Soil microbial communities and reciprocal distributions of nitrogen-cycling proxies beneath legume and non-legume patches in a disturbed orchard

Daniel Tang
daniel.tang0173@gmail.com
Langley High School, McLean, Virginia, USA

Abstract

The belowground interactions between plants and soil microbial communities play an important role in the recovery of historically disturbed land. However, these relationships remain poorly understood at fine spatial scales. In this exploratory study, we characterized bacterial, archaeal, and eukaryotic soil communities beneath five plant types (red clover, white clover, fescue, chicory, and common ragweed). The site was a historically disturbed orchard undergoing restoration at a regenerative farm in Virginia, USA. 16S and 18S rRNA gene amplicon sequencing was used to characterize soil communities. Out of 13 samples collected, all were retained for 16S analysis, eleven samples were retained for 18S analysis, while rarefied 18S diversity analyses retained ten samples. Alpha diversity did not differ significantly among plant groups in either 16S or 18S analyses. Beta diversity indicated some variation in 16S bacterial and archaeal community organization among plant groups (PERMANOVA $R^2 = 0.473$, $p = 0.020$). However, these interpretations might be limited because of unequal dispersion and unbalanced sampling. No significant association in beta diversity was found between the plant groups and the 18S community compositions ($R^2 = 0.381$, $p = 0.081$). A targeted exploratory analysis examined taxonomic proxies for microbial lineages that may be associated with nitrogen cycling. Archaeal ammonia-oxidizer proxies were found to be less abundant beneath legumes than beneath non-legumes (1.6% versus 11.9%, $q = 0.008$), while rhizobial lineage proxies showed the opposite pattern (1.6% versus 0.6%, $q = 0.008$). Overall, these findings indicate that plant functional type is associated with reciprocal distributions of nitrogen-cycling taxonomic proxies. The findings identify promising directions for future research on plant-microbe relationships, specifically whether the observed taxonomic patterns correspond to differences in nitrogen cycling during soil restoration.

Keywords: amplicon sequencing; ammonia-oxidizing archaea; legumes; nitrogen cycling; rhizobia; soil microbial community

Introduction

Restoration of historically disturbed soils involves reestablishing vegetation and recovering belowground biological processes which regulate nutrient availability, organic matter turnover, and interactions between plants and soil organisms. Soil microorganisms contribute to processes including decomposition and nutrient transformations. Microbial communities are known to change during the recovery of agricultural land, with community patterns associated with soil age, pH, and nutrient status (Kuramae et al., 2010; Gellie et al., 2017).

Plants, on the other hand, may contribute to the recovery of belowground microbial communities

by creating different biological and chemical environments in the surrounding soil. Plants supply organic substrates through root exudates and litter, alter nutrient availability, and interact directly with microorganisms. For example, controlled studies have demonstrated that root-exudate chemistry and microbial substrate preferences can influence the assembly of rhizosphere bacterial communities (Haichar et al., 2008; Zhalnina et al., 2018). Plant functional-group composition and abiotic soil properties have also been associated with differences in bulk-soil microbial communities, including distinct archaeal composition in legume plots (Dassen et al., 2017). However, associations between individual plant patches and soil microbial communities remain poorly characterized at fine spatial scales in recovering disturbed soils.
Among the plant-microbiome functional interactions, nitrogen cycling is especially relevant to restoration, because the availability and transformation of nitrogen can influence plant establishment and ecosystem productivity. Legumes are a functional group for examining these relationships. For example, red clover and white clover can form root nodules with rhizobia, in which bacterial symbionts reduce atmospheric nitrogen and provide biologically available nitrogen to the plant (Lindström and Mousavi, 2020). Ammonia-oxidizing archaea and bacteria perform the first step of nitrification, and their relative importance can vary with nitrogen availability, soil pH, and plant traits (Prosser and Nicol, 2012; Thion et al., 2016). Comparing legumes with nonlegumes therefore provides an opportunity to examine whether plant functional type is associated with different components of the nitrogen-cycling microbial community during restoration.

Here, we characterized bacterial, archaeal, and eukaryotic soil communities beneath red clover, white clover, fescue, chicory, and common ragweed within a historically disturbed orchard undergoing restoration at a regenerative farm in Fauquier County, Virginia. We investigated whether plant-associated patches differed in microbial alpha diversity and community composition and, more specifically, whether plant functional type was associated with taxonomic proxies for nitrogen-cycling microorganisms. We compared rhizobial and ammonia-oxidizer proxies between legumes and nonlegumes. Because common ragweed was a volunteer species of practical interest at the site, we also examined it as an exploratory case for microbial patterns potentially relevant to restoration. The 16S and 18S amplicon data characterize taxonomic composition instead of biological activity. Therefore, these analyses identify candidate microbial relationships rather than directly measuring nitrogen transformations or restoration success.

Methods

Sampling

The study was conducted at Hidden Creek Farm in Fauquier County, Virginia, United States (38.94080° N, 77.96127° W; approximately 202 m above sea level; U.S. Geological Survey, 2022). The site has a humid subtropical climate (Köppen-Geiger Cfa; Beck et al., 2023). The 1991 to 2020 mean annual temperature at the Warrenton 3 SE weather station was 12.7 °C, and the mean annual precipitation at the Plains 2 NNE station was 1171 mm (Palecki et al., 2021). The soil map unit at the site coordinates was identified from SSURGO data as the Tankerville Purcellville complex with rocky slopes of 7 to 15 percent (Soil Survey Staff, 2025). The soil map unit and slope are map-based estimates and not actual measurements from the sampled locations. No soil physicochemical properties were measured at the site.

Sampling was confined to an approximately 0.35 ha historically disturbed orchard undergoing restoration. The area had previously been used for informal material disposal before restoration management began. The site had been intermittently grazed by sheep approximately two to three times per year for around five years preceding sampling. Sampling locations in the orchard were identified and marked on the same day as collection. Locations were selected purposively where one plant group had the highest stem count within an approximately 30 cm radius, and competing vegetation was minimal. All sampling locations were separated by at least 3 m. Samples were collected on the 10th and 11th of July 2025.

At each location, a single vertical soil sample extending from the surface to approximately 45 cm depth was extracted using a 45 cm spade. Material from the full depth was combined into one integrated bulk soil sample of approximately 4 L, and placed in unused plastic bags. Between locations, the collection spade was wiped until no visible soil remained. However, the spade was not chemically disinfected. Thirteen samples were collected in total beneath red clover (*Trifolium pratense*; $n = 3$), chicory (*Cichorium intybus*; $n = 2$), fescue (*Festuca* species; $n = 4$), common ragweed (*Ambrosia artemisiifolia*; $n = 3$), and white clover (*Trifolium repens*; $n = 1$). Fescue was identified only to genus.

Samples were labeled and stored at approximately -18 °C. Approximately two weeks after collection, each sample was thawed and mixed thoroughly within its bag. Following homogenization, mixed samples were transferred into sterile tubes using a metal spoon cleaned with ethanol between each sample. The tubes were sealed with Parafilm, shipped to the sequencing facility in dry ice, and arrived the following day.

Molecular and bioinformatic analyses

DNA extraction, amplicon library preparation, quality assessment, and sequencing were performed by Novogene Co., Ltd. Genomic DNA was extracted from each soil sample using the NucleoSpin Soil Kit, according to the manufacturer's protocol. Amplicon libraries were prepared by PCR amplification of the target marker regions, using sample-specific barcodes. The V4 region of the 16S rRNA gene was amplified using the 515F primer (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R primer (5′-GGACTACHVGGGTWTCTAAT-3′), which produced an expected amplicon of approximately 300 bp (Caporaso et al., 2011).

The V4 region of the eukaryotic 18S rRNA gene was amplified using the 528F primer (5′-GCGGTAATTCCAGCTCCAA-3′) and 706R primer (5′-AATCCRAGAATTTCACCTCT-3′), producing an expected amplicon of approximately 350 bp (Cheung et al., 2010). However, a fungal-specific internal transcribed spacer (ITS) amplicon was not sequenced, and fungi are therefore described at the taxonomic resolution supported by the 18S marker. The amplicon libraries were sequenced in paired-end mode on an Illumina NovaSeq 6000 platform. One chicory sample and the white clover yielded insufficient 18S amplicon material for sequencing and were therefore omitted from the 18S dataset. The sequencing provider performed demultiplexing and merging of paired-end reads, and provided the resulting merged FASTQ files for downstream analysis. The raw reads are deposited in the Sequence Read Archive of the National Center for Biotechnology Information under the accession number PRJNA1508743.

The 16S and 18S rRNA gene datasets were processed separately using QIIME 2 release 2023.5 (Bolyen et al., 2019). The merged single-end sequences were quality-filtered using fastp version 0.23.4, with a minimum Phred quality score of 20, and a minimum sequence length of 100 bp (Chen et al., 2018). Primer trimming was performed using Cutadapt version 4.4 and allowed for a maximum error rate of 0.1 and minimum overlaps of 3 bp for 16S reads and 10 bp for 18S reads (Martin, 2011). Amplicon sequence variants were inferred using the QIIME 2 DADA2 plugin version 2023.5.0, based on DADA2 version 1.26.0 (Callahan et al., 2016). Denoising was performed with no fixed-length truncation, a maximum expected-error threshold of 3, and a truncation quality threshold of 2. Chimeric sequences were removed during DADA2 denoising. Taxonomic classification was performed using the QIIME 2 feature-classifier plugin version 2023.5.0 and the classify-sklearn method, with scikit-learn version 0.24.1 (Bokulich et al., 2018; Pedregosa et al., 2011). The 16S ASVs were classified against the SILVA release 138 SSURef NR99 database using a full-length classifier (Quast et al., 2013; Yilmaz et al., 2014). The 18S ASVs were classified against PR2 version 5.1.0 using a classifier trained for the 528F/706R region (Guillou et al., 2013; Vaulot et al., 2025). A confidence threshold of 0.7 was applied for 16S classification, and a threshold of 0.5 was applied for 18S classification. Before 18S diversity and differential abundance analyses, ASVs with fewer than 10 total reads or present in less than 5% of samples were removed, reducing the table from 5,076 to 2,675 ASVs. Taxonomic composition was summarized from the initial unfiltered 18S table, and diversity and differential abundance analyses used the filtered table.

Diversity analyses used rarefied data (Schloss, 2024). The 16S ASV table was rarefied to 42,440 reads per sample, retaining all 13 samples, whereas the 18S ASV table was rarefied to 51,572 reads per sample, excluding Ragweed2 and retaining 10 samples. The 16S alpha diversity values were averaged across 100 rarefactions performed without replacement using a dedicated NumPy random number generator initialized with seed 42. Cosmetic plotting offsets were generated using a separate random number generator and did not affect the analytical random number stream. The 18S alpha and beta diversity analyses used one rarefaction with random seed 1. The archived 16S rarefied Bray-Curtis distance matrix was used for beta diversity analysis. Unrarefied tables were converted to relative abundances and used to summarize taxonomic composition at the phylum and genus levels using the phyloseq package (McMurdie and Holmes, 2013). Beta diversity was assessed using Bray-Curtis dissimilarity and visualized with principal coordinates analysis in the phyloseq and vegan packages. Alpha diversity was calculated as observed richness along with the Shannon and Simpson indices. Differences among plant groups were tested using the Kruskal-Wallis test, with *p*-values adjusted across the three indices using the Benjamini-Hochberg method. Associations between plant group and community composition were evaluated using PERMANOVA on Bray-Curtis dissimilarities calculated from relative abundances derived from the rarefied tables. 9,999 permutations were used for 16S, and 999 permutations for 18S (Anderson, 2001). Two sensitivity analyses repeated the 16S PERMANOVA using the same settings after excluding white clover and after excluding both white clover and chicory. Homogeneity of group dispersions for the overall plant group analyses was assessed using betadisper, with significance evaluated using permutest with 9,999 permutations for 16S and analysis of variance for 18S (Anderson, 2006). The 16S alpha diversity values and Kruskal-Wallis tests were calculated in Python version 3.12.13 using NumPy version 2.5.2 and SciPy version 1.18.0. All remaining statistical analyses were conducted in R version

4.2.3 using vegan version 2.6.10 and phyloseq version 1.42.0 (McMurdie and Holmes, 2013).

Nitrogen-related taxonomic proxies were assigned from SILVA 138 labels at genus rank or family rank when classification did not resolve further. Ammonia-oxidizing archaea included the family-rank labels Nitrososphaeraceae and Nitrosotaleaceae, together with the genera *Candidatus Nitrososphaera*, *Candidatus Nitrocosmicus* and *Candidatus Nitrosotalea*; ammonia-oxidizing bacteria comprised *Nitrosomonas* and *Nitrosospira*; and rhizobial lineage proxies were represented by *Bradyrhizobium*, *Mesorhizobium* and the composite SILVA label *Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium*. For proxy abundance, the relative abundance of the constituent taxa in each sample was summed using the unrarefied tables. Differences between the four legume and nine nonlegume samples were tested using the Wilcoxon rank sum test, and *p*-values were adjusted across the three proxy comparisons using the Benjamini-Hochberg method.

Examining common ragweed as an exploratory case, a binary PERMANOVA compared the three ragweed samples with the ten remaining 16S samples using the same rarefied Bray-Curtis distance matrix used for the main plant group analysis. The test used adonis2 with 9,999 permutations and random seed 20260801. For the ragweed comparison, homogeneity of group dispersions was assessed using betadisper and permutest with 9,999 permutations. For the exploratory *Plectosphaerella* analysis, the filtered unrarefied 18S count table was agglomerated at genus rank. Plant groups represented by fewer than two samples were excluded from differential abundance testing, leaving ragweed (n = 3), red clover (n = 3), and fescue (n = 4). Overall differences were evaluated using ANCOM-BC2 version 2.0.2 (Lin and Peddada, 2024) and a DESeq2 version 1.38.0 likelihood ratio test (Love et al., 2014), with Benjamini-Hochberg correction applied within each method. The single chicory sample was kept only for descriptive comparison.

Results

Sequencing and Distribution of Taxa

After quality filtering and denoising, a total of 938,887 16S reads were retained across the 13 samples. Sequencing depth ranged from 42,440 to 89,235 reads per sample, and was rarefied to 42,440 reads per sample for diversity analyses. Denoising yielded 12,447 16S ASVs. Across all samples, bacterial and archaeal communities were dominated by the phyla Actinobacteriota, Proteobacteria, Acidobacteriota, Crenarchaeota, and Chloroflexi (Figure 1A). The relative abundances of these dominant phyla varied descriptively among the five plant groups (Figure 1A).

Initial 18S denoising produced 999,054 reads and 5,076 ASVs across 11 samples. After abundance and prevalence filtering, a total of 988,023 18S reads and 2,675 ASVs were retained across 11 samples, ranging from 36,373 to 141,367 reads per sample. No usable 18S data were retained for the white clover sample or one of the two chicory samples, leaving chicory represented only once. The dataset was rarefied to 51,572 reads per sample for diversity analyses, excluding Ragweed2 because its sequencing depth was under this threshold, leaving 10 samples. Based on the unrarefied relative abundance table, eukaryotic reads were dominated by fungi at 72.6%, followed by land plants, classified as Streptophyta, at 9.0%, Cercozoa at 8.5%, and Metazoa at 3.0%. Ascomycota represented 86.7% of fungal reads and 62.9% of all eukaryotic reads, whereas Basidiomycota represented 4.6% of fungal reads and 3.3% of all

eukaryotic reads. The relative abundance of Ascomycota ranged descriptively from 69.2% under ragweed to 52.6% in the single chicory sample (Figure 1B).

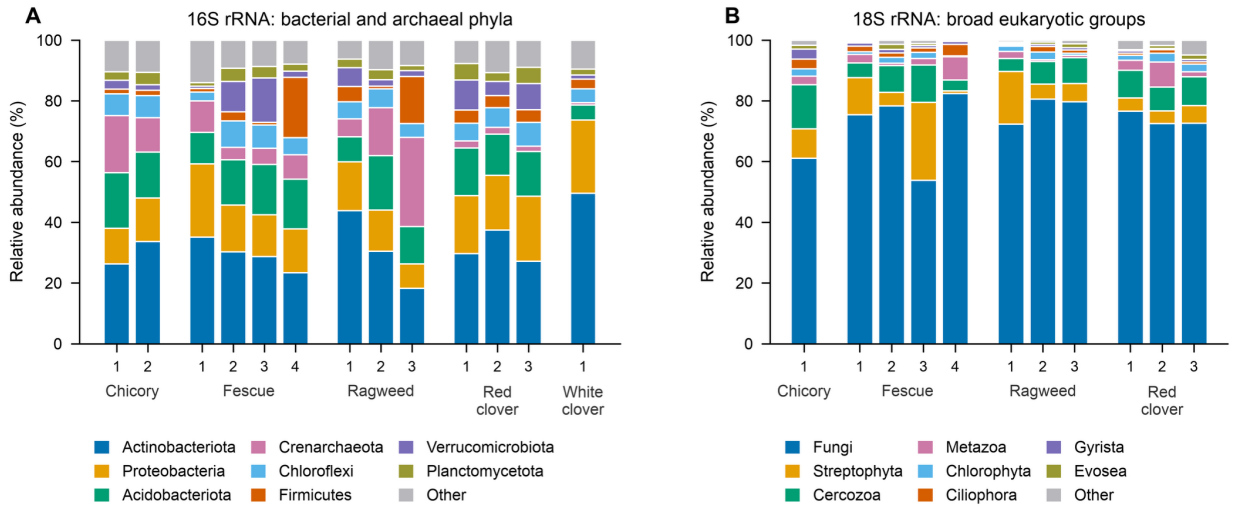

Figure 1. The relative taxonomic composition of soil microbial communities beneath the sampled plant groups. Each stacked bar represents one soil sample, and the colors indicate the relative read abundance of the taxonomic groups shown in the legend. (A) Dominant bacterial and archaeal phyla in the unrarefied 16S rRNA gene dataset (n = 13: chicory, 2; fescue, 4; ragweed, 3; red clover, 3; white clover, 1). (B) Broad eukaryotic groups in the initial unfiltered, unrarefied 18S rRNA gene dataset (n = 11: chicory, 1; fescue, 4; ragweed, 3; red clover, 3). No usable 18S data were available for white clover or one chicory sample. Taxa not displayed separately were combined in the category "Other".

Alpha diversity

For 16S, no significant difference in observed richness was detected among the five plant groups (p = 0.394; Figure 2A). Observed richness ranged from 2,236 ± 131 ASVs in red clover to 1,766 ASVs in the single white clover sample. No significant difference was detected for Shannon diversity (p = 0.113; Figure 2B), which ranged from 6.89 ± 0.02 in red clover to 6.15 ± 0.70 in ragweed. Simpson diversity also showed no significant difference (p = 0.109; Figure 2C), ranging from 0.9977 ± 0.0002 in red clover to 0.9795 ± 0.0231 in ragweed. All three indices remained nonsignificant after Benjamini-Hochberg correction within the 16S dataset: observed richness, q = 0.394; Shannon diversity, q = 0.169; and Simpson diversity, q = 0.169.

For 18S, no significant difference in observed richness was detected among the four represented plant groups (p = 0.100; Figure 3A). Observed richness ranged from 852 ± 33 ASVs in red clover to 607 ± 209 ASVs in fescue. No significant difference was detected for Shannon diversity (p = 0.086; Figure 3B), which was highest in the single chicory sample at 4.58 and was 4.55 ± 0.08 in red clover, the highest value among groups represented by multiple samples. Shannon diversity was lowest in fescue at 3.79 ± 0.63. Simpson diversity also showed no significant difference (p = 0.161; Figure 3C), ranging from 0.9670 in the single chicory sample and 0.9534 ± 0.0132 in red clover to 0.9102 ± 0.0476 in fescue. All three indices remained nonsignificant after Benjamini-Hochberg correction within the 18S dataset: observed richness, q = 0.150; Shannon diversity, q = 0.150; and Simpson diversity, q = 0.161.

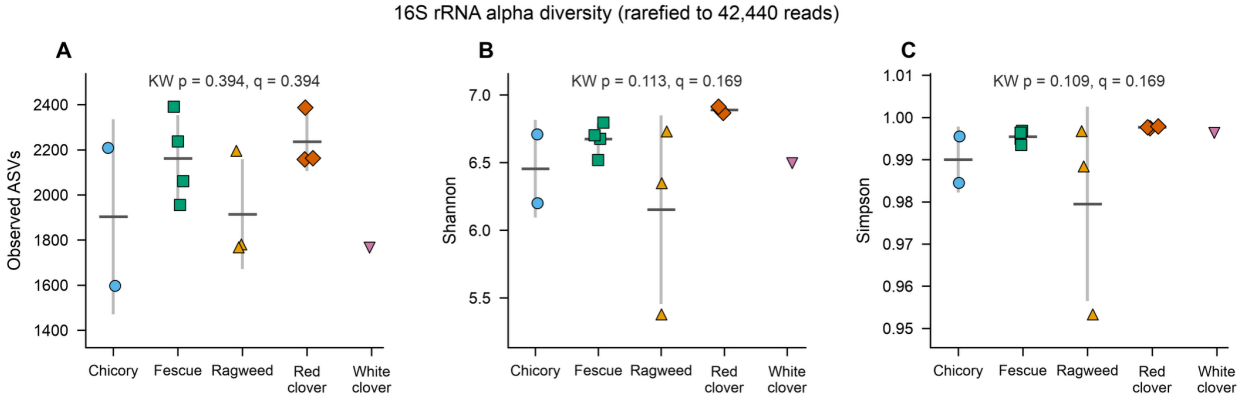

Figure 2. Alpha diversity of bacterial and archaeal communities among plant groups. (A) Observed amplicon sequence variants, (B) Shannon diversity, and (C) Simpson diversity. Each point represents one soil sample. Horizontal lines show group means, and vertical lines show one standard deviation for groups containing multiple samples. Sample sizes were chicory, $n = 2$; fescue, $n = 4$; common ragweed, $n = 3$; red clover, $n = 3$; and white clover, $n = 1$. Displayed p-values are from Kruskal-Wallis tests, and q-values were adjusted across the three diversity indices using the Benjamini-Hochberg method.

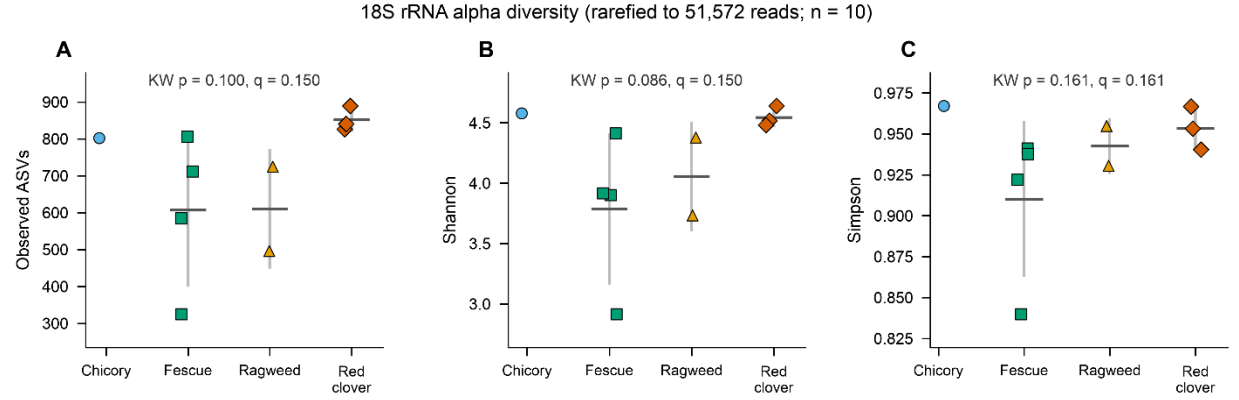

Figure 3. Alpha diversity of eukaryotic communities among plant groups. (A) Observed amplicon sequence variants, (B) Shannon diversity, and (C) Simpson diversity. Each point represents one soil sample. Horizontal lines show group means, and vertical lines show one standard deviation for groups containing multiple samples. Sample sizes were chicory, $n = 1$; fescue, $n = 4$; common ragweed, $n = 2$; and red clover, $n = 3$. No usable 18S data were available for the white clover sample or one chicory sample. One ragweed sample was excluded because its read count was below the rarefaction depth. Displayed p-values are from Kruskal-Wallis tests, and q-values were adjusted across the three diversity indices using the Benjamini-Hochberg method.

Beta diversity

Bray-Curtis principal coordinates analysis of the 13 bacterial and archaeal communities showed partial visual clustering according to plant group, with the first two axes explaining 37.6% and 13.8% of the total variation, respectively (Figure 4A). Samples collected beneath the same plant group had a lower mean pairwise Bray-Curtis dissimilarity of 0.566, compared with 0.646 between samples from different plant groups. PERMANOVA reflected significant differences in 16S community composition among plant groups ($R^2 = 0.473$, pseudo $F = 1.80$, $p = 0.020$). However, group dispersions also differed significantly (betadisper, $p < 0.001$), meaning that the PERMANOVA result may reflect differences in community composition, within-group

variability, or both. Sensitivity analysis showed that the association remained near the conventional significance threshold after excluding white clover ($R^2 = 0.402$, $p = 0.048$) but was not significant after also excluding chicory ($R^2 = 0.280$, $p = 0.182$).

For the 10 eukaryotic communities, the first two PCoA axes explained 21.4% and 15.1% of the total variation, respectively (Figure 4B). Mean pairwise Bray-Curtis dissimilarities were comparable within and between plant groups at 0.605 and 0.623, respectively. PERMANOVA did not detect significant differences in eukaryotic community composition among plant groups ($R^2 = 0.381$, pseudo $F = 1.23$, $p = 0.081$). Group dispersions nevertheless differed significantly (betadisper, $p = 0.012$).

An exploratory comparison of the three ragweed samples with the ten remaining samples did not detect a significant difference in overall 16S community composition (PERMANOVA $R^2 = 0.069$, pseudo $F = 0.811$, $p = 0.581$). Dispersion did not differ significantly between the two groups (betadisper, $p = 0.307$).

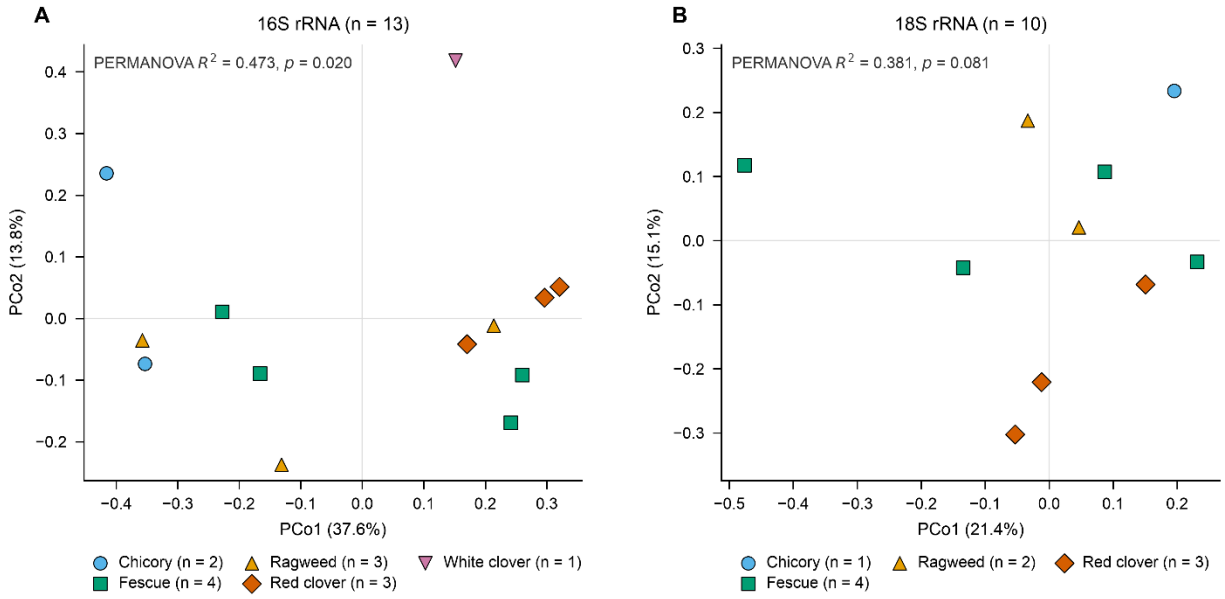

Figure 4. Principal coordinates analysis of Bray-Curtis dissimilarities among soil microbial communities beneath the sampled plant groups. (A) Bacterial and archaeal communities from the rarefied 16S rRNA gene dataset (n = 13). (B) Eukaryotic communities from the filtered, rarefied 18S rRNA gene dataset (n = 10). Each point represents one soil sample, and colors and symbols identify plant groups. Axis labels show the variation represented by each coordinate. Displayed R^2 and p-values are from PERMANOVA. Group dispersions differed significantly for both datasets, so that apparent separation may reflect community composition, within-group variability, or both.

Plant-Associated Patterns in Nitrogen-Cycling Microbial Proxies

Taxonomic proxies for ammonia-oxidizing archaea were overall more abundant than those for ammonia-oxidizing bacteria. Across samples, archaeal ammonia-oxidizer proxies, consisting mainly of Nitrososphaeraceae and related genera, had a mean relative abundance of 8.7%, compared with 0.06% for bacterial ammonia-oxidizer proxies represented primarily by *Nitrosomonas* and *Nitrosospira*. This corresponded to an approximate relative read abundance ratio of 148:1. Most archaeal ammonia-oxidizer proxy reads were classified only to family. Of the 8.73 percentage points attributed to this proxy, 8.41 came from the Nitrososphaeraceae label

rather than a named genus. The relative abundance of ammonia-oxidizing archaeal proxies descriptively varied among plant groups, ranging from 16.9% under ragweed and 14.6% under chicory to 1.9% under red clover and 0.5% under the single white clover sample. In comparison, ammonia-oxidizing bacterial proxies remained relatively uniform across samples, ranging from 0.01% to 0.14%. Mean bacterial ammonia-oxidizer proxy abundance did not differ significantly between the four legume and nine nonlegume samples (0.052% versus 0.062%, respectively; Wilcoxon rank sum test, $W = 15$, raw $p = 0.710$, $q = 0.710$, Figure 5C).

Grouped by plant functional type, the mean relative abundance of ammonia-oxidizing archaeal proxies was lower beneath the four legume samples than beneath the nine nonlegume samples, at 1.6% and 11.9%, respectively, with no overlap between the two sets of samples (Wilcoxon rank sum test, $W = 0$, raw $p = 0.00280$, $q = 0.00839$; Figure 5A). Taxonomic proxies for genera containing known nitrogen-fixing rhizobia showed the reverse pattern, reaching 1.85% under the single white clover sample and 1.57% under red clover, compared with 0.35% to 0.74% under the nonlegumes. Their mean relative abundance was 1.6% beneath legumes and 0.6% beneath nonlegumes (Wilcoxon rank sum test, $W = 35$, raw $p = 0.00559$, $q = 0.00839$; Figure 5B). Together, these taxonomic proxies showed opposite associations with plant functional type. Rhizobial proxies were more abundant beneath legumes, whereas archaeal ammonia-oxidizer proxies were more abundant beneath nonlegumes.

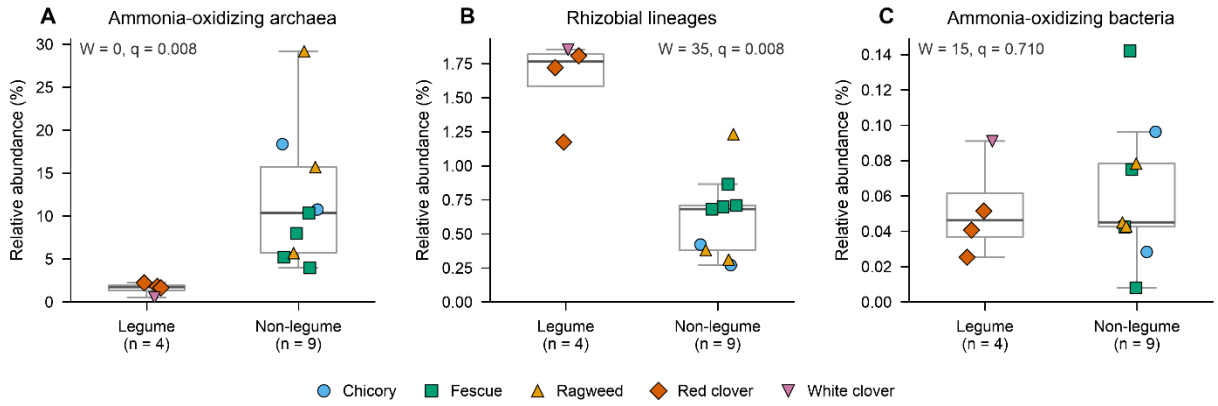

Figure 5. Relative read abundance of nitrogen-related taxonomic proxies beneath legumes and nonlegumes. (A) Ammonia-oxidizing archaea, (B) rhizobial lineages, and (C) ammonia-oxidizing bacteria. Each point represents one soil sample, with colors and symbols identifying plant groups. Boxes show the interquartile range, center lines show medians, and whiskers extend to the most extreme values within 1.5 times the interquartile range. Displayed W values are from Wilcoxon rank sum tests, and *q*-values were adjusted across the three comparisons using the Benjamini-Hochberg method. These taxonomic proxies do not directly measure nitrogen transformation activity.

Exploratory Plectosphaerella analysis

Using the filtered unrarefied 18S data, mean relative abundance of *Plectosphaerella* was 0.785% beneath ragweed ($n = 3$), 0.228% beneath red clover ($n = 3$), and 0.141% beneath fescue ($n = 4$), with no overlap among these three replicated groups. The single chicory sample, which was excluded from differential abundance testing, had a relative abundance of 5.72%. ANCOM-BC2 supported an overall difference among the three replicated groups after Benjamini-Hochberg correction ($q = 0.00144$), whereas the DESeq2 likelihood ratio test did not ($q = 0.654$; Figure 6).

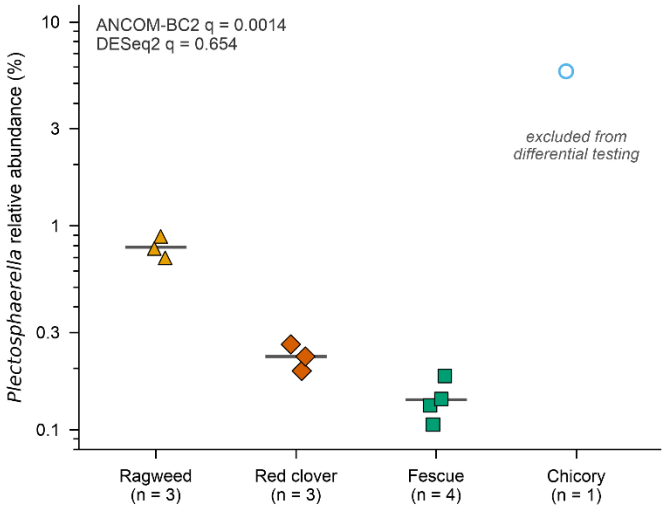

Figure 6. Relative read abundance of the 18S lineage assigned to Plectosphaerella among plant groups. Values were calculated from the filtered, unrarefied 18S dataset. Each point represents one soil sample, and horizontal lines show group means for groups with multiple samples. Chicory is shown descriptively but was excluded from differential abundance testing because only one sample was available. ANCOM-BC2 supported a difference among ragweed, red clover, and fescue after Benjamini-Hochberg correction, whereas DESeq2 did not. The y-axis uses a logarithmic scale.

Discussion

This study examined bacterial, archaeal, and eukaryotic soil communities beneath five plant groups within a historically disturbed orchard area undergoing restoration. The dominant 16S phyla were Actinobacteriota, Proteobacteria, Acidobacteriota, Crenarchaeota and Chloroflexi. Several of these phyla are commonly prominent in soils, indicating that the dominant phylum-level profile resembled patterns commonly reported from soils (Janssen, 2006). However, this resemblance does not demonstrate ecological recovery because no earlier time point was sampled. Actinobacteriota include lineages involved in the decomposition of complex plant residues, Proteobacteria are abundant and phylogenetically diverse in soils, and Acidobacteriota include many slow-growing lineages associated with the limited presence of resources (Fierer et al., 2007; Spain et al., 2009; Ward et al., 2009; Bao et al., 2021).

Fungi accounted for most 18S reads, with Ascomycota dominating the fungal community. Ascomycota include important decomposers of plant residues, relevant to restoration processes involving organic matter turnover (Ma et al., 2013). Importantly, the present data only describe taxonomic composition rather than decomposition activity or rate. Cercozoa were the most abundant recovered protist group. However, abundance alone does not identify their ecological roles. Determining those roles would require finer taxonomic resolution and supporting trait data (Dumack et al., 2020).

The most distinct targeted taxonomic pattern was observed for ammonia-oxidizer proxies. Archaeal ammonia-oxidizer proxies accounted for 8.73% of 16S reads, compared with 0.059% for bacterial ammonia-oxidizer proxies. Archaeal lineages were frequently favored in low ammonium availability and bacterial lineages responded more strongly to higher nitrogen inputs, although pH and other environmental elements also influence their distribution (Verhamme et al., 2011; Prosser and Nicol, 2012; Carey et al., 2016). Further, archaeal ammonia-oxidizer proxies showed a strong association with plant functional type, occurring at consistently lower relative

abundances beneath legumes than beneath nonlegumes. Rhizobial proxies showed the opposite pattern, with greater relative abundance beneath clovers than beneath nonlegumes. This association is biologically consistent with the capacity of legumes to form nitrogen-fixing symbioses with rhizobia. However, the detected genera include both symbiotic and nonsymbiotic lineages, and 16S classification does not resolve the functional genes required for nodulation or nitrogen fixation. The observed enrichment therefore represents a clover-associated rhizobial taxonomic signal, but not evidence of enhanced biological nitrogen fixation. The reciprocal ordering of archaeal ammonia-oxidizer and rhizobial proxies provides a hypothesis for further study. Soil ammonium, nitrate, pH, moisture, and root activity were not measured. The present finding should therefore not be generalized as a universal effect of legumes.

Ragweed is a volunteer plant associated with a practical restoration question, and it had the highest mean archaeal ammonia-oxidizer proxy abundance. The exploratory comparison of common ragweed did not support a distinct overall 16S community, relative to the other vegetation groups, indicating that its presence was not associated with broad restructuring of the bacterial and archaeal community. A more specific 18S pattern was observed for the 18S lineage assigned to *Plectosphaerella*, which had a higher mean relative read abundance beneath ragweed than beneath red clover or fescue. However, this difference was supported by ANCOM-BC2 but not DESeq2, and the single chicory sample contained an even greater relative abundance. Because ragweed and chicory both belong to *Asteraceae*, the observation raises the possibility of a more extensive plant-family association hypothesis requiring replicated testing. The ecological function of the detected *Plectosphaerella* cannot be resolved from the 18S data, making this a hypothesis for future study instead of evidence for a beneficial or detrimental role of ragweed in restoration.

Overall, the survey shows that microbial composition varied among plant-associated patches within one recovering area, while alpha diversity did not differ significantly among plant groups. Diversity values alone do not establish whether soil is healthy or successfully recovering (Fierer et al., 2021). The 16S community result supports an association with plant group, but unequal dispersion, unbalanced replication, sensitivity to the smallest groups, and unmeasured microsite conditions prevent it from being interpreted as a clean plant effect. The 18S whole community result was inconclusive.

The study identifies potential plant-microbe relationships for further investigation rather than demonstrating restoration outcomes. The clearest pattern was the contrasting distribution of nitrogen-related microbial groups beneath legumes and nonlegumes, with rhizobial proxies more abundant beneath legumes and archaeal ammonia-oxidizer proxies more abundant beneath nonlegumes. An additional exploratory pattern was observed for *Plectosphaerella*, although this finding requires confirmation with greater and more balanced replication.

Limitations

This study was conducted within one historically disturbed orchard area at one timepoint over two consecutive days. No undisturbed reference area or earlier timepoint was sampled, so the data cannot establish the site's restoration status or determine how its microbial communities have changed through time. The modest and unequal sample sizes also limit plant group-specific conclusions. Plant group therefore cannot be separated from preexisting microsite conditions.

The samples integrated bulk soil from the surface to approximately 45 cm depth and should not be interpreted as rhizosphere samples.

Methodologically, fungi were characterized only through the general eukaryotic 18S rRNA marker rather than a dedicated fungal ITS barcode; therefore, species-level fungal resolution was limited (Schoch et al., 2012; Tedersoo et al., 2015). Although 18S can recover *Glomeromycetes*, no arbuscular mycorrhizal fungi (AMF) specific primer set was used; therefore, AMF communities were not comprehensively characterized, and their low recovered proportion should not be interpreted as evidence of low abundance (George et al., 2019). As with all amplicon surveys, the selected markers and primers carry known taxonomic biases (Tedersoo et al., 2015; George et al., 2019), and the data reflect relative read abundances rather than absolute organism abundances or measured biological activity (Gloor et al., 2017).

Soil pH, ammonium, nitrate, moisture, organic matter, and other physicochemical properties were not measured directly. Consequently, observed differences in archaeal ammonia-oxidizer proxies cannot be disentangled from variation in soil chemistry, nitrogen availability or moisture across sampling locations. Soil pH is particularly important because it strongly influences the distributions of archaeal and bacterial ammonia-oxidizers (Prosser and Nicol, 2012; Carey et al., 2016).

Two features of the beta diversity analysis warrant caution. Group dispersions differed significantly for both markers, and the unequal group sizes make differences in community composition difficult to separate from differences in group dispersion (Anderson, 2006; Anderson and Walsh, 2013). The 16S result was sensitive to the inclusion of the smallest plant groups and should therefore be considered suggestive rather than conclusive.
Intermittent sheep grazing may also have produced localized variation in nitrogen, as livestock return nitrogen unevenly in dung and urine (Haynes and Williams, 1993; Moir et al., 2011). Grazing locations and waste patches were not mapped or measured, so this potential source of variation cannot be separated from plant group associations in archaeal ammonia-oxidizer proxies.

Conclusions

This pilot observational study characterized soil microbial communities within a historically disturbed orchard undergoing restoration and identified plant-associated taxonomic patterns potentially relevant to belowground biological function. Although overall bacterial, archaeal, and eukaryotic alpha diversity did not differ significantly among plant groups, differences emerged in community composition, notably in microbial taxonomic proxies associated with nitrogen cycling. Archaeal ammonia-oxidizer proxies were consistently less abundant beneath legumes than beneath nonlegumes, whereas rhizobial lineage proxies showed the opposite pattern. These reciprocal associations suggest that plant functional type may be linked to differences in the relative read abundance of taxonomic proxies involved in soil nitrogen transformations during restoration. However, because the amplicon data represent relative taxonomic abundance rather than functional activity, these patterns do not demonstrate differences in nitrification, nitrogen fixation, or restoration success. Exploratory analyses identified additional plant-associated patterns, including variation in *Plectosphaerella*, but these findings were less consistent and call for further investigation. Overall, the results suggest that biologically relevant differences among

plant-associated patches may be reflected more clearly in the representation of specific microbial functional groups than in overall microbial diversity. Future studies should combine balanced, spatially matched, and longitudinal sampling with measurements of soil physicochemical properties, functional genes, and microbial process rates to determine whether these taxonomic patterns correspond to changes in nitrogen cycling and other biological functions during soil restoration.

Acknowledgements
This descriptive study was conducted as part of an internship at Hidden Creek Farm, which organized the project and provided site access and continued assistance throughout. We thank Andrea Young, who jointly manages Hidden Creek Farm, for helping develop the project's direction and coordinating its implementation at the farm. We thank Liu Jihong for providing periodic guidance with the bioinformatic processing of the sequencing data. Amplicon sequencing was carried out by Novogene Co., Ltd.

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, *62*(1), 245–253. https://doi.org/10.1111/j.1541-0420.2005.00440.x

Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs*, *83*(4), 557–574. https://doi.org/10.1890/12-2010.1

Bao, Y., Dolfing, J., Guo, Z., Chen, R., Wu, M., Li, Z., Lin, X., & Feng, Y. (2021). Important ecophysiological roles of non-dominant Actinobacteria in plant residue decomposition, especially in less fertile soils. *Microbiome*, *9*, Article 84. https://doi.org/10.1186/s40168-021-01032-x

Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A. I. J. M., & Miralles, D. G. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, *10*, Article 724. https://doi.org/10.1038/s41597-023-02549-6

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Caporaso, J. G. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*, Article 90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible

microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Suppl. 1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Carey, C. J., Dove, N. C., Beman, J. M., Hart, S. C., & Aronson, E. L. (2016). Meta-analysis reveals ammonia-oxidizing bacteria respond more strongly to nitrogen addition than ammonia-oxidizing archaea. *Soil Biology and Biochemistry*, *99*, 158–166. https://doi.org/10.1016/j.soilbio.2016.05.014

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cheung, M. K., Au, C. H., Chu, K. H., Kwan, H. S., & Wong, C. K. (2010). Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *The ISME Journal*, *4*(8), 1053–1059. https://doi.org/10.1038/ismej.2010.26

Dassen, S., Cortois, R., Martens, H., de Hollander, M., Kowalchuk, G. A., van der Putten, W. H., & De Deyn, G. B. (2017). Differential responses of soil bacteria, fungi, archaea and protists to plant species richness and plant functional group identity. *Molecular Ecology*, *26*(15), 4085–4098. https://doi.org/10.1111/mec.14175

Dumack, K., Fiore-Donno, A. M., Bass, D., & Bonkowski, M. (2020). Making sense of environmental sequencing data: Ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Molecular Ecology Resources*, *20*(2), 398–403. https://doi.org/10.1111/1755-0998.13112

Fierer, N., Bradford, M. A., & Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology*, *88*(6), 1354–1364. https://doi.org/10.1890/05-1839

Fierer, N., Wood, S. A., & Bueno de Mesquita, C. P. (2021). How microbes can, and cannot, be used to assess soil health. *Soil Biology and Biochemistry*, *153*, Article 108111. https://doi.org/10.1016/j.soilbio.2020.108111

Gellie, N. J. C., Mills, J. G., Breed, M. F., & Lowe, A. J. (2017). Revegetation rewilds the soil bacterial microbiome of an old field. *Molecular Ecology*, *26*(11), 2895–2904. https://doi.org/10.1111/mec.14081

George, P. B. L., Creer, S., Griffiths, R. I., Emmett, B. A., Robinson, D. A., & Jones, D. L. (2019). Primer and database choice affect fungal functional but not biological diversity findings in a national soil survey. *Frontiers in Environmental Science*, *7*, Article 173. https://doi.org/10.3389/fenvs.2019.00173

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, Article 2224. https://doi.org/10.3389/fmicb.2017.02224

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(D1), D597–D604. https://doi.org/10.1093/nar/gks1160

Haichar, F. Z., Marol, C., Berge, O., Rangel-Castro, J. I., Prosser, J. I., Balesdent, J., Heulin, T., & Achouak, W. (2008). Plant host habitat and root exudates shape soil bacterial community structure. *The ISME Journal*, *2*(12), 1221–1230. https://doi.org/10.1038/ismej.2008.80

Haynes, R. J., & Williams, P. H. (1993). Nutrient cycling and soil fertility in the grazed pasture ecosystem. *Advances in Agronomy*, *49*, 119–199. https://doi.org/10.1016/S0065-2113(08)60794-4

Janssen, P. H. (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*, *72*(3), 1719–1728. https://doi.org/10.1128/AEM.72.3.1719-1728.2006

Kuramae, E. E., Gamper, H. A., Yergeau, E., Piceno, Y. M., Brodie, E. L., DeSantis, T. Z., Andersen, G. L., van Veen, J. A., & Kowalchuk, G. A. (2010). Microbial secondary succession in a chronosequence of chalk grasslands. *The ISME Journal*, *4*(5), 711–715. https://doi.org/10.1038/ismej.2010.11

Lin, H., & Peddada, S. D. (2024). Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nature Methods*, *21*(1), 83–91. https://doi.org/10.1038/s41592-023-02092-7

Lindström, K., & Mousavi, S. A. (2020). Effectiveness of nitrogen fixation in rhizobia. *Microbial Biotechnology*, *13*(5), 1314–1335. https://doi.org/10.1111/1751-7915.13517

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), Article 550. https://doi.org/10.1186/s13059-014-0550-8

Ma, A., Zhuang, X., Wu, J., Cui, M., Lv, D., Liu, C., & Zhuang, G. (2013). Ascomycota members dominate fungal communities during straw residue decomposition in arable soil. *PLoS ONE*, *8*(6), Article e66146. https://doi.org/10.1371/journal.pone.0066146

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*(4), Article e61217. https://doi.org/10.1371/journal.pone.0061217

Moir, J. L., Cameron, K. C., Di, H. J., & Fertsak, U. (2011). The spatial coverage of dairy cattle urine patches in an intensively grazed pasture system. *The Journal of Agricultural Science*, *149*(4), 473–485. https://doi.org/10.1017/S0021859610001012

Palecki, M., Durre, I., Applequist, S., Arguez, A., & Lawrimore, J. (2021). *U.S. Climate Normals 2020: U.S. Annual/Seasonal Climate Normals (1991–2020)* [Data set]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/4ek7-fk11

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Prosser, J. I., & Nicol, G. W. (2012). Archaeal and bacterial ammonia-oxidisers in soil: The quest for niche specialisation and differentiation. *Trends in Microbiology*, *20*(11), 523–531. https://doi.org/10.1016/j.tim.2012.08.001

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Schloss, P. D. (2024). Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. *mSphere*, *9*(2), Article e00354-23. https://doi.org/10.1128/msphere.00354-23

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., & Fungal Barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(16), 6241–6246. https://doi.org/10.1073/pnas.1117018109

Soil Survey Staff. (2025). *Soil Survey Geographic (SSURGO) Database* [Data set]. Natural Resources Conservation Service, U.S. Department of Agriculture. Retrieved 8/10/26, from https://websoilsurvey.nrcs.usda.gov/

Spain, A. M., Krumholz, L. R., & Elshahed, M. S. (2009). Abundance, composition, diversity and novelty of soil Proteobacteria. *The ISME Journal*, *3*(8), 992–1000. https://doi.org/10.1038/ismej.2009.43

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, H., Hildebrand, F., Bork, P., & Abarenkov, K. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*, *10*, 1–43. https://doi.org/10.3897/mycokeys.10.4852

Thion, C. E., Poirel, J. D., Cornulier, T., De Vries, F. T., Bardgett, R. D., & Prosser, J. I. (2016). Plant nitrogen-use strategy as a driver of rhizosphere archaeal and bacterial ammonia oxidiser abundance. *FEMS Microbiology Ecology*, *92*(7), Article fiw091. https://doi.org/10.1093/femsec/fiw091

U.S. Geological Survey. (2022). *The National Map: 3D Elevation Program (3DEP)* [Data set]. U.S. Department of the Interior. Retrieved 8/10/26, from https://apps.nationalmap.gov/

Vaulot, D., del Campo, J., Pohl, N., Salomaki, E., Gu, H., Bonacolta, A., Morard, R., Krabberød, A. K., Arañó Ansola, J., & Seliuk, A. (2025). *PR2 version 5.1.0 – data* (Version 5.1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15129782

Verhamme, D. T., Prosser, J. I., & Nicol, G. W. (2011). Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. *The ISME Journal*, *5*(6), 1067–1071. https://doi.org/10.1038/ismej.2010.191

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., Xie, G., Haft, D. H., Sait, M., Badger, J., Barabote, R. D., Bradley, B., Brettin, T. S., Brinkac, L. M., Bruce, D., Creasy, T., Daugherty, S. C., Davidsen, T. M., DeBoy, R. T., … Kuske, C. R. (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology*, *75*(7), 2046–2056. https://doi.org/10.1128/AEM.02294-08

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., & Glöckner, F. O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, *42*(D1), D643–D648. https://doi.org/10.1093/nar/gkt1209

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., Cho, H., Karaoz, U., Loqué, D., Bowen, B. P., Firestone, M. K., Northen, T. R., & Brodie, E. L. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology*, *3*(4), 470–480. https://doi.org/10.1038/s41564-018-0129-3