yielded significantly different alpha diversity. Particularly, 515F-806R resulted in higher richness of microorganisms compared to 515F-926R, as reflected by higher numbers of ASVs (Figure 1A). This result contradicted the *in silico* assessment using the most updated SILVA database, in which 515F-926R appeared to cover higher diversity of total prokaryotes. The incongruence between *in silico* and *in vitro* results has been observed previously (Fischer et al., 2016; Pausan et al., 2019). As the primers were used to generate profiles of highly complex microbial communities in soils, such inconsistency is most likely due to the various degrees of primer-template mismatches. For instance, the binding regions of 16S rRNA genes of some major prokaryotic taxa might have less mismatches to 926R than to 806R, resulting in over-representation of certain taxa during PCR amplification and consequently an overall decline in richness. Indeed, the presence of a single primer-template mismatch can affect the efficiency of PCR amplification and lead to underestimation of gene numbers by more than 1,000 times, depending on the primer and position of the mismatch (Bru et al., 2008). This will particularly affect the characterization of rare taxa in soils, if they have a high number

of mismatches, or the mismatches are particularly enriched near the primer extension sequence (Bru et al., 2008). Rare taxa are pervasive in microbial biosphere and often consist of a large portion of microbial diversity in various environments (Curtis et al., 2002; Lynch and Neufeld, 2015), such as soils used in this study. We cannot rule out the possibility that the differences in the hypervariability of V4 and V4-V5 regions covered by these two primers might resulted in different taxonomic resolutions and thus affected the classification accuracy (Guo et al., 2013; Yang et al., 2016; Kerrigan et al., 2019).

Compared to using mock community, the biggest challenge for primer evaluation using environmental sample is the approach to producing reference results. Methods such as fluorescence *in situ* hybridization (Fadeev et al., 2021) and shotgun metagenome sequencing (Tremblay et al., 2015; Eloë-Fadrosh et al., 2016; Fischer et al., 2016) have been used previously to establish reference datasets. This study used shotgun metagenome-derived 16S rRNA gene data as reference which should generate a more accurate microbial profile than amplicon-based sequencing approaches (Brooks et al., 2015). Through the comparison between amplicon and metagenome results in five differently managed plots (Figure 5), we generalized that: (i) For most major phyla the bias was more pronounced by either 515F-806R or 515F-926R; only for *MethyloMirabilota*, *Cyanobacteria*, and *Latescibacterota* the bias was similar between these two primer pairs; (ii) Both primer pairs consistently overestimated relative abundances of phyla *Acidobacteriota*, *Bacteroidota*, *Crenarchaeota*, *Firmicutes*, *Nitrospirota*, and *Entotheonellaeota*, and underestimated *Actinobacteriota*; (iii) The phylum-level evaluation did not always apply to the affiliated genera. The discrepancy between phylum and genus-level results thus emphasizes the necessity of preliminary test of newly

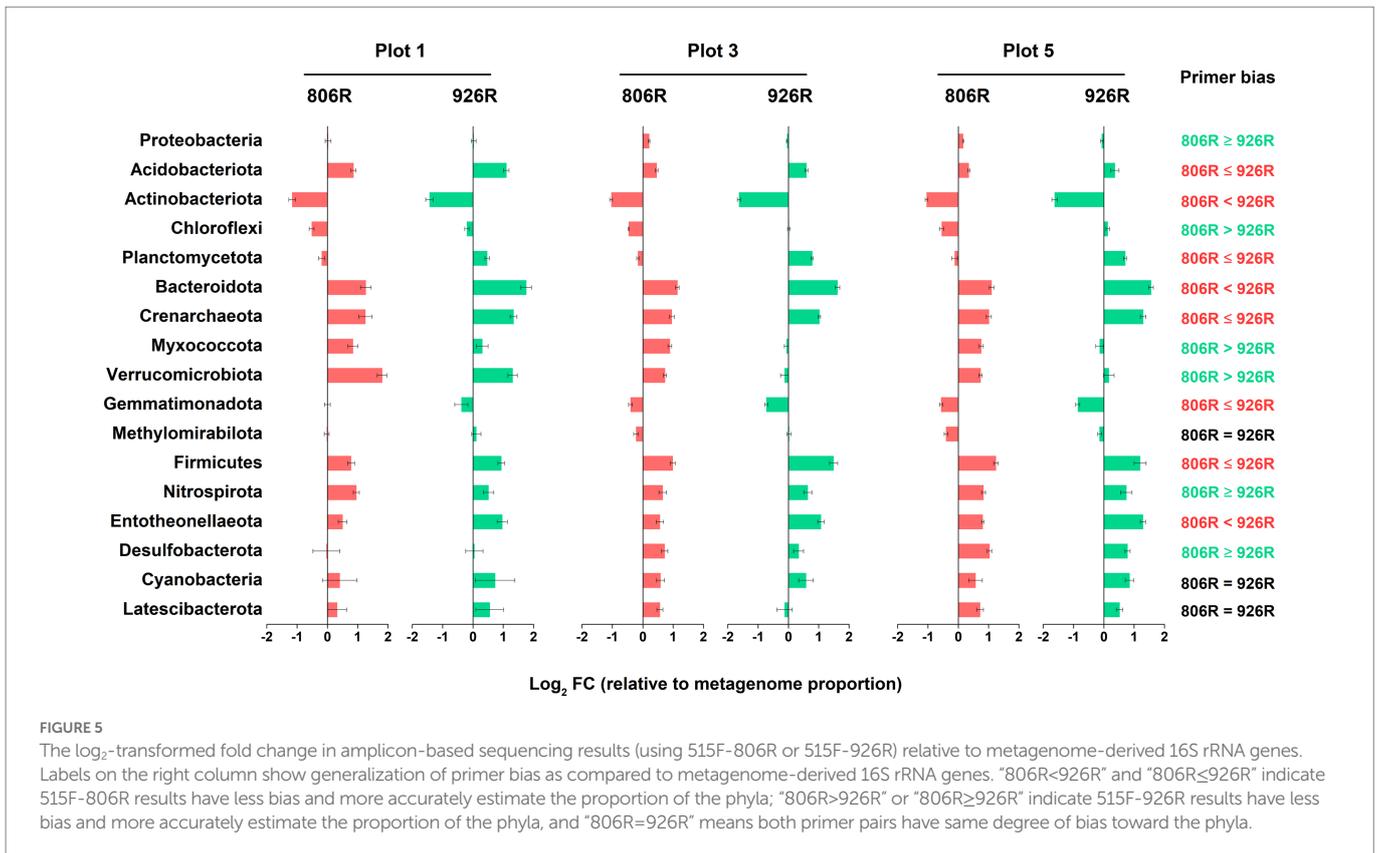
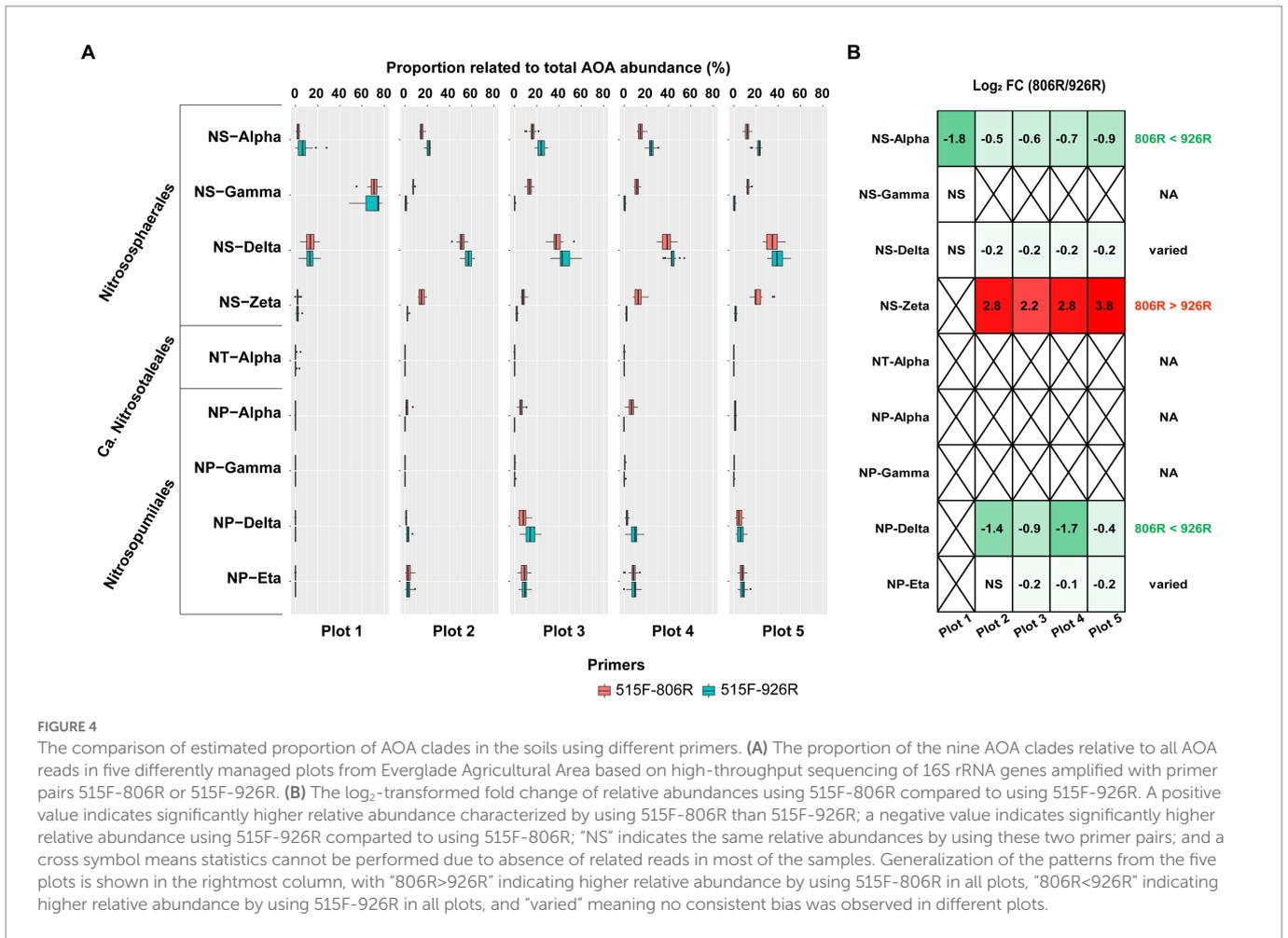


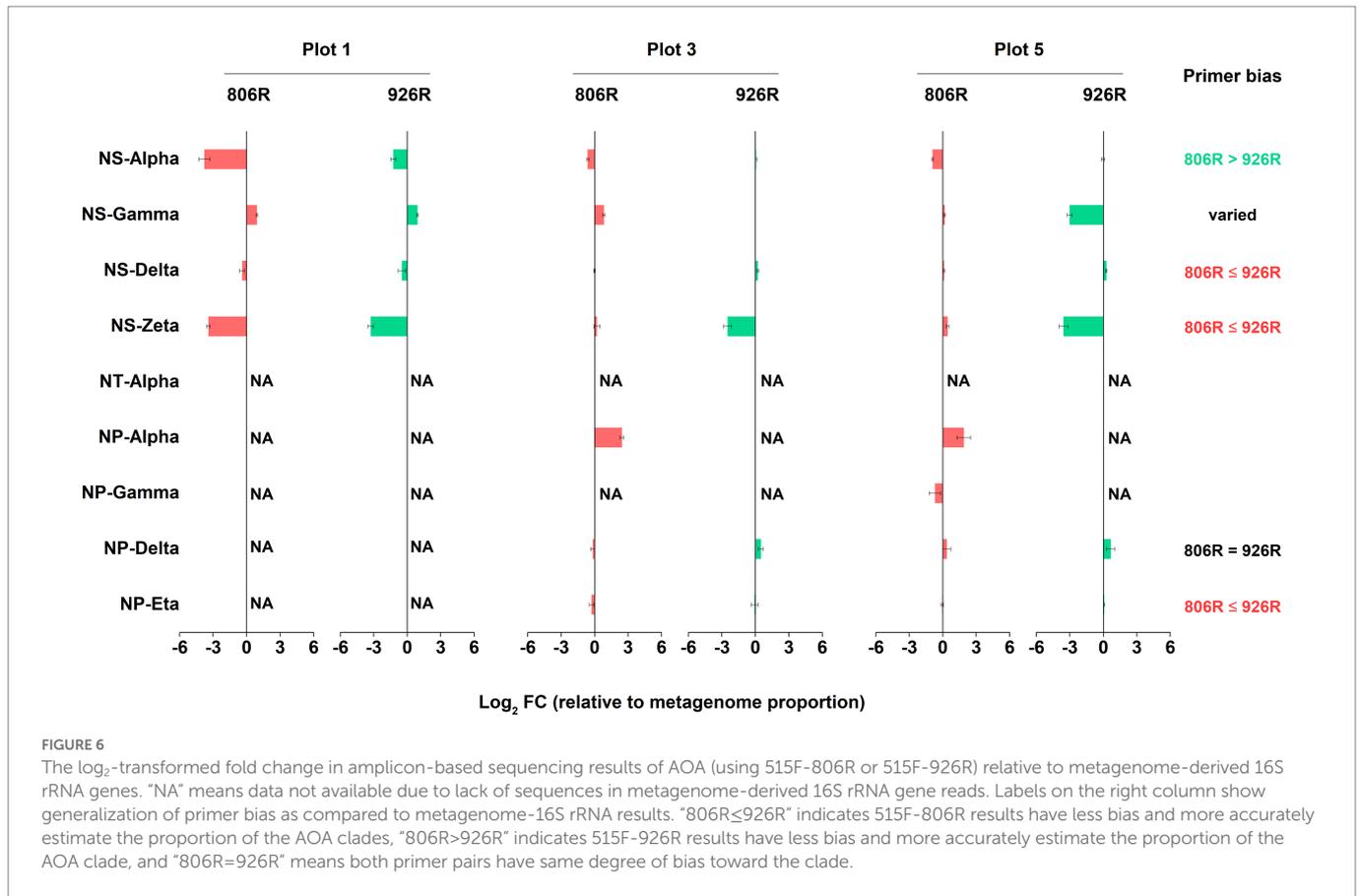
acquired environmental samples for accurate quantification of the proportion of specific taxa.

Although the amplicon sequencing of field samples generated inconsistent results between *in silico* evaluation and metagenome-based data for some specific taxa, the use of different primer pairs did not affect the generalization of overall trends of microbial community variations associated with different land uses. For instance, both primer pairs revealed significantly higher richness of soil prokaryotes in agriculturally managed soils (plot 2–5) than in the unmanaged native soil (plot 1, Figure 1A). We also observed significant distinction in microbial composition between different plots irrespective of primer used, especially between unmanaged soil and agricultural soils (Figure 2). Despite primer-induced variation in relative abundance of many taxa, the general patterns of the major phyla abundance can still be well reflected. For instance, both primer pairs showed that the 17 major phyla characterized in our soils collectively predominated the microbial community abundances, accounted for 93.0–99.3% and 94.9–99.5% of the total prokaryotic 16S rRNA reads, respectively (Figure 3). For AOA analysis, both primer pairs showed the highest abundance of NS-Gamma in unmanaged soil, while NS-Delta, NS-Alpha, NP-Delta and NP-Eta co-dominated AOA abundance in agricultural soil (Figure 4).

The conversion of native high-organic matter Everglades soil for agricultural use led to changes in soil physiochemistry and microbiology, and the choice of primers did not affect the revelation of the major environmental factors associated with the microbial

community changes. Soil pH is characterized as the primary factor influencing the whole community composition (Figure 2). The unmanaged soil had a pH value between 5.0–5.8 all year round, while the agricultural soils all showed more neutral pH of 6.5–7.8 due to previous liming, leading to drastic difference in prokaryotic community composition between these two types of land uses (Figure 2). This was consistent with previous meta-analysis showing soil pH as the major determinant for microbial assembly process (Tripathi et al., 2018). Soil pH was also recognized as the main driver determining the niche specialization of terrestrial AOA at global scale (Gubry-Rangin et al., 2011) in addition to bacterial nitrifiers (Tao et al., 2021a), and might similarly affect the AOA composition in our soils. Indeed, we found that the unmanaged soil with low pH contained much higher proportion of NS-Gamma AOA clade (Figure 4), which has been classified as acido-neutral clade and can well adapt to acidic soil (Wang et al., 2019; Zhao et al., 2021). As expected, we observed increased proportions of NS-Delta and NS-Alpha in agricultural soils with higher pH, which represent the major AOA clades in soils globally and are adapted to neutral-alkaline pH soil (Gubry-Rangin et al., 2011; Alves et al., 2018; Zhao et al., 2021). Surprisingly, a fair proportion of *Nitrosopumilales*-related AOA clades were also detected in agricultural soils, mostly from clades of NP-Delta and NP-Eta. These two AOA clades are represented by *Nitrosopumilus*-, *Ca. Nitrosoarchaeum*-, and *Ca. Nitrosotenuis*-affiliated strains and were more frequently detected in oligotrophic environments such as marine and deep oligotrophic soil (Mosier et al., 2012; Jung et al., 2014; Tolar





et al., 2020). Their presence in the fertile agricultural soils suggested that some of the phylotypes from these clades can adapt well to environments with high nutrient availability. However, the absence of *Nitrosopumilales* AOA in the unmanaged soil implied high sensitivity of related clades to low soil pH.

5. Conclusion

This study comprehensively evaluated two commonly used primer pairs for characterization of complex microbial communities in soils, in comparison to *in silico* evaluation and shotgun metagenome-based results. As previously observed, the *in silico* evaluation cannot be confidently used to predict the *in vitro* results of PCR amplification, and we observed lower richness of prokaryotic microorganisms from 515F-926R which theoretically covers higher prokaryotic diversity. It is thus our recommendation that 515F-806R is better for characterization of microbial diversity in organic-rich agricultural soils similar to those used in this study but a preliminary test should be conducted for analyzing microbes in other types of soils (e.g., forest, sediments). Despite dissimilar results by using different primer pairs, the main patterns of microbial community structure (diversity and association to environmental factors) between different soils can still be consistently characterized regardless of primer used. Therefore, it is methodologically rational to integrate amplicon sequencing data from different studies for comparative studies or meta-analyses if the primers were identical, and results should be interpreted with caution if different primer pairs or sequencing pipelines were applied.

Data availability statement

The datasets presented in this study can be found in online at: <https://www.ncbi.nlm.nih.gov/genbank/> under Bioproject PRJNA831877, and <https://www.ncbi.nlm.nih.gov/sra> under accessions SRP258623, SRP258632, SRP258633, SRP258644, SRP258647, SRP258661, SRP258662, SRP258690, SRP258692, SRP258713, SRP258714, SRP258717, SRP258719, SRP258720, SRP258807, SRP258809, SRP258810, SRP270154.

Author contributions

JZ and WM-H designed the study. JZ and JR conducted the experiments. JZ drafted the manuscript with input from JR and WM-H. All authors contributed to the article and approved the submitted version.

Funding

Funding for this project was provided by the Florida Agricultural Experiment Station Hatch project FLA-FTL-005680, UF IFAS Early Career Award, and USDA NIFA award #2022-67019-36501, and Department of Energy Joint Genome Institute Community Sequencing Program Project #503337 to WM-H. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Acknowledgments

We thank Everglades Research and Education Center farm manager Leonard Fox for help with plot selection. Seemanti Chakrabarti, Jennifer Cooper, Eunkyung Choi, Maryory Palacios, and Rachelle Berger are acknowledged for help during sampling.

Conflict of interest

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

References

- Alves, R. J. E., Minh, B. Q., Urich, T., Von Haeseler, A., and Schleper, C. (2018). Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. *Nat. Commun.* 9:1517. doi: 10.1038/s41467-018-03861-1
- Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbiol. Ecol.* 75, 129–137. doi: 10.3354/ame01753
- Bai, X., Hu, X., Liu, J., Gu, H., Jin, J., Liu, X., et al. (2022). Evaluation of four primer sets for analysis of comammox communities in black soils. *Front. Microbiol.* 13:944373. doi: 10.3389/fmicb.2022.944373
- Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of archaea. *Nat. Microbiol.* 5, 887–900. doi: 10.1038/s41564-020-0715-z
- Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J.* 5, 908–917. doi: 10.1038/ismej.2010.171
- Bickel, S., and Or, D. (2020). Soil bacterial diversity mediated by microscale aqueous-phase processes across biomes. *Nat. Commun.* 11:116. doi: 10.1038/s41467-019-13966-w
- Bower, C. E., Holm-Hansen, T. J. C. J. O. F., and Sciences, A. (1980). A salicylate-hypochlorite method for determining ammonia in seawater. *Can. J. Fish. Aquat. Sci.* 37, 794–798. doi: 10.1139/f80-106
- Brooks, J. P., Edwards, D. J., Harwich, M. D. Jr., Rivera, M. C., Fettweis, J. M., Serrano, M. G., et al. (2015). The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* 15:66. doi: 10.1186/s12866-015-0351-6
- Bru, D., Martin-Laurent, F., and Philippot, L. (2008). Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Appl. Environ. Microbiol.* 74, 1660–1663. doi: 10.1128/AEM.02403-07
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Caporaso, J. G., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.0510-335
- Crowther, T. W., Van Den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A. D., Mo, L., et al. (2019). The global soil community and its influence on biogeochemistry. *Science* 365:eaav0550. doi: 10.1126/science.aav0550
- Curtis, T. P., Sloan, W. T., and Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10494–10499. doi: 10.1073/pnas.142680199
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x
- Domeignoz-Horta, L. A., Shinfuku, M., Junier, P., Poirier, S., Verrecchia, E., Sebag, D., et al. (2021). Direct evidence for the role of microbial community composition in the formation of soil organic matter composition and persistence. *ISME Commun.* 1:64. doi: 10.1038/s43705-021-00071-7
- Eloe-Fadros, E. A., Ivanova, N. N., Woyke, T., and Kyrpides, N. C. (2016). Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* 1:15032. doi: 10.1038/nmicrobiol.2015.32
- Fadeev, E., Cardozo-Mino, M. G., Rapp, J. Z., Bienhold, C., Salter, I., Salman-Carvalho, V., et al. (2021). Comparison of two 16S rRNA primers (V3-V4 and V4-V5) for studies of arctic microbial communities. *Front. Microbiol.* 12:637526. doi: 10.3389/fmicb.2021.637526
- Fischer, M. A., Gullert, S., Neulinger, S. C., Streit, W. R., and Schmitz, R. A. (2016). Evaluation of 16S rRNA gene primer pairs for monitoring microbial community structures showed high reproducibility within and low comparability between datasets generated with multiple archaeal and bacterial primer pairs. *Front. Microbiol.* 7:1297. doi: 10.3389/fmicb.2016.01297

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1140487/full#supplementary-material>

- Fouhy, F., Clooney, A. G., Stanton, C., Claesson, M. J., and Cotter, P. D. (2016). 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol.* 16:123. doi: 10.1186/s12866-016-0738-z
- Frintde, K., Pape, R., Werner, K., Löffler, J., and Knief, C. (2019). Temperature and soil moisture control microbial community composition in an arctic-alpine ecosystem along elevational and micro-topographic gradients. *ISME J.* 13, 2031–2043. doi: 10.1038/s41396-019-0409-9
- García-Robledo, E., Corzo, A., and Papaspyrou, S. (2014). A fast and direct spectrophotometric method for the sequential determination of nitrate and nitrite at low concentrations in small volumes. *Mar. Chem.* 162, 30–36. doi: 10.1016/j.marchem.2014.03.002
- Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B. C., James, P., et al. (2011). Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc. Natl. Acad. Sci. U. S. A.* 108, 21206–21211. doi: 10.1073/pnas.1109000108
- Gubry-Rangin, C., Kratsch, C., Williams, T. A., Mchardy, A. C., Embley, T. M., Prosser, J. I., et al. (2015). Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proc. Natl. Acad. Sci. U. S. A.* 112, 9370–9375. doi: 10.1073/pnas.1419329112
- Gubry-Rangin, C., Nicol, G. W., and Prosser, J. I. (2010). Archaea rather than bacteria control nitrification in two agricultural acidic soils. *FEMS Microbiol. Ecol.* 74, 566–574. doi: 10.1111/j.1574-6941.2010.00971.x
- Guo, F., Ju, F., Cai, L., and Zhang, T. (2013). Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PLoS One* 8:e76185. doi: 10.1371/journal.pone.0076185
- Hink, L., Gubry-Rangin, C., Nicol, G. W., and Prosser, J. I. (2018). The consequences of niche and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. *ISME J.* 12, 1084–1093. doi: 10.1038/s41396-017-0025-5
- Huang, L., Chakrabarti, S., Cooper, J., Perez, A., John, S. M., Daroub, S. H., et al. (2021). Ammonia-oxidizing archaea are integral to nitrogen cycling in a highly fertile agricultural soil. *ISME Commun.* 1:19. doi: 10.1038/s43705-021-00020-4
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1:16048. doi: 10.1038/nmicrobiol.2016.48
- Jung, M. Y., Park, S. J., Kim, S. J., Kim, J. G., Sinnighe Damste, J. S., Jeon, C. O., et al. (2014). A mesophilic, autotrophic, ammonia-oxidizing archaeon of thaumarchaeal group I.1a cultivated from a deep oligotrophic soil horizon. *Appl. Environ. Microbiol.* 80, 3645–3655. doi: 10.1128/AEM.03730-13
- Kerrigan, Z., Kirkpatrick, J. B., and D'Hondt, S. (2019). Influence of 16S rRNA hypervariable region on estimates of bacterial diversity and community composition in seawater and marine sediment. *Front. Microbiol.* 10:1640. doi: 10.3389/fmicb.2019.01640
- Klimasmith, I. M., and Kent, A. D. (2022). Micromanaging the nitrogen cycle in agroecosystems. *Trends Microbiol.* 30, 1045–1055. doi: 10.1016/j.tim.2022.04.006
- Kopylova, E., Noe, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611
- Liu, J., Yu, Z., Yao, Q., Sui, Y., Shi, Y., Chu, H., et al. (2019). Biogeographic distribution patterns of the archaeal communities across the black soil zone of Northeast China. *Front. Microbiol.* 10:23. doi: 10.3389/fmicb.2019.00023
- Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *Microbial Syst.* 3:e00055-18. doi: 10.1128/mSystems.00055-18
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, B., et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

- Lynch, M. D., and Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* 13, 217–229. doi: 10.1038/nrmicro3400
- Mosier, A. C., Lund, M. B., and Francis, C. A. (2012). Ecophysiology of an ammonia-oxidizing archaeon adapted to low-salinity habitats. *Microb. Ecol.* 64, 955–963. doi: 10.1007/s00248-012-0075-1
- Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* 67, 437–457. doi: 10.1146/annurev-micro-092412-155614
- Oton, E. V., Quince, C., Nicol, G. W., Prosser, J. I., and Gubry-Rangin, C. (2016). Phylogenetic congruence and ecological coherence in terrestrial *Thaumarchaeota*. *ISME J.* 10, 85–96. doi: 10.1038/ismej.2015.101
- Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023
- Pausan, M. R., Csorba, C., Singer, G., Till, H., Schopf, V., Santigli, E., et al. (2019). Exploring the Archaeome: detection of archaeal signatures in the human body. *Front. Microbiol.* 10:2796. doi: 10.3389/fmicb.2019.02796
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pedros-Alí, C., and Manrubia, S. (2016). The vast unknown microbial biosphere. *Proc. Natl. Acad. Sci. U. S. A.* 113, 6585–6587. doi: 10.1073/pnas.1606105113
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Quince, C., Lanzen, A., Davenport, R. J., and Turnbaugh, P. J. (2011). Removing noise from Pyrosequenced amplicons. *BMC Bioinformatics* 12:18. doi: 10.1186/1471-2105-12-38
- Rappé, M. S., and Giovannoni, S. J. (2003). The uncultured microbial majority. *Annu. Rev. Microbiol.* 57, 369–394. doi: 10.1146/annurev-micro.57.030502.090759
- Ren, M., Feng, X., Huang, Y., Wang, H., Hu, Z., Clingenpeel, S., et al. (2019). Phylogenomics suggests oxygen availability as a driving force in *Thaumarchaeota* evolution. *ISME J.* 13, 2150–2161. doi: 10.1038/s41396-019-0418-8
- Rillig, M. C., Ryo, M., and Lehmann, A. (2021). Classifying human influences on terrestrial ecosystems. *Glob. Chang. Biol.* 27, 2273–2278. doi: 10.1111/gcb.15577
- Rillig, M. C., Ryo, M., Lehmann, A., Aguilar-Trigueros, C. A., Buchert, S., Wulf, A., et al. (2019). The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science* 366, 886–890. doi: 10.1126/science.aay2832
- Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340–1351. doi: 10.1038/ismej.2010.58
- Sheridan, P. O., Raguideau, S., Quince, C., Holden, J., Zhang, L., Thames, C., et al. (2020). Gene duplication drives genome expansion in a major lineage of *Thaumarchaeota*. *Nat. Commun.* 11:5494. doi: 10.1038/s41467-020-19132-x
- Sun, X., Zhao, J., Zhang, L., Zhou, X., Xia, W., Zhao, Y., et al. (2022). Effects of agricultural land use on the differentiation of nitrifier communities and functional patterns from natural terrestrial ecosystems. *Sci. Total Environ.* 835:155568. doi: 10.1016/j.scitotenv.2022.155568
- Sun, X., Zhao, J., Zhou, X., Bei, Q., Xia, W., Zhao, B., et al. (2021). Salt tolerance-based niche differentiation of soil ammonia oxidizers. *ISME J.* 16, 412–422. doi: 10.1038/s41396-021-01079-6
- Tao, R., Li, J., Hu, B., and Chu, G. (2021a). Ammonia-oxidizing bacteria are sensitive and not resilient to organic amendment and nitrapyrin disturbances, but ammonia-oxidizing archaea are resistant. *Geoderma* 384:114814. doi: 10.1016/j.geoderma.2020.114814
- Tao, R., Li, J., Hu, B., and Chu, G. (2021b). Mitigating N₂O emission by synthetic inhibitors mixed with urea and cattle manure application via inhibiting ammonia-oxidizing bacteria, but not archaea, in a calcareous soil. *Environ. Pollut.* 273:116478. doi: 10.1016/j.envpol.2021.116478
- Tao, R., Zhang, H., Gu, X., Hu, B., Li, J., and Chu, G. (2021c). Di-(2-ethylhexyl) phthalate (DEHP) exposure suppressed the community diversity and abundance of ammonia-oxidizers and mitigated N₂O emissions in an alkaline soil. *Ecotoxicol. Environ. Saf.* 227:112910. doi: 10.1016/j.ecoenv.2021.112910
- Thijs, S., Op De Beeck, M., Beckers, B., Truyens, S., Stevens, V., Van Hamme, J. D., et al. (2017). Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. *Front. Microbiol.* 8:494. doi: 10.3389/fmicb.2017.00494
- Thion, C., and Prosser, J. I. (2014). Differential response of nonadapted ammonia-oxidizing archaea and bacteria to drying-rewetting stress. *FEMS Microbiol. Ecol.* 90, 380–389. doi: 10.1111/1574-6941.12395
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi: 10.1038/nature24621
- Tolar, B. B., Reji, L., Smith, J. M., Blum, M., Pennington, J. T., Chavez, F. P., et al. (2020). Time series assessment of Thaumarchaeota ecotypes in Monterey Bay reveals the importance of water column position in predicting distribution–environment relationships. *Limnol. Oceanogr.* 65, 2041–2055. doi: 10.1002/lno.11436
- Toole, D. R., Zhao, J., Martens-Habbena, W., and Strauss, S. L. (2021). Bacterial functional prediction tools detect but underestimate metabolic diversity compared to shotgun metagenomics in Southwest Florida soils. *Appl. Soil Ecol.* 168:104129. doi: 10.1016/j.apsoil.2021.104129
- Torsvik, V., and Øvreås, L. (2002). Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.* 5, 240–245. doi: 10.1016/S1369-5274(02)00324-7
- Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., et al. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* 6:771. doi: 10.3389/fmicb.2015.00771
- Tripathi, B. M., Stegen, J. C., Kim, M., Dong, K., Adams, J. M., and Lee, Y. K. (2018). Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* 12, 1072–1083. doi: 10.1038/s41396-018-0082-4
- Verhamme, D. T., Prosser, J. I., and Nicol, G. W. (2011). Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. *ISME J.* 5, 1067–1071. doi: 10.1038/ismej.2010.191
- Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2016). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *Microbial Systems* 1:e00009-15. doi: 10.1128/mSystems.00009-15
- Wang, H., Bagnoud, A., Ponce-Toledo, R. I., Kerou, M., Weil, M., Schleper, C., et al. (2021). Linking 16S rRNA gene classification to amoA gene taxonomy reveals environmental distribution of ammonia-oxidizing archaeal clades in peatland soils. *mSystems* e0054621:e0054621. doi: 10.1128/mSystems.00546-21
- Wang, B., Qin, W., Ren, Y., Zhou, X., Jung, M. Y., Han, P., et al. (2019). Expansion of *Thaumarchaeota* habitat range is correlated with horizontal transfer of ATPase operons. *ISME J.* 13, 3067–3079. doi: 10.1038/s41396-019-0493-x
- Wang, B., Zhao, J., Guo, Z., Ma, J., Xu, H., and Jia, Z. (2015). Differential contributions of ammonia oxidizers and nitrite oxidizers to nitrification in four paddy soils. *ISME J.* 9, 1062–1075. doi: 10.1038/ismej.2014.194
- Wear, E. K., Wilbanks, E. G., Nelson, C. E., and Carlson, C. A. (2018). Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environ. Microbiol.* 20, 2709–2726. doi: 10.1111/1462-2920.14091
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer Verlag.
- Yang, B., Wang, Y., and Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 17:135. doi: 10.1186/s12859-016-0992-y
- Yang, Y., Zhang, C., Lenton, T. M., Yan, X., Zhu, M., Zhou, M., et al. (2021). The evolution pathway of ammonia-oxidizing archaea shaped by major geological events. *Mol. Biol. Evol.* 38, 3637–3648. doi: 10.1093/molbev/msab129
- Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330
- Zhang, L.-M., Hu, H.-W., Shen, J.-P., and He, J.-Z. (2012). Ammonia-oxidizing archaea have more important role than ammonia-oxidizing bacteria in ammonia oxidation of strongly acidic soils. *ISME J.* 6, 1032–1045. doi: 10.1038/ismej.2011.168
- Zhang, L. M., Offre, P. R., He, J. Z., Verhamme, D. T., Nicol, G. W., and Prosser, J. I. (2010). Autotrophic ammonia oxidation by soil thaumarchaea. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17240–17245. doi: 10.1073/pnas.1004947107
- Zhao, J., Meng, Y., Drewer, J., Skiba, U. M., Prosser, J. I., and Gubry-Rangin, C. (2020). Differential ecosystem function stability of ammonia-oxidizing archaea and bacteria following short-term environmental perturbation. *Microbial Systems* 5:5. doi: 10.1128/mSystems.00309-20
- Zhao, J., Wang, B., and Jia, Z. (2015). Phylogenetically distinct phylotypes modulate nitrification in a paddy soil. *Appl. Environ. Microbiol.* 81, 3218–3227. doi: 10.1128/AEM.00426-15
- Zhao, J., Wang, B., Zhou, X., Alam, M. S., Fan, J., Guo, Z., et al. (2021). Long-term adaptation of acidophilic archaeal ammonia oxidisers following different soil fertilisation histories. *Microb. Ecol.* 83, 424–435. doi: 10.1007/s00248-021-01763-2
- Zhou, Z., Wang, C., and Luo, Y. (2020). Meta-analysis of the impacts of global change factors on soil microbial diversity and functionality. *Nat. Commun.* 11:3072. doi: 10.1038/s41467-020-16881-7