An Exploration into the Delicate Balance of Soil Microbial Community Interactions

Presented in Partial Fulfillment of
the Requirements for the Degree of

Master of Science

with a Major in

Statistical Science

in the

College of Graduate Studies

University of Idaho

by

Kristen Renae Petersen

Major Professor
Michelle Wiest, Ph.D.

Committee
Stephen Lee, Ph.D.
Amy Ross-Davis, Ph.D.
Matthew Settles, Ph.D.

Department Administrator
Christopher Williams, Ph.D.

August 2016

Authorization to Submit Thesis

This thesis of Kristen Renae Petersen, submitted for the degree of Master of Science with a Major in Statistical Science and titled "An Exploration into the Delicate Balance of Soil Microbial Community Interactions," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor:	Michelle Wiest, Ph.D.	Date
Committee Members:	Stephen Lee, Ph.D.	Date
	Amy Ross-Davis, Ph.D.	Date
	Matthew Settles, Ph.D.	Date
Department Administrator:	Christopher Williams, Ph.D.	Date

Abstract

This study assesses the bacterial and fungal communities within the Priest River Experimental Forest in northern Idaho. We employed multiple methods to test for significant impacts of season, habitat type (i.e. moisture level), and sample depth on the soil microbial community compositions, as well as explore how members within these communities interact. Abundance data were obtained using two different loci for both bacterial and fungal taxa, therefore, we were able to compare the communities captured. Results suggest that the bacterial component of the forest soil biome differs with habitat type and sampling depth below the forest floor, while the fungal component differs with habitat type. Hierarchical clustering was performed to identify microbes with similar interactions; however, there were no apparent patterns in ecological functionality to explain the clustering. These findings indicate that further analysis is required to enhance our grasp of the microbial interactions and to provide insights into which microbes are aiding the spread, or suppression, of different diseases. Conclusions also suggest that future analyses should use both loci from the respective taxa to obtain a more complete snapshot of the community.

Acknowledgements

I thank my major professor, Dr. Michelle Wiest, for all the time she invested in helping me complete this thesis. Her expertise, insights, encouragement, and experiences not only helped me think critically about the academics associated with this degree, but also the options that will follow.

I thank my committee members: Dr. Stephen Lee, for always making time when I got stuck on a problem; Dr. Amy Ross-Davis, for letting me analyze her data further, and for sharing her enthusiasm in forest ecology; and Dr. Matthew Settles, for the time he invested in being a great mentor and expressing support in every direction that came during the course of this degree. Each of my committee members brought different insights and expertise to learn from and I am forever grateful to each one of them.

Additionally, I would like to thank the faculty, staff, and fellow students at the University of Idaho. Without your advice, encouragement, and friendship, my time at the university would not have been as enjoyable.

Dedication

To my family
For their unconditional love,
support, and unceasing prayers

Table of Contents

List of Tables

List of Figures

chapter 1

Introduction

Forests are critical to the overall health of the planet, and the health of the forest itself is largely dependent on soil quality. Astonishingly, within a single gram of soil there can be upwards of thousands of microbes present (Fierer *et al.*, 2012; Torsvik *et al.*, 1990). These microorganisms are diverse and play critical roles in the carbon and nitrogen cycles, as well as breaking down other nutrients to be used by plants (Fierer *et al.*, 2012). While some microorganisms are crucial to the flourishing of plants, others can sometimes be detrimental to the host, resulting in a forest becoming vulnerable to disease outbreak. For instance, *Armillaria* root disease is a diverse fungal disease that effects a broad range of trees and shrubs (Worrall, 2007). In 2008, it was revealed by the Forest Health Protection and State Forestry Organizations that this disease, notably found in northern Idaho and western Montana forests, is potentially causing large areas of timber-producing sites to be destroyed (Hagle, 2008). In addition to the microbes being diverse and potentially harmful, the overall soil composition is exceedingly diverse from region to region, due to the associated environmental factors that are location specific (Talbot *et al.*, 2014). Therefore, the importance of soil quality for the overall health of a forest, along with the high complexity of the ecosystem, makes soil communities a widely studied topic.

While the diversity within soil microbial communities is a well studied topic, advances in biotechnologies have only recently begun to allow soil ecologists to determine the full extent of the community captured within a sample. Many of the previous methods of measuring microbial diversity have been culture-dependent experiments. Approximately 99% of bacterial microbes, along with many fungal microbes, are not able to be cultured with the standard laboratory protocols (Kirk *et al.*, 2004). Therefore, various methods were developed to overcome the non-culturable nature of the microbes. Some of these methods include biochemical-based methods (e.g. plate counts, community level physiological profiling, and fatty acid methyl ester analysis) and molecular-based methods (e.g. guanine plus cytosine, DNA microarrays and DNA

hybridization, and denaturing and temperature gradient gel electrophoresis) (Kirk *et al.*, 2004). Each of these methods have associated limitations, such as favoring fast growing microbes, or requiring large amounts of DNA; in addition each method still only provides a partial snapshot of the community (Kirk *et al.*, 2004).

The limitation of only identifying part of the community captured may never be fully overcome, but through the use of metagenomics the other limitations listed can be addressed. Metagenomics is becoming a widely used method because it allows for both the bacterial and fungal communities to be identified without the loss of non-cultrable material (Sharpton, 2014). In order to examine the microbes within a soil sample, researchers begin by classifying the microbial sequences to determine the respective community compositions; this is widely done through amplicon sequencing. The basic protocol is to extract taxonomically informative genomic markers common to all organisms of interest within your collected sample, amplify it through polymerase chain reaction (PCR), then sequence using modern technologies (e.g. Illumina sequencers) (Sharpton, 2014). Common genomic markers used within bacterial microbes are the variable regions within the 16S ribosomal RNA (rRNA) locus, and the internal transcribed spacer (ITS) regions, as well as the large-subunit (LSU) and small-subunit (SSU) rRNA are used to identify the fungal microbes present (Fierer *et al.*, 2012; Sharpton, 2014). While these loci are present in all bacteria or fungi, respectively, the hypervariable regions allow for the taxonomic variation within in the communities to be captured. In addition, the multiple regions allow for the identification of pseudo-replicate communities from the same sample. Therefore, despite metagenomic approaches still being relatively new, it is a culture-independent method that allows for advances in understanding of members within the bacterial and fungal communities.

Once the DNA has been extracted and sequenced, researchers begin to determine the respective community compositions by classifying the sequences. The most extensively used method of sequence classification is by clustering the sequences into operational taxonomic units (OTUs) based on sequence similarity, with a percent similarity usually between 90-100% (Navas-Molina *et al.*, 2013; Větrovský and Baldrian, 2013). Previous research has provided a solid foundation for soil microbial analysis using OTUs for classification and determining functionality from the clusters. Still, a limita-

tion to this approach is the number of OTUs is often reported, but this value provides little to no information to how each sequence was grouped (Sun *et al.*, 2012). There is also significant variation between the estimates derived from different methods (e.g. QIIME, and mothur), due to differences in the underlying clustering algorithms (Sun *et al.*, 2012). Nevertheless, an advantage of using OTUs is the ability to assess diversity with such indices as species richness, intra-variation (alpha diversity), and inter-variation (beta diversity). OTUs were utilized in the study of a Tibetan forest to explore the relationship between alpha diversity and the environmental factors, such as pH, moisture, and organic matter, within the fungal community. In this study it was observed that the alpha diversity was considerably affected by the pH level (Wang *et al.*, 2015). While Wang *et al.* determined pH significantly affected alpha diversity, Coince *et al*. (2013) examined the fungal community observed in a French beech forest and determined that pH did not have a significant impact on the community structure within a plot. Coince *et al.* did, however, conclude the fungal composition was spatially correlated by depth, meaning there were distinct differences in the compositions at different sample depths from the same plot. This conclusion further confirmed previous findings of autocorrelation within the fungal community determined by Lilleskov *et al*. (2004), and Toljander *et al*. (2006).

Although the previous studies found relationships between environmental factors and the fungal community, these studies did not investigate the relationships within the bacterial community. Yet, since the bacterial and fungal communities coexist, it makes sense to explore interactions between bacteria and fungi within the forest soil ecosystem. By examining the interactions between community microbes insights into which microbes aide, or suppress, different detrimental diseases will be established. Recent studies have thus begun to examine the two communities as a whole (Dematheis *et al.*, 2012; He *et al.*, 2008; Urbanová *et al.*, 2015; Young *et al.*, 2014). In 2008, He *et al.* reported that the two communities were correlated and there was a significant difference in taxonomic diversity between summer and winter observations. This conclusion confirms that differences in the overall soil composition may result from the environmental factor of changes in seasonality. Therefore, these studies have provided independent evidence that distinctions within soil composition can be attributed to

changes in sample depth and season, however, these studies did not consider the two factors together.

Another factor to take in to consideration when analyzing microbial communities is that the communities are sparsely distributed and therefore using the raw abundance values is problematic. Using the raw abundance community data is problematic because it can result in microbes with high abundance masking the underlying community patterns. In 2007, Lozupone *et al.* compared a qualitative method to a quantitative method and reported noticeable differences in the conclusions regarding beta diversity, which is defined as the ratio between the regional (gamma) and local (alpha) diversities (Whittaker, 1960; Jost, 2007). Qualitative measures are useful when the communities differ based on global environmental factors, or phylogenetic differences; whereas quantitative measures are effective in determining differences due to changes in relative abundances (Lozupone *et al.*, 2007). The present study will focus on exploring what the effect of different quantitative standardization techniques is on the resulting conclusions. Some of the standardization techniques used are: frequency, normalization, presence/absence, χ^2, and hellinger distance. The frequency technique divides each microbe in the sample by the sample maximum; the resulting value is then multiplied by the total number of microbes present in the sample (Oksanen, 1983). Normalization standardizes the values so the sum of squares across samples for each microbe is equal to one; presence/absence scales the data to be 1 or 0 whether the microbe is present or absent, respectively (Oksanen *et al.*, 2016). The χ^2 method divides the microbe sum by the sample sum, this value is then adjusted by the square root of the sum for the entire community (Legendre and Gallagher, 2001). Finally, the hellinger distance standardizes the community data by calculating the square root of the relative microbe frequency (Legendre and Gallagher, 2001). The χ^2 and hellinger distance methods are commonly used within the field of community ecology. Each of these techniques will be examined in an exploratory manner to examine a measure of evenness for this data.

The aforementioned conclusions do confirm that the soil composition is diverse from region to region; also suggesting that there are other factors effecting the composition, e.g. changes in climate. In 2008, He *et al.* investigated the bacterial and

fungal communities together and concluded the two communities within that particular study site were correlated. The two communities were also examined across two seasons (summer and winter) and significant differences between the seasons were recorded. This conclusion confirms that there are some differences between the compositions; yet, this does not signify if there are differences between fall and spring.

While there has been evidence of significant differences in soil composition from environmental factors, as well as sample depth, these differences have been determined on independent study sites. This present study will aim to determine the effect of seasonal changes and moisture level on the composition of the soil within one study site, the Priest River Experimental Forest in northern Idaho. The study will also identify if significant differences are observed at various depths. Finally, the interplay between the bacterial and fungal communities present will be examined, and potentially determine if environmental factors explain any significant interactions.

CHAPTER 2

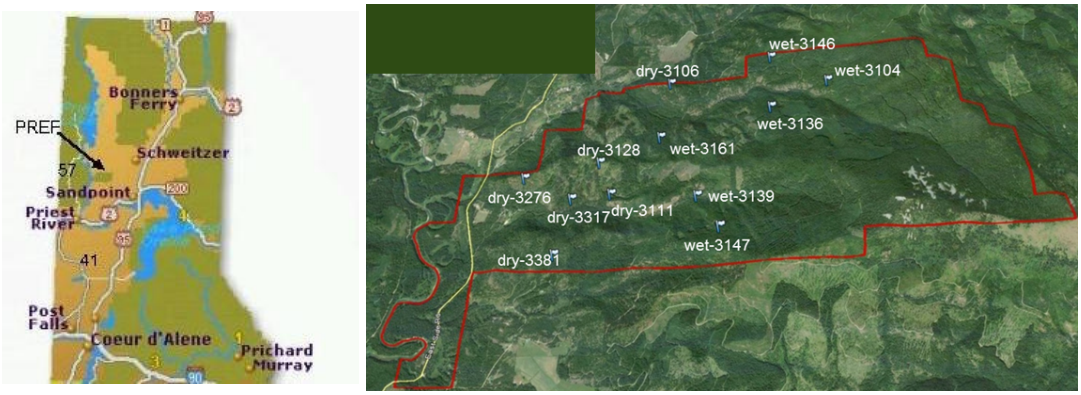
Methods

2.1 Motivation and experimental design

This study was motivated by the lack in knowledge of how the bacterial and fungal communities differ across seasons, habitat type series, and with depth below the forest floor, within the Priest River Experimental Forest (PREF) in northern Idaho. As mentioned, *Armillaria* root disease is a diverse fungal disease that affects a wide range of plants and trees, however, the presence of the fungi does not mean that it will be detrimental to the host, which may be the result of cross community interactions. The associations across communities related to season, habitat type series, and depth below the forest floor were explored through the implementation of metagenomic approaches to identify the microbial communities without losing information about microbes that are not able to be cultured. However, since soil microbial communities are a global topic of interest, many different techniques have been developed. Each method implements a slightly different algorithm, producing unique limitations that were explored within this dataset. Therefore, this study was not only motivated by the need to delve into the associations between communities, microbial and environmental, but into the methods implemented as well.

Samples were collected by Dr. Ross-Davis from sites previously established within the forest by the Forest Inventory and Analysis (FIA). The sites were established in two different habitat type series: western red cedar (*Thuja pilcata*)/western hemlock (*Tsuga heterophylla*), and Douglas-fir (*Pseudotsuga menziesii*)/grand fir (*Abies grandis*). These two habitat type series provide a comparison of soil moisture levels; the western red cedar/western hemlock habitat type can be attributed to a wetter environment, while the Douglas-fir/grand fir habitat type can be attributed to a dryer environment. In this study, data were collected from 12 sites: six in the western red cedar/western hemlock (wet environment) and six in the Douglas-fir/grand fir (dry environment) (see Figure 2.1). Mineral soil core samples were collected from each plot at the midpoint between

the bole and drip line of the tree nearest the center of the plot. Four 0.25g subsamples were collected from each soil core (three from different depths below the forest floor [0 cm, 7.5 cm, and 15 cm], and a composite sample). Sampling began fall 2013 and was repeated in spring 2014. In addition to exploring the associations related to season, moisture level, and sample depth (which will aide Dr. Ross-Davis's current analysis), the data collected was used in this study to characterize the interactions between the different microbial communities present within the PREF.



(A) (B)

FIGURE 2.1: **Priest River Experimental Forest (PREF).** The location of the PREF in northern Idaho (A), and the location of the sampling sites and their associated environment (B)[1].

2.2 MICROBIAL COMMUNITY CHARACTERIZATION ACROSS SAMPLES

In order to begin to address these interactions, the microbial composition across samples needed to be characterized. There are five loci studied to determine what is present within the bacterial and fungal communities: variable regions of the 16S ribosomal RNA (rRNA) gene (V1-V3 and V4-V5), the first and second internal transcribed spacer regions (ITS1 and ITS2), and the large subunit rRNA gene (LSU). The 16S V1-V3 and V4-V5 regions are frequently used to examine regional composition and diversity of a bacterial community because the gene is ubiquitous in bacteria and the sequences vary across bacterial species (Schloss *et al.*, 2016; Smets *et al.*, 2016). The ITS1 and

[1] Images were obtained from http://forest.moscowfsl.wsu.edu/ef/pref/ and Dr. Amy Ross-Davis.

ITS2 regions are commonly used to identify fungal communities because of the high variability within the regions; the spacers are also present in all fungi, making it similar to the 16S rRNA gene (Porter *et al.*, 2016; Žifčáková *et al.*, 2016). LSU sequences are conserved, and therefore can be used as an alternative to the two hypervariable ITS regions to determine the fungal community captured (Brown *et al.*, 2014). The four subsamples from each site were stored in a buffer and held on ice until the DNA was isolated from each using the PowerLyzer® PowerSoil® DNA Isolation Kit (MoBio Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer's recommended protocol with the following exceptions: samples were heated to 65°C for 10 min in a water bath prior to bead beating using a FastPrep® FP120 cell disruptor (Qbiogene Inc., Carlsbad, CA, USA) for 45 sec at speed 4.5 and centrifuged at 10,000 x 1g for 3 min at room temperature. DNA yields were quantified via a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Using double-barcoded amplicons generated from replicated polymerase chain reaction (PCR), the microbial communities for each loci described were obtained from each of the samples and utilized in this analysis.

Once the DNA from the respective regions has been extracted and sequenced, one of two methods is commonly implemented to identify the microbes present. These methods include clustering the sequences based on similarity into operational taxonomic units (OTUs), or classifying each sequence separately by utilizing a database of known reference sequences. Currently, there are three widely used applications in microbial community analysis. Two of these (QIIME (Navas-Molina *et al.*, 2013) and mothur (Schloss *et al.*, 2009)) are OTU based methods and provide information about clusters and classification, whereas the other, RDP (Vilo and Dong, 2012), is used for classification purposes only. Determining the functionality of the sequences classified was out of the scope of the present study, therefore the RDP classifier was utilized through the implementation of *dbcAmplicons* (`https://github.com/msettles/dbcAmplicons`). *dbcAmplicons* is a custom Python pipeline that was developed at the University of Idaho in the Genomic Resources Core; this pipeline was used to characterize the two bacterial and three fungal communities. Read sequences were first classified using the RDP classifier, then abundance tables were generated by aggregat-

ing all the sequences with the same classifications. The resulting abundance tables allowed for the assessment of community diversity and correlation metrics, as well as verification that the two respective regions do indeed represent the same community composition. Should the two respective communities differ, the non-redundant classifications between the two regions provide evidence supporting the use of both regions to obtain a more complete depiction of the community.

2.3 IDENTIFICATION OF ENVIRONMENTAL DIFFERENCES

As previously mentioned, detecting if changes in seasonality, moisture level, or depth of the sample, had a significant effect on the soil composition was a question of interest. Since communities are often dominated by few relatively abundant taxa, the raw abundance data needed to transformed to accurately weight the observations, in order to reduce the potential for misleading influence when drawing conclusions. However, choosing a transformation technique to precisely depict the differences is not a trivial task. For instance, within the commonly used R package, vegan, there are 10 different standardization methods for community ecology (Oksanen *et al.*, 2016). The methods include: *freq*, *pa* (presence/absence), *hellinger distance*, *normalize*, *total*, *max*, *range*, *standardize*, χ^2, and *log*. The *freq* method standardizes the data by dividing each taxa in a sample by the sample maximum and multiplies the resulting values by the number of taxa present within the sample (Oksanen, 1983). The *pa* method scaled the microbes identified to be represented by a 1 or 0 depending on if the microbe was present or absent, respectively, within each sample (Oksanen *et al.*, 2016). The *range* standardization technique transforms the values within a sample to range from 0 to 1 (Oksanen *et al.*, 2016). The *hellinger distance* and χ^2 methods are widely used within ecological community analysis and were published by Legendre and Gallagher (2001). The *hellinger distance* is calculated by taking the square root of the relative frequencies, whereas the χ^2 distance is the sample sum divided by the microbial sum, adjusted for the total sum (Legendre and Gallagher, 2001). *Normalize* transforms the data so the sum of squares for each taxa are equal to one, and the *standardize* technique standardizes the values to have mean 0 and variance 1 (Oksanen *et al.*, 2016). Finally,

the *log* technique transforms the data by calculating the log of the non-zero elements and adding 1; this method was suggested by Anderson *et al*. (2006). A representative subset of the methods (*hellinger distance*, χ^2, *normalize*, and *log*) were implemented in an exploratory capacity by comparing Pielou's evenness J measure (McCune *et al*., 2002).

We first examined the effects of standardization on a commonly used evenness measure within community analysis, Pielou's evenness J. Pielou's evenness J is calculated by dividing Shannon's diversity index, H, by the log of the total number of species (Oksanen *et al*., 2016). Shannon's diversity index is one of the most commonly used diversity indices, and is calculated using the following equation:

$$H = -\sum_{i=1}^{S} p_i log(p_i)$$

where p_i is the proportion of microbe i, and S is the number of microbes to meet the constraint $\sum_{i=1}^{S} p_i = 1$ (Hill, 1973). Using the *diversity* and *specnumber* functions within the vegan package, Pielou's evenness J was easily calculated. The number of microbes is unaffected by the standardization method; therefore, this exploratory analysis examined the effects on the diversity measure. To test for significant differences within the measures, the evenness distributions from each standardization method was tested against the mean evenness measure from the other methods using a students t-test. Due to the fact that the composite samples contain information from the various depths, these samples were removed from further analyses to detect independent variations.

In order to detect compositional differences and remove weighted influence from high abundant microbes, the data were first transformed using the *pa* method. Using the resulting matrix, the samples were condensed into 12 sample categories based on group (season and environment) and depth, i.e. Fall from a wet environment at depth 0. In this case, the samples are explained by the categorical predictor variables: group, X_{i1}, and depth, X_{i2}, and the response vector, $\boldsymbol{y_i} = (y_{i1}, \ldots, y_{ir})^T$, which consists of the microbes present, $j = 1, \ldots, r$, for $i = 1, \ldots, 12$ categories. Finally, we then supposed that $\boldsymbol{y_i}$ followed a multinomial distribution with probabilities, $\pi_j(\boldsymbol{x})$, and implemented a multinomial logistic regression model to determine the odds of being in one category

relative to the designated baseline category, j^* (Agresti, 2002). The logistic model equation is then

$$log\left(\frac{\pi_j(\boldsymbol{x})}{\pi_{j^*}(\boldsymbol{x})}\right) = \alpha_j + \boldsymbol{\beta}_j'\boldsymbol{x}$$

where $j \neq j^*$, α_j = intercept coefficient, β' = vector of estimated coefficients given j^*, and x = vector of predictor variable indicators. The exponentiated log-odds, $log\left(\frac{\pi_j(\boldsymbol{x})}{\pi_{j^*}(\boldsymbol{x})}\right)$, are interpreted as the odds of a particular taxa relative to the baseline taxa within a set combination of predictor values. The multinomial logistic regression model was implemented for each community abundance table and inferences within, and across, communities were identified.

The compositional differences were also tested by implementing the *adonis* function within the vegan package. This method is a permutational multivariate analysis of variance that utilizes distance matrices to determine significant differences within the data (Oksanen *et al.*, 2016). In this study *adonis* was executed using the Bray-Curtis dissimilarity distance metric to test three main hypotheses. The Bray-Curtis index of dissimilarity is calculated using the equation

$$d_{kl} = \frac{\sum |x_{rk} - x_{rl}|}{\sum (x_{rk} + x_{rl})}$$

where r = the microbes classified, and k and l = the samples such that $k \neq l$, and is the default metric within the *vegdist* function (Oksanen *et al.*, 2016). Under default conditions, the hypotheses that the associations related to group, depth, and/or their interaction do not effect the composition (H_{01}), the associations related to group independently do not effect the composition (H_{02}), and the associations related to depth independently do not effect the composition (H_{03}) were tested. These hypotheses, however, negate the fact plots are nested within an environment. The use of the *strata* parameters within the function tested the hypotheses taking plot into account by restricting the permutation within the strata (i.e. plot). At the same time, this also restricted the samples to a certain environment, thus not measuring the variable associations. Lastly, since each group is defined by a season and environment/moisture level, hypotheses to test if season, moisture level, or both, have no significant associations to differences within sample composition were tested.

2.4 QUANTIFY THE STRENGTH OF COMMUNITY INTERACTIONS

Recent studies have only begun to study the interactions between the bacterial and fungal communities. Identifying the interactions could however, provide insights into which microbes are key to aiding the spread, or suppression, of different diseases, e.g. *Armillaria* root disease. In this method of analysis the samples are not considered independent; therefore, to investigate microbial interactions from independent samples, the communities were grouped by season and depth, e.g Fall at depth 7.5, and visualized through heatmaps. The correlations between the pairwise bacterial and fungal communities were first calculated using the spearman rank-correlation analysis technique.

Hierarchical clustering was then implemented on each individual community, using *hclust* in R. This clustering aided in identifying the members of the respective community that commonly appear with the same correlation values across the microbial community, and are potentially influential in driving the community correlations. This analysis method, however, again brought up the issue of which distance clustering algorithm to use, and where to cut the hierarchical tree, or dendrorgram, to determine the clusters. There are eight different clustering approaches within *hclust*, therefore, to test the strength of the determined clusters four widely used methods from the list were implemented: Ward's minimum variance (ward.D2), single-linkage, average-linkage, and complete-linkage. These four methods provide a representative subset of the clustering algorithms within the function. Each of these methods is an agglomerative approach, meaning every microbe is its own cluster and the tree is built from the bottom-up. Ward's minimum variance method combines clusters by determining the increase in the sum of squares should the two clusters merge and merging the two clusters with the smallest increase (Rencher and Christensen, 2012). The single-linkage method merges clusters based on the shortest distance between two single microbes, whereas the complete-linkage approach consolidates clusters based on the shortest maximum distance between two elements of the existing clusters (Rencher and Christensen, 2012). While the average-linkage approach combines clusters based on the average distance between all the elements of two existing clusters (Rencher and Christensen, 2012).

After the implementation of each of these methods, elbow plots were generated and examined to determine the optimal number of clusters; then using that optimal number, the four resulting trees were cut and the clusters were examined. The elbow plots were generated using the function *fviz_nbclust* in the factoextra R package (Kassambara and Mundt, 2016). This function calculates the total within sum of squares and where the values start to taper off, or a bend in the graph, is determined to be the optimal number of clusters (Kassambara and Mundt, 2016). For the purposes of the this study, the euclidean distances were calculated, using *dist*, for each step in the clustering algorithms and correlation heatmaps were constructed and the interactions were investigated further.

chapter 3

Results

3.1 microbial community characterization across samples

There were 1,024 and 1,026 microbes classified within the bacterial community using the 16S V1-V3 and 16S V4-V5 regions, respectively, and 605 and 711 within the fungal community using ITS1 and ITS2, after the implementation of *dbcAmplicons*. The data was cleaned to remove possible classification errors; this was done through the removal of microbes with relative frequencies below a certain threshold. In order to increase the probability of obtaining accurate classifications, for a microbe to be retained there needed to be at least two samples where the relative frequency was above 1%, or at least one sample with a relative frequency greater than 5%. These criteria reduced the chances of uncertain classifications due to sequencing errors. After this reduction technique was executed there were 79, 79, 61, and 68 total microbial classifications, and 77, 78, 70, 82, and 129 genus microbial taxonomic classifications within the respective bacterial and fungal communities (see Table 3.2). Another aspect of the reduction technique included removing those samples with minimal information due to a lack in read sequences, which may be the result of sampling errors. If a sample contained fewer than 3,000 read sequences, it was removed from downstream analysis. The samples that were removed from each region can be seen in Tables 3.3-3.7. When examining the LSU region, 14 of the 96 samples were removed, therefore further analysis will focus on the 16S V1-V3, 16S V4-V5, ITS1, and ITS2 regions.

Once the data had been cleaned the microbial classifications of each region were examined further. After additional investigation into the bacterial microbes classified to the genus level, inconsistencies were identified. Of the 77 and 78, genus level bacterial taxonomic classifications, there were 26 and 27 classification distinctions between the 16S V1-V3 and V4-V5 regions, respectively. Similarly, of the 70 and 82 genus level fungal classifications, there were 13 and 25 classification distinctions

between the ITS1 and ITS2 regions, respectively. These non-redundant classifications within the respective communities therefore provide evidence for continued use of both regions to gain a more complete bacterial, or fungal, community composition, and analysis of the community.

TABLE 3.1: **Raw Community Classifications by Level.** The number of taxonomic classifications identified at each region broken down by taxonomic level, *before* the data were reduced.

Level	Bacterial		Fungal		
	16S V1-V3	16S V4-V5	ITS1	ITS2	LSU
Phylum	38	37	5	5	16
Class	88	91	28	30	47
Order	151	154	81	97	128
Family	265	276	189	227	292
Genus	824	819	464	554	730

TABLE 3.2: **Reduced Community Classifications by Level.** The number of taxonomic classifications identified at each region broken down by taxonomic level, *after* the data were reduced.

Level	Bacterial		Fungal		
	16S V1-V3	16S V4-V5	ITS1	ITS2	LSU
Phylum	12	12	4	4	4
Class	27	25	12	13	16
Order	40	46	29	34	36
Family	56	63	54	60	68
Genus	77	78	70	82	129

TABLE 3.3: **16s V1-V3 all taxonomic level totals per sample.** Using the reduced set of the taxonomic classifications, and the data from all taxonomic levels, the table below describes the number of taxa in each sample for the 16s V1-V3 region. The maximum number of taxa per sample is 79. Samples may have a variety of taxa present, however, if a sample had fewer than 3,000 read sequences, it was removed from downstream analysis due to the possibility of sampling error (red).

Environment	Plot	Season	Depth			
			0	7.5	15	C
Dry	3106	FA	73	79	77	4
		SP	77	78	79	78
	3111	FA	69	73	70	75
		SP	77	76	75	76
	3128	FA	70	76	67	8
		SP	74	75	77	75
	3276	FA	76	76	78	76
		SP	74	77	75	78
	3317	FA	76	75	77	75
		SP	75	76	77	77
	3381	FA	77	79	79	79
		SP	76	79	79	79
Wet	3104	FA	73	77	77	4
		SP	76	78	77	79
	3136	FA	74	74	72	78
		SP	77	78	78	79
	3139	FA	73	68	76	78
		SP	76	4	78	79
	3146	FA	71	75	78	77
		SP	75	77	76	78
	3147	FA	74	64	74	76
		SP	73	69	76	75
	3161	FA	75	75	76	75
		SP	73	73	79	78

TABLE 3.4: **16s V4-V5 all taxonomic level totals per sample.** Using the reduced set of the taxonomic classifications, and the data from all taxonomic levels, the table below describes the number of taxa in each sample for the 16s V4-V5 region. The maximum number of taxa per sample is 79. Samples may have a variety of taxa present, however, if a sample had fewer than 3,000 read sequences, it was removed from downstream analysis due to the possibility of sampling error (red).

Environment	Plot	Season	Depth			
			0	7.5	15	C
Dry	3106	FA	76	77	78	79
		SP	78	79	78	78
	3111	FA	76	78	78	76
		SP	77	77	77	77
	3128	FA	76	77	73	77
		SP	75	78	79	77
	3276	FA	78	79	78	79
		SP	77	79	78	79
	3317	FA	78	77	78	78
		SP	78	78	78	78
	3381	FA	78	78	79	77
		SP	78	79	79	79
Wet	3104	FA	76	77	77	79
		SP	76	78	78	78
	3136	FA	74	76	69	78
		SP	76	10	78	79
	3139	FA	73	77	76	78
		SP	78	77	78	53
	3146	FA	75	78	78	79
		SP	75	78	76	79
	3147	FA	75	31	78	54
		SP	76	72	76	75
	3161	FA	77	76	78	74
		SP	77	75	76	78

TABLE 3.5: **ITS1 all taxonomic level totals per sample.** Using the reduced set of the taxonomic classifications, and the data from all taxonomic levels, the table below describes the number of taxa in each sample for the ITS1 region. The maximum number of taxa per sample is 61. Samples may have a variety of taxa present, however, if a sample had fewer than 3,000 read sequences, it was removed from downstream analysis due to the possibility of sampling error (red).

Environment	Plot	Season	Depth			
			0	7.5	15	C
Dry	3106	FA	42	35	48	46
		SP	42	40	42	39
	3111	FA	45	42	46	44
		SP	41	42	40	46
	3128	FA	25	19	28	47
		SP	36	38	38	41
	3276	FA	39	39	40	39
		SP	46	48	36	43
	3317	FA	40	39	40	44
		SP	45	41	41	47
	3381	FA	50	48	49	49
		SP	49	47	44	53
Wet	3104	FA	35	39	44	43
		SP	46	40	42	45
	3136	FA	43	35	29	49
		SP	39	44	43	45
	3139	FA	39	32	38	43
		SP	47	44	46	48
	3146	FA	26	42	45	46
		SP	45	49	44	46
	3147	FA	43	35	43	42
		SP	43	33	43	44
	3161	FA	44	32	43	42
		SP	42	40	41	46

TABLE 3.6: **ITS2 all taxonomic level totals per sample.** Using the reduced set of the taxonomic classifications, and the data from all taxonomic levels, the table below describes the number of taxa in each sample for the ITS2 region. The maximum number of taxa per sample is 88. Samples may have a variety of taxa present, however, if a sample had fewer than 3,000 read sequences, it was removed from downstream analysis due to the possibility of sampling error (red).

Environment	Plot	Season	Depth			
			0	7.5	15	C
Dry	3106	FA	54	46	63	61
		SP	54	56	55	65
	3111	FA	66	61	59	65
		SP	61	59	57	67
	3128	FA	51	30	45	64
		SP	51	62	57	57
	3276	FA	67	62	60	63
		SP	67	58	55	62
	3317	FA	58	62	60	69
		SP	66	54	60	64
	3381	FA	76	69	72	72
		SP	71	70	63	38
Wet	3104	FA	50	53	63	54
		SP	40	58	52	65
	3136	FA	57	45	30	64
		SP	60	64	60	64
	3139	FA	52	55	62	56
		SP	67	58	60	64
	3146	FA	34	62	59	56
		SP	54	61	62	68
	3147	FA	58	42	57	57
		SP	59	51	60	66
	3161	FA	62	50	62	63
		SP	64	60	57	66

TABLE 3.7: **LSU all taxonomic level totals per sample.** Using the reduced set of the taxonomic classifications, and the data from all taxonomic levels, the table below describes the number of taxa in each sample for the LSU region. The maximum number of taxa per sample is 114. Samples may have a variety of taxa present, however, if a sample had fewer than 3,000 read sequences, it was removed from downstream analysis due to the possibility of sampling error (red).

Environment	Plot	Season	Depth			
			0	7.5	15	C
Dry	3106	FA	5	61	89	77
		SP	69	60	73	73
	3111	FA	65	74	71	73
		SP	68	78	70	80
	3128	FA	34	11	47	70
		SP	61	74	66	74
	3276	FA	11	68	69	70
		SP	75	73	62	78
	3317	FA	11	61	68	68
		SP	72	62	66	73
	3381	FA	67	70	74	1
		SP	80	86	69	87
Wet	3104	FA	6	49	72	71
		SP	77	65	65	78
	3136	FA	84	42	31	78
		SP	71	74	69	24
	3139	FA	58	43	59	75
		SP	84	71	78	79
	3146	FA	54	74	79	75
		SP	69	82	68	86
	3147	FA	66	52	63	67
		SP	74	67	60	63
	3161	FA	22	57	67	67
		SP	74	66	41	74

3.2 COMPOSITIONAL DIFFERENCES WITHIN A COMMUNITY

In order to detect if season, moisture level, and/or depth had an impact on the microbial compositions, the data first needed to be transformed. Previously shown, the standardization method implemented can significantly affect the diversity measure calculations (Lozupone *et al.*, 2007). Therefore, using the *decostand*, *diversity*, and *specnumber* functions within the vegan R package, a representative subset of four different standardization methods were implemented, and the resulting Pielou's evenness J measures were tested for a significant difference. The standardization techniques implemented were the *hellinger distance*, χ^2, *normalize*, and *log*. When the Peilou's evenness J values were tested for significant differences using a student's t-test, all values were statistically significant except for the *normalize* technique when tested against the raw evenness measure (Table 3.8). Due to the clear-cut calculation of the *hellinger distance*, and that it is among the more commonly used methods within soil community analysis, this standardization technique was chosen for downstream analyses within this study.

TABLE 3.8: **Standardization method effects on Pielou's evenness J.** The mean evenness measure for each region. The *normalize* method resulted in the same evenness measure as using the raw data for each region.

	16S V1-V3	16S V4-V5	ITS1	ITS2
Raw	0.836	0.848	0.563	0.627
Hellinger	0.948	0.954	0.820	0.854
χ^2	0.907	0.918	0.660	0.664
Normalize	0.836	0.848	0.563	0.627
Log	0.987	0.990	0.954	0.958

We then began to explore how the community composition differed with season, habitat type series, and depth below the forest floor. These differences were investigated using two methods: multinomial logistic regression and a permutational multivariate analysis of variance. To implement the multinomial logistic regression model the data needed to be standardized into a *presence/absence (pa)* matrix, and for the permutational multivariate analysis of variance the data was standardized using the *hellinger distance*.

The multinomial logistic regression was first implemented using the *vglm* within the vgam R package to select the best fit model (Yee, 2016). The *vglm* function is used to fit vector generalized linear models, and the multinomial model is specified through the parameter *family*, as can be seen in the command below:

$$vglm(data\_frame \sim group + depth + group : depth, data = data\_frame, family = multinomial)$$

In order to select the model to be used, an analysis of deviance was conducted to compare the significance of the each variable. The deviance G^2 for each fitted model within each region can be found in Table 3.9 and were calculated using *deviance*. Each model was then tested using Pearson's χ^2 test, and the model with the main effects from group and depth (group + depth) was determined to be the best fit model for each region.

TABLE 3.9: **Deviance calculations for model selection.** The deviance G^2 and degrees of freedom (df) for each model fitted for each region. These values were used to implement Pearson's χ^2 test. The main effects model was selected for further analysis in each region.

(A) 16S V1-V3

Model	G^2	df
Group + Depth	47.226	468
Group	78.876	624
Depth	65.355	702
Null	96.844	858

(B) 16S V4-V5

Model	G^2	df
Group + Depth	31.552	468
Group	52.607	624
Depth	51.288	702
Null	72.213	858

(C) ITS1

Model	G^2	df
Group + Depth	102.083	360
Group	160.004	480
Depth	274.088	540
Null	332.938	660

(D) ITS2

Model	G^2	df
Group + Depth	174.358	522
Group	250.323	696
Depth	424.435	783
Null	501.213	957

When fitting the main effects model for the two fungal communities, the model did not converge using the default number of iterations (100), therefore, the iterations were increased to 200, and the model converged after ~160 iterations. The estimated coefficients and odds ratios were then examined using *multinom* within the nnet package (Ripley and Venables, 2016). The odds ratios were calculated by exponentiating the coefficient values and are displayed in Figures 3.1-3.4. First investigating the changes within the two bacterial communities (Figures 3.1 and 3.2), there are many microbes that did not have an increase or decrease in odds relative to the baseline taxa (*Ganoderma*) across the various levels of the two predictor variables. However, there were a few microbes that a one-unit increase in a category appeared to yield a significant increase, or decrease, in the odds for that microbe. For example, in the 16S V1-V3 region, there was an increase in the odds of *Thermoactinomyces* relative to *Ganoderma* within samples at 7.5 and 15 cm below the forest floor. Examining the 16S V4-V5 region, the presence of *Thermoactinomyces* appeared to have similar associations relative to changes in depth, but a decrease in odds with changes in group. However, after investigating the 95% confidence intervals, in both bacterial communities, only the increased odds at depth 7.5 cm below the forest floor were significant. Further, *Segetibacter* appeared to have a decreased odds relative to *Ganoderma* across the different levels of both variables, but this decrease was inconclusive due to non-significance. While there was some overlap in potentially influential microbes between the two communities, most were distinct, providing further evidence for the use of both regions to gain a more complete understanding of the bacterial composition with the PREF.

We next examined the changes in the two fungal communities (Figures 3.3 and 3.4). While again the odds of a microbe did not appear to increase or decrease relative to the baseline taxa (*Ganoderma*), across the two regions there were microbes with significant increases or decreases, suggesting potentially noteworthy associations; these include *Diversispora*, *Tylospora*, and *Bevicellicium*. However, each of the 95% confidence intervals spanned across one, making the apparent associations inconclusive. Similarly to the bacterial regions, within the fungal communities potentially influential microbes overlapped between the two regions, within the exception of *Diversispora* in the ITS1

community. While the associations between microbes relative to *Ganoderma* were largely inconclusive, the distinctions within the community provides further evidence for the use of both fungal regions as well.

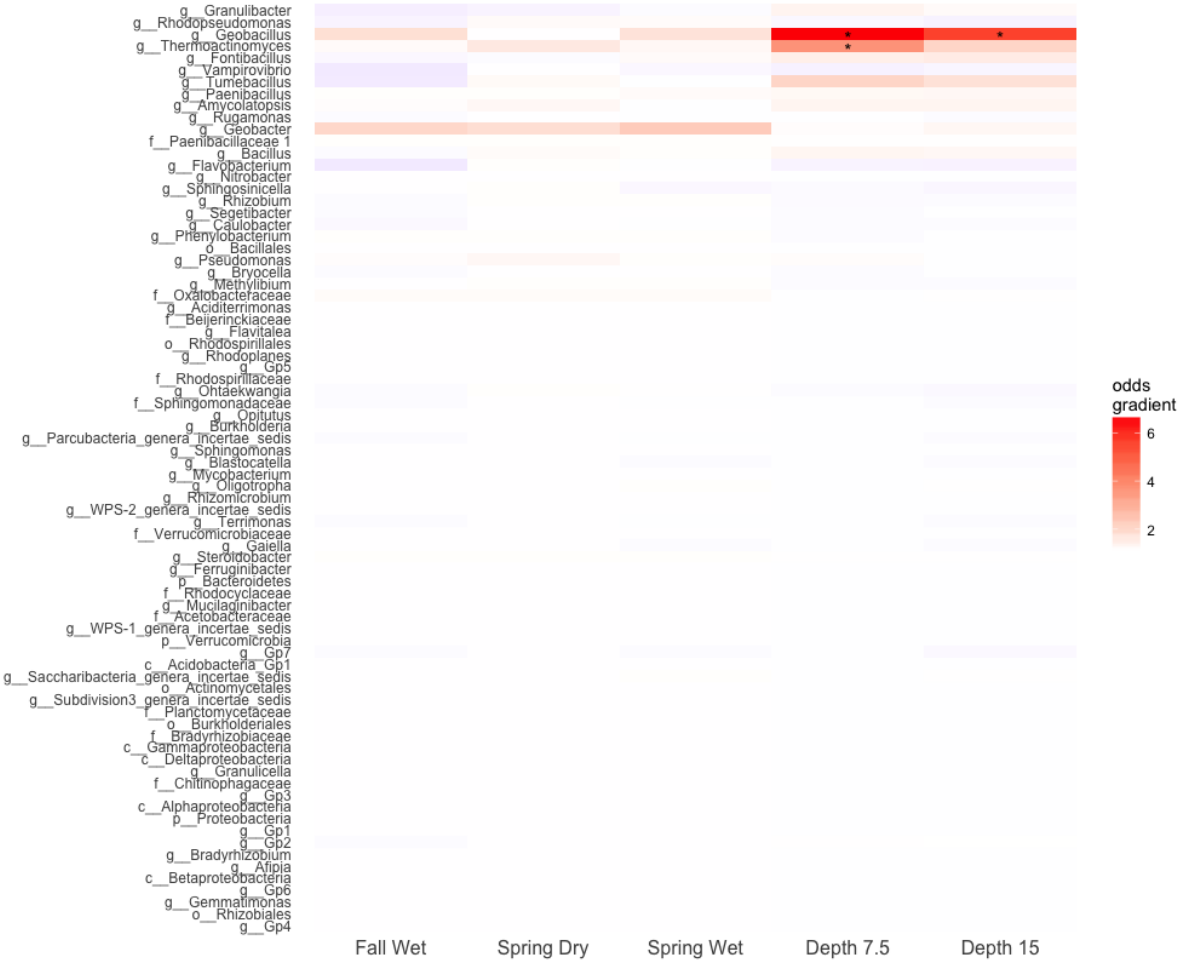


FIGURE 3.1: **Odds ratios for the 16S V1-V3 community.** The heatmap indicates there are a few microbes with odds increased, or decreased, across the variables examined. The microbes with decreased odds relative to the baseline taxa (*Ganoderma* are blue, and odds with increased odds are red. The significant odds ratios with $p < 0.05$ are signified with a *.

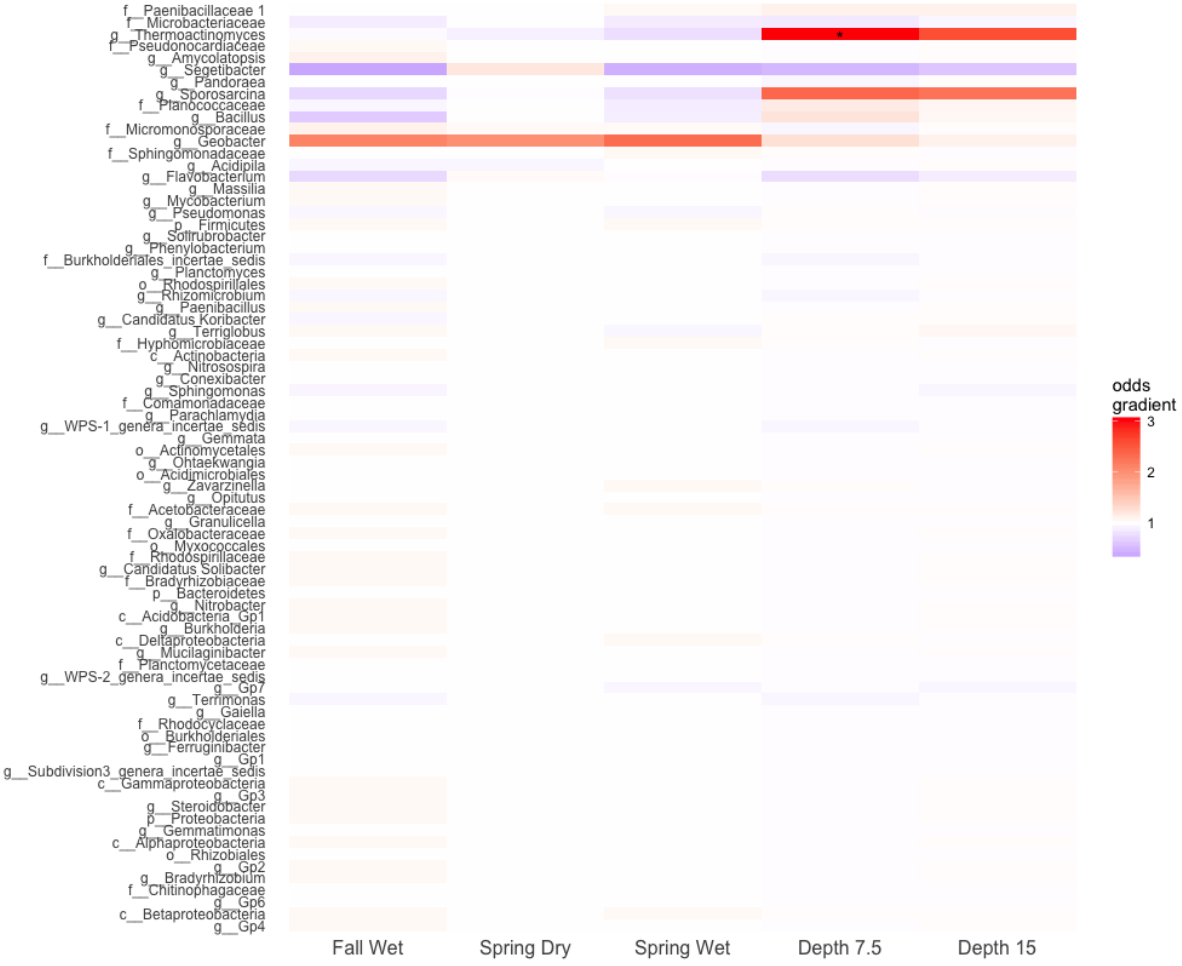


FIGURE 3.2: **Odds ratios for the 16S V4-V5 community.** The heatmap indicates there are a few microbes with odds increased, or decreased, across the variables examined. The microbes with decreased odds relative to the baseline taxa (*Ganoderma* are blue, and odds with increased odds are red. The significant odds ratios with $p < 0.05$ are signified with a *.

FIGURE 3.3: **Odds ratios for the ITS1 community.** The heatmap indicates there are a few microbes with odds increased, or decreased, across the variables examined. The microbes with decreased odds relative to the baseline taxa (*Ganoderma* are blue, and odds with increased odds are red. The significant odds ratios with $p < 0.05$ are signified with a *.

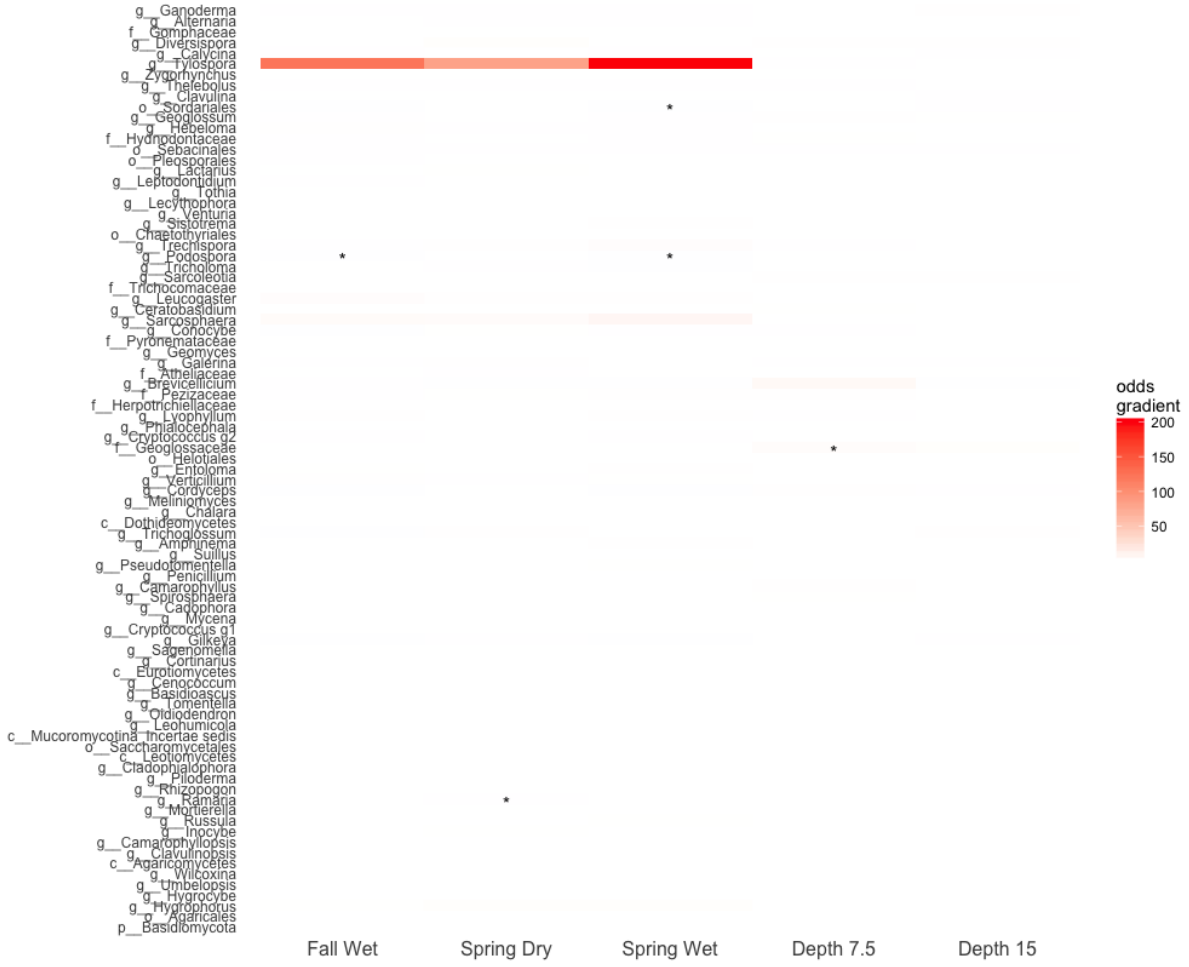


FIGURE 3.4: **Odds ratios for the ITS2 community.** The heatmap indicates there are a few microbes with odds increased, or decreased, across the variables examined. The microbes with decreased odds relative to the baseline taxa (*Ganoderma* are blue, and odds with increased odds are red. The significant odds ratios with $p < 0.05$ are signified with a *.

As mentioned, a permutational multivariate analysis of variance (*adonis*) was also implemented to detect differences within the samples. The three hypotheses were tested using the following commands:

H_{01}: Associations related to the interaction between group and depth do not effect the community composition

$$adonis(data\_frame \sim group * depths, data = data\_frame).$$

H_{02}: Associations related to group do not effect the community composition

$$adonis(data\_frame \sim group , data = data\_frame).$$

H_{03}: Associations related to depth do not effect the community composition

$$adonis(data\_frame \sim depths, data = data\_frame).$$

The results in Table 3.10 indicate group and depth both have significant associations within the bacterial sample composition, while the fungal sample compositions were only significantly associated to differences in the group (see Table 3.11). Even though the associations related to both group and depth were determined significant through H_{01} for both bacterial communities, the associations were still tested independently. Tables 3.12 and 3.13 show that the associations were significant with $p < 0.05$. When group and depth were independently tested for the fungal communities, the associations related to depth were calculated to be significant with $p < 0.05$ within the ITS2 fungal community, and remained non-significant within the ITS1 fungal community. Therefore, these results provide evidence that both the group and from what depth the sample was collected had a significant impact on the sample composition, in three out of the four communities examined. However, the associations related to the interaction of these factors did not result in a significant difference.

Due to the fact that the associations related to group were significant in all four communities, we used the same procedure to examine the independent associations related to season and moisture level. Interestingly, in all four communities the interaction was non-significant. In addition the associations due to season were non-significant in all communities except the 16S V1-V3 community. These results suggest that the microbes present are impacted by the moisture level within the soil, rather

than the time of year. Therefore, in the case of the bacterial communities, composition was impacted mainly by moisture level and depth of the sample. Whereas the fungal community composition was mainly impacted by moisture level.

TABLE 3.10: **Group & depth associations in the bacterial communities.** The results for the *adnois* permutation analysis for the associations related to group and depth on the sample composition (H_{01}) when examining the two bacterial communities. The tables show there is a significant association ($p < 0.05$) from group and depth, but the interaction is not significant.

(A) 16S V1-V3

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Group	3	0.244	0.081	2.680	0.098	**0.003**
Depths	2	0.398	0.199	6.552	0.160	**0.001**
Group:Depths	6	0.175	0.029	0.958	0.070	0.538
Residuals	55	1.671	0.030		0.672	
Total	66	2.488			1	

(B) 16S V4-V5

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Group	3	0.137	0.046	2.683	0.100	**0.001**
Depths	2	0.223	0.111	6.535	0.163	**0.001**
Group:Depths	6	0.090	0.015	0.885	0.066	0.627
Residuals	54	0.919	0.017		0.671	
Total	65	1.369			1	

TABLE 3.11: **Group & depth associations in the fungal communities.** The results for the *adnois* permutation analysis for the associations related to group and depth on the sample composition (H_{01}) when examining the two fungal communities. The tables show there is only significant associations ($p < 0.05$) from group, but associations related to depth and the interaction between group and depth were not significant.

(A) ITS1

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Group	3	0.729	0.243	2.057	0.096	**0.002**
Depths	2	0.238	0.119	1.007	0.031	0.408
Group:Depths	6	0.400	0.067	0.564	0.052	0.999
Residuals	53	6.262	0.118		0.821	
Total	64	7.629			1	

(B) ITS2

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Group	3	1.141	0.380	2.913	0.128	**0.001**
Depths	2	0.360	0.180	1.378	0.040	0.061
Group:Depths	6	0.513	0.086	0.655	0.057	0.999
Residuals	53	6.921	0.131		0.775	
Total	64	8.936			1	

TABLE 3.12: **Independent associations related to group & depth within the 16S V1-V3 community.** The results for the independent associations related to group (H_{02}) and depth (H_{03}). The associations from group and depth remained significant at the $p < 0.05$ level.

(A) H_{02} Group effect

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Group	3	0.182	0.061	2.452	0.109	**0.001**
Residuals	60	1.481	0.025		0.891	
Total	63	1.663			1	

(B) H_{03} Depth effect

	Df	Sum of Sqs	Mean Sq	F.Model	R2	Pr(>F)
Depths	2	0.283	0.142	6.262	0.170	**0.001**
Residuals	61	1.379	0.023		0.830	
Total	63	1.663			1	

TABLE 3.13: **Independent associations related to group & depth within the 16S V4-V5 community.** The results for the associations related to group (H_{02}) and depth (H_{03}). The significant association based on group and depth previously concluded remained significant at the $p < 0.05$ level.

(A) H_{02} Group effect

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Group	3	0.137	0.046	2.298	0.100	**0.002**
Residuals	62	1.232	0.020		0.900	
Total	65	1.369			1	

(B) H_{03} Depth effect

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Depths	2	0.222	0.111	6.087	0.162	**0.001**
Residuals	63	1.148	0.018		0.838	
Total	65	1.369			1	

TABLE 3.14: **Independent associations related to group & depth within the ITS1 community.** The results for the associations related to group (H_{02}) and depth (H_{03}). The associations based on depth were examined by restricting the samples to the correct groups, and remained non-significant.

(A) H_{02} Group effect

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Group	3	0.729	0.243	2.148	0.096	**0.002**
Residuals	61	6.900	0.113		0.904	
Total	64	7.629			1	

(B) H_{03} Depth effect

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Depths	2	0.242	0.121	1.016	0.032	0.341
Residuals	62	7.387	0.119		0.968	
Total	64	7.629			1	

TABLE 3.15: **Independent associations related to group & depth within the ITS2 community.** The results for the associations related to group (H_{02}) and depth (H_{03}). The associations based on depth were examined by restricting the samples to the correct groups, and were calculated to be significant with $p < 0.05$.

(A) H_{02} Association based on group

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Group	3	1.141	0.380	2.978	0.128	**0.001**
Residuals	61	7.794	0.128		0.872	
Total	64	8.936			1	

(B) H_{03} Association based on depth

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Depths	2	0.359	0.180	1.298	0.040	**0.048**
Residuals	62	8.577	0.138		0.960	
Total	64	8.936			1	

TABLE 3.16: **Season & moisture level associations in the bacterial communities.** The results for the *adnois* permutation analysis for the significance of season and moisture on the sample composition when examining the two bacterial communities. Interestingly, the tables show within the 16S V1-V3 community there was a significant association ($p < 0.05$) from season and moisture level. However, within the 16S V4-V5 community there was only significant associations due to moisture level.

(A) 16S V1-V3

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Season	1	0.070	0.070	2.763	0.039	**0.012**
Moisture	1	0.097	0.097	3.852	0.055	**0.001**
Season:Moisture	1	0.022	0.022	0.854	0.012	0.533
Residuals	63	1.590	0.025		0.894	
Total	66	1.778			1	

(B) 16S V4-V5

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Season	1	0.028	0.028	1.351	0.019	0.225
Moisture	1	0.099	0.099	4.830	0.068	**0.001**
Season:Moisture	1	0.011	0.011	0.544	0.008	0.804
Residuals	64	1.311	0.020		0.905	
Total	67	1.448			1	

TABLE 3.17: **Season & moisture level associations in the fungal communities.** The results for the *adnois* permutation analysis for the significance of season and moisture level on the sample composition when examining the two fungal communities. The tables show there was only a significant association ($p < 0.05$) from the moisture level.

(A) ITS1

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Season	1	0.131	0.131	1.158	0.017	0.275
Moisture	1	0.451	0.451	3.986	0.059	**0.002**
Season:Moisture	1	0.147	0.147	1.302	0.019	0.167
Residuals	61	6.900	0.113		0.904	
Total	64	7.629			1	

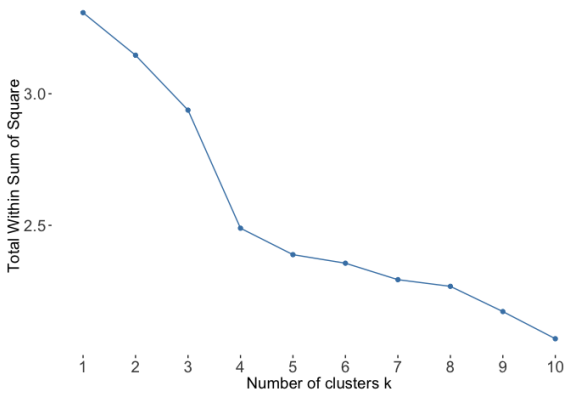
(B) ITS2

	Df	Sum of Sq	Mean Sq	F.Model	R2	Pr(>F)
Season	1	0.205	0.205	1.605	0.023	0.058
Moisture	1	0.809	0.809	6.328	0.090	**0.001**
Season:Moisture	1	0.128	0.128	0.999	0.014	0.422
Residuals	61	7.794	0.128		0.872	
Total	64	8.936			1	

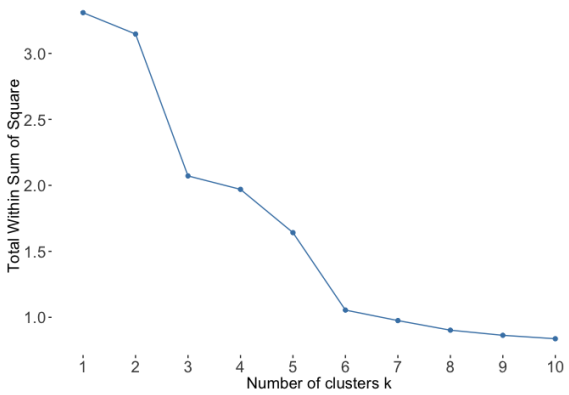
3.3 COMMUNITY ASSOCIATIONS

Using the *hclust* function in R, four agglomerative clustering methods (single-linkage, average-linkage, complete-linkage, and ward.D2) were implemented to determine the differences associated with the clustering approach. Before we could compare the clustering methods, the optimal number of clusters was first determined from elbow plots (Kassambara and Mundt, 2016). When examining the elbow plots for each method, the general trend of single-, average-, and complete-linkage resulted in a more prominent bend at an optimal number of clusters (Figures 3.5a,b, and c); this was consistent across regions. Ward's minimum variance method, however, generally resulted in a smooth curve, making the determination of an optimal number of clusters exceedingly subjective (Figure 3.5d). Since there was a prominent bend at four clusters using both single- and complete-linkage for this dataset, this was the value used to compare the dendrograms from each clustering method. Figure 3.6 displays the four resulting dendrograms with clusters highlighted. From these results it appeared that the Ward minimum variance and complete-linkage methods performed similarly in terms of cluster size and members, with one large cluster and three smaller clusters (Figures 3.6c and d). The average-linkage method resulted in four relatively even cluster sizes (Figure 3.6b). However, the single-linkage method resulted in one large cluster and three single microbe clusters; this is a result of the sequential (nearest neighbor) approach of the single-linkage method (Figure 3.6a). Due to the straight forward and stringent nature of the algorithm, complete-linkage was implemented for the community interaction analysis. Since the general clustering trend was consistent across communities and regions, only the results from fall at depth 0 using the 16S V1-V3 region are shown in Figures 3.5 and 3.6.

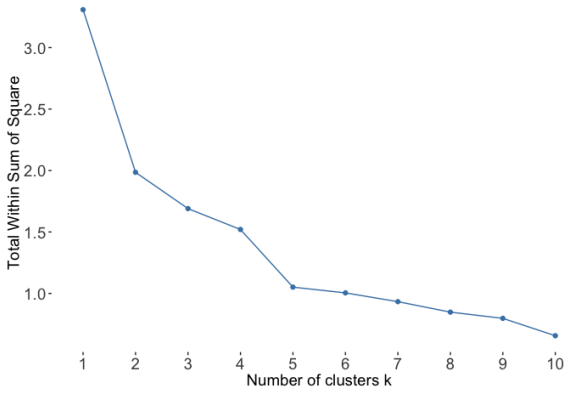
While the optimal number of clusters was detected to be four from the elbow plot, after examination of the correlation heatmaps, the number of clusters that appeared to optimally highlight the correlations consistently well across the regions was six. In order to ensure that appropriate microbial interactions were observed, the data for each region was first divided into season and depth, i.e. Fall at depth 0. Spearman's rank correlations were then calculated for each pairwise combination, at each depth, using the *cor* function in R. Using the resulting correlation matrix the microbial clusters for



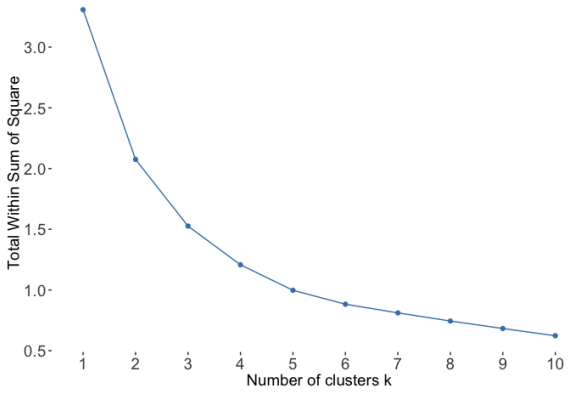
(A) Single-linkage clustering method



(B) Average-linkage clustering method

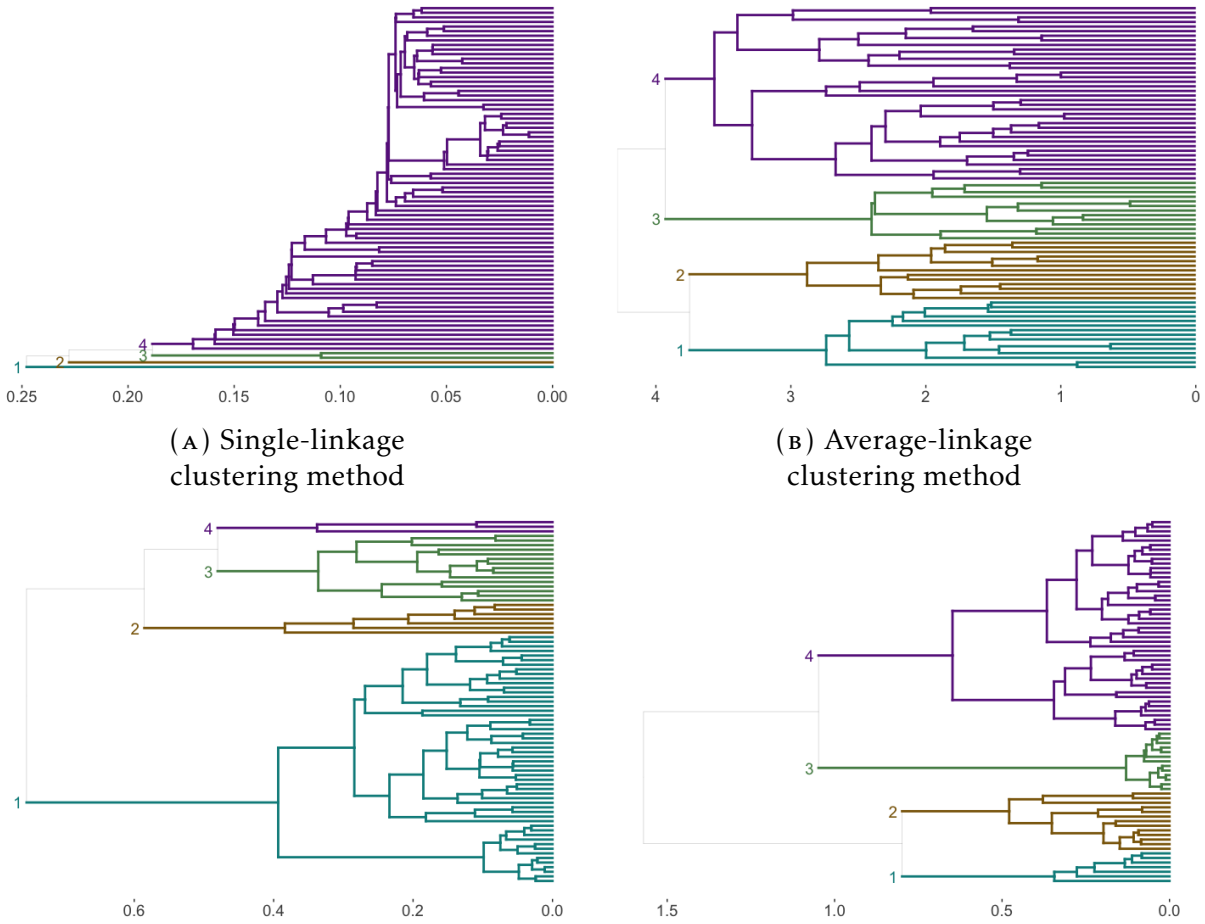


(C) Complete-linkage clustering method



(D) Ward minimum variance clustering method

FIGURE 3.5: **Elbow plots for 16S V1-V3 Fall at depth 0 samples.** The four elbow plots used to determine the optimal number of clusters for the 16S V1-V3 Fall at depth 0 samples. There is a more prominent bend in total within sum of squares at $k = 4$ clusters when using the single-linkage methods; the bend is prominent at $k = 3$ using the average-linkage method. Within the complete-linkage method there are two bends in the graph at $k = 2, 4$. However, the curve is smooth in the case of Ward's minimum variance, which resulted in determining the optimal number of clusters exceedingly subjective.



(A) Single-linkage clustering method

(B) Average-linkage clustering method

(C) Complete-linkage clustering method

(D) Ward minimum variance clustering method

FIGURE 3.6: **Hierarchical clusterings for 16S V1-V3 Fall at depth 0 samples.** The four resulting dendrograms for the 16S V1-V3 Fall at depth 0 samples. The single-linkage method resulted in one large cluster with three outlying microbes, whereas the complete-linkage and ward minimum variance methods clustered similarly in terms of size and membership. The average-linkage method resulted in four relatively equal clusters.

each community were determined through complete-linkage hierarchical clustering. Figures 3.7 and 3.9 display the correlation heatmap for the interaction between the 16S V1V3 and ITS1 regions for the fall and spring at depth 0, and the rest of the interaction heatmaps are located in Appendix A.

In Figures 3.7 and 3.9 there were clusters of microbes identified that were all correlated similarly; this trend was consistent across region and depth. After further examination into these correlations, it was seen that for some microbial combinations there was one sample where the two microbes were present and abundant, but in most samples either one or the other was present (see Figure 3.11d). The microbes contained in the four clusters displayed in Figures 3.8 and 3.10 were further examined for patterns explaining why the interactions were clustered. After further investigation there were no readily apparent patterns related to overall ecological functions within genera, however, there were still some commonalities among the clusters across region. For instance within the two bacterial regions, *Bacillus*, *Gaiella*, and *Opitutus* tended to be clustered together, as well as *WPS-1_genera_incertae_sedis*, *WPS-2_genera_incertae_sedis*, and *Subdivision3_genera_incertae*. There were also sets of fungal microbes commonly clustered across the two regions and the different combinations of variables examined. These include *Penicilium*, *Cryptococcus_g1*, and *Cryptococcus_g2*, and *Piloderma* and *Lactarius*. However, there were distinctions between the microbes classified between the two communities. Thus within the ITS1 community, *Armillaria* tended to be clustered with *Spirosphaera*, or *Tylospora*, and within the ITS2 community, *Russula* was associated with *Suillus*. Nevertheless, while not consistently clustered, within the fungal community there were multiple microbes associated with edible coral (e.g. *Inocybe* and *Clavulinopsis*), as well as genera associated with toxic and non-toxic mushrooms (e.g. *Lacterius* and *Tricholoma*) clustered together. Further research is required to interpret the commonalities in clustering across community interaction combinations, along with the diversity in the ecological functions associated with each genera.

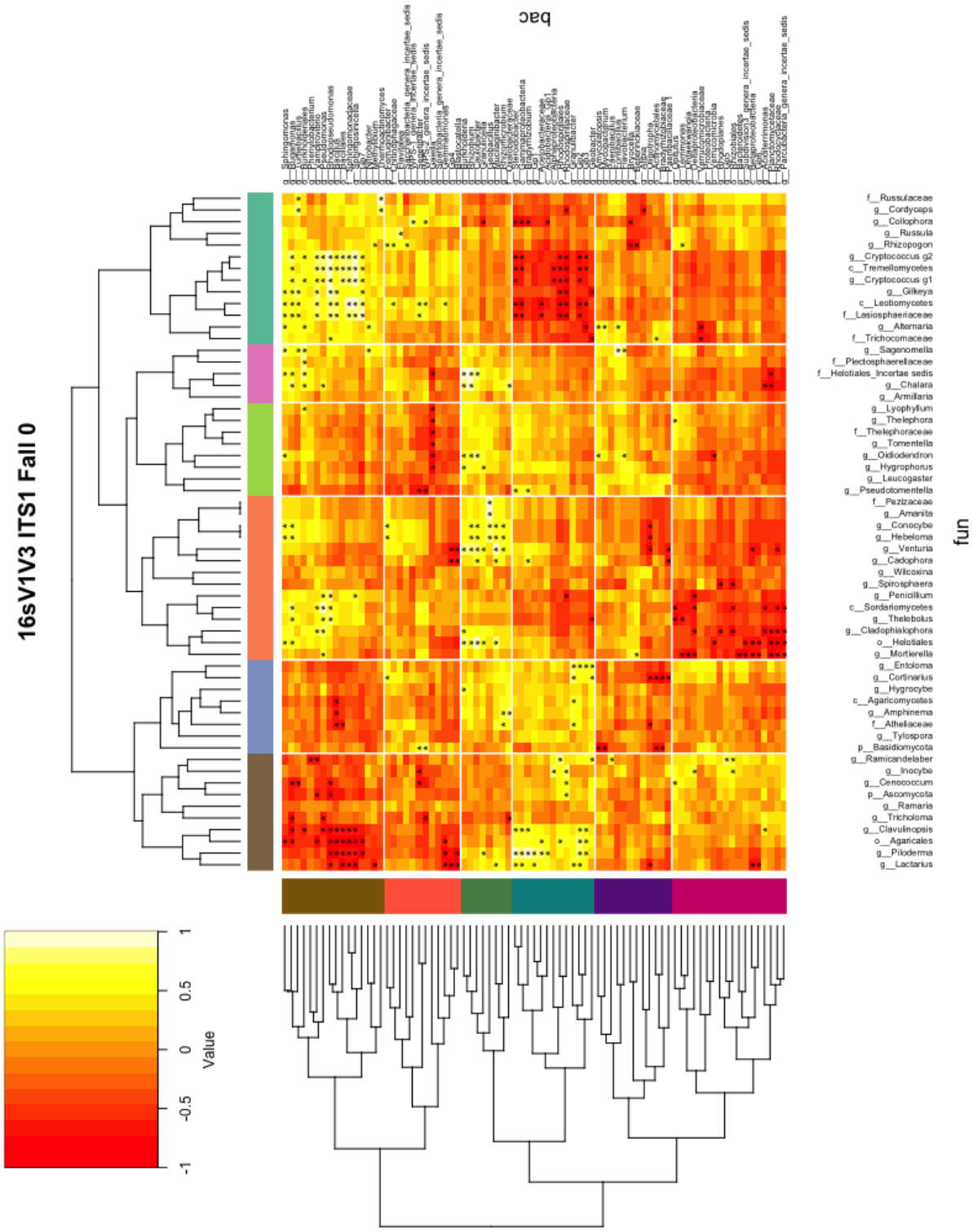
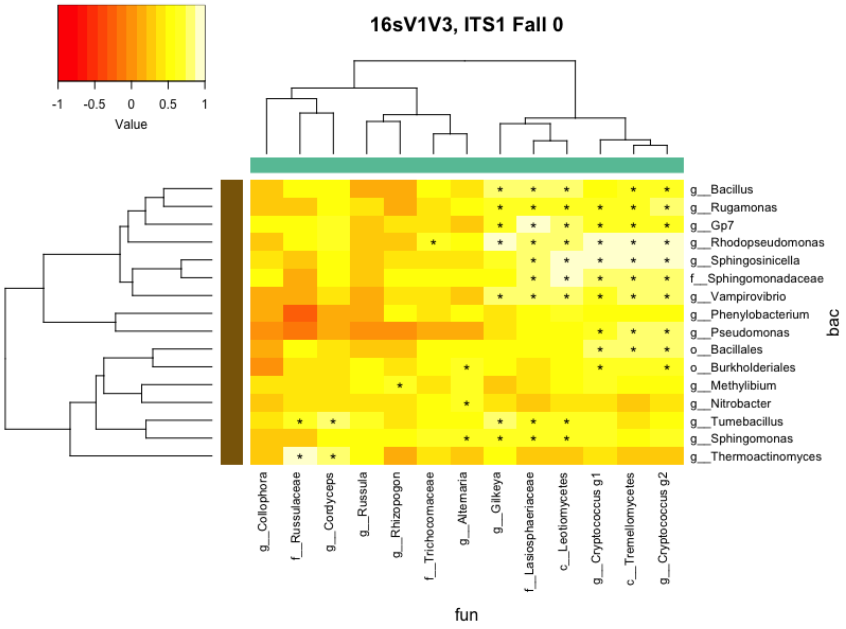
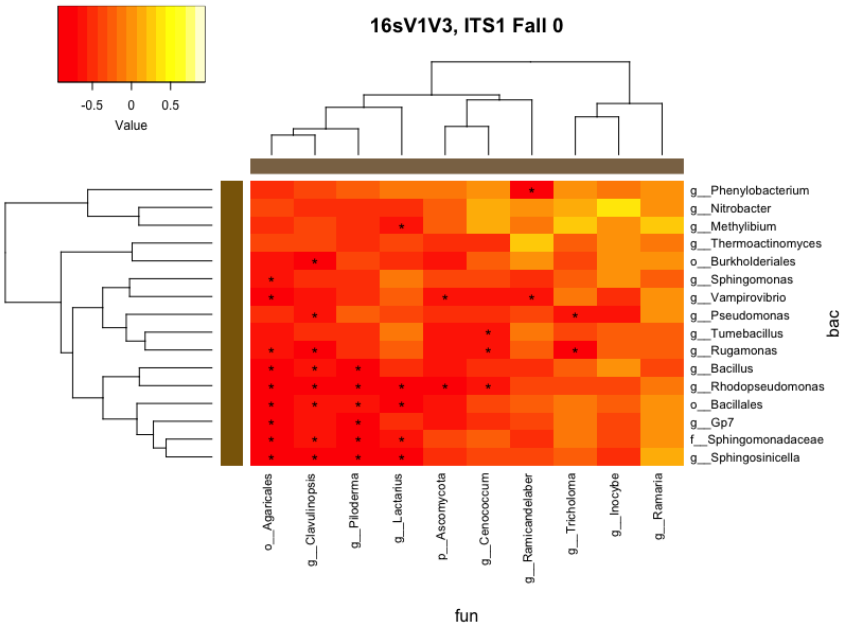


FIGURE 3.7: **Correlations between 16S V1-V3 and ITS1 for Fall at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white, with significant correlations ($p < 0.05$) designated by *.



(A) Cluster of microbes with generally positive correlations



(B) Cluster of microbes with generally negative correlations

FIGURE 3.8: **Specific clusters from Figure 3.7.** These two clusters of microbes were pulled out of Figure 3.7 to allow for further examination of the respective members, since the entire cluster generally had the same correlation. The bacterial (row) and fungal (column) cluster colors are directly related to the cluster colors within the full heatmap, with significant correlations ($p < 0.05$) designated by *.

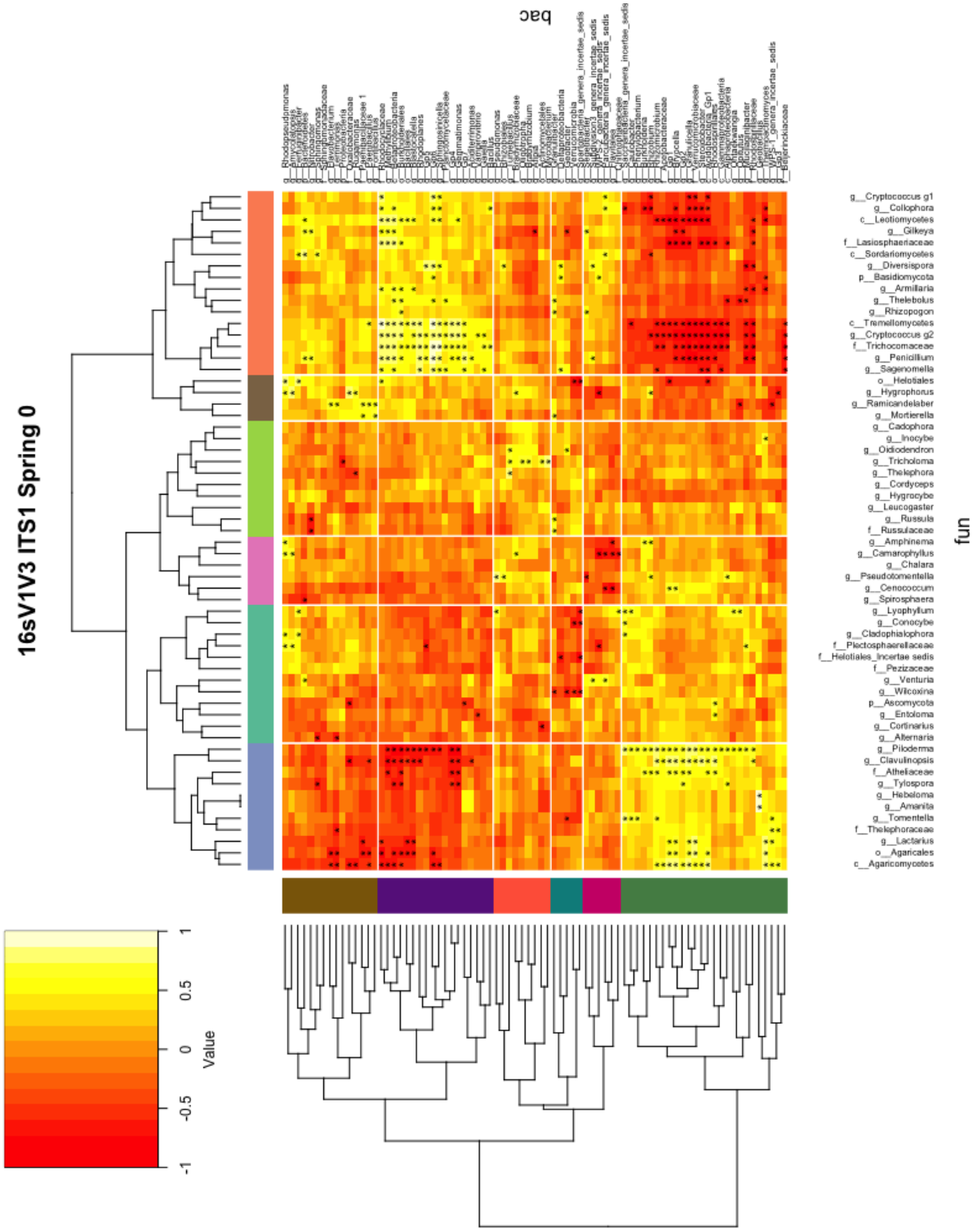
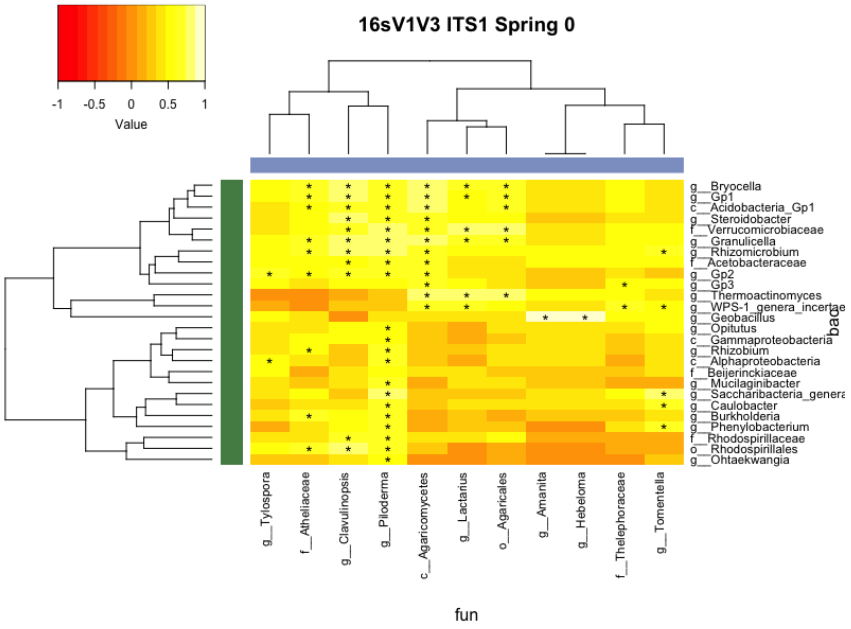
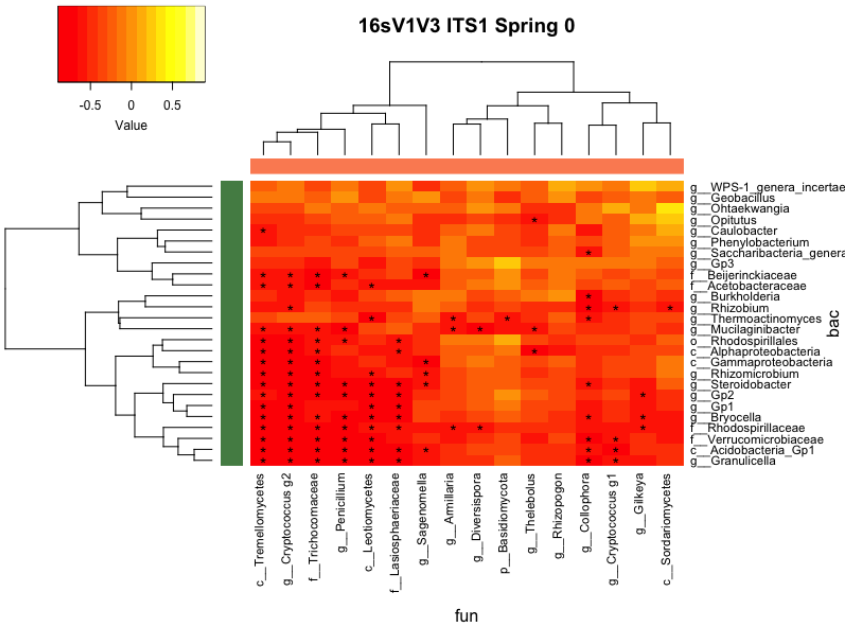


FIGURE 3.9: **Correlations between 16S V1-V3 and ITS1 for Spring at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white, with significant correlations ($p < 0.05$) designated by *.

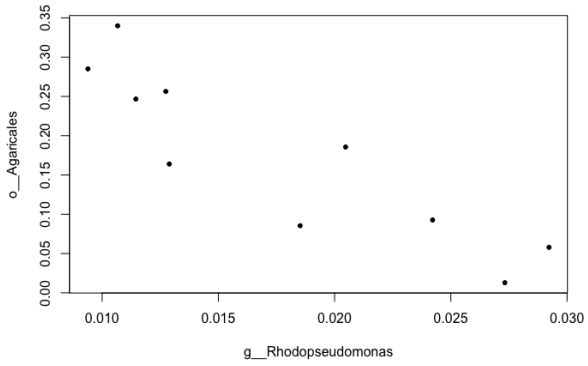


(A) Cluster of microbes with generally positive correlations

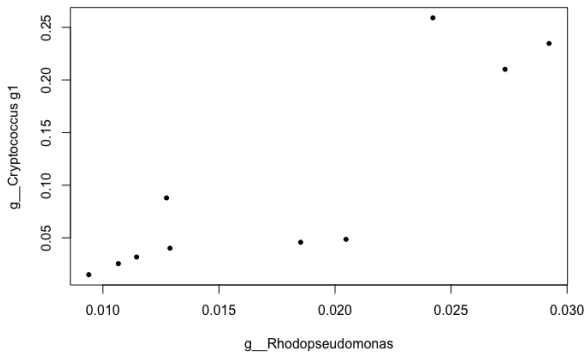


(B) Cluster of microbes with generally negative correlations

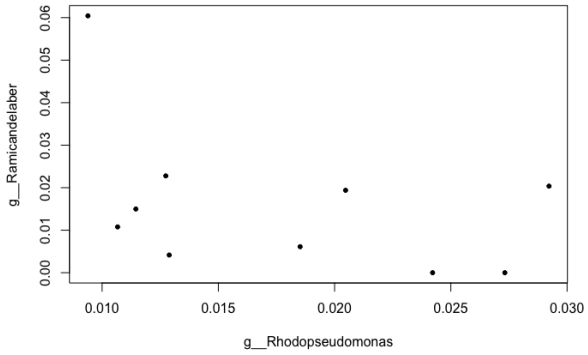
FIGURE 3.10: **Specific clusters from Figure 3.9.** These two clusters of microbes were pulled out of Figure 3.9 to allow for further examination of the respective members, since the entire cluster generally had the same correlation. The bacterial (row) and fungal (column) cluster colors are directly related to the cluster colors within the full heatmap, with significant correlations ($p < 0.05$) designated by *.



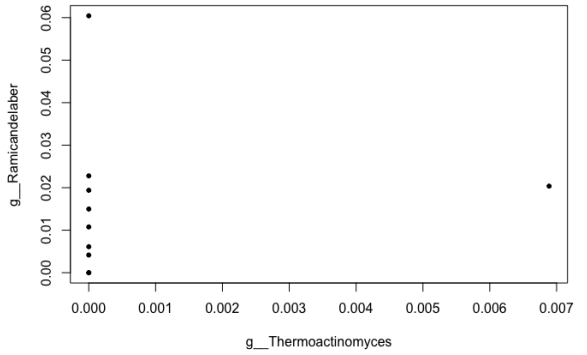
(A) Negative correlation between microbes



(B) Positive correlation between microbes



(C) No, or low, correlation between microbes



(D) No correlation, bacteria not present in most samples

FIGURE 3.11: **Interaction scatterplot correlations.** The four types of correlation influences that were consistent across regions. The plots display the hellinger standardized abundances for the specified bacteria (x) and fungi (y) for the Fall samples at depth 0.

chapter 4

Conclusion

Soil communities are well studied around the globe and are used to determine the health of the area. As the climate continues to change the health of forests are increasing at risk for disease, which are often diverse and difficult to detect. For example, *Armillaria* root disease is a diverse fungal disease which effects a broad range of trees and shrubs and can be found in the forests in northern Idaho, specifically Priest River Experimental Forest (PREF) (Worrall, 2007). Recent studies showed that while the diseases are diverse, the general microbial communities are also complex from region to region. Therefore, samples from the PREF were analyzed to study the environmental correlates on the bacterial and fungal communities, as well as microbial interactions.

After the communities were cleaned, 14 of the 96 samples were removed from the LSU region, therefore this community was removed from further analysis. The quantitative and qualitative methods compared by Lozupone *et al.* (2007) were the weighted and unweighted UniFrac distances; these methods are implemented in QIIME and mothur (two OTU based methods) and can be used to test for phylogenetic differences, or cluster samples. Therefore, the significant results from the exploration of standardization methods implies that the standardization method does affect conclusions about the evenness measure, as well as phylogenetic differences. However, it is infeasible to test all standardization methods for every dataset, thus for future analyses consistency is key.

Analysis indicated that this study was unable to detect statistically significant interactions between season and moisture level, and depth within the sites samples. In each community there were microbes that appeared to be significantly associated with changes in the predictor variables. Within the two fungal communities, there were taxa that resulted in a significant increase in odds when compared to the baseline taxa (*Ganoderma*); while this increase suggests a strong association, further examination of the 95% confidence intervals suggested uncertainty within the data, since each interval spanned across 1. This implies that more data may be needed to determine if the

environmental variables examined do explain variation within taxa present. When the variables were tested using the permutational analysis of variance, there was significance in three out of the four communities, implying that the variables do explain some of the variation within the compositions. Although group had a significant association, the significance was a result from the moisture level (or habitat types series) rather than season. This implies that correlates associated with season (e.g. temperature) may be too similar between fall and spring to explain variation.

Due to limitations in knowledge and understanding of the ecological importance of identified microbes, the clusters with members that all interacted similarly were briefly investigated. Although there were no apparent patterns to how the microbes clustered, there were some commonalities among the members. For instance, fungi associated with edible coral were clustered and had similar interactions with the bacterial community. However, determining the specific species would be beneficial to the understanding of the community, as well as management plans when investigating a genera with species known to be detrimental to the host, such as *Armillaria*. *Bacillus* is a well studied bacterial microbe that has species known to be utilized in controlling insects, as well as others to be medically important. Therefore, the speciation of pivotal genera and those commonly clustered across the different variables is required as the next step to advance ecological interpretations of the community interactions.

In addition to the limitation of ecological knowledge about the specific functions of the microbes identified, there are other limitations about the study that should be considered for further research. First, although the sampling method and data collection was carefully designed and well executed, the data was only collected for a year long pilot study by Dr. Ross-Davis and her team. For this reason, information was lost from a location during the cleaning process when samples had to be removed. Therefore, in order to make definite conclusions about the environmental impacts on the compositional structure, replicate samples would need to be collected. Secondly, there are two fungal databases within the RDP classifier, for the purposes of this study only the warcup fungal ITS databased was implemented. Applying the other database (UNITE fungal ITS) may result in different community compositions, thus leading to distinct conclusions, due to differences in the microbes present in the database. A

limitation embedded within the RDP classifier is classifications are only to the genus level, causing information to be lost since there can be a wide range of ecological functions within genera. Thirdly, since soil communities are a well studied topic there are many different methods of analysis. Thus, while this study attempted to implement most of these methods, at least in an exploratory capacity, subjectivity based on understanding was employed when choosing which method to use for downstream analysis. Lastly, modifying differential expression analyses (e.g. gene ontology) accordingly may provide further insights into the community interactions.

Overall, this study sought to investigate the soil communities within the PREF, as well as provide an alternative procedure to identify microbial communities within soil samples, through the implementation of a reference database. While there was no comparison of the communities identified through the OTU based methods, it has been shown that for this dataset a reference based approach sufficed. However, it is recommended to examine the fungal communities identified using the UNITE fungal ITS database within the RDP classifier; as a result a better understanding of variable microbes within the samples will be obtained. Additionally, since there were only 12 independent samples for each season and depth, it is recommended for future research to increase the sample size, or take replicate samples at each current plot. After further ecological analysis of the microbes identified our grasp of the microbial structure between communities would be enhanced, thus leading to a more thorough understanding of the community interactions.

Despite the aforementioned limitations, this exploration provided valuable knowledge into the delicate balance of soil microbial communities within the PREF. Through this analysis differences that were observed previously on multiple sites were tested within a single site. This study also provided support for the use of metagenomic sequencing to analyze soil microbes. Specifically, using multiple amplicon regions for each community validated the community identified, as well as provided a more complete snapshot of the captured community. Thus providing significant evidence for the continued use of each region in future analyses. While the environmental differences based on specific taxa were inconclusive, the analysis of variance did show moisture level and depth had a significant association with differences in the soil

composition. Therefore, with replicate sampling these specific taxa differences may become more apparent. The loss of classification to the species level masked some of the associations between bacterial and fungal microbes, as well as within community associations. Nevertheless, the exploration of the associations provided a baseline for the balance between fungi and bacteria. In conclusion, the communities below the forest floor are vast and complex, and this exploration only scratched the surface of soil microbial communities within the PREF.

Bibliography

Agresti A. 2002. Logit Models for Multinomial Responses, in Categorical Data Analysis. 2 ed., John Wiley & Sons, Inc., Hoboken, NJ.

Anderson M.J., Ellingsen K.E., and McArdle B.H. 2006. Multivariate dispersion as a measure of beta diversity. Ecology Letters 9:683–693.

Brown S.P., Rigdon-Huss A.R., and Jumpponen A. 2014. Analyses of ITS and LSU gene regions provide congruent results on fungal community responses. Fungal Ecology 9:65–68.

Coince A., Caël O., Bach C., Lengellé J., Cruaud C., Gavory F., Morin E., Murat C., Marçais B., and Buée M. 2013. Below-ground fine-scale distribution and soil versus fine root detection of fungal and soil oomycete communities in a French beech forest. Fungal Ecology 6:223–235.

Dematheis F., Kurtz B., Vidal S., and Smalla K. 2012. Microbial communities associated with the larval gut and eggs of the Western corn rootworm. PloS one 7:e44685.

Fierer N., Leff J.W., Adams B.J., Nielsen U.N., Bates S.T., Lauber C.L., Owens S., Gilbert J.A., Wall D.H., and Caporaso J.G. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences 109:21390–21395.

Hagle S.K. 2008. Management guide for armillaria root disease. `http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5187208.pdf`.

He X.Y., Wang K.L., Zhang W., Chen Z.H., Zhu Y.G., and Chen H.S. 2008. Positive correlation between soil bacterial metabolic and plant species diversity and bacterial and fungal diversity in a vegetation succession on Karst. Plant and Soil 307:123–134.

Hill M.O. 1973. Diversity and Evenness: A Unifying Notation and Its Consequences. Ecology 54:427–432.

Jost L. 2007. Partitioning Diversity into Independent Alpha and Beta Components. Ecology 88:2427–2439.

Kassambara A. and Mundt F. 2016. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.3.

Kirk J.L., Beaudette L.A., Hart M., Moutoglis P., Klironomos J.N., Lee H., and Trevors J.T. 2004. Methods of studying soil microbial diversity. Journal of Microbiological Methods 58:169–188.

Legendre P. and Gallagher E.D. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280.

Lilleskov E.A., Bruns T.D., Horton T.R., Taylor D., and Grogan P. 2004. Detection of forest stand-level spatial structure in ectomycorrhizal fungal communities. FEMS microbiology ecology 49:319–32.

Lozupone C.A., Hamady M., Kelley S.T., and Knight R. 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology 73:1576–1585.

McCune B., Grace J.B., and Urban D.L. 2002. Analysis of Ecological Communities. MjM Software design, Gelenden Beac, OR.

Navas-Molina J.A., Peralta-Sánchez J.M., González A., McMurdie P.J., Vázquez-Baeza Y., Xu Z., Ursell L.K., Lauber C., Zhou H., Song S.J., Huntley J., Ackermann G.L., Berg-Lyons D., Holmes S., Caporaso J.G., and Knight R. 2013. Advancing our understanding of the human microbiome using QIIME. Methods in enzymology 531:371–444.

Oksanen J. 1983. Ordination of Boreal Heath-like Vegetation with Principal Component Analysis, Correspondence Analysis and Multidimensional Scaling. Vegetatio 52:181–189.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., and Wagner H. 2016. vegan: Community Ecology Package. R package version 2.3-5.

Porter T.M., Shokralla S., Baird D., Brian Golding G., and Hajibabaei M. 2016. Ribosomal DNA and plastid markers used to sample fungal and plant communities from wetland soils reveals complementary biotas. PLoS ONE 11:1–18.

Rencher A.C. and Christensen W.F. 2012. Methods of Multivariate Analysis. 3 ed., John Wiley & Sons, Inc., Hoboken, NJ.

Ripley B. and Venables W. 2016. Feed-Forward Neural Networks and Multinomial Log-Linear Models. R package version 7.3-12.

Schloss P.D., Jenior M.L., Koumpouras C.C., Westcott S.L., and Highlander S.K. 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. PeerJ 4:e1869.

Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Lesniewski R.A., Oakley B.B., Parks D.H., Robinson C.J., Sahl J.W., Stres B., Thallinger G.G., Van Horn D.J., and Weber C.F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537–41.

Sharpton T.J. 2014. An introduction to the analysis of shotgun metagenomic data. Frontiers in plant science 5:209.

Smets W., Leff J.W., Bradford M.A., McCulley R.L., Lebeer S., and Fierer N. 2016. A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. Soil Biology and Biochemistry 96:145–151.

Sun Y., Cai Y., Huse S.M., Knight R., Farmerie W.G., Wang X., and Mai V. 2012. A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. Briefings in bioinformatics 13:107–21.

Talbot J.M., Bruns T.D., Taylor J.W., Smith D.P., Branco S., Glassman S.I., Erlandson S., Vilgalys R., Liao H.L., Smith M.E., and Peay K.G. 2014. Endemism and functional convergence across the North American soil mycobiome. Proceedings of the National Academy of Sciences 111:6341–6346.

Toljander J.F., Eberhardt U., Toljander Y.K., Paul L.R., and Taylor A.F.S. 2006. Species composition of an ectomycorrhizal fungal community along a local nutrient gradient in a boreal forest. New Phytologist 170:873–884.

Torsvik V., Goksøyr J., and Daae F.L. 1990. High diversity in DNA of soil bacteria. Applied and environmental microbiology 56:782–7.

Urbanová M., Šnajdr J., and Baldrian P. 2015. Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. Soil Biology and Biochemistry 84:53–64.

Větrovský T. and Baldrian P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PloS one 8:e57923.

Vilo C. and Dong Q. 2012. Evaluation of the RDP Classifier Accuracy Using 16S rRNA Gene Variable Regions. Metagenomics 1:1–5.

Wang J.T., Zheng Y.M., Hu H.W., Zhang L.M., Li J., and He J.Z. 2015. Soil pH determines the alpha diversity but not beta diversity of soil fungal community along altitude in a typical Tibetan forest ecosystem. Journal of Soils and Sediments 15:1224–1232.

Whittaker R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs 30:279–338.

Worrall J. 2007. Armillaria rood disease. `http://www.forestpathology.org/dis_arm.html`.

Yee T.W. 2016. VGAM: Vector Generalized Linear and Additive Models. R package version 1.0-2.

Young J.M., Weyrich L.S., and Cooper A. 2014. Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. Forensic Science International: Genetics 0:176–184.

Žifčáková L., Větrovský T., Howe A., and Baldrian P. 2016. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. Environmental Microbiology 18:288–301.

APPENDIX A

Community Interaction Heatmaps

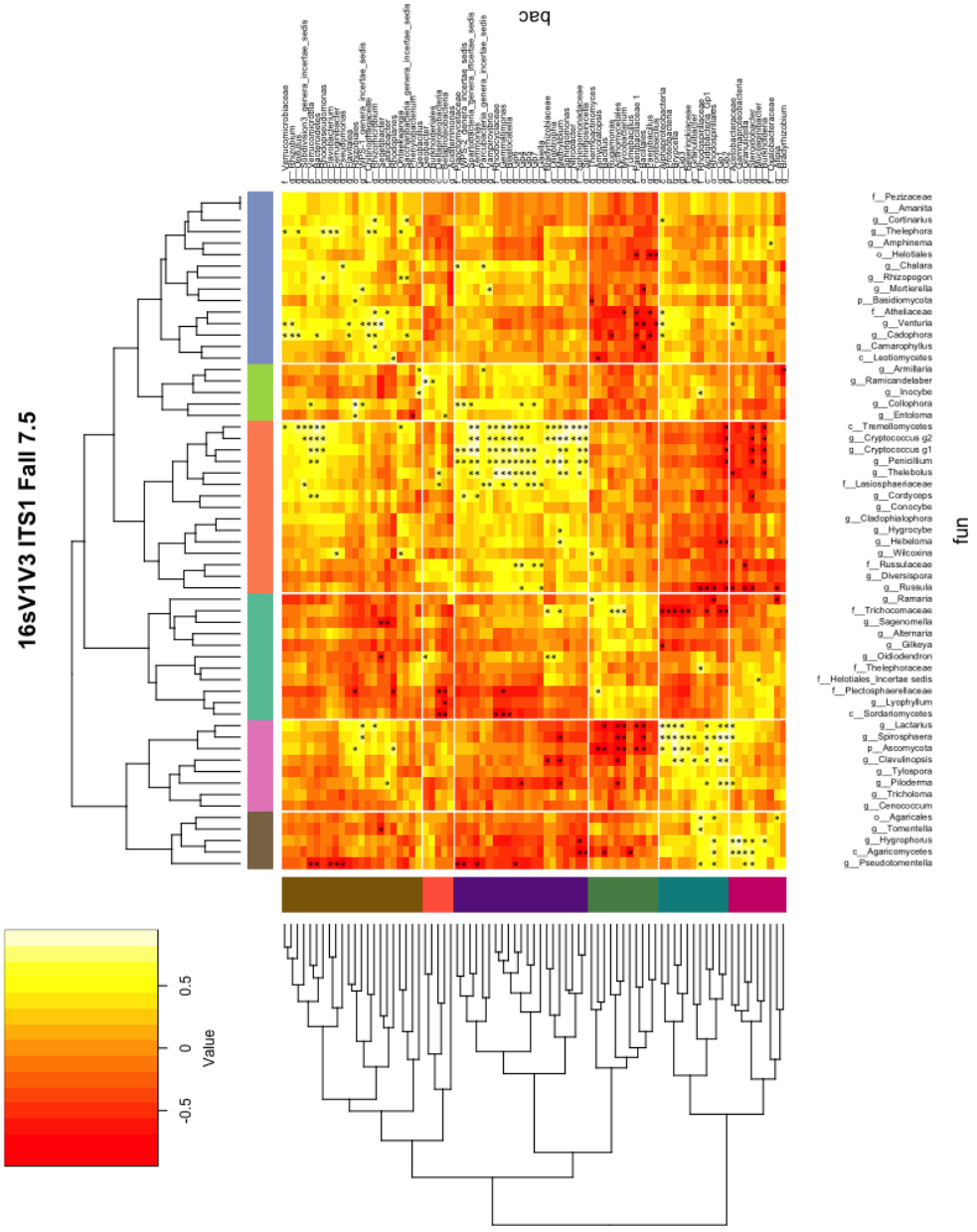


FIGURE A.1: **Correlations between 16S V1-V3 and ITS1 for Fall at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

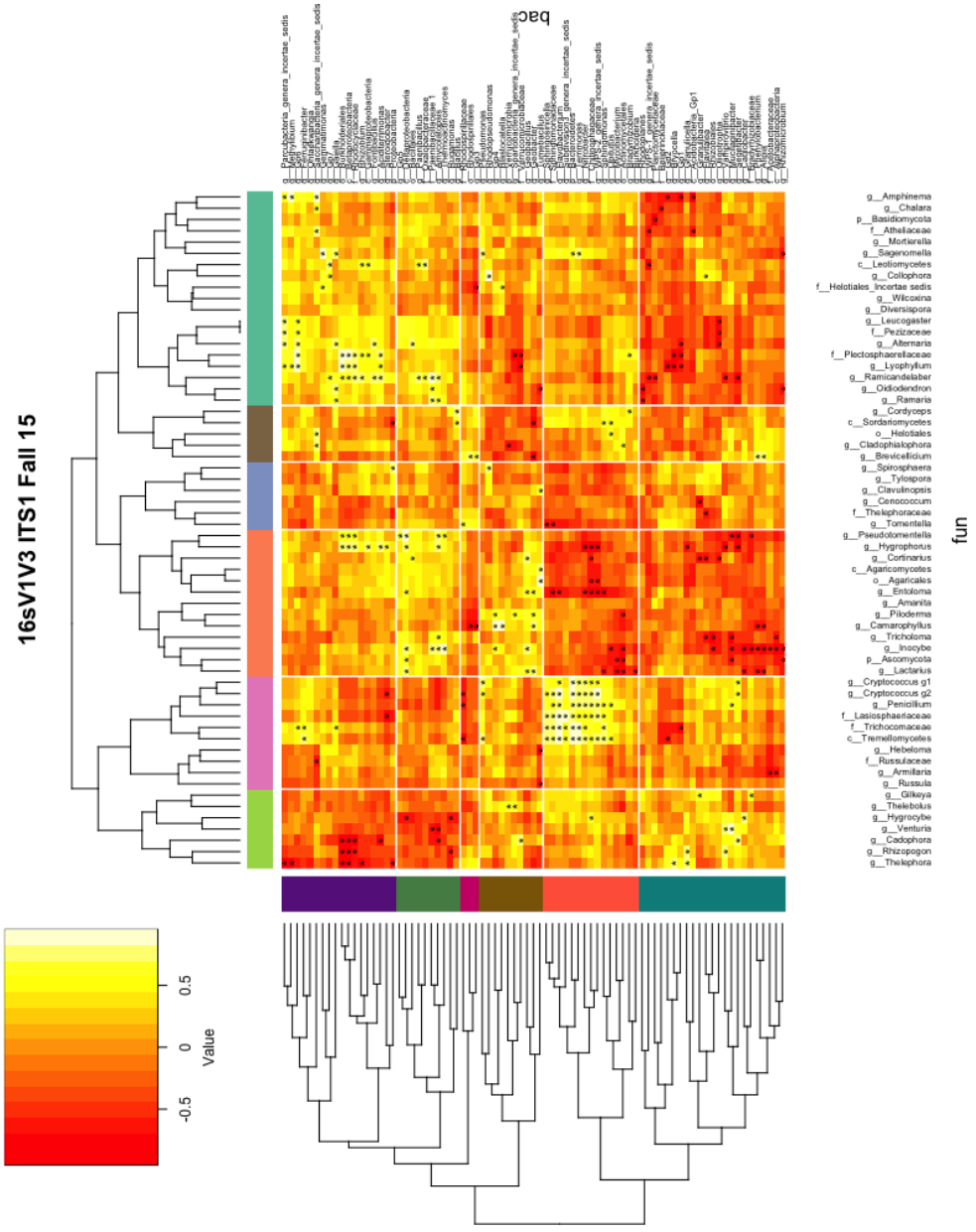


FIGURE A.2: **Correlations between 16S V1-V3 and ITS1 Fall at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

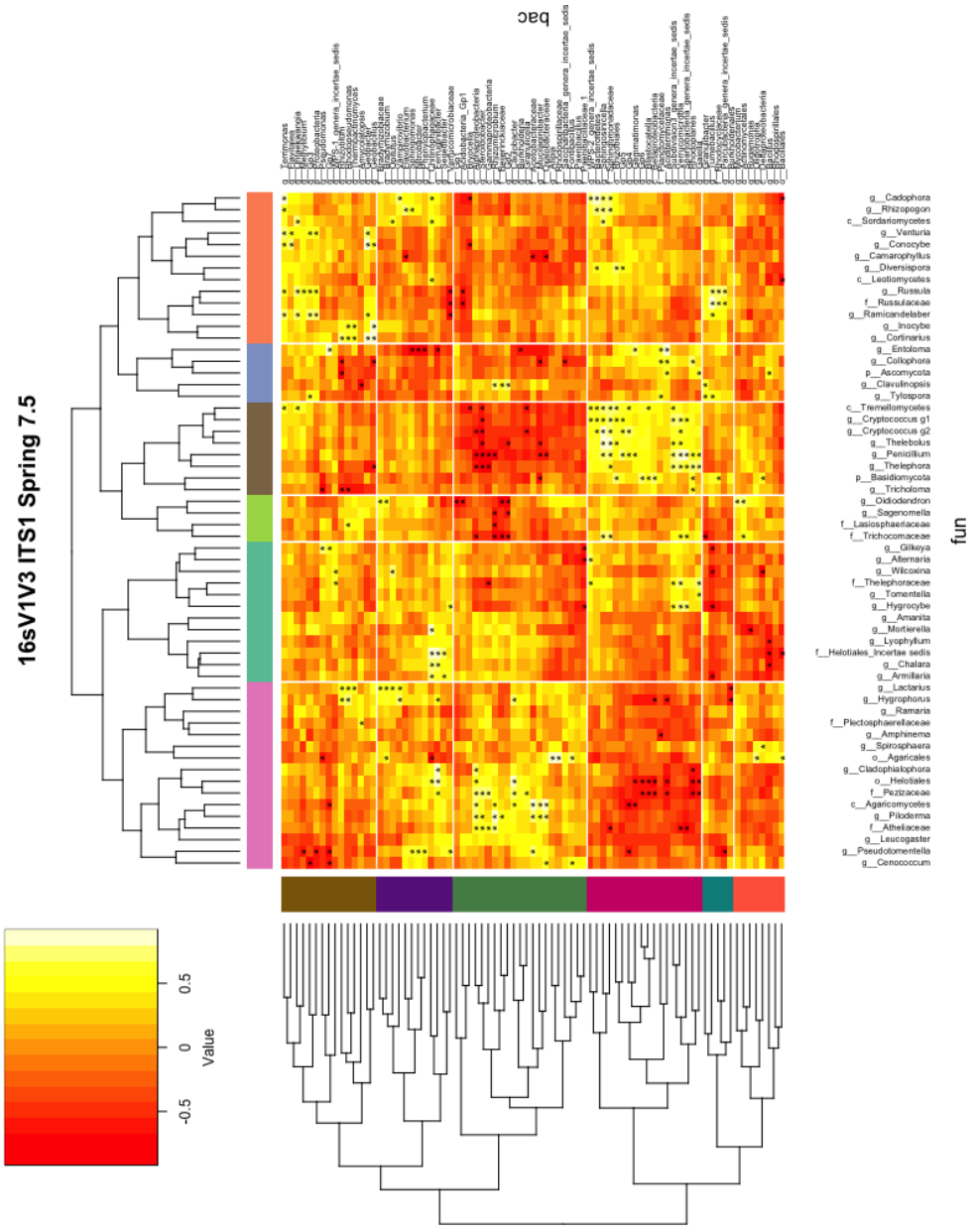


FIGURE A.3: **Correlations between 16S V1-V3 and ITS1 for Spring at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

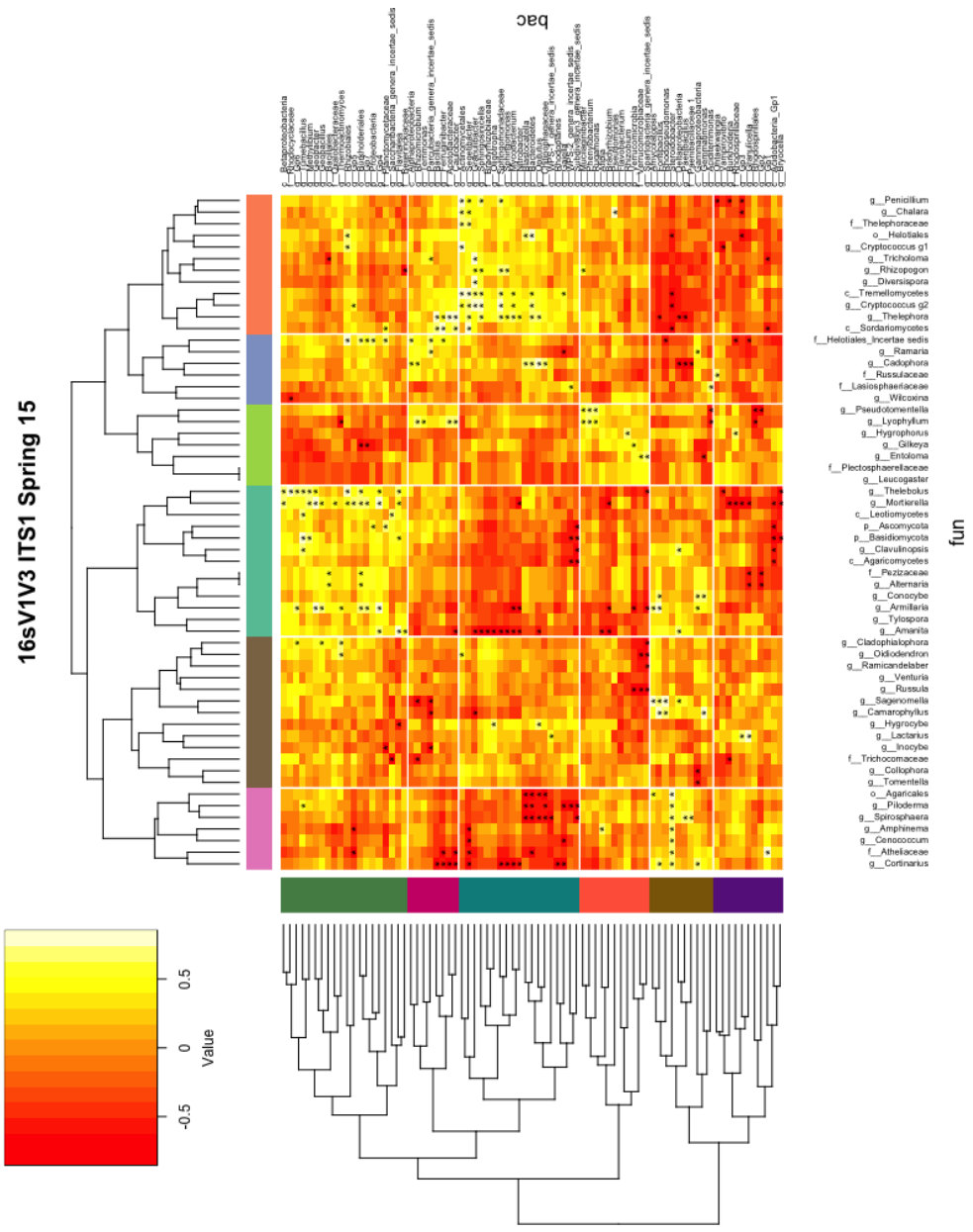


FIGURE A.4: **Correlations between 16S V1-V3 and ITS1 for Spring at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

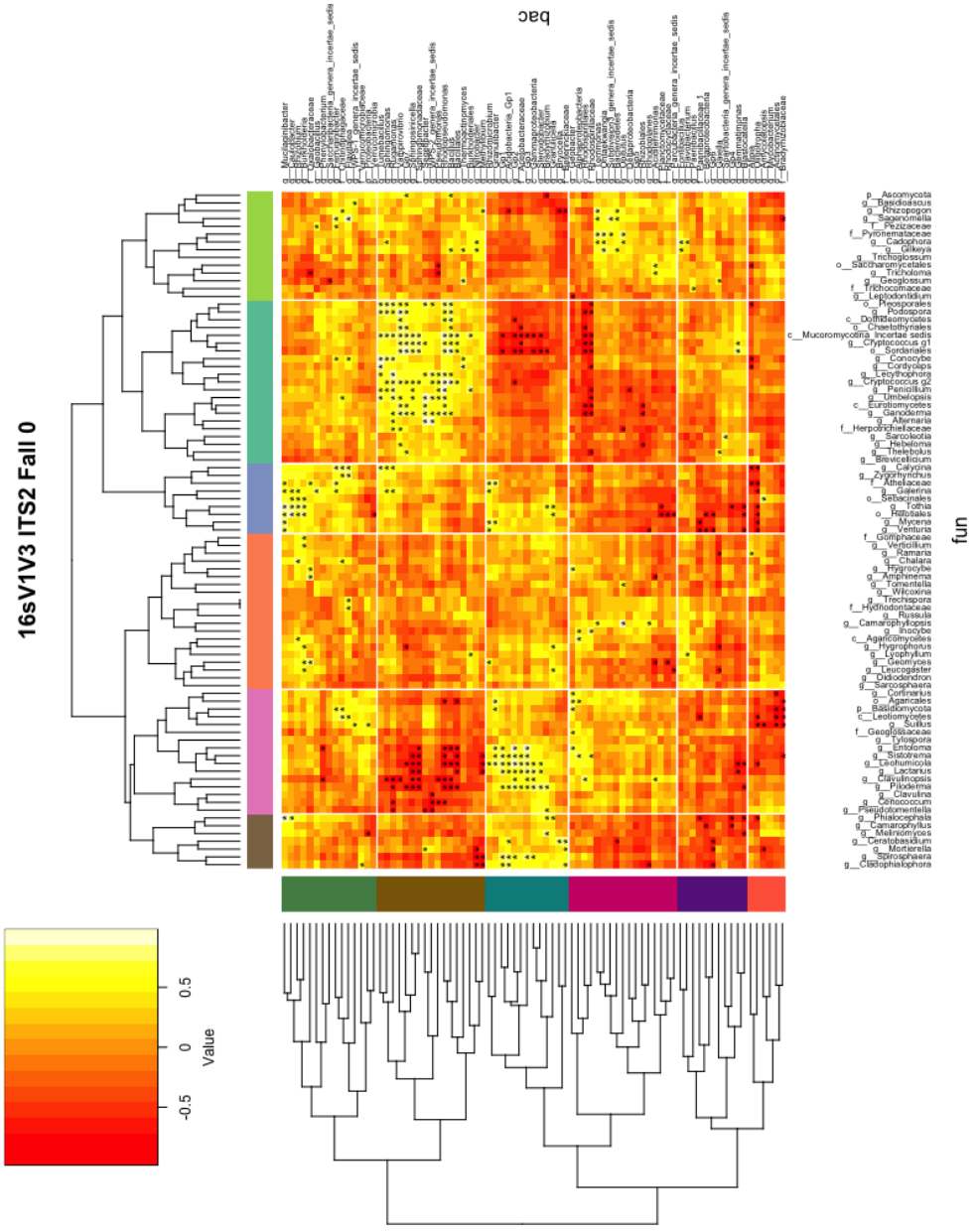


FIGURE A.5: **Correlations between 16S V1-V3 and ITS2 for Fall at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

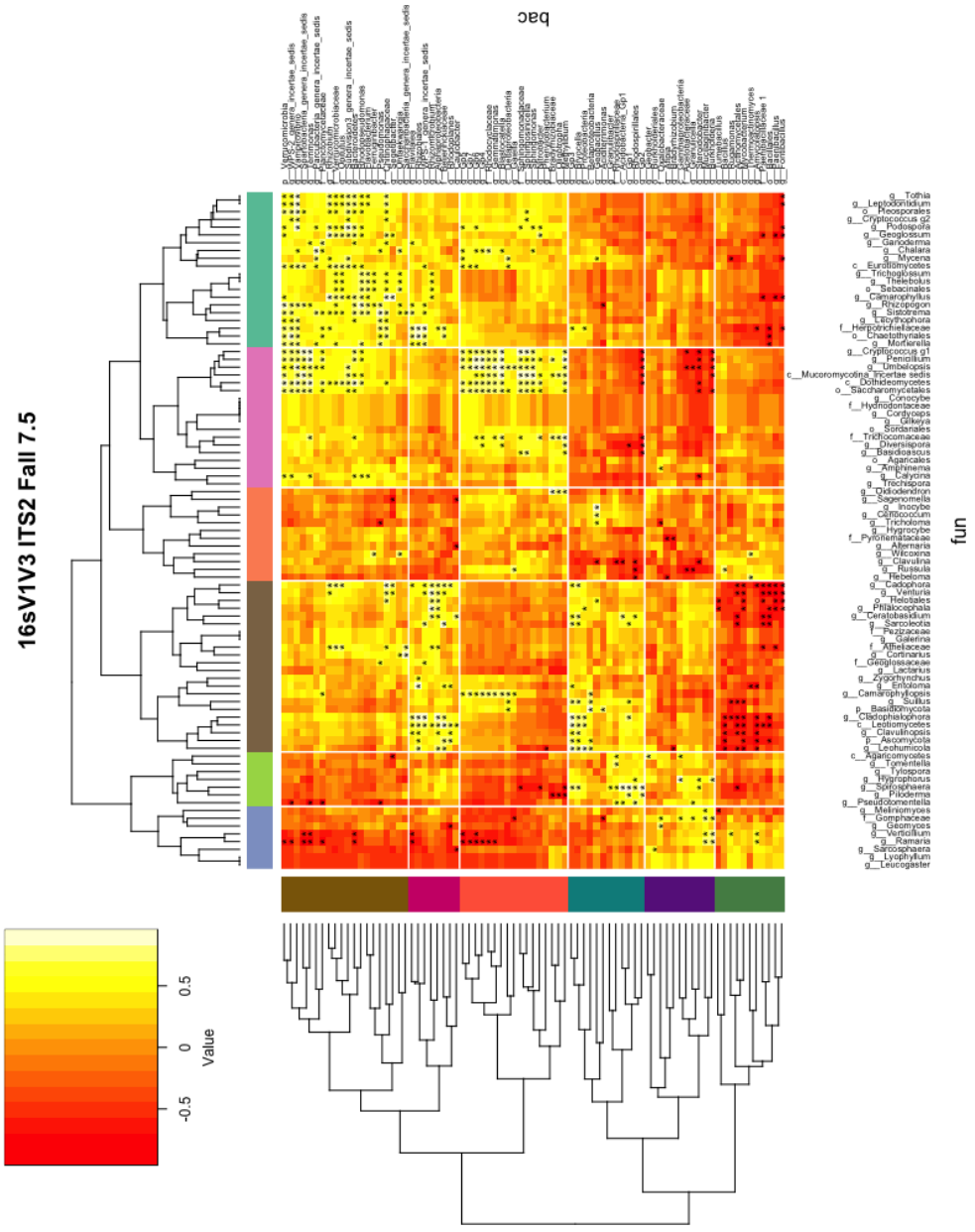


FIGURE A.6: **Correlations between 16S V1-V3 and ITS2 for Fall at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

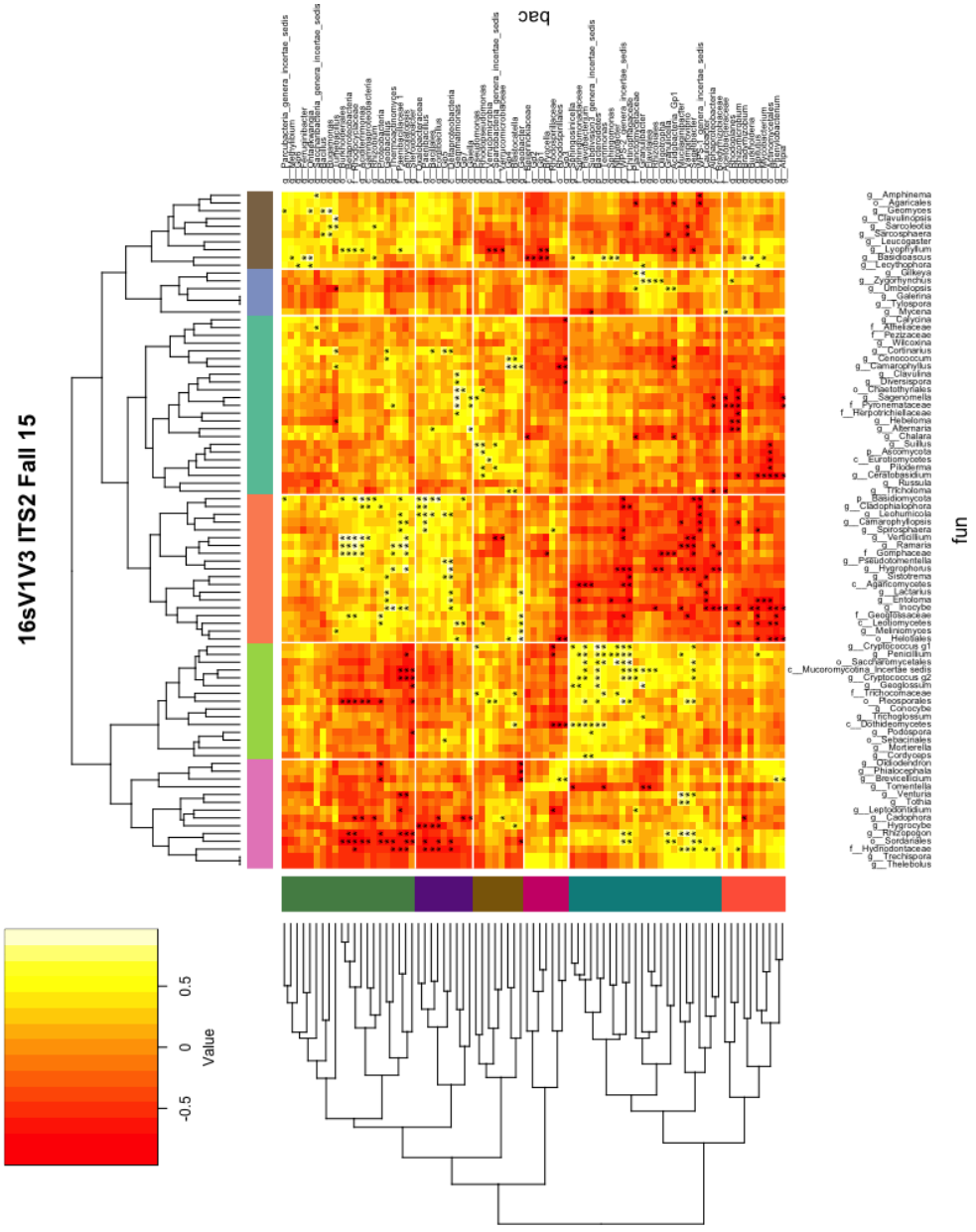


FIGURE A.7: **Correlations between 16S V1-V3 and ITS2 Fall at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

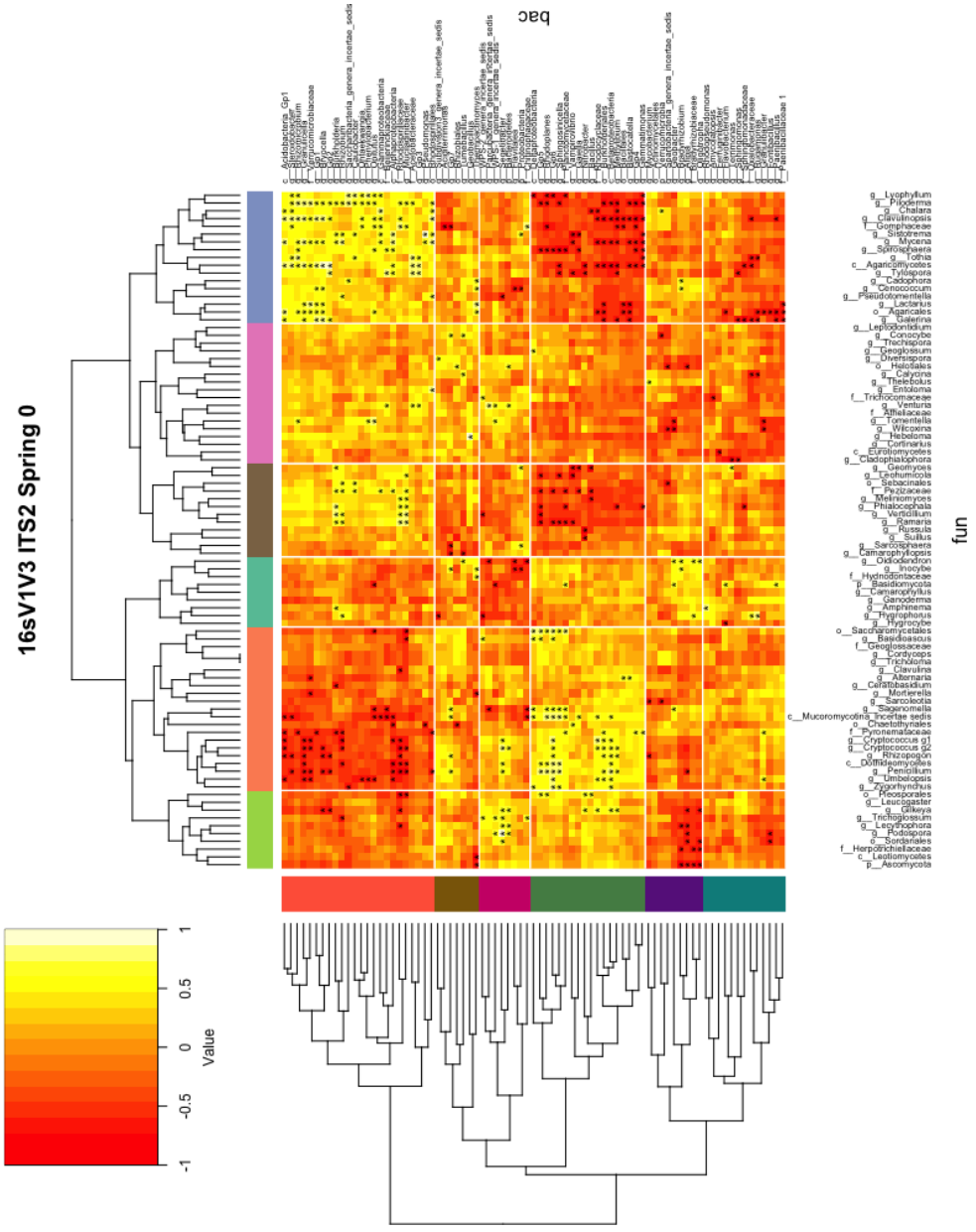


FIGURE A.8: **Correlations between 16S V1-V3 and ITS2 for Spring at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

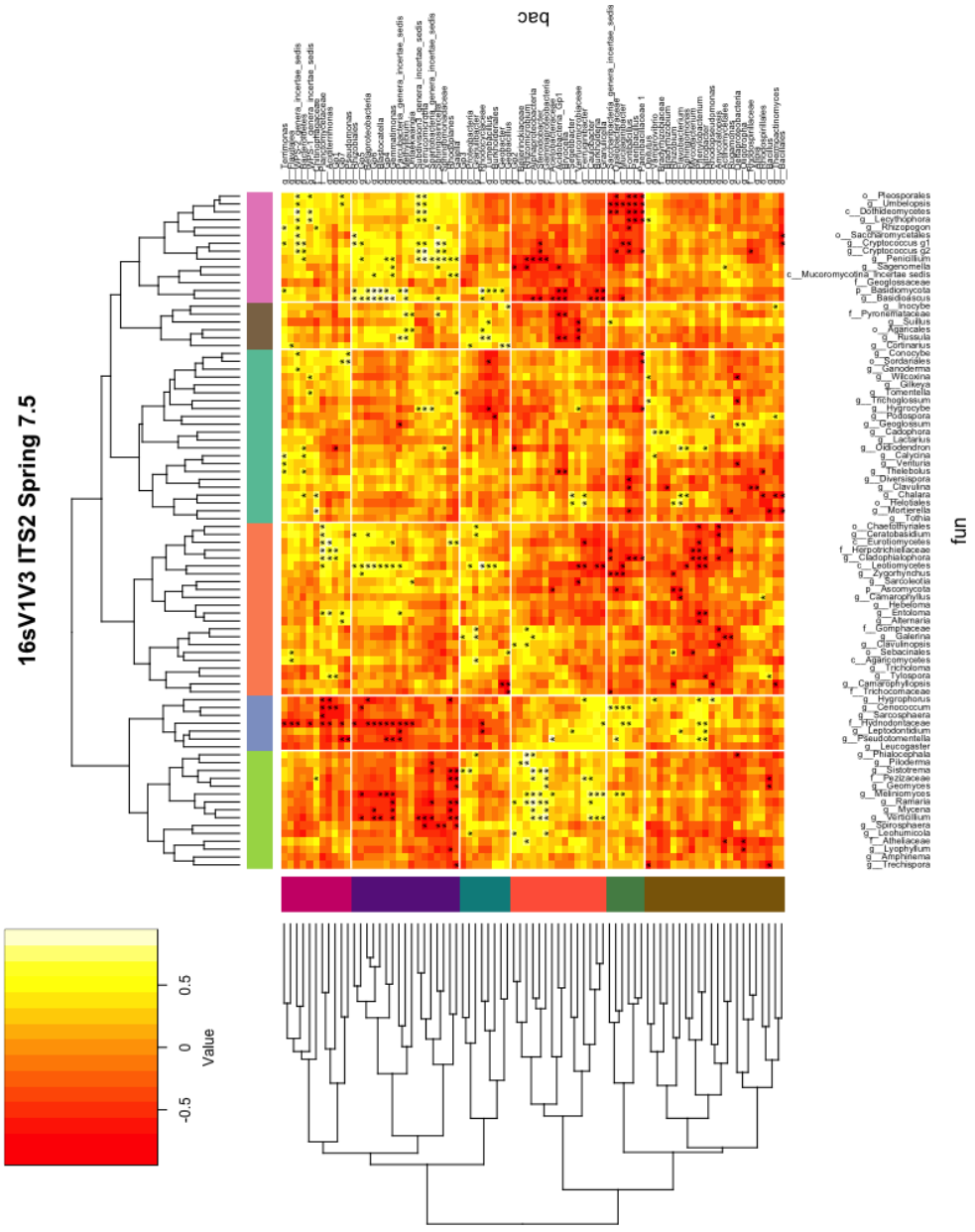


FIGURE A.9: **Correlations between 16S V1-V3 and ITS2 for Spring at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

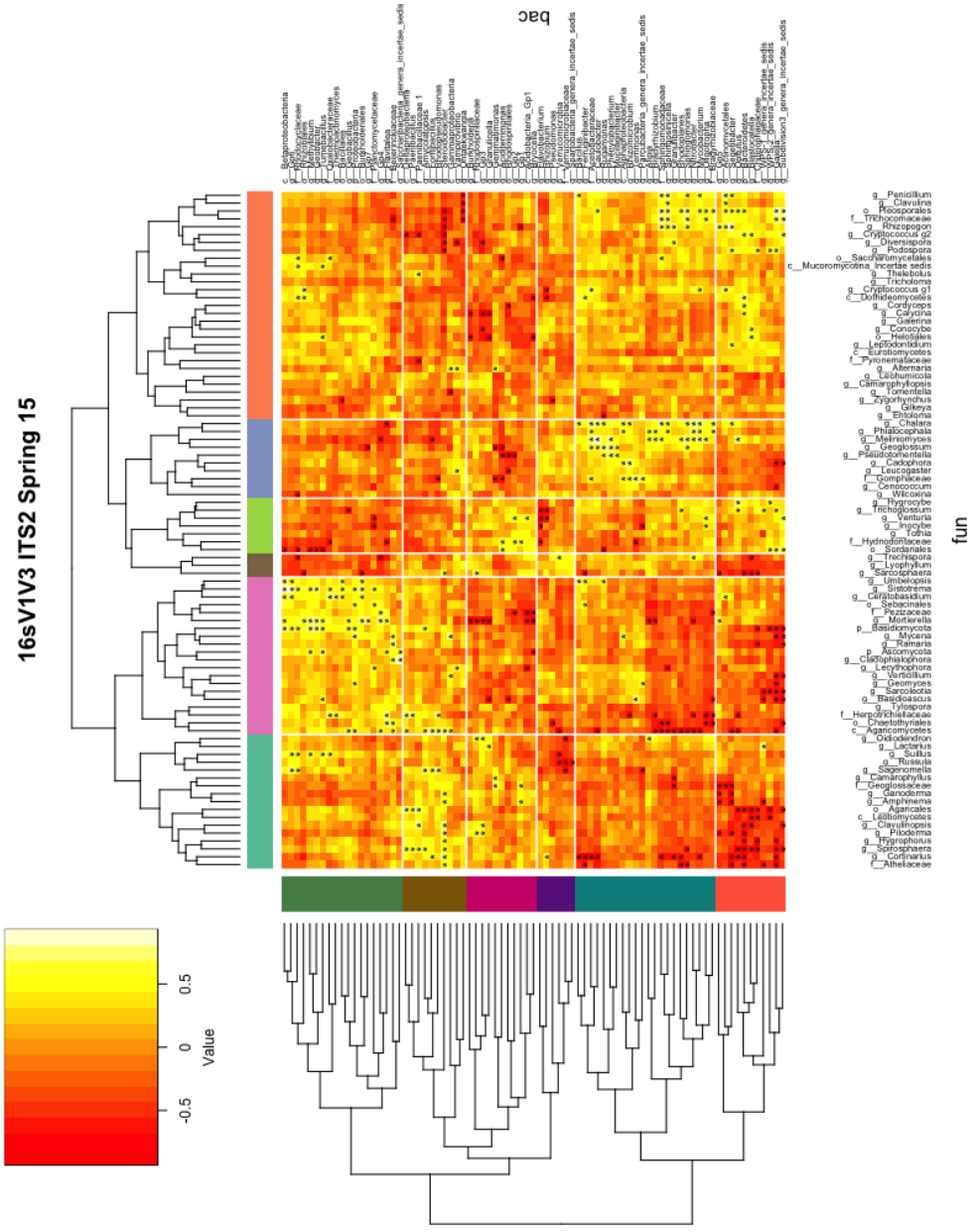


FIGURE A.10: **Correlations between 16S V1-V3 and ITS2 for Spring at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

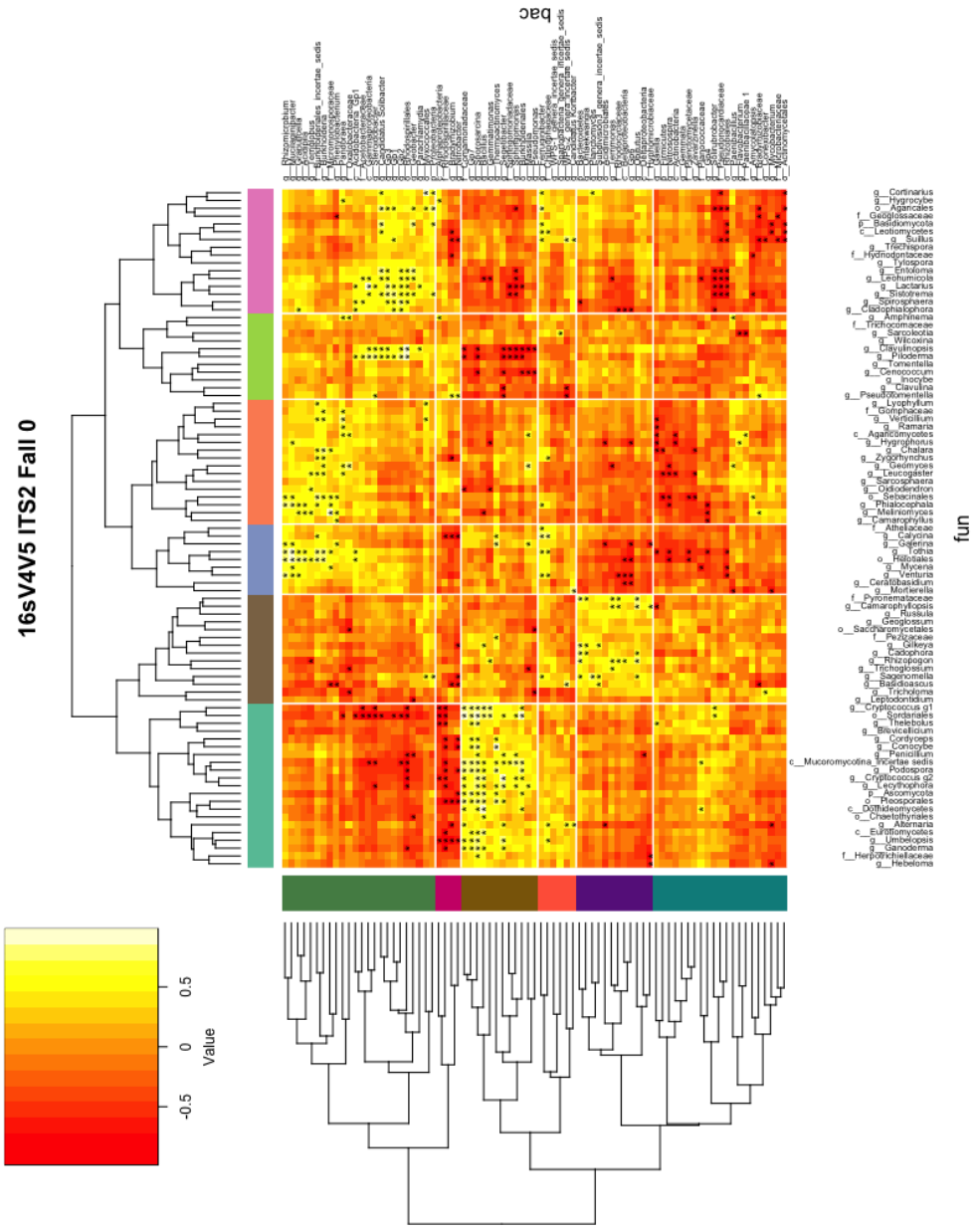


FIGURE A.11: **Correlations between 16S V4-V5 and ITS1 for Fall at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

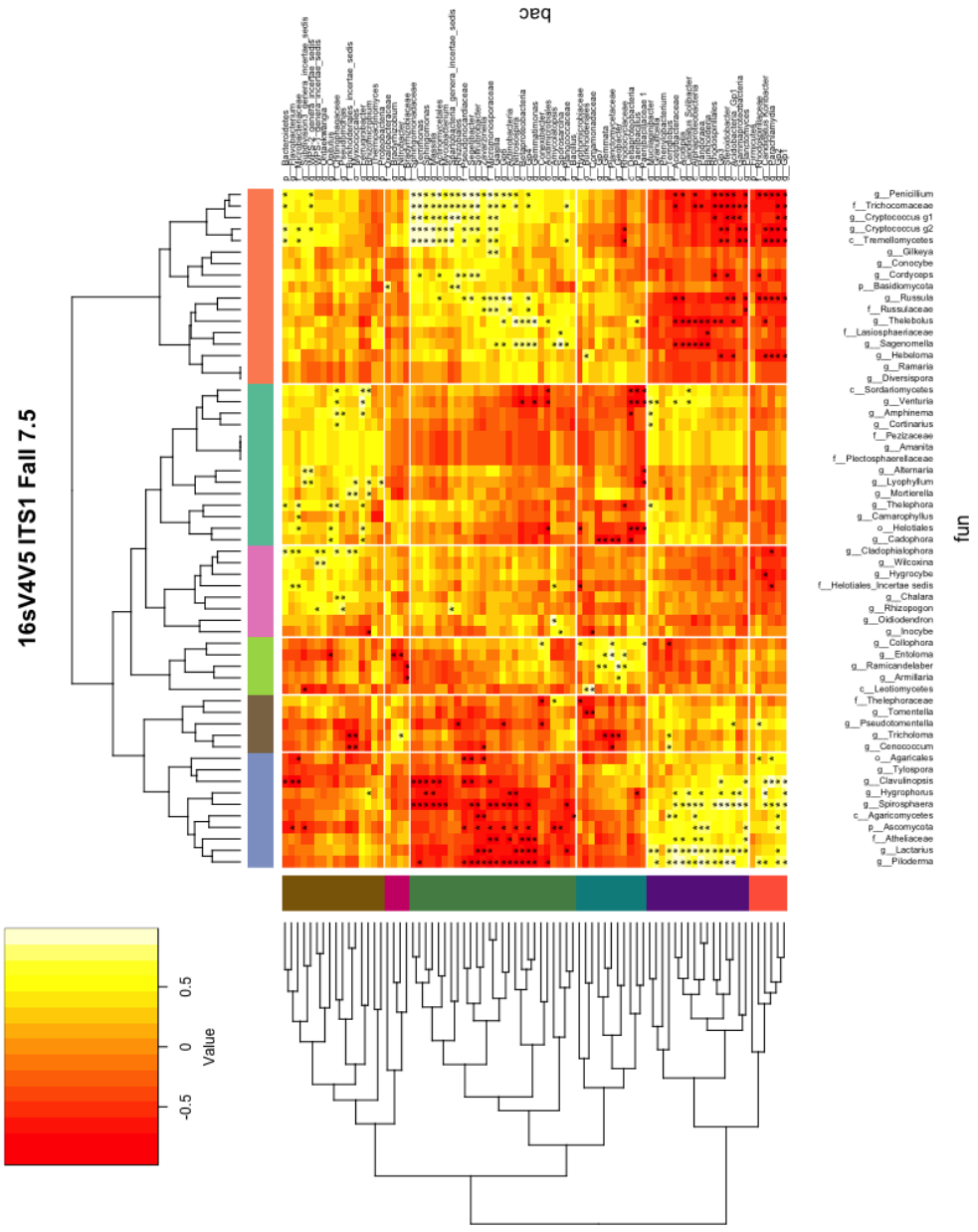


FIGURE A.12: **Correlations between 16S V4-V5 and ITS1 for Fall at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

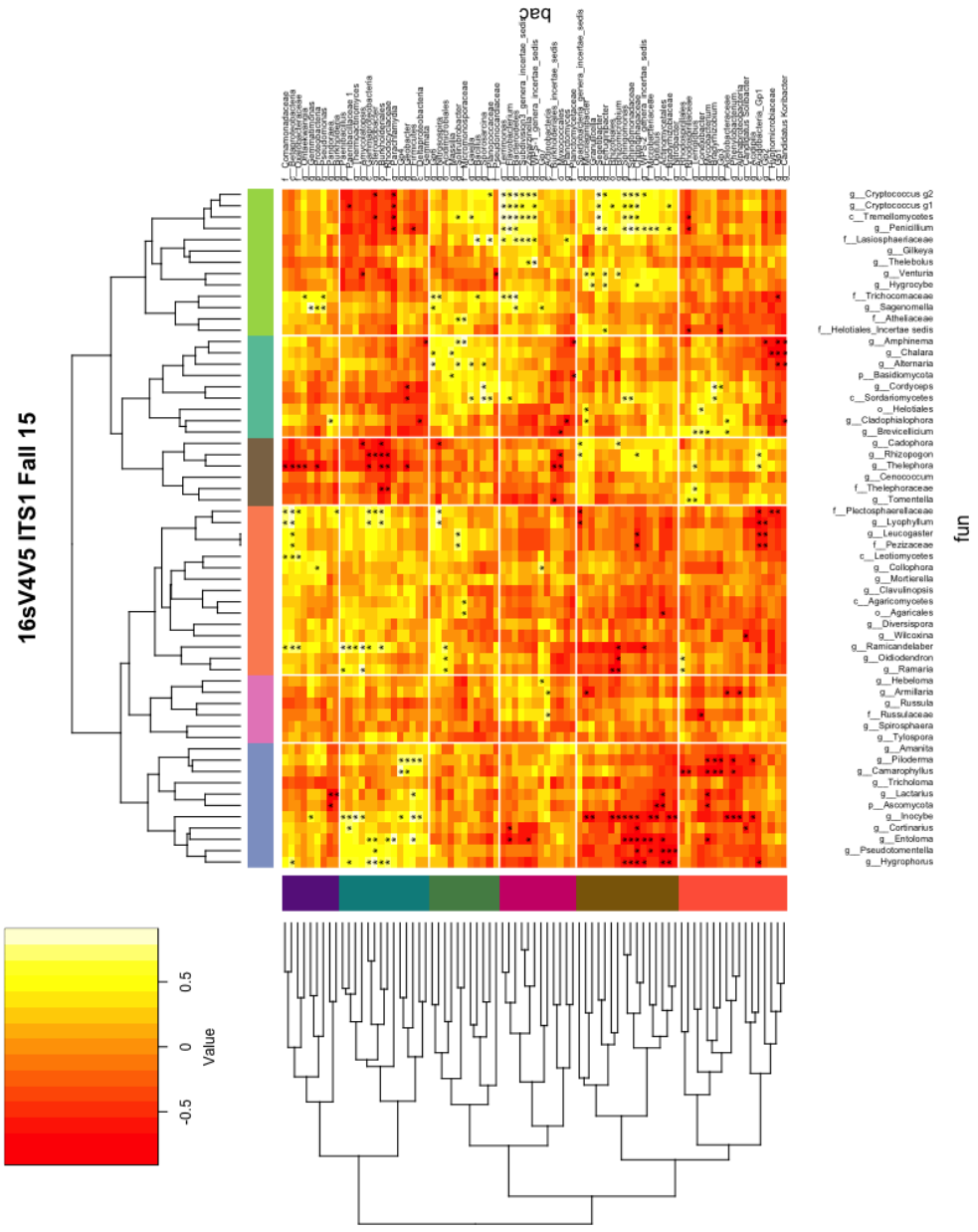


FIGURE A.13: **Correlations between 16S V4-V5 and ITS1 Fall at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

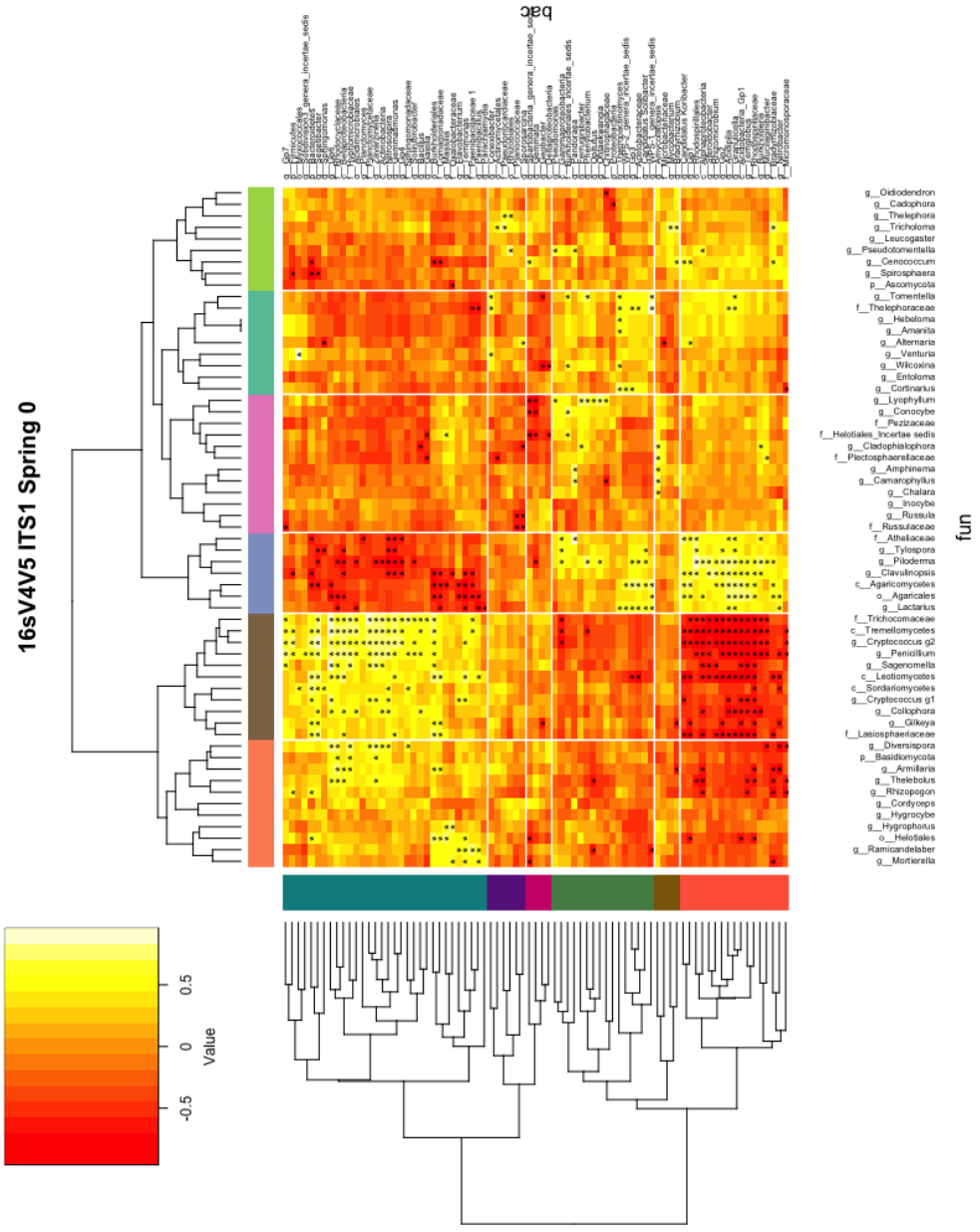


FIGURE A.14: **Correlations between 16S V4-V5 and ITS1 for Spring at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

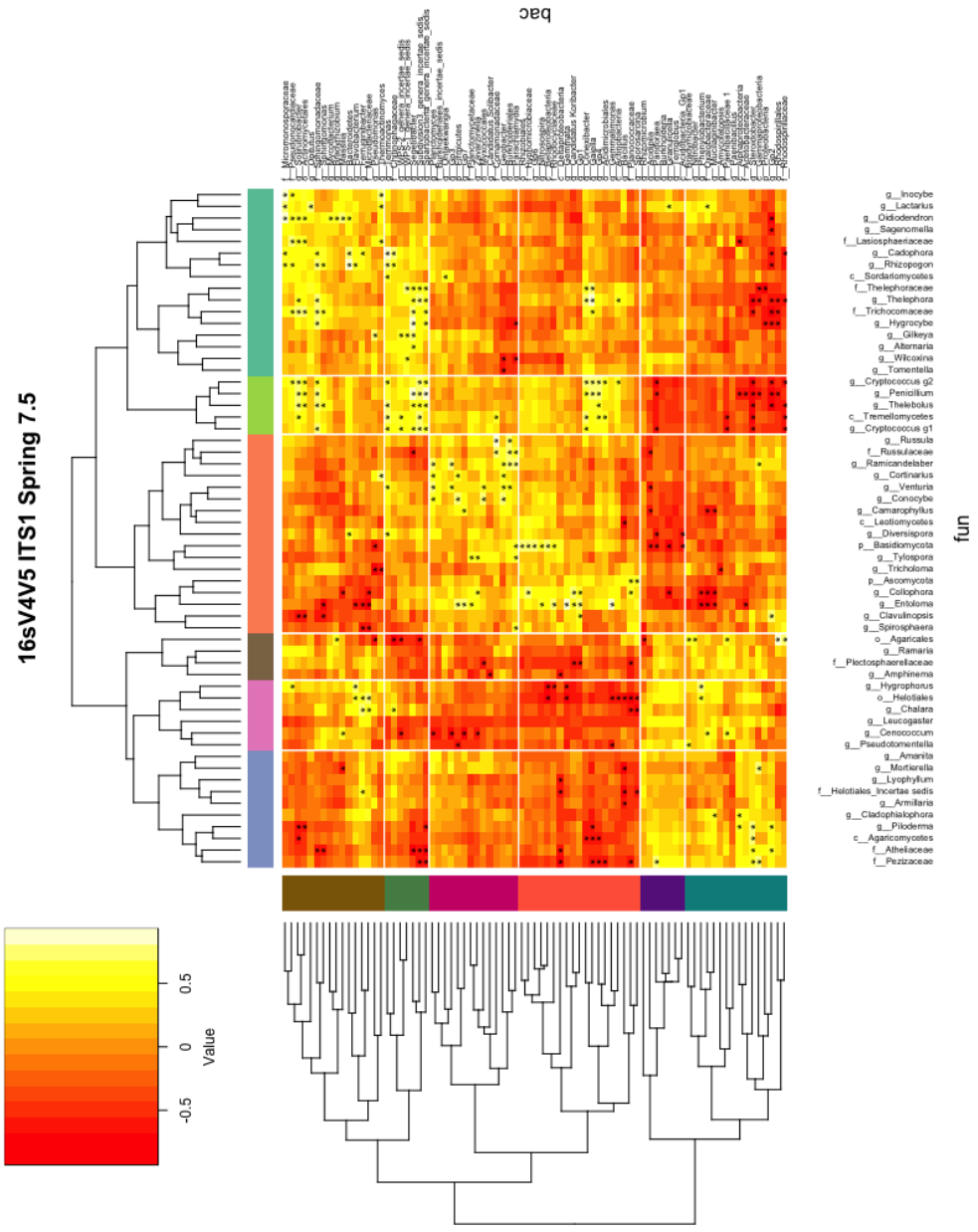


FIGURE A.15: **Correlations between 16S V4-V5 and ITS1 for Spring at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

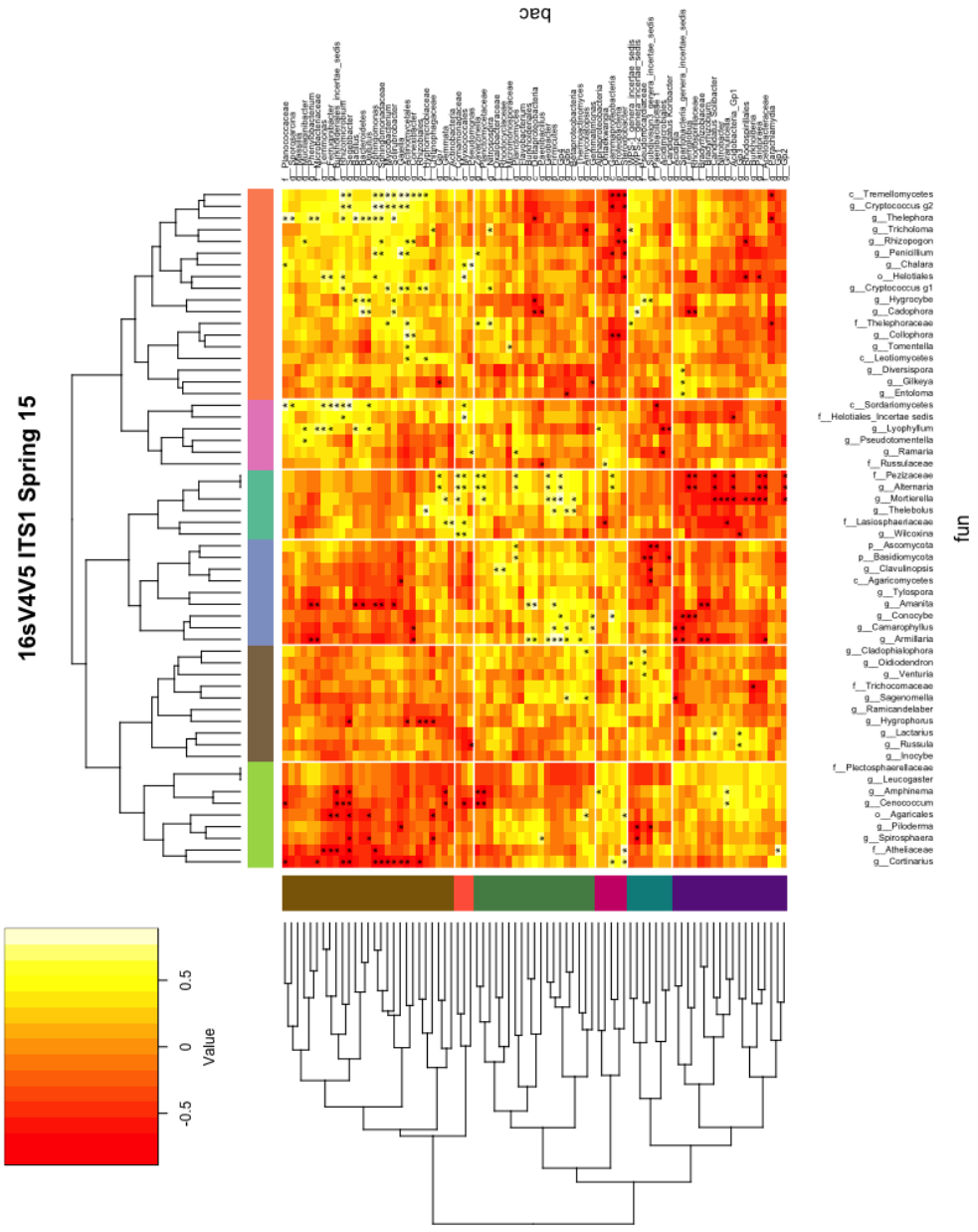


FIGURE A.16: **Correlations between 16S V4-V5 and ITS1 for Spring at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

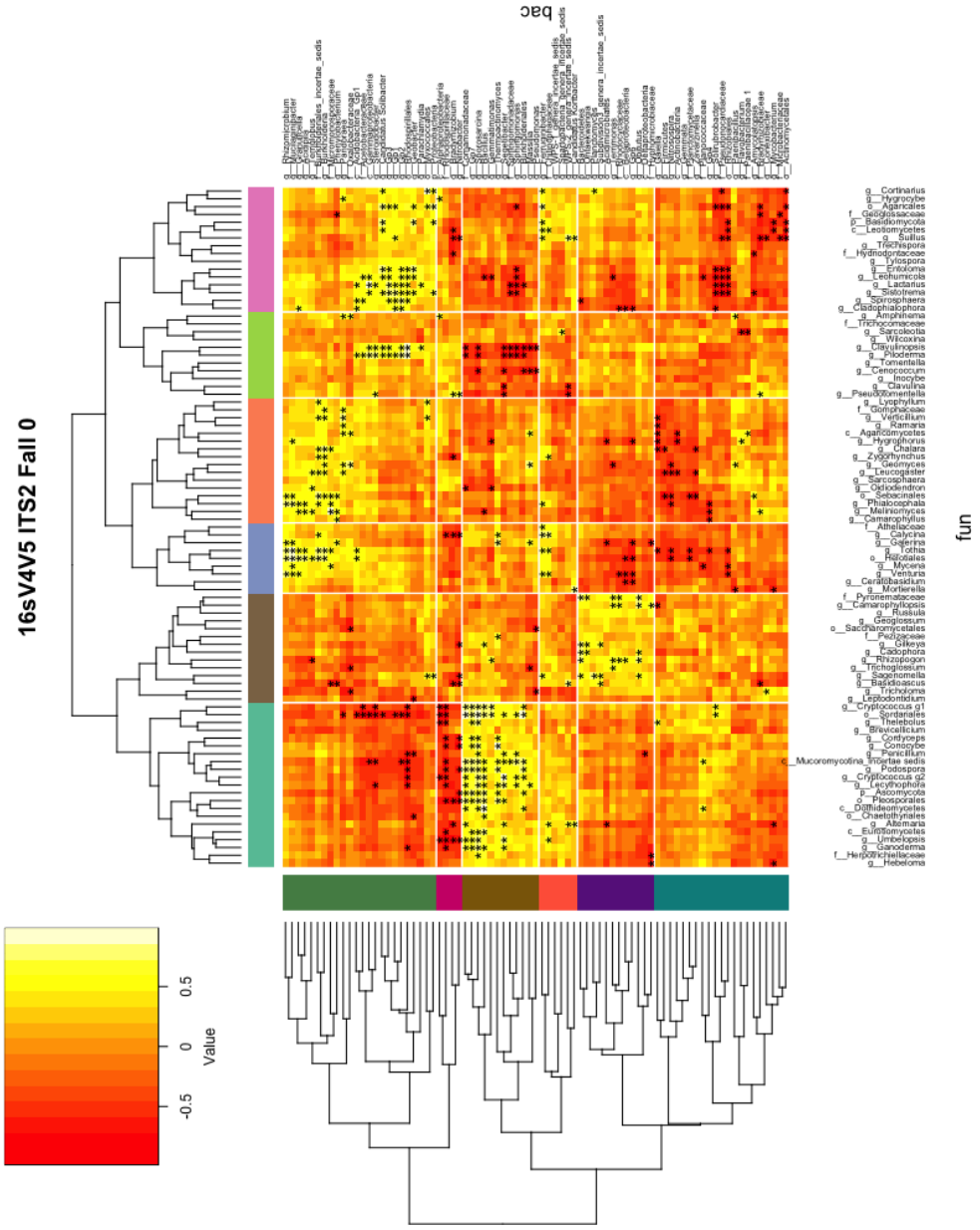


FIGURE A.17: **Correlations between 16S V4-V5 and ITS2 for Fall at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

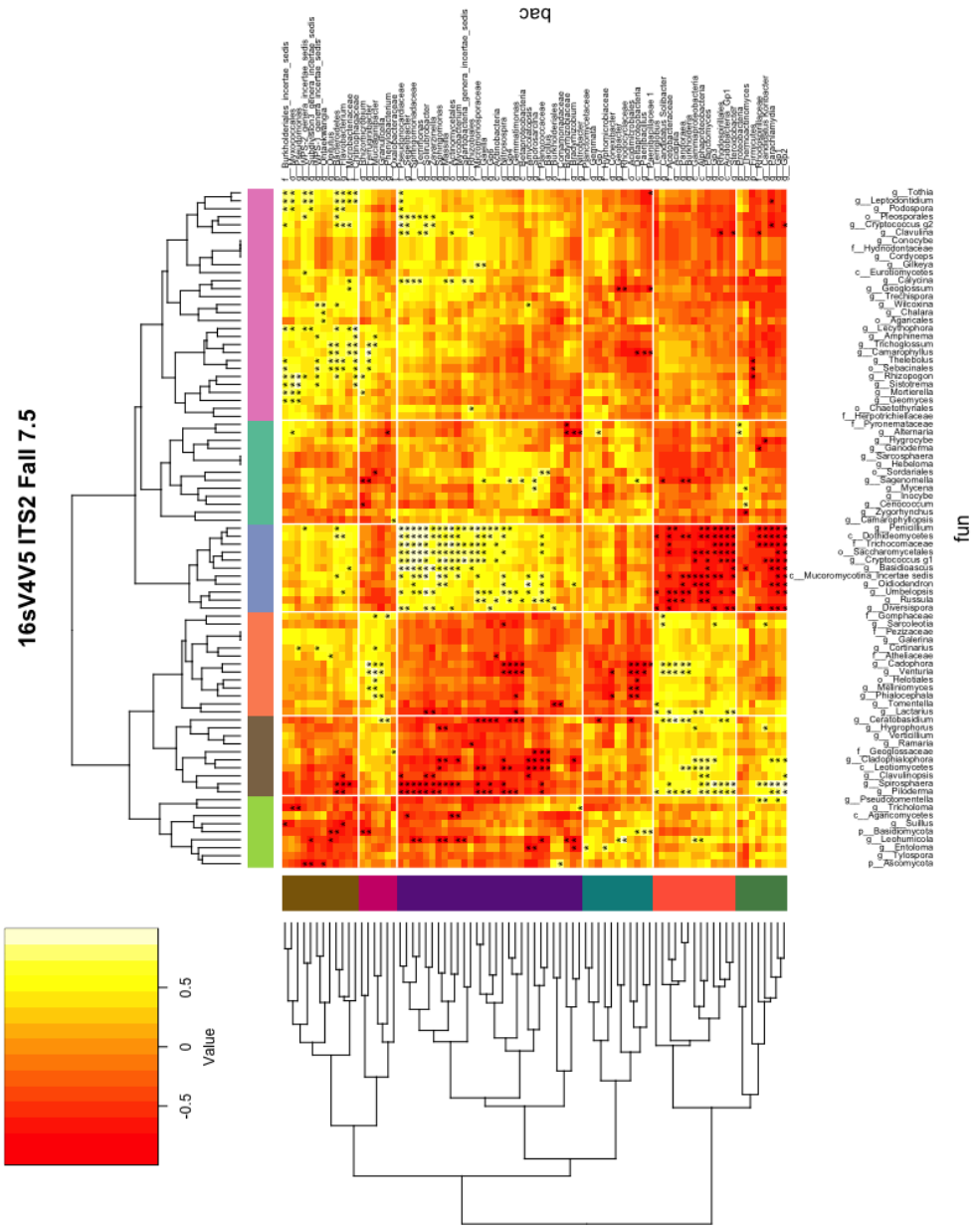


FIGURE A.18: **Correlations between 16S V4-V5 and ITS2 for Fall at depth 7.5.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

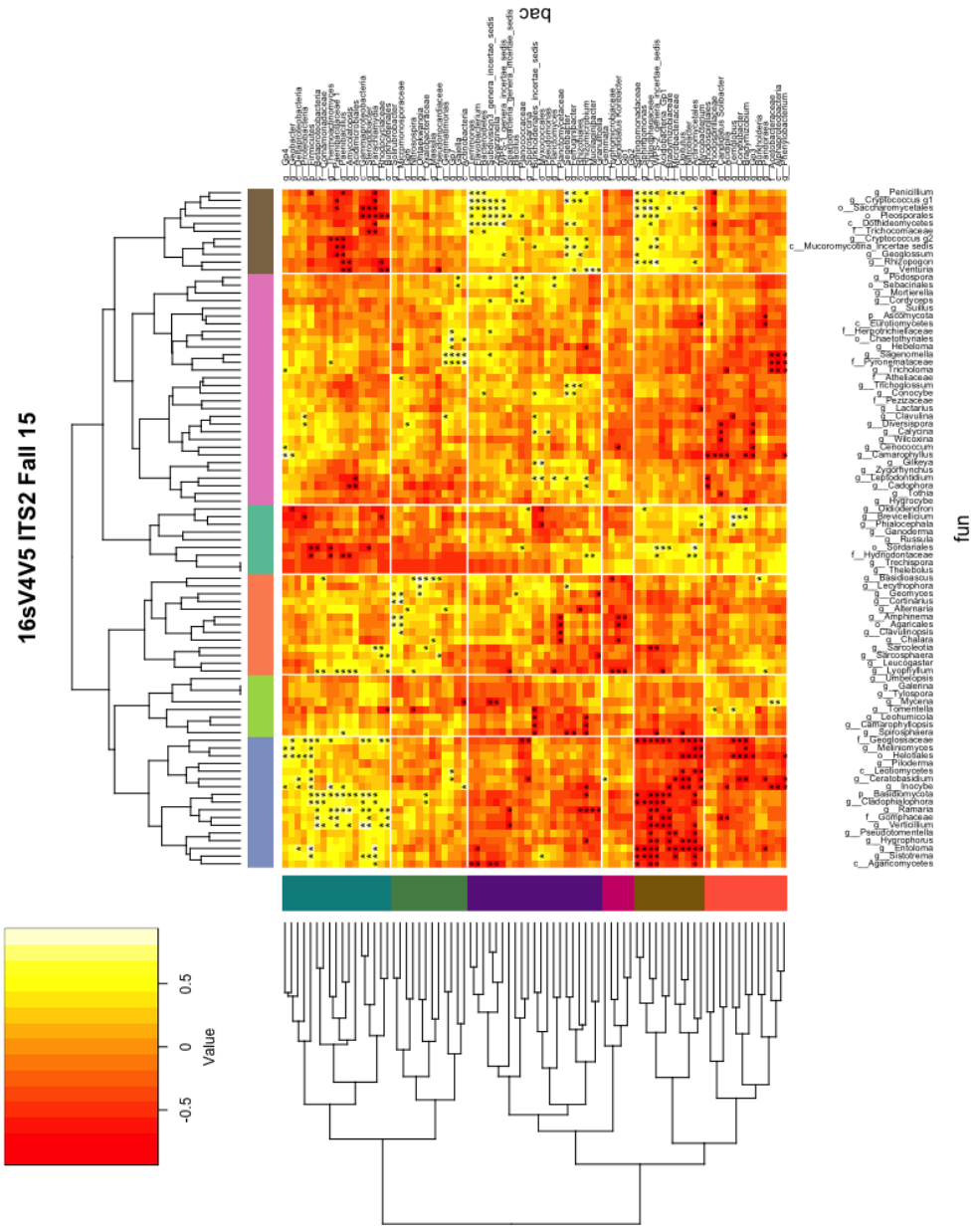


FIGURE A.19: **Correlations between 16S V4-V5 and ITS2 for Fall at depth 15.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

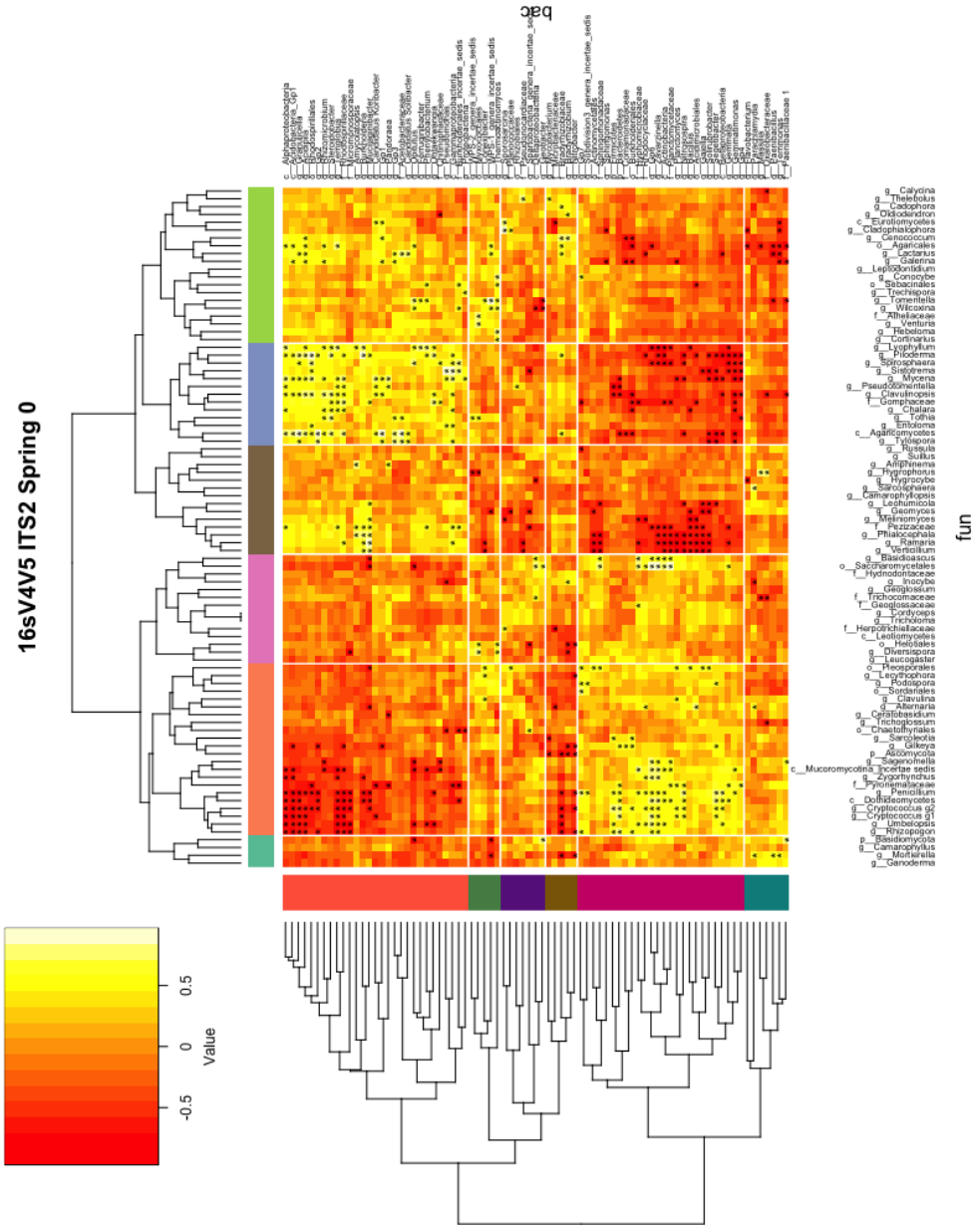


FIGURE A.20: **Correlations between 16S V4-V5 and ITS2 for Spring at depth 0.** Hierarchical clustering was performed on the bacterial (row) and fungal (column) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

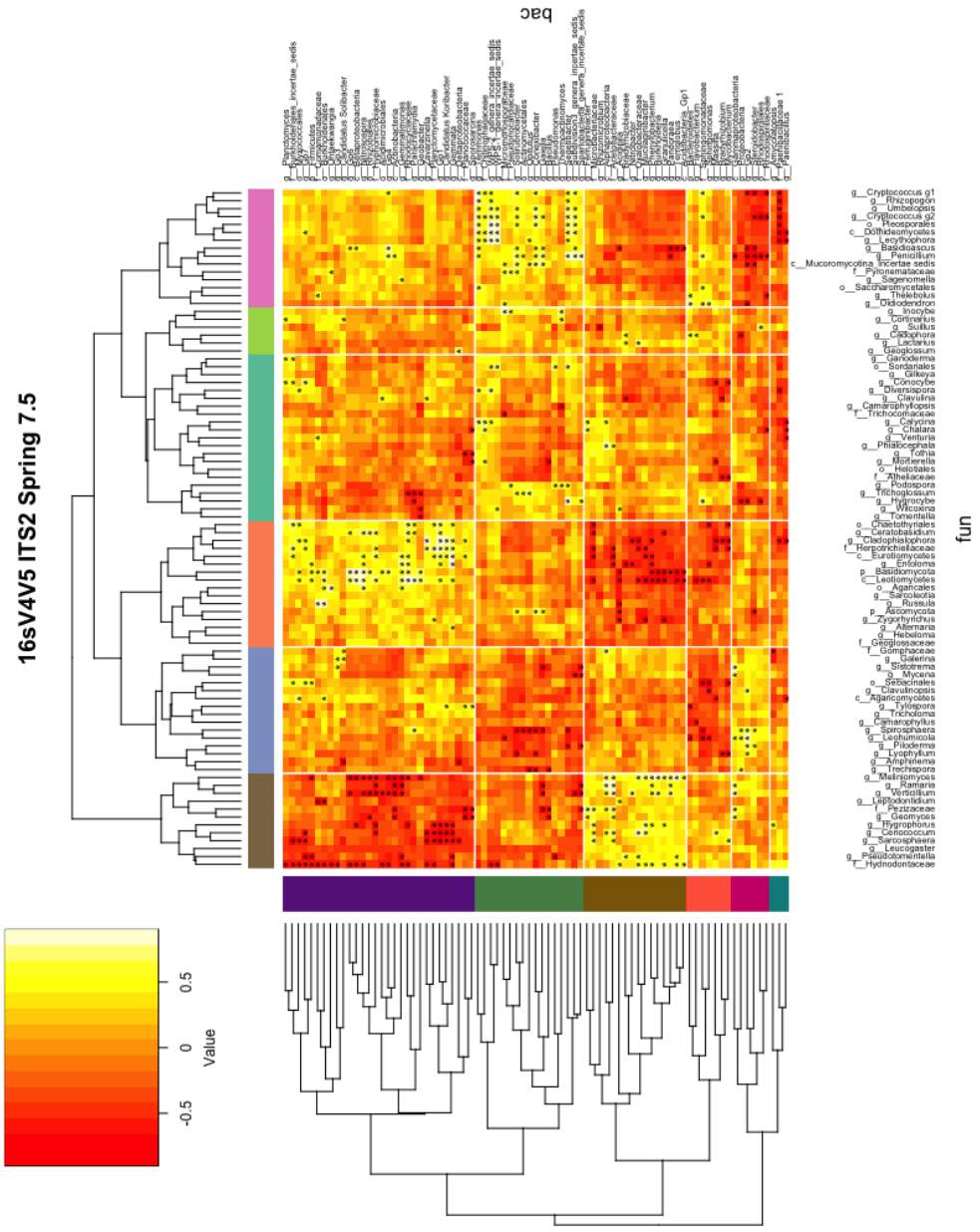


FIGURE A.21: **Correlations between 16S V4-V5 and ITS2 for Spring at depth 7.5.** Hierarchical clustering was performed on the bacterial (vertical axis) and fungal (horizontal axis) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.

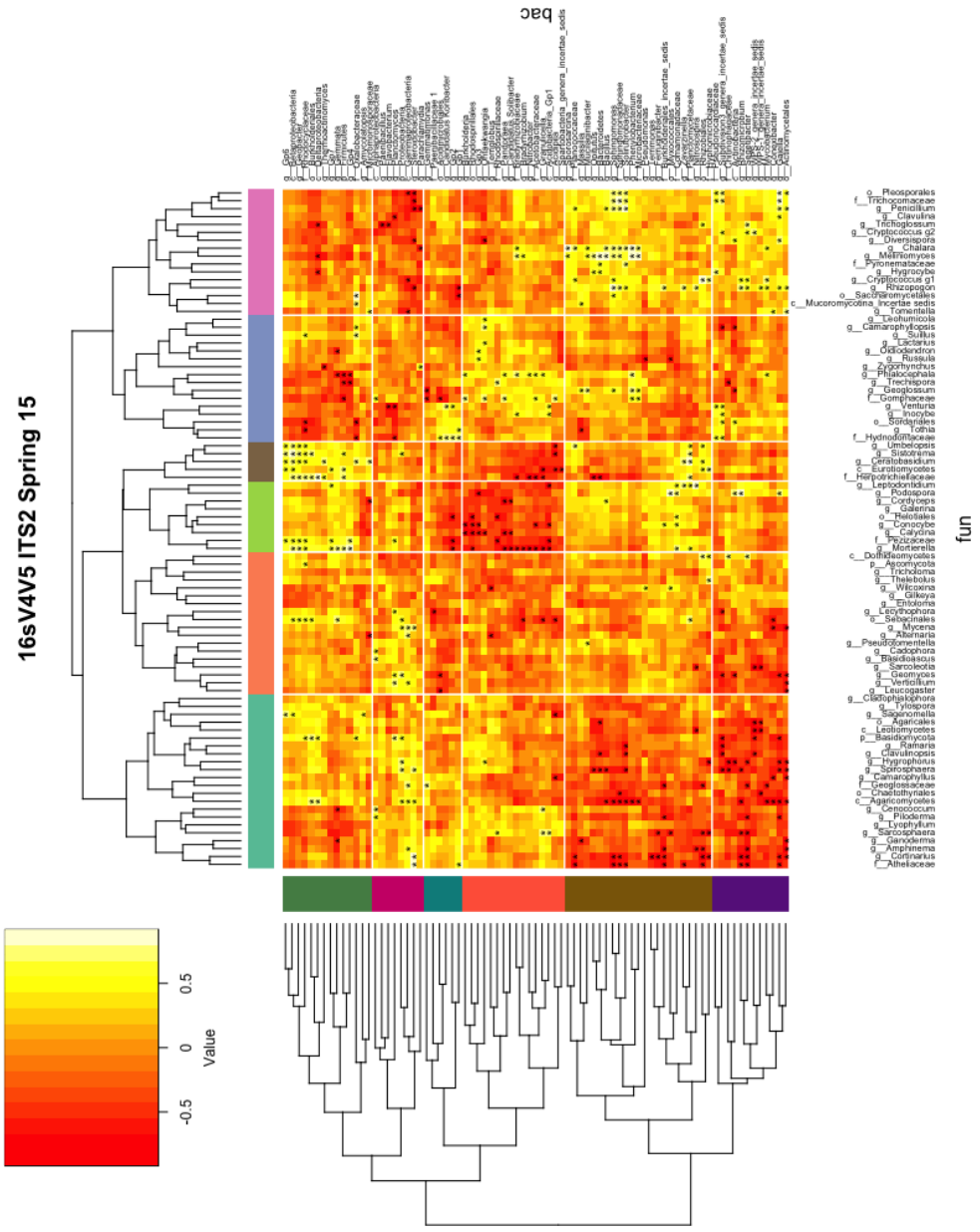


FIGURE A.22: **Correlations between 16S V4-V5 and ITS2 for Spring at depth 15.** Hierarchical clustering was performed on the bacterial (vertical axis) and fungal (horizontal axis) communities independently. The six clusters on each axis signify the microbes that commonly appear with the same correlations across the community. Strong negative correlations (-1) are red, while strong positive correlations (1) are white.