UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Influences of Geospatial Location and Depth on Soil Microbial Community Composition

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Genetics, Genomics, and Bioinformatics

by

Keshav Arogyaswamy

June 2020

Thesis Committee:

Dr. Emma Aronson, Chairperson
Dr. James Borneman
Dr. Samantha Ying

The Thesis of Keshav Arogyaswamy is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I have had a great deal of support, advice, and guidance from many people, but I would particularly like to thank my committee members who have been with me from the beginning, Emma Aronson, James Borneman, and Samantha Ying. I am also grateful for the guidance of Holly Bik and Thomas Girke. I am incredibly thankful for the support and advice of Brooke Pickett, without whom I wouldn't be in this place now. In addition, I owe a great deal to my friends and colleagues who have always been eager to talk about problems and share excitement about new findings: Chelsea Carey, Mia Maltz, Denise Jackson, Hannah Shulman, Hannah Freund, Geoffrey Logan, Karisa Caso, and Rachelle Ruble. And of course, last but in no way least, my friends and family who listened patiently while I babbled on about my research without any context, but still unfailingly responded with enthusiasm, support, and love.

To my Mom, who has given me constant support and always pushed me to fulfill my potential. To my sister, whose relentless love and validation have kept me going. And to my family, friends, and colleagues who have helped me along the way.

ABSTRACT OF THE THESIS

The Influences of Geospatial Location and Depth on Soil Microbial Community Composition

by

Keshav Arogyaswamy

Master of Science, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, June 2020
Dr. Emma Aronson, Chairperson

Soil microbial communities are vital to a wide range of ecosystems, but they are poorly understood. Advances in high-throughput amplicon sequencing have enabled a greater understanding of how communities differ from each other. Previous studies have both found important variations with depth within soil profiles and across geospatial location, but few studies have examined both depth and location simultaneously. This study utilizes the Critical Zone Observatory Network to assay a total of 18 sites at a continental scale, each sampled to approximately 100 cm in 10-cm increments. Each of the approximately 180 biologically distinct samples was assessed for a variety of pedological variables including pH, particle size distribution, and nutrient composition. In addition, 16S (prokaryotic) and fungal ITS1 amplicons were sequenced to examine microbial community composition. Although previous literature has seen a marked decrease in diversity with increasing depth for 16S communities, this study shows that such soil profiles form a distinct subset, and that in many sites, there is no significant decrease in diversity with depth (termed "gradient" and "uniform" profiles here, respectively). Futhermore, while site-to-site variations tend to be

important drivers of community differences, there are also strong similarities between extremely geographically distant communities that share similarities in ecological setting. In particular, soil taxonomic order and dominant vegetation types show strong correspondences with similarities between communities, but the tendency towards distinction between such groupings is overshadowed by the gradient-uniform dichotomy. This study, therefore, suggests that future work in comparisons of microbial communities across geospatial gradients needs to account for the ecological setting of selected sites, in particular accounting for soil taxonomy and vegetation. Although there are hints that the gradient-uniform dichotomy is linked to changes in soil texture and the relative availability of carbon and nitrogen, further research is required to robustly explain this distinction.

Contents

List of Figures

Introduction

Soil microbial communities perform a vast variety of ecological functions. They demonstrate a great degree of diversity both phylogenetically and functionally (Torsvik and Øvreås, 2002). Much of this diversity is only beginning to be understood with the advent of high-throughput sequencing technology, because the vast majority of soil microbes have not been successfully cultured (Torsvik and Øvreås, 2002; Daniel, 2005). While it has been shown that microbes can have major impacts on the cycling of nutrients such as nitrogen and carbon(Rothman et al., 2014; Placella et al., 2012; Bardgett et al.; Conrad, 2009), much remains to be understood about the ways in which microbial communities and their activities vary between terrestrial environments. Previous studies have demonstrated that these communities show strong differences with increasing depth in soil profiles (Gabor et al., 2014), and have found greater differences across depth than between locations (Eilers et al., 2012). Studies at larger scales have found strong correlations between the environment and microbial community composition (Thompson et al., 2017), but did not investigate the effect of depth within soil profiles. Differences in the soil microbial communities' composition and function between ecologically distinct areas can be important for better understanding the relationship of soil microbes and their macro-environment.

The Critical Zone

The Critical Zone (CZ) is defined as the three-dimensional part of the Earth situated between the highest treetops and the lowest groundwater (Anderson, 2015).In other words, it is the zone of greatest heterogeneity and the highest concentration of life, encompassing most terrestrial life, soil, regolith, and groundwater, as well as the vegetation-atmosphere interface, and the bedrock weathering front. The primary objective of CZ science is to characterize and understand how the reciprocal interactions among rock, soil, water, air, and terrestrial organisms influence the Earth's surface and near-surface environment. The

CZ perspective incorporates multiple spatial scales, both through a vertical profile, as well as horizontally across the large environmental gradient of the Earth's surface. Further, CZ science examines interactions over multiple time scales, ranging from eons to seconds as it incorporates geological and ecological time scales.

Generally, the Critical Zone Observatories (CZOs) have incorporated a watershed approach because watersheds can be treated as a cohesive system. Within that context, various locations of the former CZOs offer a wide spectrum of tectonic, hydrological, biological, climatic, and anthropogenic processes driving both CZ and landscape evolution. In the Southern Sierra CZO in California, there is a strong climate gradient and rain to snow transition. Bedrock type drives critical differences at Luquillo CZO in Puerto Rico. At the Jemez-Catalina CZO in Arizona and New Mexico, the focus has been on how subsurface biogeochemistry determines changes in the CZ. The Calhoun CZO in South Carolina is in a region that faced serious environmental degradation and recent reforestation. Finally, Reynolds Creek CZO in Idaho has focused on how various drivers of the variation in CZ properties impact soil carbon transformations and balance. For logistical reasons, the Eel River CZO was not included in this study.

Soil microbial communities have been seen to vary strongly by depth (Fierer et al., 2003), more so than by location. Previous studies have shown only modest differences between microbial communities based on geospatial location, but those studies have either focused on a smaller geospatial extent (Fierer et al., 2003; Eilers et al., 2012) or have focused on surficial soil samples (Thompson et al., 2017; Keller et al., 2008; Seaton et al., 2020), rather than examining the microbial communities that are unique to deeper soils.

Related analysis of these samples (Brewer et al., 2019) showed that geospatial location had a strong influence on overall 16S community composition by site. A small subset of phyla were found to increase in relative abundance with increasing depth, including a relatively undescribed phylum, *Dormibacteraeota*, which had previously been described only in

Antarctic soils and thawing permafrost (Ji et al., 2017; Woodcroft et al., 2018), and a general shift in taxonomy with depth. However, while that analysis was focused on assembly and characterization of *Dormibacteraeota*, it also reported a general decrease in diversity with depth, as has been observed in previous studies (Fierer et al., 2003; Eilers et al., 2012). The current study, although using the same underlying prokaryotic 16S sequencing data, highlights differences between sites concerning the trend of diversity with depth—differences that were not readily apparent from shotgun metagenomic sequencing and a solely amplicon sequence variant (ASV)-based analysis of the 16S data. This study incorporates prokaryotic 16S and fungal ITS amplicon sequencing—the former being analyzed using both legacy operational taxonomic unit (OTU)-analysis of 16S communities—as well as phospholipid fatty acid (PLFA) analysis to examine microbial community diversity more than individual taxa. As outlined by Tripathi et al. (2018), there are still significant gaps in microbial ecology databases for non–host-associated microbiomes, suggesting that environmental microbiology studies, particularly in poorly-described habitats like terrestrial soils, should use community-based descriptions rather than try to draw conclusions about community functions, given the ubiquity of undescribed or poorly described taxa at the ASV or even species level.

Furthermore, to properly address soil microbial community activities, DNA-based studies can only imply potentials, rather than actual activities. Relic DNA from dead microbes whose DNA is not yet degraded can persist in dry, terrestrial environments for long periods (Carini et al., 2016). Use of PLFA-based analysis can help mitigate this effect for calculations of diversity because PLFAs degrade more quickly than DNA; thus PLFAs are more likely to represent living organisms than is DNA. However, this approach is unable to account for expression and activity of functional genes. Functional activity for some processes can be captured for soils, and is being conducted for the samples in this study (Dove et al., 2020), but is limited to processes for which established assays exist.

Methods

Soil Collection

Soil pits were excavated by hand at 20 sites within the United States (including two in Puerto Rico) representing the 10 Critical Zone Observatories (Lin et al., 2011; Critical Zone Observatory Network, 2018), shown in Table 1 and Figure 1. At most sites, soil pits were dug to at least 100 cm or refusal. In addition, if a particular profile extended beyond 100 cm, an optional sample of the last 10 cm to refusal were also collected. The sites from the Intensively Managed Landscape CZO and Christina River Basin CZO were sampled by core auger due to personnel constraints (Christina was composited from 4 cores). For all but two sites, soil from the pit or core was collected sterilely using either a soil knife or a coring auger inserted into the pit wall horizontally, integrating soil in 10-cm increments (i.e., 0–10 cm, 10–20 cm, etc.). In contrast, the Luquillo CZO sites were collected at every 10-cm interval, covering plus-or-minus 3 cm, except for the surface which was sampled for the first 6 cm (i.e., 0–6 cm, 7–13 cm, 17–23 cm, etc.).

Once collected, samples were sent to the University of California, Riverside by overnight delivery without freezing, but kept cool with ice packs to minimize microbial community changes and activity after collection.

Collection dates were planned based on local peak greenness as estimated from Normalized Difference Vegetation Index and Enhanced Vegetation Index measured by NASA's MODIS (MODerate-resolution Imaging Spectroradiometer) instrument aboard the *Terra* satellite.

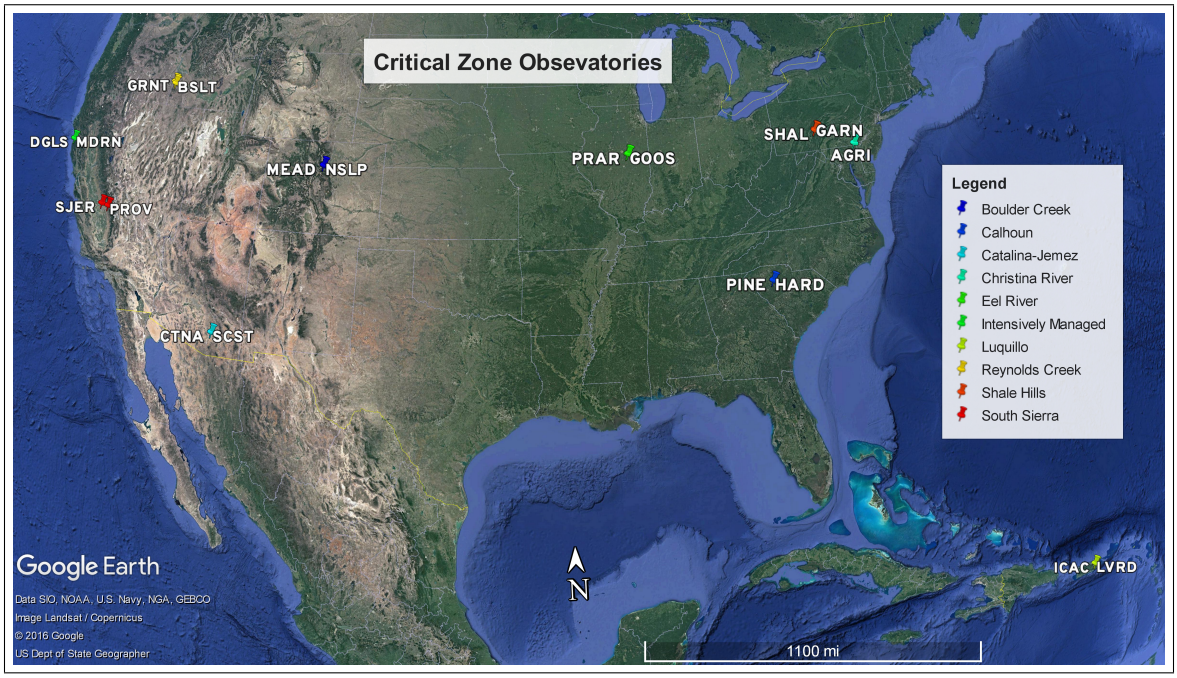

Figure 1: A map of the sampling sites used. The twenty sites were selected by collaborators at each of the ten Critical Zone Observatories, and represent a variety of biomes and soil types.

CZO	**Site Description**	Code	Location
Boulder Creek	Gordon Gulch Meadow	MEAD	40.02, -105.47
	Gordon Gulch North-Facing Slope	NSLP	40.01, -105.46
Calhoun	Hardwood Forest (R7H1)	HARD	34.60, -81.72
	Pine Forest (R7P1)	PINE	34.60, -81.72
Catalina/Jemez	Catalina—Granite Profile	CTNA	32.43, -110.77
	Catalina—Schist Profile	SCST	32.43, -110.76
Christina	White Clay Creek (Flood Plain)	FLUD	39.86, -75.78
	White Clay Creek (Agricultural Site)	AGRI	39.86, -75.78
Intensively Managed	Goose Creek (Corn field)	GOOS	40.43, -88.55
	Sangamon River Forest Reserve	PRAR	40.42, -88.60
Luquillo	Icacos	ICAC	18.28, -65.79
	El Verde	LVRD	18.32, -65.81
Reynolds Creek	Whisky Hill (Granite Profile)	GRNT	43.19, -116.81
	S. Upper Sheep Creek (Basalt Profile)	BSLT	43.11, -116.72
Shale Hills	Shale Hill	SHAL	40.66, -77.90
	Garner Run	GARN	40.69, -77.91
South Sierra	San Joaquin Experimental Reserve	SJER	37.10, -119.73
	Providence	PROV	37.06, -119.19

Table 1: This table shows a list of all of the 18 sites included in this project. Local names and descriptions are provided where available.

Soil Processing

Upon arrival to UC Riverside, a portion of each field sample was sieved (2 mm opening, ASTM No. 10), and the sieved portion was homogenized by shaking and/or mixing the sample within the sterile sample bag. The sieved, homogenized samples were divided into subsamples for further analysis, some stored at 4°C, −20°C, and −80°C. For some soils (particularly some wet, finely textured depth intervals), sieving was not practical under field-fresh conditions. These samples were homogenized by mixing, but during the subsampling process, larger root fragments and granules were excluded by visual observation. In addition, SHAL samples from 70–100 cm consisted almost entirely of medium-sized rocks ("cobbles" in USGS nomenclature); soil was collected both by manually crushing rocks to pass through the sieve and by scraping soil from the faces of the cobbles.

Soil profile classification

Due to personnel constraints, in situ characterization of soil profiles was not possible, so the identity of the observed soil series was inferred using a combination of soil survey data and empirically obtained data. Geospatial location was used to identify the soil map unit as described in the Natural Resources Conservation Service and the National Cooperative Soil Survey (Survey, 2018; Staff et al., 2018). Soil map units which contained multiple co-occurring series were more precisely identified by comparing soil texture, total profile depth, and landform position. The NCSS official series descriptions were used to infer soil taxonomic order, dominant vegetation type, primary biome, and moisture/temperature regimes.

DNA Extraction

DNA was extracted using the MoBio PowerLyzer PowerSoil DNA Purification Kit with modifications to increase yield based on the assumption that some sites and depths would

have a relatively low microbial biomass. Specifically, 0.25 g of soil was weighed in triplicate (i.e., $3 \times 0.25\text{g} = 0.75\text{g}$ total soil per sample) from a frozen aliquot of the sieved soil (from a subsample reserved only for DNA extractions). Extractions on the replicates proceeded in parallel until the stage when DNA was adsorbed to the spin filter; replicates were pooled at this point onto a single filter, and the extraction proceeded from this point as a single sample. In addition, the final step of elution of the DNA from the filter in solution C6 was done with 50 µL instead of 100 µL, and the initial flow-through was reapplied to the filter and passed through a second time to further increase yield.

16S Sequencing Analysis–OTU-based approach

Paired-end reads were joined using a 25% maximum difference in overlapping sequences. The pair-joined sequences were then quality filtered, using default parameters in QIIME (truncating sequences after three or more base calls below a Q4 Phred score, while also requiring at least 75% of the reads in the overall sequence to be above that threshold. Operational Taxonomic Unit (OTU) picking was conducted in QIIME using an open-reference strategy. Briefly, the first step involves a closed-reference OTU picking against the Greengenes 13.8 database at 97% identity clustering. A random subset of the sequences that fail to match the database are then clustered de novo, with cluster centroids being used as a new reference database for an additional round of closed-reference OTU picking. The remaining unclustered sequences are then clustered de novo, forming the final set of OTUs. OTUs that were observed only once in the entire dataset (singletons) were discarded as probable artifacts. While the main analysis used un-rarefied tables, the resulting OTU tables were rarefied at various depths and analyzed in parallel to confirm that read count was not a biasing factor in our analyses.

During this analysis, it was observed that a subset of sites exhbited a sharp decrease in diversity as measured by the Exponent of Shannon Diversity metric described by Jost

(2006), while the remaining sites showed no statistically significant decrease in diversity or showed no statically significant separation in ordinations or by permutational ANOVA. For the purposes of this study, such sites were classified as "gradient" or "uniform" profiles, respectively.

16S Sequencing Analysis–ASV-based approach

Separately, raw reads were processed following the DADA2 1.16 pipeline as implemented in the dada2 R package (Callahan et al., 2016). Since read quality profiles did not show strong decreases in quality, no pre-trimming was applied before the DADA2 denoising algorithm using all other recommended settings. Paired-end reads were joined requiring at least 12 overlapping bases with identical sequences. Taxonomy was assigned through the dada2 package, using Silva version 138 as the reference database.

ITS Sequencing Analysis–ASV-based approach

Sequence analysis of fungal ITS amplicons proceeded largely as described above, following the DADA2 1.16 pipeline. However, pre-trimming was required on reverse-orientation reads, and the threshold of expected errors in the quality filtering step was raised to 4 per read for both forward and reverse reads. Taxonomy was assigned using the UNITE 8.2 Fungi reference database. Many samples showed very few reads per sample, with almost no reads surviving quality filtering. This contrasted with the 16S sequencing, so for community comparisons a threshold of 8000 reads was used for a sample to be included (Figure 2).

Community analysis

Community analysis utilized a variety of packages in R (R Core Team, 2020), primarily vegan (Oksanen et al., 2019), ape (Paradis and Schliep, 2018), ggpubr (Kassambara, 2020), and DESeq2 (Love et al., 2014), with additional support from packages including

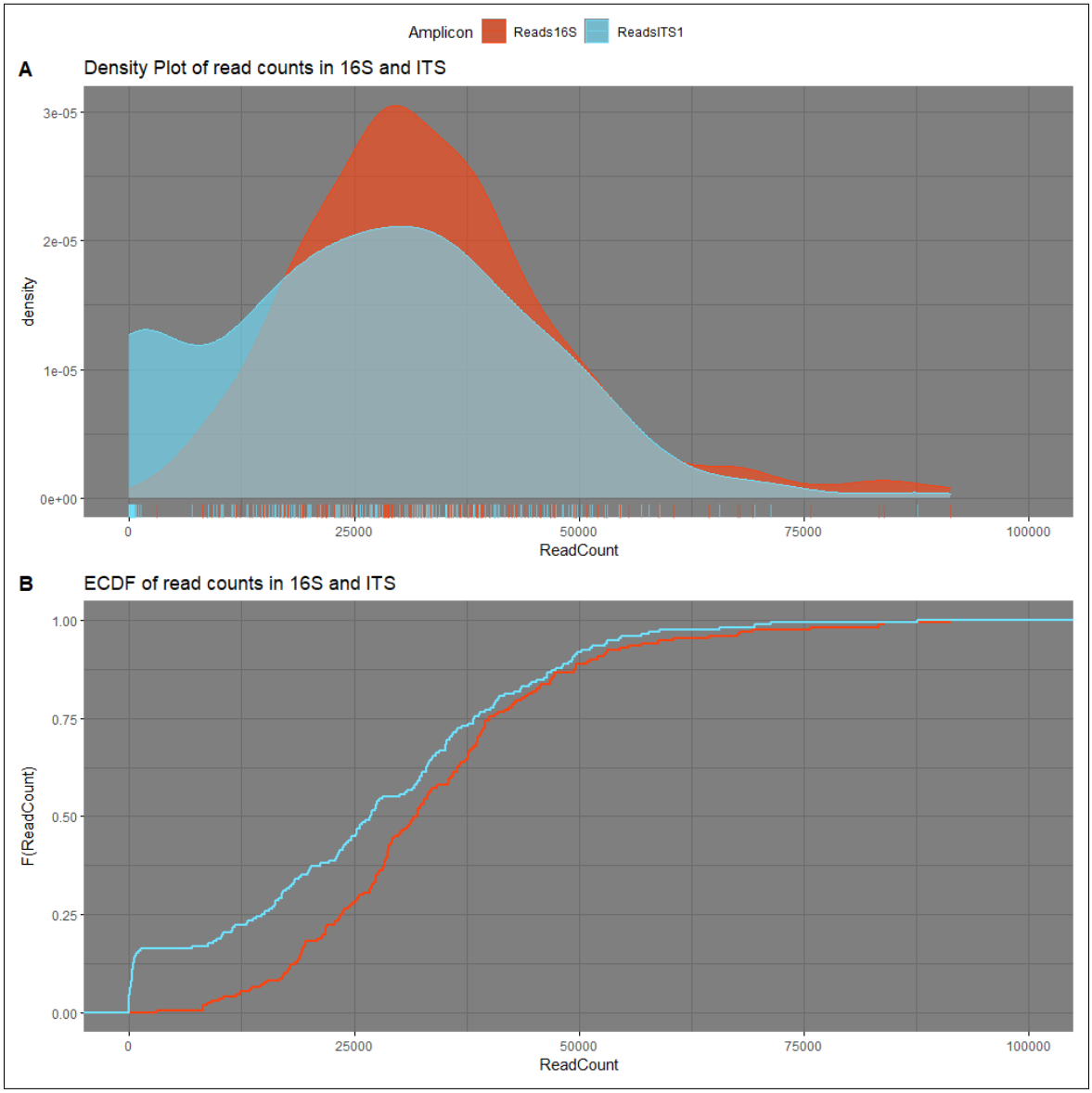

Figure 2: In contrast to 16S sequencing, the ITS amplicon showed a large subset of samples with very low read counts, requiring exclusion of some samples.

tidyr (Wickham and Henry, 2020), aqp (Beaudette et al., 2013), and soiltexture (Beaudette et al., 2013). Community diversity was calculated for each sample using the exponent of Shannon diversity metric as described by Jost (2006). Beta diversity between samples was estimated using Bray-Curtis dissimilarities for 16S communities, but for ITS communities, which showed a much sparser distribution across samples, the Clark distance metric was used, as implemented in the vegan package. Principal coordinates analysis was implemented via the ape package for visual representation of community dissimilarities. To assess the effects of variable read counts between samples, a variance stabilizing transformation was applied from the DESeq2 package, followed by a Euclidean distance metric. Symmetric Procrustes analysis between the normalized Euclidean and un-normalized Bray-Curtis distance matrices for 16S sequences showed a high degree of overlap ($R > 0.75, p < 0.001$), and no differences were seen in significance levels between the dissimilarity measures. Visually, however, principal coordinates analysis of the Bray-Curtis dissimilarities were clearer, so Bray-Curtis derived PCoA plots are used throughout for 16S comparisons. For ITS sequences, on the other hand, a strong difference was observed between Clark distances and VST-normalized Euclidean distances, possibly due to the sparseness of ITS feature table, so normalized-Euclidean distances are used.

Statistical analysis

Permutational ANOVAs were used to assess statistical differences between categories of sites, implemented in R using the vegan package. Procrustes analysis was also implemented via vegan. Differentially abundant taxa were assessed using the DESeq2 package, using a fold-change threshold of 2, and an adjusted alpha of 0.01. For clustering, communities were compared based on the ASV abundances. Using the ape package and base R, complete linkage clustering, centroid-based clustering, and the BIONJ clustering algorithm (Gascuel, 1997) were all used on dissimilarity matrices based on un-normalized Bray-Curtis distances

and normalized Euclidean distances, but none of these showed differences in the high-level clustering. BIONJ trees based on Bray-Curtis dissimilarities (for consistency with PCoA plots) were exported in Newick format and loaded into the Interactive Tree of Life online tool (Letunic and Bork, 2019) to generate figures. Procrustes analysis ($\alpha = 0.01$) was used to compare dissimilarity matrices to determine if they represented similar groupings of samples within ordination spaces. Correlation coefficients were calculated using Pearson's correlation method as implemented in the ggplot2 R package.

Results

16S diversity analysis

Initial analysis of the OTU-based sequences showed that according to both Bray-Curtis dissimilarity and normalized Euclidean distances, 16S communities of most soil profiles were fairly similar, some samples below 40 cm showed a strong divergence from this cluster (Figure 3A). As seen in Figure 3B, this occurred in only a subset of sites (5 of 18). This divergence coincided with a stronger decrease in diversity below 40 cm for the five sites with a decrease in diversity (for this paper, termed "gradient" sites), as seen in Figure 3 (C–D). While the bowing or arching pattern seen in Figures 3A and B can often indicate extremely high beta diversity that cannot be captured by the ordination, the key feature is that the samples that were subsequently classified as deep, gradient soils are highly similar to each other. Notably, the bowing pattern does not form a "horseshoe" that could indicate that highly dissimilar communities appear graphically close to each other due to constraints of a two-dimensional ordination. This dichotomy was not readily apparent under ASV-based analysis using the same methodology. Instead, under both Bray-Curtis and normalized Euclidean dissimilarity metrics (not shown), samples were grouped by soil taxonomic order (Figure 4) and by dominant vegetation type (Figure 5). In permutational

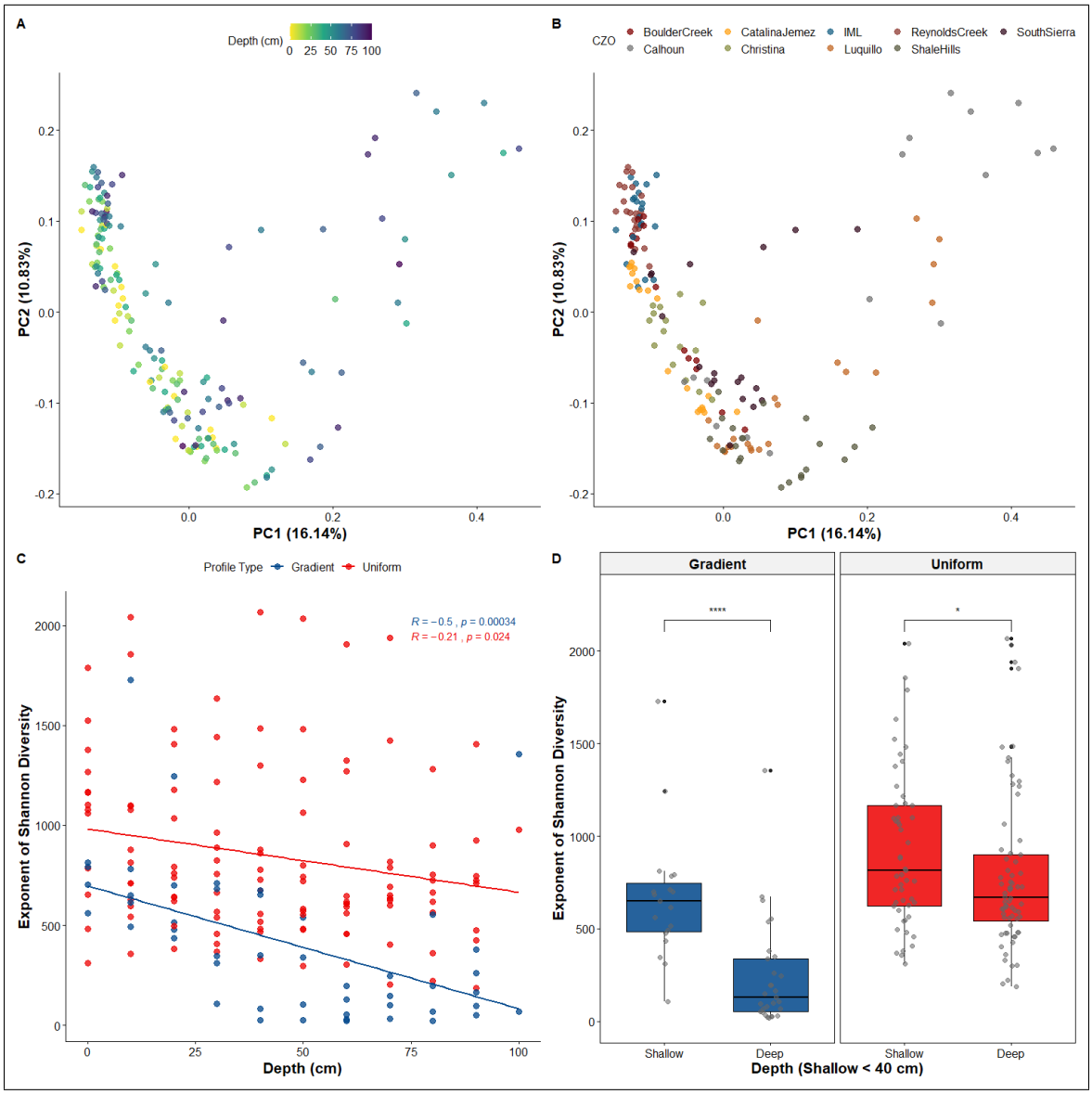

Figure 3: Some sites show a strong decrease in diversity with depth based on OTU analysis. In (A), it is clear that all samples shallower than approximately 40 cm are much more similar to each other than soils below 40 cm, which show a broader distribution, indicating greater dissimilarity between communities. In (B), it is apparent that only a subset of sites exhibit this trend (colors are collaposed by CZO for visual identification; only the South Sierra CZO had two sites with different patterns). In (C) and (D), the sites that were shown to exhibit a shift in community composition as observed in panels A and B are classified as gradient sites, while others are termed uniform. This distinction corresponds with a drop in diversity that is seen both linearly (C) and when using 40 cm as a cutoff between deep and shallow portions of the soil profiles (D).

ANOVA (PERMANOVA) analysis, each factor independently yielded a strong explanatory value ($R^2 > 0.15$, $p < 0.001$), and the interaction of both soil order and vegetation type showed an $R^2 > 0.32$, $p < 0.0001$.

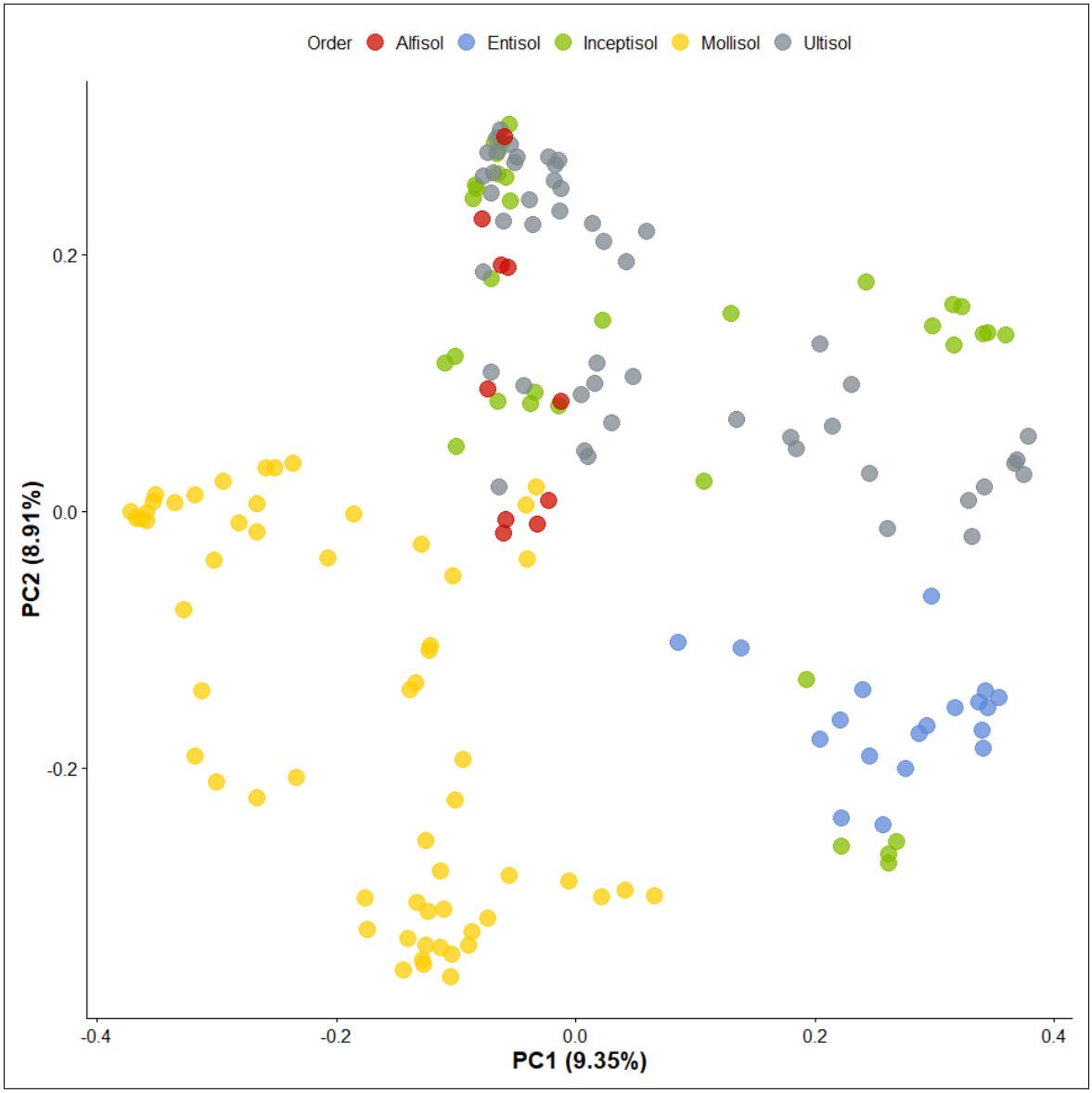

Figure 4: Principal Coordinates Analysis of 16S communities by soil taxonomic order. Distances are calculated using the Bray-Curtis dissimilarity method.

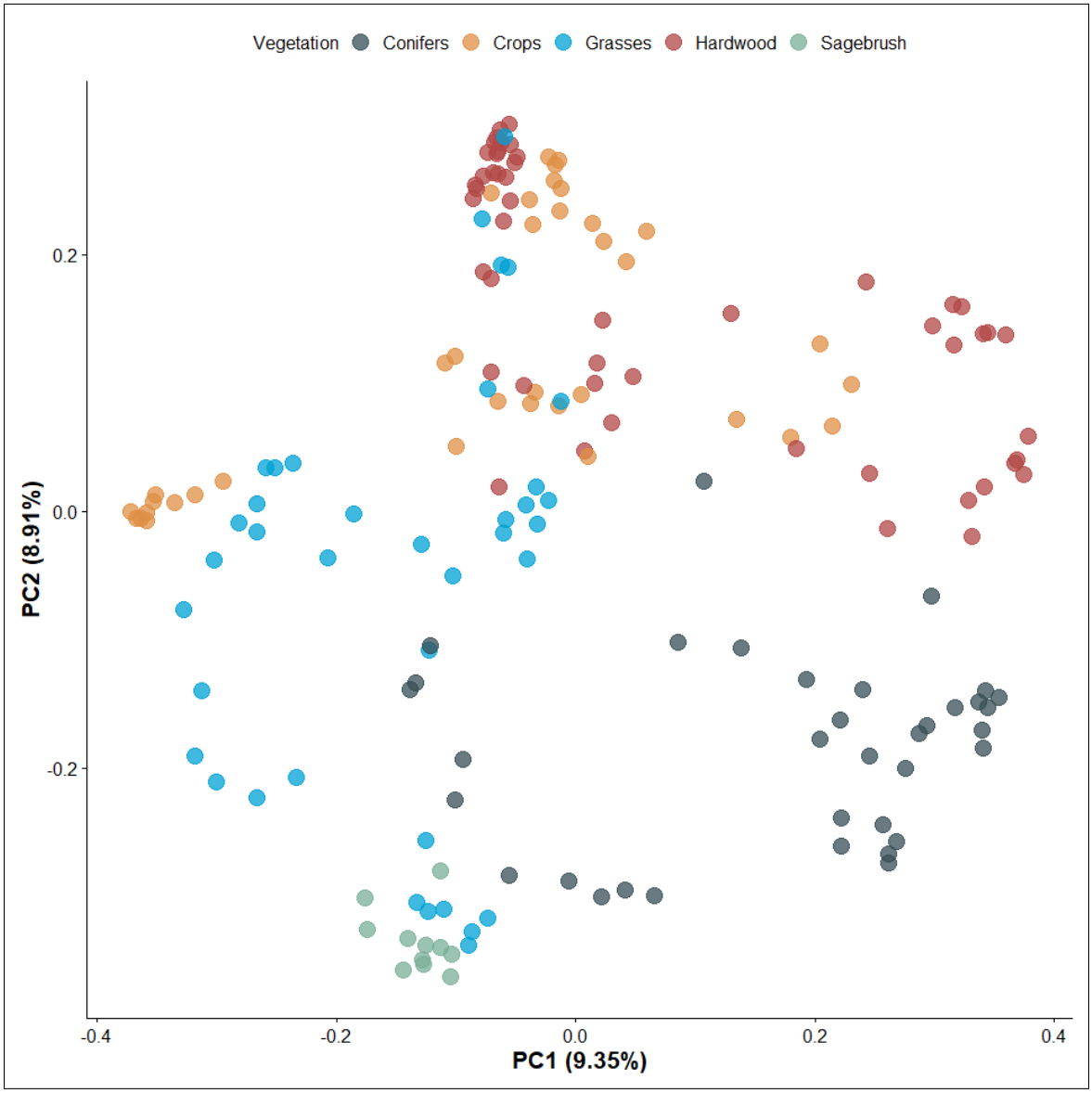

Figure 5: Principal Coordinates Analysis of 16S communities by inferred dominant vegetation type. Distances were calculated using the Bray-Curtis dissimilarity method.

Hierarchical cluster analysis

Clustering of 16S communities based on ASVs showed that the "deep" (> 40 cm) soil samples from the sites categorized as "gradient" sites under OTU-based analysis formed a unique cluster that included some of the "shallow" samples from the gradient sites, but notably no samples from any depth of the "uniform" sites (Figure 6). In this way, using an independent analysis method of the underlying sequencing data, the gradient-uniform dichotomy first derived from OTU-based analysis was recaptured in the ASV-based analysis using different analytical methods. The hierarchical clustering also reinforced the finding that 16S communities are strongly linked to soil taxonomic order. In particular, soil profiles classified as mollisols formed a unique cluster, while the rest of the soil orders show less differentiation from each other (Figure 7). This reinforces the Bray-Curtis spatial separation seen in Figure 4 for ASV-based community analysis.

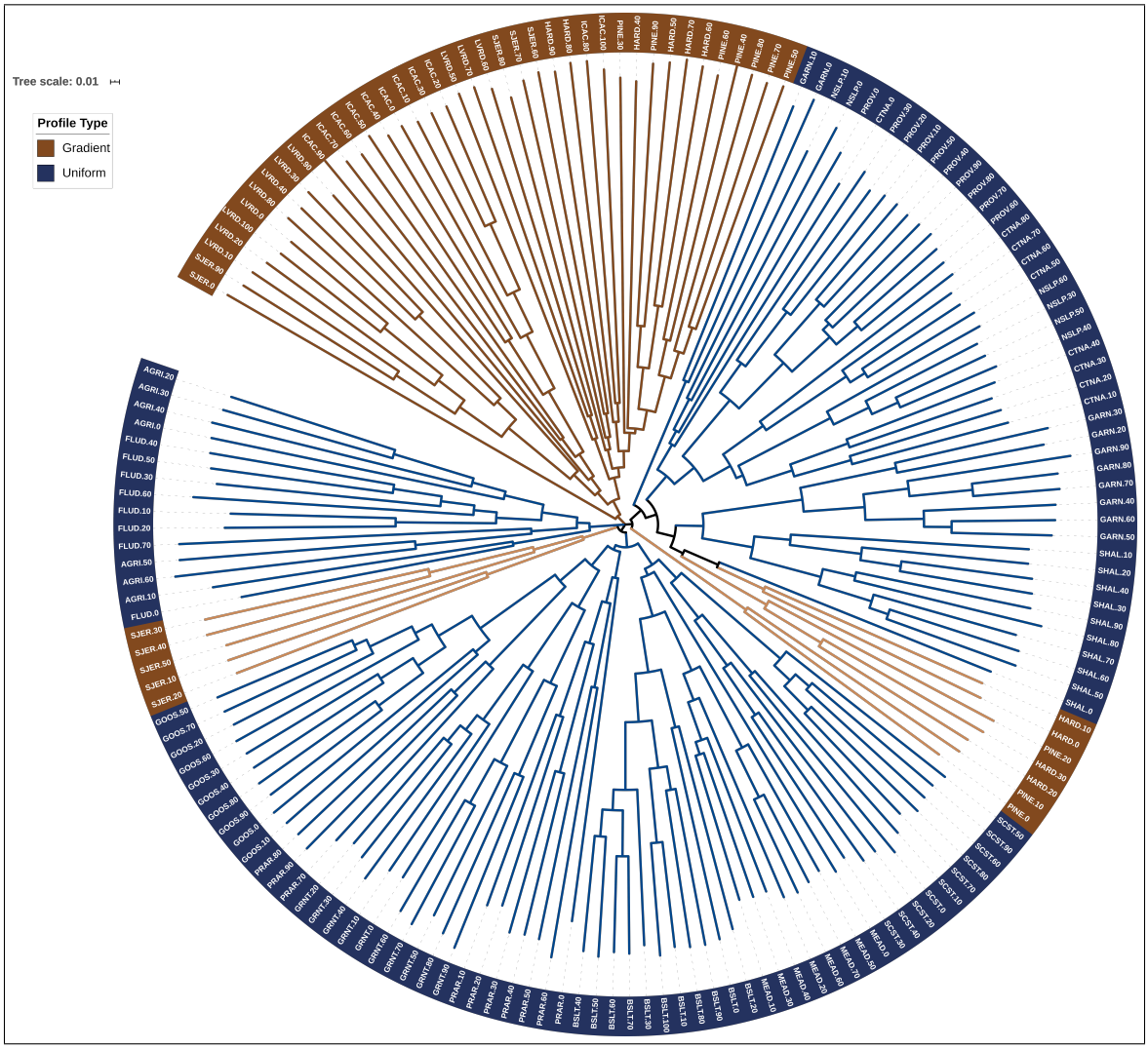

Figure 6: Hierarchical clustering tree comparing 16S communities across all samples. Clustering used the BIONJ algorithm, and coloring reflects profile type according to the gradient-uniform dichotomy initially observed in OTU-based analysis.

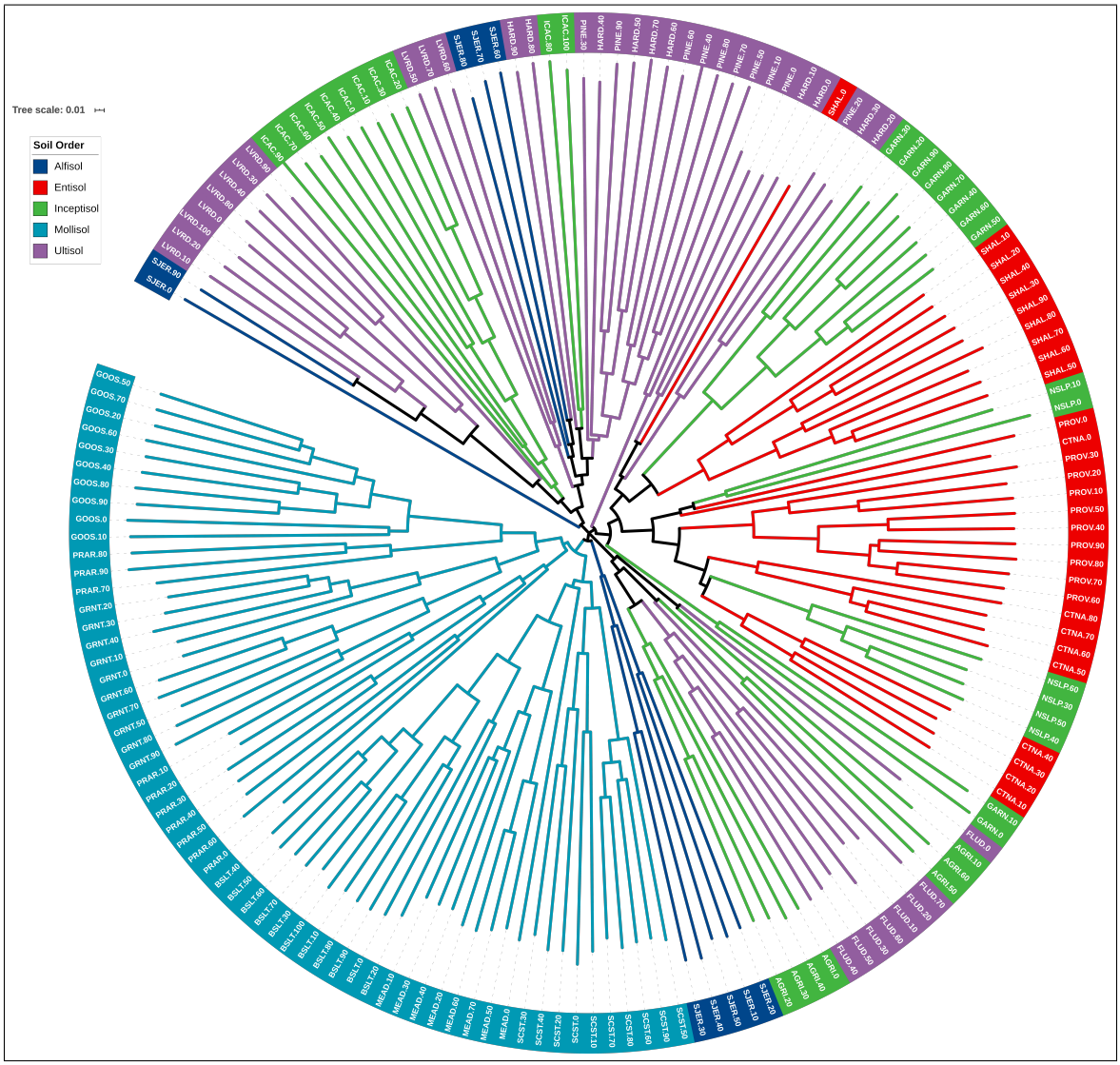

Figure 7: Hierarchical clustering tree comparing 16S communities across all samples. Clustering used the BIONJ algorithm, and coloring reflects inferred soil taxonomic order.

Phospholipid fatty acid analysis

The gradient-uniform dichotomy was also seen independently in phospholipid fatty acid (PLFA) analysis (Figure 8). Note that the bowing or arching effect seen is not necessarily a distortion based on high beta (between-sample) diversity between all samples; instead, it appears to be caused by a strong differentiation between the deep samples from gradient profiles from all others. Biomass, as measured by PLFA analysis, indicates that while there is a positive relationship between total soil biomass and diversity ($R = 0.42, p << 0.0001$), that effect is almost totally driven by the stronger relationship between biomass and diversity in gradient-profile soils ($R = 0.53, p < 0.0001$). This relationship is statistically absent in uniform-profile soils ($R = 0.0051, p = 0.96$) (Figure 9).

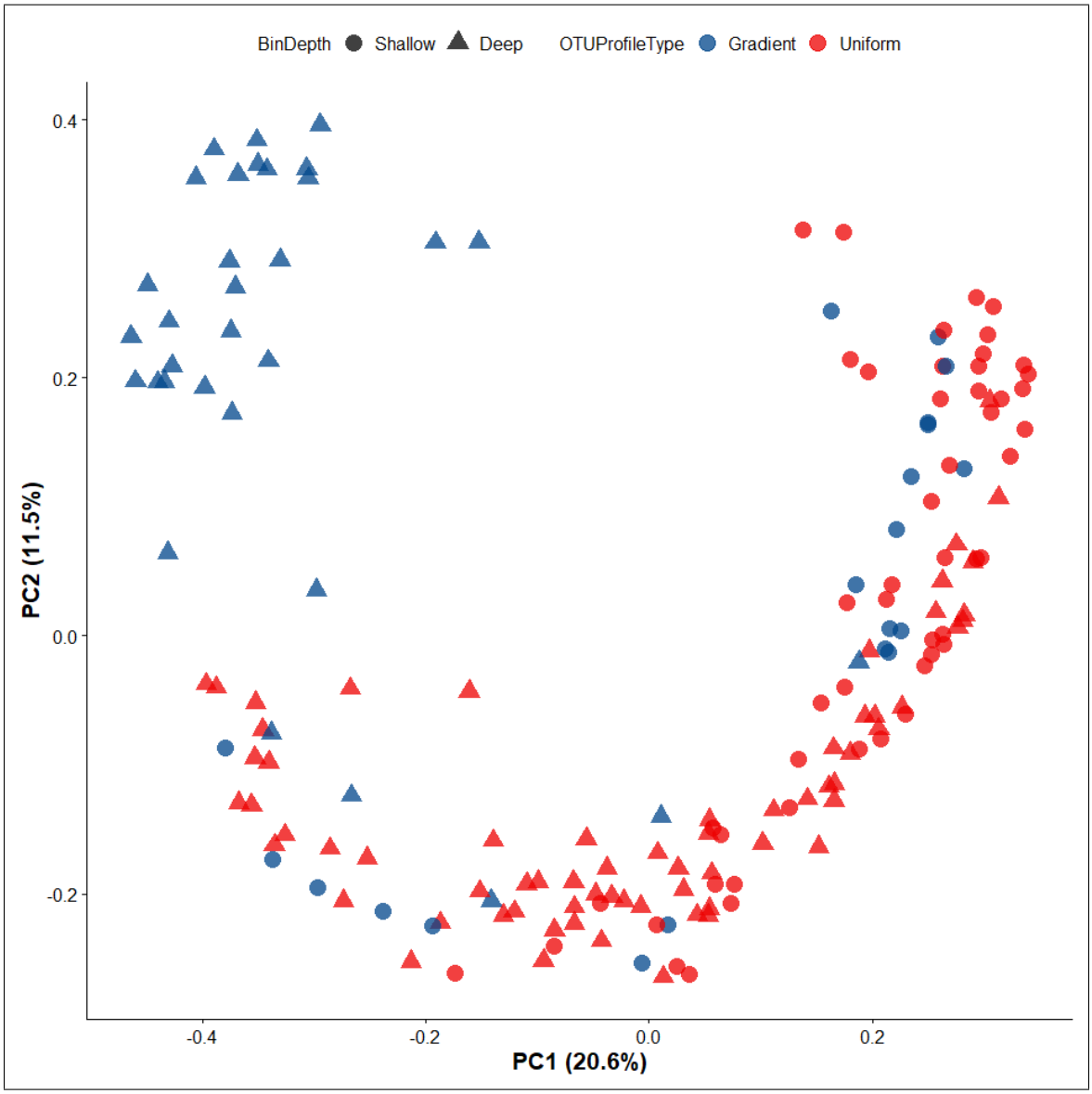

Figure 8: Principal Coordinates Analysis of total phospholipid fatty acid content. The binned depth is defined such that deep soils are those below 40 cm, and shallow soils are those from the surface to 40 cm. The gradient-uniform dichotomy is derived from patterns first observed in the independent OTU-based 16S sequencing analysis. Distances were calculated using Clark distance metric to account for the sparseness in the PLFA feature table.

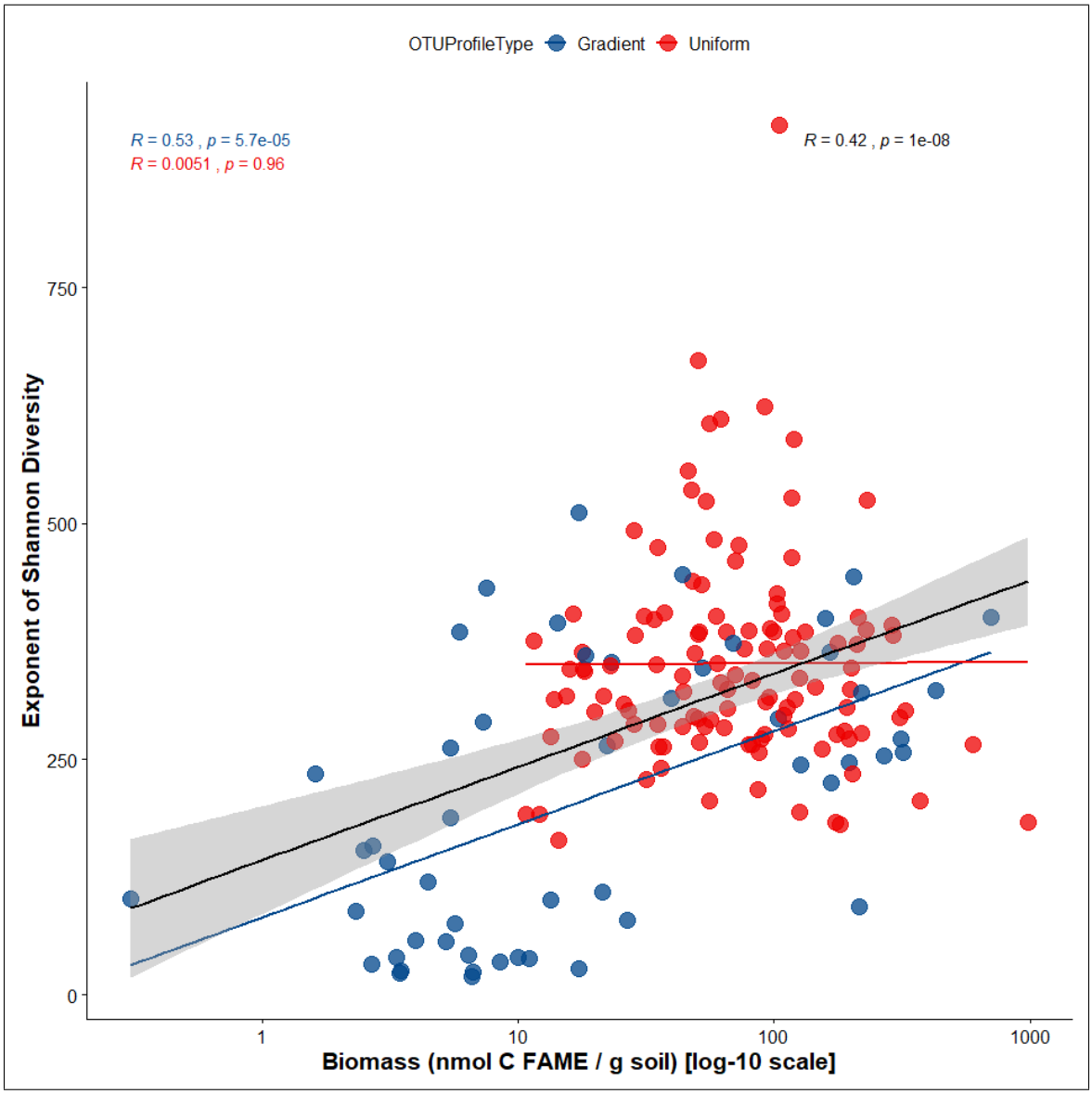

Figure 9: Exponent of Shannon diversity is positively correlated to log_{10}(biomass). Biomass is reported in units of nanomoles of carbon in fatty acid methyl esters (FAME) per gram of soil. Shaded area shows 95% confidence interval for the global correlation. Coefficients are for Pearson's correlations.

ITS and 16S community diversity

A comparison of ITS and 16S community (ostensibly fungal and prokaryotic, respectively) diversity showed a strong positive correlation ($R = 0.47, p\ ll 0.0001$). This was robust across both gradient and uniform soil types (Figure 10), as well as across vegetation type and both gradient uniform soils (not shown).

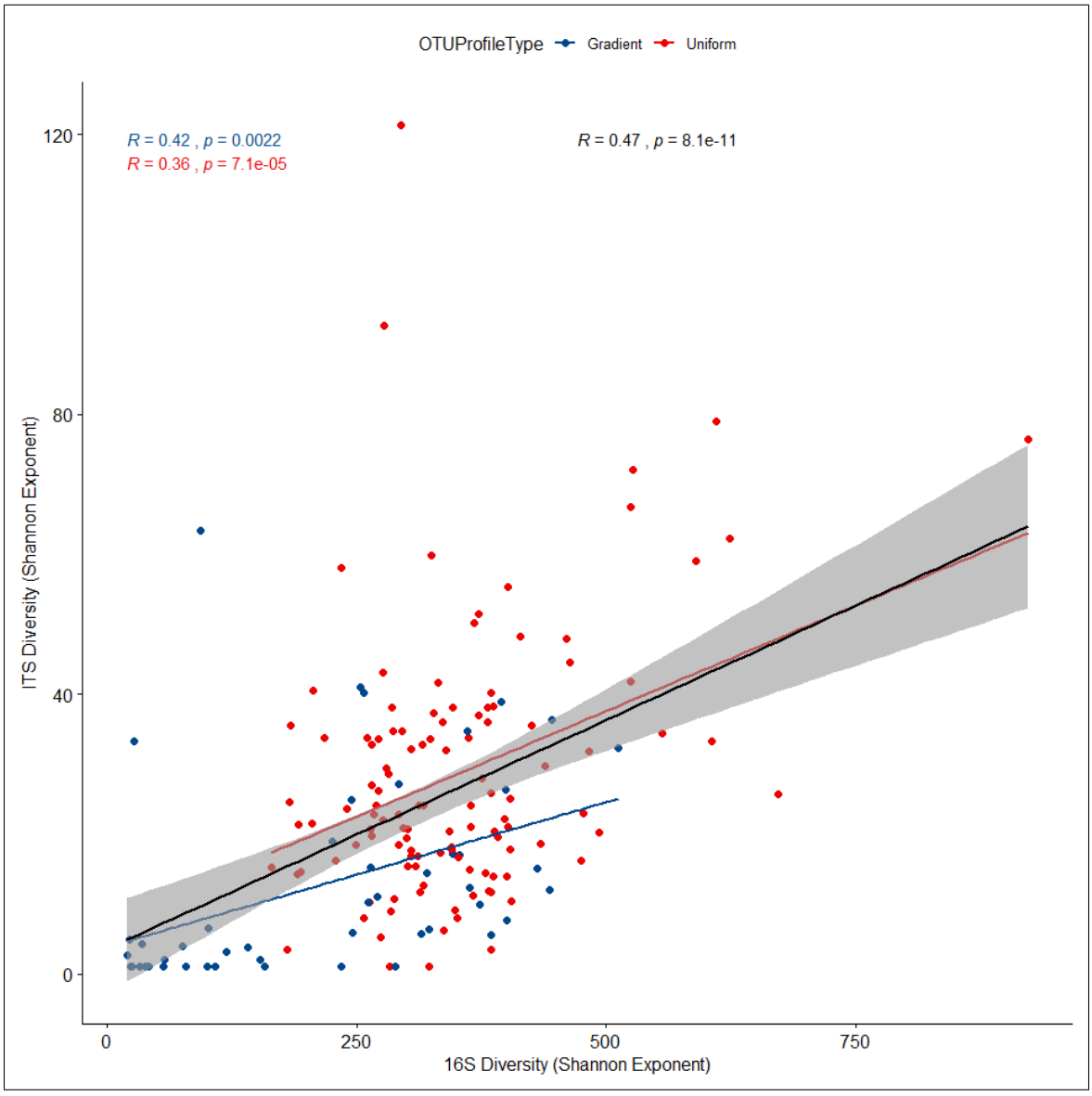

Figure 10: Comparison of ITS (fungal) and 16S (bacterial/prokaryotic) diversity. There is a strong correlation between the diversity of ITS and 16S communities across a wide range of soil types.

Discussion

Gradient-uniform dichotomy

Although gradient-uniform dichotomy observed in this study was initially observed in an OTU-based analysis, which is no longer considered the gold standard for microbiome analysis (Knight et al., 2018), the validity of that classification is reinforced by a variety of other findings. In particular, the PLFA-based results (Figure 8) are independent of DNA-based biases. Although it is surprising that the dichotomy is not as readily observed in ASV-based analysis of 16S or ITS sequences by ordination, it is still readily apparent in clustering analysis (Figure 6) and by similar comparisons as used in OTU-based analysis (see Figures S2 & S3). While none of the mollisols or entisols were classified as gradient profiles, it is not readily apparent what sets the two groups apart—i.e., why do the gradient soils show a decrease in diversity and biomass? Potential factors that have been recognized as important factors related to soil microbial community composition such as pH, water content, and soil texture were not found to be correlated with any diversity indicators when accounting for between-site differences. This is distinct from previous studies that have found such factors to be significant drivers of 16S communities (Thompson et al., 2017; Seaton et al., 2020), but did not examine differences between communities observed at differing depths.

Contrast with other studies

While neither soil textural composition at the sample level nor textural class at the profile level were associated with changes in diversity of 16S or ITS communities, or for PLFA-defined communities, a recent study by Seaton et al. (2020) found that soil textural heterogeneity was a strong determinant of fungal diversity, but not of bacterial diversity. While that finding can't be precisely compared because texture in this study was determined gravi-

metrically (rather than by laser granulometry), the two studies may not even be generally comparable because of vastly different soil textural properties between both studies. While the Seaton et al. (2020) study of soils in Wales examined highly silty topsoils, this study had an abundance of extremely sandy soils (Supplemental Figure S1). While this project draws different conclusions from earlier studies (Eilers et al., 2012) that found stronger differences by depth than between sites, most such studies were focused on smaller geospatial displacements.

Caveats

For a project of this type, due to the importance of soil taxonomy and soil profile characterization in general, pedological expertise is vital to future efforts. Personnel trained to characterize soil profiles should be involved during soil pit excavation. When that is not possible, extensive photographic documentation of the site and pit wall must be undertaken. When possible, site selection should be managed such that for vital criteria such as soil taxonomic order and dominant vegetation type, statistically relevant analyses can be performed.

This study underlines the conclusions noted by Tripathi et al. (2018) that microbial community studies, particularly in terrestrial biomes require extensive advances in databases. One of the key limitations of this study was the lack of annotated 16S sequences. Notwithstanding the diffiiculties in ascribing actual relative abundances to PCR-dependent amplicon-based sequencing plans (Knight et al., 2018), 99.2% of observed ASVs could not be assigned to the species level, presenting a major challenge for inferring functional potentials from such DNA-based surveys.

Conclusion

As these challenges are met by intensive shotgun sequencing programs, this study should provide some parameters for the breadth of terrestrial soil samples to be included in such programs. In particular, variations in soil taxonomy and dominant vegetation type need to be accounted for. While previous studies have seen important changes in overall microbial communities with depth, this study suggests that at broader (continental) scales, microbial community composition is driven more by ecological setting than by variances of depth within soil profiles. Further research across a more focused set of soil profiles can provide important insights into the environmental factors that drive the distinction between gradient and uniform profiles. The importance of the gradient-uniform dichotomy, in particular, as well as soil taxonomic order, must be accounted for in studies that seek to examine how microbial communities are structured by depth.

References

R. S. Anderson. Pinched topography initiates the critical zone. *Science*, 350(6260):506–507, 2015.

R. D. Bardgett, C. Freeman, and N. J. Ostle. Microbial contributions to climate change through carbon cycle feedbacks. 2(8):805–814. ISSN 1751-7362. doi: 10.1038/ismej.2008.58. URL `http://www.nature.com/ismej/journal/v2/n8/full/ismej200858a.html`.

D. E. Beaudette, P. Roudier, and A. O'Geen. *Algorithms for quantitative pedology: A toolkit for soil scientists*, 2013. URL `http://ncss-tech.github.io/AQP/`.

T. E. Brewer, E. L. Aronson, K. Arogyaswamy, S. A. Billings, J. K. Botthoff, A. N. Campbell, N. C. Dove, D. Fairbanks, R. E. Gallery, S. C. Hart, et al. Ecological and genomic attributes of novel bacterial taxa that thrive in subsurface soil horizons. *MBio*, 10(5): e01318–19, 2019.

B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. Dada2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, 13:581–583, 2016. doi: 10.1038/nmeth.3869.

P. Carini, P. J. Marsden, J. W. Leff, E. E. Morgan, M. S. Strickland, and N. Fierer. Relic dna is abundant in soil and obscures estimates of soil microbial diversity. *Nature microbiology*, 2(3):1–6, 2016.

R. Conrad. The global methane cycle: recent advances in understanding the microbial processes involved. *Environmental Microbiology Reports*, 1(5):285–292, 2009.

Critical Zone Observatory Network. National critical zone observatory, 02 2018. URL `http://criticalzone.org/`.

R. Daniel. The metagenomics of soil. *Nature Reviews Microbiology*, 3(6):470–478, 2005. ISSN 1740-1526. doi: 10.1038/nrmicro1160. URL `http://www.nature.com/nrmicro/journal/v3/n6/abs/nrmicro1160.html`.

N. C. Dove, K. Arogyaswamy, S. A. Billings, J. K. Bothoff, C. J. Carey, J. L. DeForest, D. Fairbanks, N. Fierer, R. E. Gallery, J. P. Kaye, K. A. Lohse, M. R. Maltz, E. Mayorga, J. Pett-Ridge, W. H. Yang, S. C. Hart, and E. L. Aronson. Continental-scale patterns of extracellular enzyme activity in the subsoil: an overlooked reservoir of microbial activity. *in review*, 2020.

K. G. Eilers, S. Debenport, S. Anderson, and N. Fierer. Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biology and Biochemistry*, 50:58–65, 2012.

N. Fierer, J. P. Schimel, and P. A. Holden. Variations in microbial community composition through two soil depth profiles. *Soil Biology and Biochemistry*, 35(1):167–176, 2003.

R. S. Gabor, K. Eilers, D. M. McKnight, N. Fierer, and S. P. Anderson. From the litter layer to the saprolite: Chemical changes in water-soluble soil organic matter and their correlation to microbial community composition. *Soil Biology and Biochemistry*, 68:166–176, 2014.

O. Gascuel. Bionj: an improved version of the nj algorithm based on a simple model of sequence data. *Molecular biology and evolution*, 14(7):685–695, 1997.

M. Ji, C. Greening, I. Vanwonterghem, C. R. Carere, S. K. Bay, J. A. Steen, K. Montgomery, T. Lines, J. Beardall, J. Van Dorst, et al. Atmospheric trace gases support primary production in antarctic desert surface soil. *Nature*, 552(7685):400–403, 2017.

L. Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.

A. Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*, 2020. URL `https://CRAN.R-project.org/package=ggpubr`. R package version 0.3.0.

M. Keller, D. S. Schimel, W. W. Hargrove, and F. M. Hoffman. A continental strategy for the national ecological observatory network. *Frontiers in Ecology and the Environment*, 6(5):282–284, 2008.

R. Knight, A. Vrbanac, B. C. Taylor, A. Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciolek, L.-I. McCall, D. McDonald, et al. Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(7):410–422, 2018.

I. Letunic and P. Bork. Interactive tree of life (itol) v4: recent updates and new developments. *Nucleic acids research*, 47(W1):W256–W259, 2019.

H. Lin, J. W. Hopmans, and D. d. Richter. Interdisciplinary sciences in a global network of critical zone observatories. *Vadose Zone Journal*, 10(3):781–785, 2011.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15:550, 2014. doi: 10.1186/s13059-014-0550-8.

J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. *vegan: Community Ecology Package*, 2019. URL `https://CRAN.R-project.org/package=vegan`. R package version 2.5-6.

E. Paradis and K. Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35:526–528, 2018.

S. A. Placella, E. L. Brodie, and M. K. Firestone. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. *Proceedings of the National Academy of Sciences*, 109(27):10931–10936, 2012.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

D. H. Rothman, G. P. Fournier, K. L. French, E. J. Alm, E. A. Boyle, C. Cao, and R. E. Summons. Methanogenic burst in the end-permian carbon cycle. *Proceedings of the National Academy of Sciences*, 111(15):5462–5467, 2014.

F. M. Seaton, P. B. George, I. Lebron, D. L. Jones, S. Creer, and D. A. Robinson. Soil textural heterogeneity impacts bacterial but not fungal diversity. *Soil Biology and Biochemistry*, page 107766, 2020.

S. S. Staff, N. R. C. Service, and U. S. D. of Agriculture. Web soil survey, 2018. Available online. Accessed 2018–2020 `https://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm`.

N. C. S. Survey. National cooperative soil characterization database, 2018. Available online. Accessed 2018–2020.

L. R. Thompson, J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, et al. A communal catalogue reveals earth's multiscale microbial diversity. *Nature*, 551(7681):457–463, 2017.

V. Torsvik and L. Øvreås. Microbial diversity and function in soil: from genes to ecosystems. *Current opinion in microbiology*, 5(3):240–245, 2002.

A. Tripathi, C. Marotz, A. Gonzalez, Y. Vázquez-Baeza, S. J. Song, A. Bouslimani, D. McDonald, Q. Zhu, J. G. Sanders, L. Smarr, et al. Are microbiome studies ready for hypothesis-driven research? *Current opinion in microbiology*, 44:61–69, 2018.

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

H. Wickham and L. Henry. *tidyr: Tidy Messy Data*, 2020. URL `https://CRAN.R-project.org/package=tidyr`. R package version 1.0.2.

B. J. Woodcroft, C. M. Singleton, J. A. Boyd, P. N. Evans, J. B. Emerson, A. A. Zayed, R. D. Hoelzle, T. O. Lamberton, C. K. McCalley, S. B. Hodgkins, et al. Genome-centric view of carbon processing in thawing permafrost. *Nature*, 560(7716):49–54, 2018.

Supplemental Figures

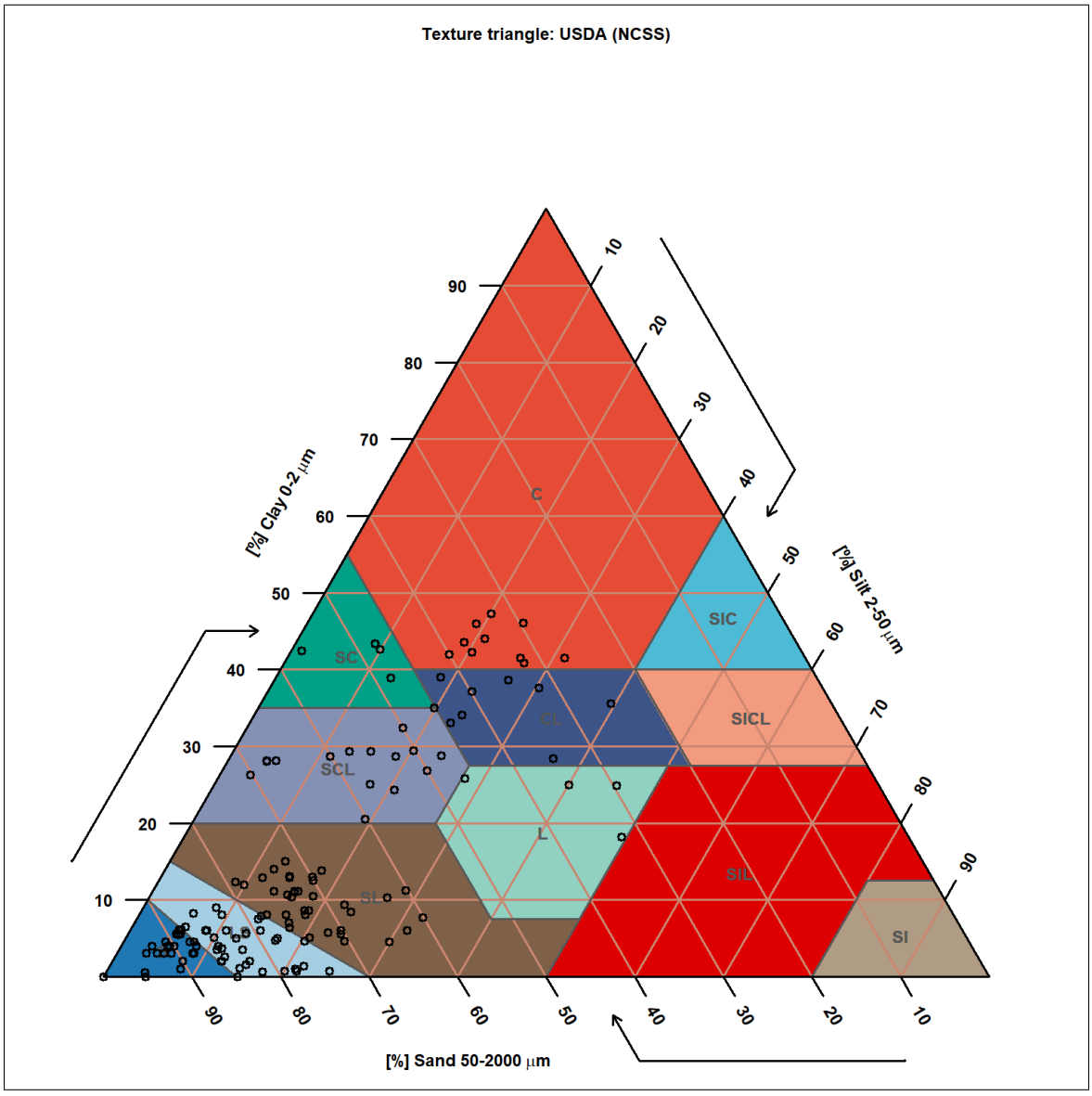

Figure S1: Particle size distribution of all samples in this study. Particle size distribution by mass contribution, classified according to USDA/NCSS descriptions. (S = Sand, C = Clay, SI = Silt, L = Loam; e.g., SL = Sandy loam, SICL = Silt clay loam)

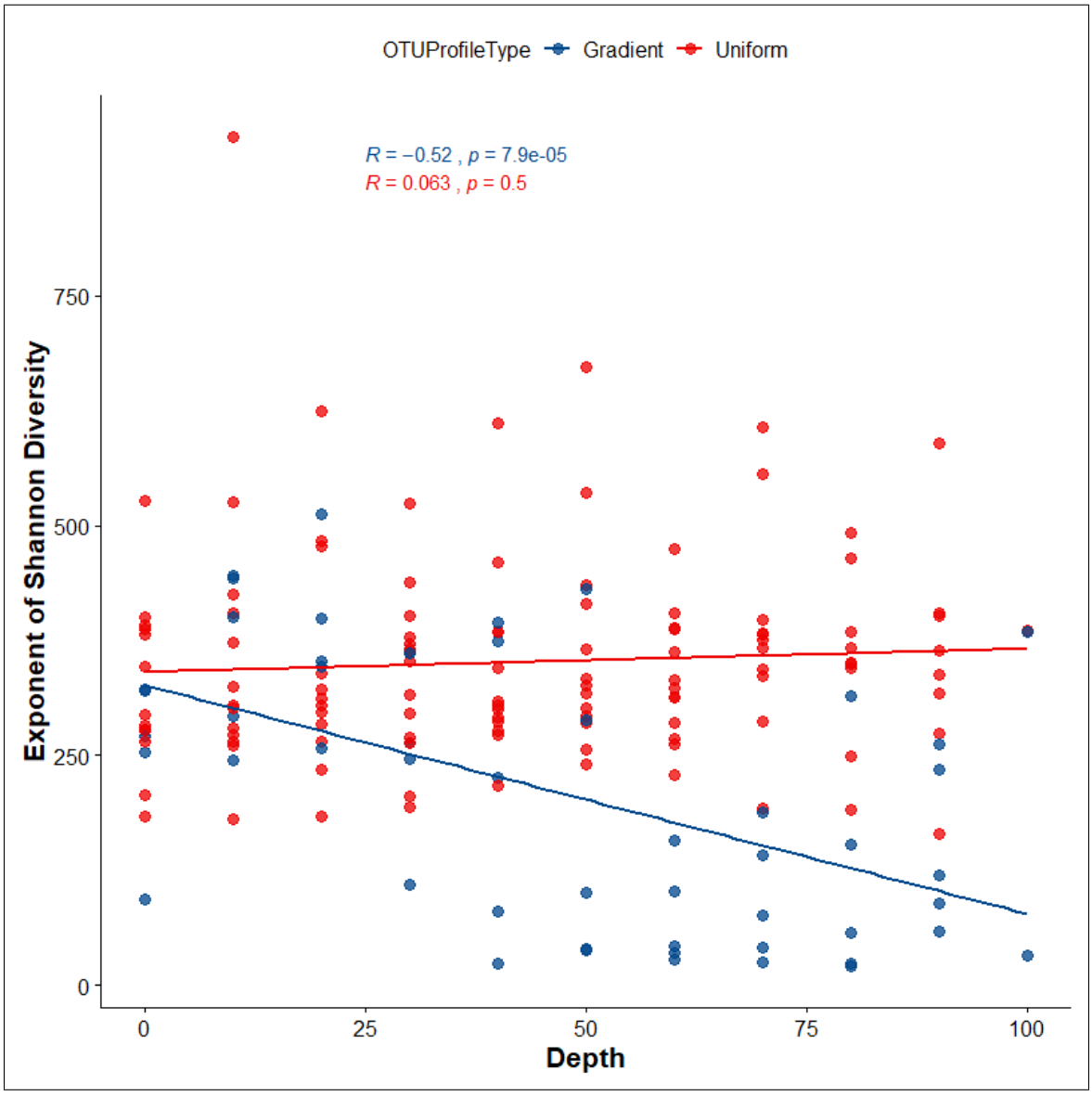

Figure S2: Increasing depth in gradient soils shows a decrease in diversity measured by the exponent of Shannon diversity. While this data represents ASV-based analysis, the gradient-uniform dichotomy is derived from OTU-based analysis.

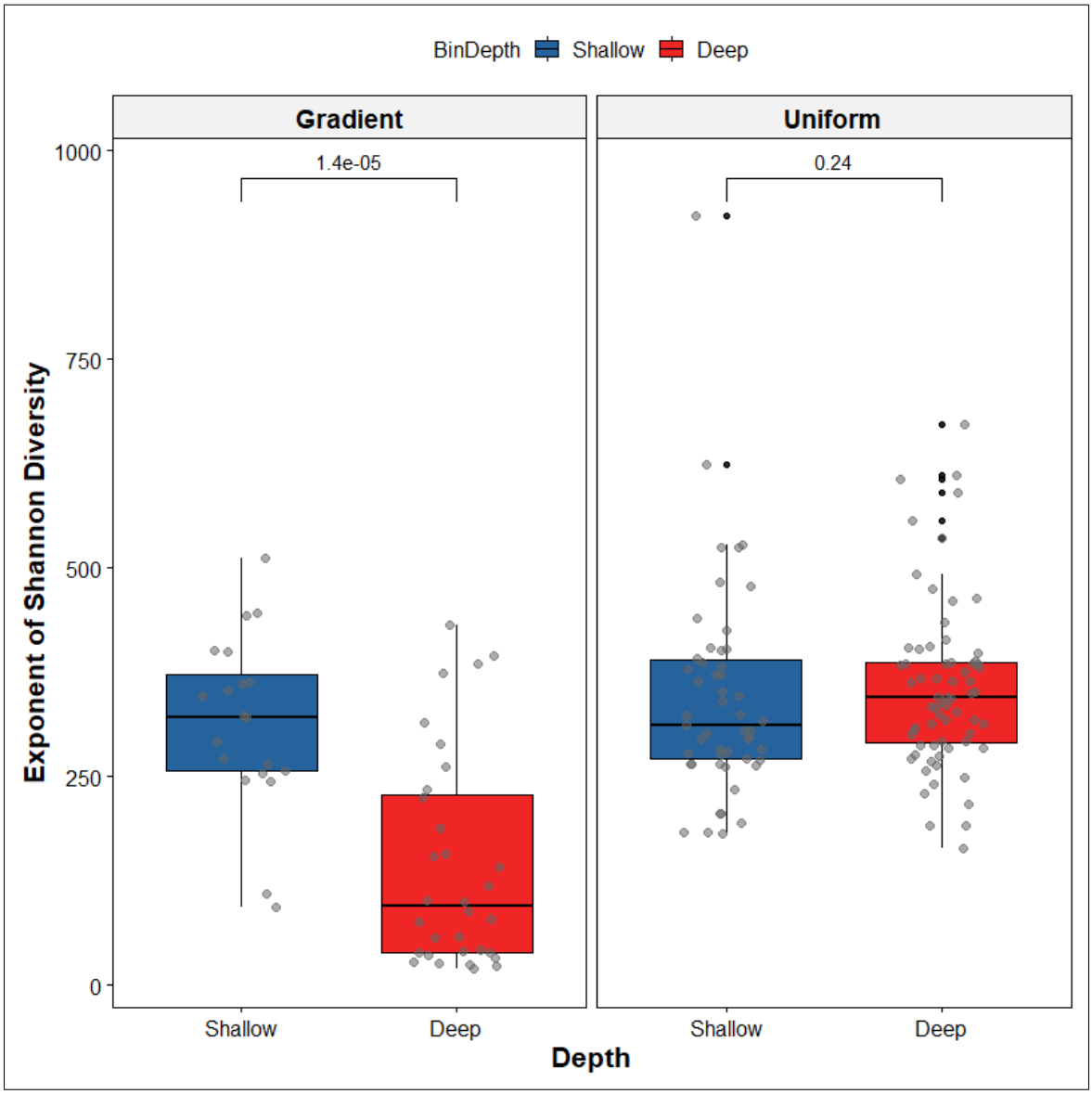

Figure S3: When binned by depth ("deep" soils being more than 40 cm from the surface), gradient soil profiles show a strong drop in diversity measured by the exponent of Shannon diversity. While this data represents ASV-based analysis, the gradient-uniform dichotomy is derived from OTU-based analysis.