EXPLORING THE NODULE MICROBIOME COMMUNITY STRUCTURE OF TRIFOLIUM SPECIES

By

Prateek Shetty

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Biology - Master of Science

2016

ABSTRACT

EXPLORING THE NODULE MICROBIOME COMMUNITY STRUCTURE OF TRIFOLIUM SPECIES

By

Prateek Shetty

Plant associated microbes have been shown to increase plant growth and production drastically, yet we are just beginning to understand the parameters that impact these interactions. Rhizobia are primary bacterial symbionts of legumes and infect root hairs to form nodules, within which, the symbiotic rhizobia fix atmospheric nitrogen into biologically available forms in exchange for carbon from the host. The aim of this project is to understand the community structure and diversity of the nodule microbiome, with emphasis on the less abundant members, among coexisting clover species. North American clover *Trifolium-Rhizobium* communities are a good system to study host interactions with microbiomes given the high local species diversity. We analyzed the nodule microbiome of six congeneric clover plants when they were grown in soils conditioned by members of their own species and in soils conditioned by congener species by sequencing the 16s rRNA gene. The visualized microbiomes are similar, with 96% of all reads belonging to the order Rhizobiales. The rest of the OTUs belong to rarer groups of microbes. Further, the structure of the microbiome is impacted by both the host plant species and the soil in which the host is grown in, with soil explaining a larger degree of variation. There also is a strong interaction between soil and host in structuring the microbiome. The results are similar when the microbiome is analyzed with and without its most dominant order (*Rhizobiales*).

ACKNOWLEDGEMENTS

I would like to pay my gratitude to the persons below who made my research successful and assisted me at every step of the way:

My supervisor, Dr. Maren Friesen for her constant support and assistance. Her encouragement was instrumental in the completion of my thesis.

My committee members, Dr. Yair Shachar-Hill and Dr. Ashley Shade, whose guidance during my research was invaluable.

Dr. Alan Prather, Graduate director of the Department of Plant biology, whose reminders and constant motivation encouraged me to meet the deadlines.

All the members of the Friesen Lab, interacting with whom not only improved my research but also helped me ask better questions.

Finally, I would like to thank my family members and friends, without whom I would be a whole lot lesser. They not only assisted me financially but more importantly, extended moral and emotional support during the course.

This work was funded by grant NSF 1342793 to Dr. Maren Friesen

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

Bar	*Trifolium barbigerum*
Bif	*Trifolium bifidum*
Mac	*Trifolium macraei*
Mic	*Trifolium microdon*
Wor	*Trifolium wormskioldii*
Fuc	*Trifolium fucatum*
PLFA	Phospholipid Fatty Acid Analysis
NMDS	Non-metric Multi Dimensional Scaling
CCA	Canonical Correspondence Analysis
PCoA	Principal Coordinate Analysis

Introduction

Understanding factors that contribute to the coexistence

The diversity and species composition in plant communities is thought to be regulated by several different factors such as competition between plant hosts (Tilman 1990), resource partitioning (Tilman 1982) and interaction with other organisms in the ecosystem (Bever 1997). Competition between plant species can range from beneficial to detrimental to each other. The foremost study that explored the effect of negative interaction between two ecologically similar species showed that one species invariably was led to extinction (Gause 1934). This theory was later modified by Hardin (Hardin 1960) and named as the competitive exclusion principle. Explicitly, the competitive exclusion principle states that: *Two species competing for the same resource cannot continue to exist in the same population.* Thus, coexistence will not be observed if all members within the community are competing for the same limiting resources. However, Hutchinson's study (Hutchinson 1959) on phytoplankton communities showed confounding results. Phytoplankton communities show a larger diversity than what is predicted by competitive exclusion. According to the exclusion principle, since all members of phytoplankton compete for same set of resources (nutrients, light, space); the species that acquires them most efficiently will out-compete others leading the rest to extinction. Further, under model developed by Tilman (Tilman 1982), in an environment with a limited number of resources the number of coexisting species should not exceed the number of resources they compete for. However, most natural phytoplankton communities are highly diverse and not representative of this pattern (Hutchinson 1959).

While competition and resource partitioning have been well studied as primary forces that structure natural communities, results from long term studies don't always hold up. An eleven-year study carried out to test the effect of competition on coexistence of prairie plants showed only a few communities predictably coexist, suggesting that there may be other factors that contribute to coexistence (Dybzinski 2007). In an effort to explain the discrepancy, another view was put forward by Hubbel (Hubbel 1979). According to Hubbel's neutral theory of ecology, all species are functionally equivalent, originate and fluctuate in abundance at random. All species originate from a meta-community and migrate into smaller, local pools of dispersal limited communities. Within the smaller pools of local communities, the most common species account for a large fraction of the individuals sampled while the rest of the members are present in very low abundance. If communities as structured largely through neutral processes then we should obtain a sigmoidal curve, with common high abundance species to the right and rare, low abundance members towards the left of the curve. However, the neutral theory considers that all species within the community are functionally equivalent and have the same probability of being replaced.

The neutral theory and competition exclusion theory view interaction between species to either be nonexistent or constantly negative. However, in nature interactions between species can be considered to lie on a continuum (Stachowicz 2001, Saikkonen *et al.* 1998). Thompson proposed the geographic mosaic of coevolution theory in an effort to unify importance of geographic structure and the dynamic nature of species interactions (Thompson 2005). It makes use of three distinct components:

a) Geographic selection mosaics; fitness of interacting species is impacted by locally

co-occurring species and a genotype-by-genotype-by environment interaction.

b) Co-evolutionary hotspots; interactions between locally occurring species are reciprocally affected.

c) Trait remixing; due to gene flow between species and genetic drift continually shape the genetic structure of the species in the local environment.

Another way to explore Geographic mosaic theory is using ecological niche theory (Chase and Leibold 2003). Though the term “niche” was used earlier (Grinnell 1917), it was Elton (Elton 1927) who formalized the term “ecological niche” as: *The functional role occupied by a species in a trophic level.* The ecological niche theory relates a set of environment variables to the fitness of an organism. If an organism has traits that are suitable for the environment, then the organism continues to persist and traits are fixed in the given environment.

Recapping, we can study factors that contribute to coexistence and community structure under these three broad views:

1) Neutral theory of ecology

   All species are functionally equivalent. The success of a species in a particular environment has doesn’t depend on the species or its inherent traits and interactions with other members of the community. Instead, the success of a species in a local community can be attributed solely to migration, drift and abundance in the meta-community.

2) Competition exclusion theory

   All species are continually competing for resources. Members of the same species are more likely to compete for similar resources. Thus the only species

that will persist within the community are those that can outcompete other members for the limiting resource. Hence, community structure and member abundance at equilibrium will be directly proportional to the number of limited resources in the environment.

3) Geographic mosaic of coevolution

   All species are randomly distributed into large meta-communities. Species migrate into local, communities at random and depending on their abundance in the meta-community. Species that persist within the local communities are those that share traits with other co-occurring members and interact positively with the local environment.

The big difference between these processes is that the competition exclusion view assumes a constant type of interaction (negative, i.e. no mutualistic - parasitic interaction continuums) between members while the geographic mosaic theory is a more dynamic view, with interactions subject to change depending on the local environment and species present within that environment.

Previous studies using long term survey data to explore how species abundance changes with samples sites (Preston 1948, Bell 2001) comment on the consistency of abundant species identified. Sites sharing environmental similarity that are also in close proximity with each other tend to also share abundant species. An analysis by Preston on several years of survey data also show that such simple survey count data show a log normal distribution with an intermediate number of species present in moderate to high abundance and a large fraction of species that are detected in low abundance (Preston 1948). Microbial communities are several fold more complex than populations

of moths and birds that were typically used in survey data. Since microbial communities are more diverse, we predict that our abundance curve will be more skewed towards rare members and have fewer highly abundant members.

One reason why the competition model fails for other natural communities could be due to the omission of soil microorganisms in these studies. Soil microbial community can modulate interactions between different plant species and thereby on the plant community diversity and persistence (Moora 1996). Further, different plant species uniquely associate with microbial partners in the soil which perform a variety of different functions from nitrogen fixation to protection from pathogens.

Apart from this, the soil microbiome can play a significant role in maintaining diverse plant communities thorough processes other than competition and neutral dispersion; i.e by negative frequency dependent selection. A well-studied example is by accumulation of pathogens, also known as the Janzen-Connell hypothesis (Klironomos *et al.* 2002, Bever 2003, Fitzsimons *et al.* 2010, Bever *et al.* 2012). According to the Janzen-Connell hypothesis, individual plants accumulate species specific herbivores and pathogens over the period of its lifetime, thus reducing the success of conspecifics growing near the older plant. Thus, an understanding of the diversity of soil mutualists, and the strength and direction of interaction between the soil microbial mutualists, is essential to understand the contribution of different factors to the coexistence of plants.

Importance of microbial community on plant health

Microbial partners can improve plant health through their effects on nutrient availability as well as modulating abiotic or biotic stress. Elemental nitrogen is among the most

abundant resources on our planet and is a limiting factor in the production of protein and DNA. However, since accessible soil nitrogen reserves are poor, atmospheric nitrogen needs to be reduced to ammonia before it can be biologically usable by plants (Hardy 1968). Biologically usable nitrogen can be applied as fertilizers or be fixed by natural methods (symbiosis or lighting; Kim 1994). Biological nitrogen fixation within root nodules is carried out by members belonging to the Rhizobiaceae group (Sørensen and Sessitsch, 2007). Further, addition of beneficial Plant Growth Promoting Rhizobacteria (PGPRs) improve plant health and productivity by synergistic interactions with already present species (Cummings 2009; Guiñazú *et al*. 2009, Friesen *et al.* 2011) or by inhibiting other microbial members that are detrimental to the host (De Vleesschauwer and Höfte 2009). Apart from this, soil microbes also play an integral role in alleviating the different kinds of stress. For example, a study done on peppers showed selected isolates to improve protection under a drought like environmental condition by increased solubilization of phosphate and secreting a gel like material around the root zones to protect the root (Rolli 2015).

Generalist vs specialist selection within plant microbiome

An interesting pattern that emerges from most of these host associated microbiome studies is the presence of a small abundant community and a long tail of rare microbial members. Abundant microbes are generally found across closely related host species (Turner 2013, Dohrmann 2013). Further, abundant taxa contributions to microbial community structure do not vary wildly while rare taxa contributions vary (Dohrmann 2013). Abundant phyla generally include members like Actinomycetes and

Proteobacteria that can produce a plethora of compounds including antibacterial, nematicidal and antiviral properties (Muharram 2013, Mendes 2011). Rare microbes on the other hand are involved in more specific functions such as supplying phytohormones to plants (Karadeniz *et al.* 2006) or sulfate reduction (Pester 2010). Since members belonging to these two groups (abundant vs rare or generalist vs specialist) perform a wide array of different overrepresented and underrepresented functions, the two groups may be under different selection factors.

BNF: Biological nitrogen fixation

Biological nitrogen fixation is the process by which atmospheric nitrogen gas is incorporated into plant tissue (Hardy 1968). This can take place in two different ways: nitrogen fixation within nodule and associative nitrogen fixation. In nitrogen deficient environments, plants can form symbioses with certain members of rhizobia (associated with Fabaceae, Sprent 2001) or Frankia (associated with Rosids, Diagne 2013). These symbiotic partners use the nitrogenase enzyme to reduce atmospheric nitrogen to ammonia which is useable by the host.

Associative nitrogen fixation

Associative nitrogen fixation is yet another way by which plants can obtain the nitrogen required for their growth needs. In associative nitrogen fixation, diazotrophic bacteria are able to fix nitrogen in endophytic compartments with the help of nitrogenase (Doty *et al.* 2016). *Azospirillum* species is the best studied system for associative nitrogen fixation (Kucey 1998, Steenhoudt 2000). They are aggressive colonisers of both the root

and endophytic compartments and invade the host through undifferentiated root tips or cracks in the root tissue (Dommelen *et al.* 2007). The environment within root tissues is microaerobic and thus allows for nitrogenase to function. While associative nitrogen fixers are found within root tissues, there is no evidence yet for the occupancy of these members within root nodules.

Current literature on nodule microbiome

Currently, there are no studies that have used high throughput sequencing to profile member presence in nodules of nitrogen fixing plants. However, there has been a lot of sequence and trait data looking at microbial presence, activity and abundance within different regions of soil and root zones; specifically, the rhizoplane, rhizosphere and endophytic communities (Weinert *et al*. 2011, Inceoglu *et al.* 2011). Diversity of members within rhizosphere ranges from < 3000 OTUs (Weinert *et al.* 2011) to > 55,000 OTUs (Inceoglu *et al.* 2011). Further, most of these communities are so diverse and species rich that it is hard to reproducibly generate species richness values by subsampling reads from samples. The most abundant members are commonly from the phyla: Proteobacteria, Actinobacteria, Firmicutes, Bacteriodetes. An important pattern seen is the change in species richness value between root zones. Richness is highest in the bulk soil and then due to host selection reduces as distance to the root zone increases. Thus, diversity is typically highest at the rhizoplane zone and reduces across the rhizosphere and endosphere zones (Marilley 1998).

Soil legacy

Soil acts as a reservoir for local, diverse microbial populations, with upwards of 10^4 bacterial species found per gram of soil (Weinert *et al.* 2011). Despite the use of high throughput sequence technologies, we still fall short of completely profiling the microbial members in soil samples. Presence of microbial members in the soil depends specifically on the pH (Fierer and Jackson 2006; Lauber *et al.* 2008; Rousk *et al.* 2010), grain size and nutrient content (Faoro *et al.* 2010, Chaparro 2013). Large deviation in taxa presence and abundance is seen when soil pH is varied. Further, there is a strong positive correlation between the soil pH and diversity and composition of soil microbes (Rousk *et al.*2010). A related factor with soil legacy that plays a role in modulating the benefits from microbial partners is the exposure time with the microbial partners. Hosts that grow in soils in which members of their own species were previously grown will be able to select for the most beneficial partners and have a reproducible microbial community structure. While those grown in soils with members of different species will show a more diverse microbiome (Bulgarelli *et al* 2013).

Host species selection

While soil is the primary source, housing all microbial genetic diversity, host species act as a sink, selecting for specific members from the meta community pool. Host species release root exudates rich in sugar molecules and have shown to strongly influence the microbial community structure and abundance (Broeckling *et al.* 2008). Further, host species show specific enrichment of certain microbial taxa (Lundberg *et al.* 2012, Peiffer *et al.* 2013) and when hosts are grown with non-native microbial partners, host growth is affected with them having lower biomass as compared to hosts grown with native

microbial partners (Lou *et al.* 2014).

Techniques to study microbial populations

Culture dependent methods- Before the widespread use of NGS techniques

Traditionally, the occurrence of microorganisms in a given environment or in an industrial process has been studied by culture-based methods. While these methods were initially successful in exploring the characteristics and functions of members in the sample of interest, they were always labour intensive. Further, these methods often fail for microorganisms that require selectively enriched media (Lagier 2008). In addition, conventional culture based methods are only able to detect only half the OTUs identified by high throughput sequencing methods (Goodman *et al.* 2011, Rettedal *et al.* 2014).

Culture independent methods (Table 1)

Lipid analysis

Phospholipid fatty acid (PLFA) analysis is a biochemical technique that uses phospholipid fatty acids within the plasma membrane of bacterial cells to build a profile (Mitchell *et al.* 2016). The chemical composition differs depends on the type of bacterial organism. Thus, PLFA can be used to evaluate microbial community structure and activity. PLFAs can be extracted from the soil and their composition is analysed by gas chromatography. Changes in PLFA profiles are indicative of changes in the overall structure of microbial communities (Zelles 1999). PLFA analysis offers an advantage over culture-based techniques as it avoids the selectivity bias that is inherent to in the isolation and culture techniques (White *et al.* 1997).

Gradient Gels

Gradient gel electrophoresis methods are a step up in throughput compared to the previous techniques (Favier *et al.* 2002). The basic premise of these methods is targeted amplification of marker genes and visualization on gels. The electrophoresis will separate DNA molecules based on their shape, charge and molecular weight. Increasing concentration of denaturing agents (urea or temperature) will force the double stranded DNA molecule to melt. Depending on the nature of the sequence, different DNA strands will have different melting temperatures and thus will only melt at their corresponding denaturing gradient. The advantage of using gradient gels is that you do not need to have a reference and offers an initial view at the diversity and abundance of different groups within the sample. The technique was used to profile the oral cavity microbiome in children in order to understand if under a diseased condition (dental carries) certain microbial members are overrepresented (Ling 2010).

Hybridisation techniques - FISH, Microarray

Hybridisation techniques are a powerful tool to visualize structure and abundance of members within a community. Presence of a certain microbial taxa can be identified using corresponding oligonucleotide probes. Hybridisation methods allows for identification of even the rarest member of the community (Amann *et al.* 2008) and does not allow for any PCR based biases that previous methods inherently have. The downside of using a hybridisation technique is novel taxa can't be identified.

Targeted amplification and sequencing of a marker gene

Amplification of conserved marker genes has been the preferred methods of choice when multiple samples need to be identified and characterized. With the advent of

cheaper and high throughput strategies, sequencing multiple samples with great coverage has become progressively more affordable (Shendure *et al.* 2008). The high coverage, large range of reference databases allows users to characterize entire communities with taxonomy and in some cases function. Coupling these with the wide range of "-omics" platforms that are now available, allows one to explore the functional contribution of different taxa (Marcobal *et al.* 2013). Since its inception in 2007, the human microbiome project has generated more than 35 billion reads from 690 samples taken from 300 US based human subjects from various body sites (Turnbaugh *et al.* 2007, Ilseung *et al.* 2012) in order to explore how diet and nutritional status affect the microbiome assembly, succession of members within the microbiome and function of different microbial members.

Importance of studying native plants

Native species offer the best avenue to study established relationships between each other and their symbiotic partners. Further, native plants are typically locally-adapted to their home environment (Coleman-Derr *et al.* 2016) and are a thus a good resource to study microbial associations, which depend on the nature of the local environment (Heath & Tiffin 2007). For example, in environments where the host is not nutritionally limited, there is no selection for it to continue maintaining a costly symbiosis. Thus, such associations can be affected by fertilizer inputs or land management practices (Ding *et al.* 2016). Fields with long term fertilizer inputs have microbial populations that show reduced dependence on root exudates (Ai *et al* 2015). This may lead to fewer associations between microbes and plant hosts. Finally, invasive plant hosts can break

down associations between native hosts by changing soil chemistry through allelopathy (Cipollini *et al.* 2012), bringing in new partners, associating with other microbial partners (Putten *et al.* 2007).

About *Trifolium* species

Clovers are perennial herbs that have palmate compound, trifoliate bright green leaves. All species have a distinctive, round flower head composed of many, small, pea-like flowers. This involucre is considered to be a distinguishing feature of members belonging to this family. Flowers are typically small, reddish- purple in color with white tips. Well known members from this group are *Trifolium repens* and *Trifolium pratense.* The different species (Figure 1) that occur at the field site, Bodega Bay, in this study are named and described by Seringe (1825): 1) *Trifolium barbigerum* (here abbreviated Bar) or Bearded clover is an annual herb, native to Northern California and Oregon. The plant blooms between February to March. 2) *Trifolium bifidum* (here abbreviated Bif) or Notch leaf clover is an annual herb, native to the western region of North America (Washington to California). The plant blooms between April to June. 3) *Trifolium macraei* (here abbreviated Mac) is an annual herb, native to California but is also found in other parts of North America and the world. The plant blooms between March to May.4) *Trifolium microdon*(here abbreviated Mic) is an annual herb, endemic to California. The plant blooms between April to June. 5) *Trifolium wormskoldii* (here abbreviated Wor) is a perennial herb, native to California and found in other parts of the Western North America. The plant blooms between May to June. 6) *Trifolium fucatum* (here abbreviated Fuc) is an annual herb, native to the western North America and

California. The plant blooms between April to June.

In this project we aim to explore how selection factors specifically soil legacy type, host species, nodule size and phylogenetic distance has an impact on the microbial community composition. We used targeted amplicon sequencing of 16s rRNA to profile the nodule microbial community of our samples. Making use of 16s rRNA allows us not only to estimate abundances of unique taxa but also allows us to identify them using reference Greengenes database. However, we can't directly use the reads generated by the sequencing run. The reads must first be demultiplexed into individual samples and then quality filtered. Once the reads are cleaned, they are then ready to be clustered into representative sequencing. Such a form of data aggregation allows us to not only generate counts per representative sequence but also gives us a basic understanding on the diversity of unique clusters seen. In this study, we make use of Operational Taxonomic Units (OTUs). OTUs can be generated by using QIIME (Caparaso *et al.* 2010), MOTUR (Schloss *et al.* 2009) or UCLUST (Edgar 2010). We decided to use QIIME for read clustering as it supports the several different clustering algorithms, including MOTHUR and UCLUST. Further, QIIME also makes use of RTAX (Soergel 2012), a tool for assigning taxonomies using a reference database. RTAX makes use of mate pair information when assigning taxonomies making it ideal for our dataset with non-overlapping mate pairs.

Specifically, in this study we explore the following questions:

1) Analyze the similarity of nodule microbiome community for different *Trifolium species.*
2) What is the core nodule microbiome for the home *Trifolium* samples.

3) Do we see the presence of **Soil** and **Species** factor impacting species richness and diversity.
4) Do we see the presence of **Soil** and **Species** factor in structuring microbial communities.
5) Do neutral processes contribute to nodule microbiome community assembly.

Results

Sample read distribution table.

16s rRNA gene fragments from nodules of 6 home Californian species of *Trifolium* grown experimentally in “home” and “away” soil. A total of 227,196,520 reads were generated through paired end Illumina sequencing (Table 2). Reads were demultiplexed, filtered for quality and length. Averages read length for forward and reverse reads were 51 and 53 bases respectively. Read counts per sample varied from 6,675 to 463,075 reads per sample (Figure 2a).

OTU picking and Taxonomic Classification

OTU picking generated a total of 2,394 OTUs (Figure 3a). Close to 15% to 54.5% of reads within samples were poorly classified and taxonomically assigned to “None” or “NOHIT”. All OTUs with this label were removed. The filtered OTU table (Figure 3b) consisted of 1314 OTUs and read counts per sample varied from 6,546 to 462,142 reads per sample. Proteobacteria (representing 97.5 to 99.9% of all the matched reads per sample), Bacteroidetes (representing 0.035 to 2.04% of all the matched reads per sample) and Actinobacteria (representing 0.009 to 0.69% of all the matched reads per sample) were the top three most dominant phyla. By far the most dominant order of bacteria was identified within the Proteobacteria phylum and was called “Rhizobiales”. The order represented 86.5 to 99.6% of all the matched reads per sample and was made up of 123 OTUs. These OTUs were extracted and labeled as the “Rhizobiales” microbiome.

In order explore the nodule microbiome in greater detail, we specifically excluded all OTU's belonging to the order "Rhizobiales" and named it the rare microbiome (Figure 3c). The rare microbiome OTU table consisted of 1191 OTU's and read counts per sample varied from 163 to 10,867 reads per sample (Figure 2b, Table 2). Proteobacteria (representing 51.4 to 94.2% of all the matched reads per sample), Bacteroidetes (representing 3.8 to 43.1% of all the matched reads per sample) and Actinobacteria (representing 0.89 to 20.07% of all the matched reads per sample) were the top three most dominant phyla.

The distribution of OTUs across all samples was largely similar. There were very few OTUs that were abundant and present in high frequency across all samples. Further, there was the presence of a large rare tail of low abundant community members. This suggests that the nodule microbiome typically consists a few dominant groups that contribute to nitrogen fixation and occupancy within nodules and a large population of low abundance microbial partners that are present within the root tissues as endosymbionts may or may not contributing to active nitrogen fixation.

Core Microbiome

Presence/Absence matrix

Threshold 1: OTUs present in at least 2 of the 4 samples.

A total number of 64 OTUs (4.8% of all OTUs, Figure 5a) were identified as core. Of these OTUs, the phylum Proteobacteria was the most represented (44 OTUs) followed by members of the phylum "Bacteroidetes" (11 OTUs), "Actinobacteria" (6 OTUs) and "Chloroflexi", "Armatimonadetes" and "TM7" (1 OTU).

Threshold 2: OTUs present in at least 3 of the 4 samples.

A total number of 41 OTUs (3.1% of all OTUs, Figure 5b) were identified as core. Of these OTUs, the phylum Proteobacteria was the most represented (31 OTUs) followed by members of the phylum “Bacteroidetes” (6 OTUs), “Actinobacteria” (3 OTUs) and “Chloroflexi” (1 OTU).

Abundance and Presence/Absence matrix

Using an abundance threshold did not change the top three identified core OTUs within all 6 samples. However, making use of an abundance threshold did reduce the number of unique OTUs per sample as represented in the graph (Figure 5c and 5d).

Diversity estimate

Alpha diversity

Rarefaction curves generated do not show saturation, further most curves have large error bars and overlap each other (Figure 6a-6d). This is most probably due to the low depth of rarefaction and large diversity of OTUs seen. Rarefaction greater than 6546 reads/ sample led to loss in samples and curves still do not show saturation. Rarefaction curves remained comparable between hosts grown in home and away soil. In order to calculate species richness and impact of meta-data factors on species richness all OTU tables were rarefied to ensure samples had equal number of reads. All samples in the complete microbiome table were rarefied to 6500 reads/ sample and the rare microbiome table after dropping the “Rhizobiales” order was rarefied to 163 reads/

sample. The abundant microbiome table was rarefied to 6250 reads/sample. We carried out anova tests on the rarefied complete OTU table and abundant OTU table, with diversity metrics as the response variable. We tested to see if Soil, Species and nodule size factors had an effect on the species richness.

The three metrics used (Shannon, Observed and Chao1) showed similar results: no factor significantly affected species richness or there was a very weak, insignificant effect (Table 3a). The rarefied rare OTU table showed no consistent patterns of significant Soil and Species interaction (Table 3c).

Ordinations

Beta diversity

All samples in the complete microbiome table were rarefied to 6500 reads/ sample and the rare microbiome table after dropping the “Rhizobiales” order was rarefied to 163 reads/ sample. The abundant microbiome table was rarefied to 6250 reads/sample.

All beta diversity estimates were calculated from the rarefied, count normalized OTU table. Such an approach allows us to scale different samples within our data. Beta diversity was first visualized using ordination methods.

PCoA plot was built using Bray-Curtis dissimilarity distance matrix. Bray-curtis was used as it works better in tables with large null values. We compared the PCoA plot for all three microbiome tables: all (Figure 7a.i), abundant (Figure 7b.i)and rare(Figure 7c.i). For the rare microbiome, the first and second axis explained 12% and 10% of the total variance respectively. Further, the total variance explained by an axis increased to 55.1% and 51% as we compared the complete and abundant table. Both these plots

also showed an overlapping of samples indicating that the relationship was being driven by the abundant OTU order of “Rhizobiales”. We also plotted the eigenvalues using a scree plot to visualize the range and spread of values for the all axis. The complete and abundant tables alone showed a very large PC1 axis compared to the rare tables, further giving evidence that the order “Rhizobiales” maybe driving the relationship. In order to confirm if the strong grouping effect is driven by abundant taxa that we see in the both the complete and abundant OTU tables, we ran another ordination with NMDS. This allowed us to compare how the samples and the OTUs were related with each other (Figure 8a, Figure 8b, Figure 8c). We also visualized the dataset using CCA and constraining the axis by Species and Soil factors. Doing this allowed us to visualize how these factors contributed to the variance seen in the microbial community structure. Constraining ordinations by the Soil factor explained the largest amount of variation in the tables (Figure 9a.i: All; CCA1: 3.4% and CCA2: 2.2%. Figure 9c.i: Rare; CCA1: 2% and CCA2: 1.5%). Further we plotted the eigenvalues of the constrained samples to visualize how the factors are driving these relationships. Constraining by Soil factors showed samples belonging to Mic and Wor clustering together in the same direction, while Bif and Mac samples clustered to the opposite direction, showing that Mic and Wor have more similar community structure compared to the other samples. This relationship was constantly identified in both complete and rare tables. Further, clustering by Species factors showed no clustering with arrows pointing in opposite directions, indicating that samples coming from the same species but different soils have more diverse microbiomes.

Adonis modeling results

The rarefied OTU tables were used for multiple factor testing. Multiple factor testing on the complete OTU table showed that communities were being partitioned out by similar factors: Soil (R^2=0.07711, p=0.006) and Species (R^2=0.05961, p=0.043). Interestingly, there seems to be no significant interaction effects between Soil and Species factors (R^2=0.15806, p=0.522).

Multiple factor testing on the abundant OTU table showed that communities were being partitioned out by similar factors: Soil (R^2=0.05605, p=0.054) and Species (R^2=0.07655, p=0.006). Interestingly, there seems to be no significant interaction effects between Soil and Species factors (R^2=0.19574, p=0.079).

Multiple factor testing on the rare OTU table showed that communities were indeed partitioned out by several factors, most important of them being; Soil (R^2=0.115, p=0.001) and Species (R^2=0.07676, p=0.001). Also, interaction between Soil and Species effect (R^2=0.183, p=0.001) is highly significant too (Table 5). Both these give similar results. It is interesting to note that the model with Soil*Species interaction explains the highest amount of variance explained. These results mean that there is a strong underlying interaction that influences microbial community assembly even at the endophyte level. Another interesting result is the strength of Soil factor. In all the Adonis models, Soil seems to have a greater effect compared to any other factor.

Neutral community model

More number of OTUs were detected for away sample (578 OTUs) compared to home samples (293). We also had more number of OTUs that were selected for and against

within away samples compared to home samples. Fishers test on the number of OTUs selected for, against and neutrally distributed showed that there was indeed a significant difference between home and away samples (pvalue=3.842e-5). Both OTU tables fit the neutral distribution pretty well with a marginally higher r-squared value for away samples (rsquared= 0.918) compared to home samples (rsquared= 0.877).

Discussion

Composition of nodule microbiome

A total of 1314 OTUs were identified within the nodule microbiome of *Trifolium sp.* The nodule microbial community is less diverse compared to the bulk, rhizosphere and phyllosphere community described in previous literature (Uroz *et al.* 2010, Mendes 2011, Inceoglu *et al.* 2011, Peiffer *et al.* 2013, Chaparro *et al.* 2013). The nodule community not only has fewer OTUs but also fewer identified taxa. However, some of the important taxa groups are identified: Proteobacteria, Actinobacteria and Bacteroidetes. Members belonging to these phyla are seen as abundant members within root associated microbiome (Uroz *et al.* 2010, Mendes 2011). The large tail of rare microbial members is yet another pattern seen in most natural samples and rare microbial members are important in successfully assessing diversity metrics (Lynch & Neufeld 2015).

The top phyla identified in our study belonged to Proteobacteria, Actinobacteria and Bacteriodetes. Members belonging to these phyla are typically observed to be enriched in different rhizosphere studies (Wieland *et al.* 2001, Peiffer *et al.* 2013, Chaparro *et al.* 2013). However, the range of diversity also depends on the host: < 3000 OTUs (Weinert *et al.* 2011) to > 55,000 OTUs (Inceoglu *et al.* 2011).

The most abundant OTU belonged to the group *Agrobacterium*. Typically, members from this group are pathogenic and not associated with nitrogen fixation. The host *Sesbania* has shown to nodulate with a group of agrobacterium. These members also have a low similarity at the 16s region with rhizobium species (Cummings *et al.* 2009). Further, 57 agrobacterium members were isolated from different legume hosts. Wide

strain variation was also observed. The agrobacterium OTU identified in our samples may also be one that fixes nitrogen associatively. However, we need biochemical tests to confirm this. Phylogenetic analysis puts the abundant agrobacterium with rhizobium species (Figure 10).

While it is identified that multiple strains can colonize the same individual plant (Denison and Kiers 2004, Denison 2000), how many strains colonize a single nodule is lesser known. In lab studies there has been evidence of nodules with mixed strains. A study looking at the nodule occupancy in *Medicago sativa* showed presence of mixed infection with nitrogen fixing and non-fixing strains occurring in the same nodule (Checcucci *et al.* 2016). Our results identify multiple OTUs present in high abundance identified to the order Rhizobiales within every sample, where each sample comes from a single nodule. This could be an evidence for mixed infections within clover nodules. An oligotyping analysis would help answer this question with greater sensitivity and accuracy.

We touched upon 2 views of studying community structure and assembly. One views community assembly processes through competitive exclusion and niche theory. Competitive exclusion theory suggests that all species in nature are competing for resources. Species that are closely related will complete for the same set of resources. Further, two species competing for the same resource niche will not continue to exist as in the long term the community will be dominated by the one species that has a slight growth advantage. Thus, the number of species in the environment will be dictated by the number of limiting resources. Compared to this, is the other view a combination of neutral theory of ecology and the geographic mosaic of coevolution. The neutral theory

of ecology says that individuals are functionally equivalent and recruited from a larger meta-community pool into small local communities. Extending this with the assumptions from the geographic mosaic theory, individuals are distributed randomly across different local communities. Within local communities with members whose traits match their environment interact positively and coevolve more rapidly compared to local communities with members having dissimilar traits or interacting negatively with the environment.

The competitive exclusion theory has met with a lot of resistance primarily due to the fact that microbial communities in nature are highly diverse (Hutchinson 1959). A meta-community analysis looked at the presence of competition in different microbiome communities through phylogenetic dispersion (Koeppel *et al.* 2014). If members within a community are assembled due to competitive exclusion, then the community will be phylogenetically over dispersed with more distantly related members. In most communities they found no signal of phylogenetic overdispersion until they looked OTUs clustered at higher ranges of sequence similarity. We found several closely related OTUs occurring across all our nodules, much more than what could be expected by just the number of limiting resources. However, we would need longer read lengths to cluster OTUs at a finer scale and test them for phylogenetic over or under dispersion to infer the strength of competition in assembling nodule communities.

According to the geographic mosaic of coevolution, individuals drift or migrate into communities randomly and depending on their abundance in the large meta-community. Their survival in the new environment depends on the traits they carry. Organisms with traits that closely match their environment will survive and persist while the rest go

extinct over time. Such a process should give us a sigmoidal species abundance curve with a long tail of rare, low abundance members and a small group of commonly occurring, high abundance members. The slope of the curve will depend on the number of abundant species. The neutral model on the other hand states that member presence within a local community is solely a function of drift, migration and abundance in the meta community. The neutral model was applied to study how the gut microbiome of zebra fishes was structured over time (Burns *et al.* 2015).

The neutral mode fit better for younger ad juvenile fishes better than adult fishes, implying that random processes play a large role during the initial microbiome assembly. As the fish ages, the fit of the neutral model decreases indicating the importance of environment and interaction between local microbial members in structuring communities.

Our adonis results show that both soil and species factors play an important role in the community structure of trifolium nodule microbiome, with soil factor playing a bigger role and explaining a greater degree of variation. If we consider the soil factor to be the reservoir or the meta-community housing all the individuals from which members in our local community are derived, then this result aligns well with both the results of neutral assembly in young zebra fish gut microbiome and the view of neutral assembly process described in the introduction. Microbial species from the soil may migrate into the endophytic microbiome however such a process will only progress if the microbial species has the right set of genes to interact with the host. Further since we have several different hosts, the genetic alignment between the microbial partner and plant host is even more crucial. So while we can consider survival and presence of microbes

in soil to be a neutrally distributed process, the survival and presence within the plant host need not necessarily be neutral. Similarly, while initial colonization of microbes within juvenile zebra fishes maybe neutrally assembled, their long term presence within adult zebra fishes doesn't fit well with the neutral model due to the interaction between the microbial partners and the microenvironment offered by the host species.
Further, we would need a time series data to study how the assembly process changes over development time of the trifolium host and if soil continues playing a larger role compared to species over long term growth of the host.
We were unable to reliably estimate richness due to the sheer diversity of OTUs and unbalanced samples that we had. Anova run at different rarefactions and metrics showed no significant difference in microbial species richness.

Core microbiome

Core microbiome analysis was carried out using both presence/absence matrix and an abundance matrix. Both the methods gave us similar results. The "core" microbiome identified was primarily made of up 3 different phyla; Proteobacteria, Actinobacteria and Bacteroidetes. Members within these phyla have been previously described to be present in high abundance within plant tissues (Reinhold-Hurek et al. 2015, Turner et al. 2013, Hirsch 2012). In our results we see a very small group of core microbial members, these members also happen to be those that are abundant across all samples.

Alpha Diversity analysis

Despite running alpha diversity at multiple different rarefactions and using different species richness metrics we were unable to see any saturation in the rarefaction curves. Further we made use of an anova to look if there were any factors had a significant effect on the species richness of the community. Most factors were insignificant and those that were significant showed only borderline significance or disappeared at higher rarefactions indicating that the effect was possibly a chance observation. Calculating species richness robustly requires high read coverage within samples to run a robust, reproducible rarefaction analyses.

Ordinations

Across all tables, few consistent patterns arose. Primarily, we see that abundant OTUs drive clustering of samples. This implies that apart from the rare tail taxa, the samples share most of the abundant taxa. Further for the complete and abundant microbiome table, we see only two axis that largely contribute to the amount of variance explained, implying that there are two or three important factors impacting the structuring of microbial communities.

a) PCoA

PCoA of complete tables showed that there was a lot of similarity between the samples themselves and a large PC1 axis. The PCoA with just the rare microbiome showed separation of samples by axis. Samples belonging to Worm, bar and mic clustered together away from Bif and mac samples. This could indicate that *Trifolium* species

recruit members into the nodule microbiome community differently.

b) CCA

When we look at samples coming from the same soil but different species (constraining by soil factor) we see a more similar microbiome as evidenced by the clustering of arrows. Whereas, when we look at samples coming from different soils but same species we see a more varying microbiome. This indicates the strong effect of soil legacy.

c) NMDS

For both the complete and Rhizobiales OTU table, we see that sample clustering overlaps clustering by the abundant microbial community members. The same relationship is absent when looking at an NMDS plot generated using only the rare microbiome. Thus giving us more evidence that the grouping along a single axis is most probably driven by the dominant taxa.

Beta diversity model comparisons

We made use of adonis analysis to explore individual and interaction effects between our metadata. Both complete and rare microbiome table showed strong individual species and soil effects. This indicated that both factors play an important role in structuring the microbial community. A strong interaction seemed to be present when we looked at species*soil effects only for the rare microbiome table. This effect was not seen as a significant effect for the complete table. Further, the only time we see soil explaining a smaller variance compared to species is when we look at the abundant microbiome table (Soil R^2=0.05605 and Species R^2=0.07655). Soil playing a smaller

role in the structure of abundant community is understandable. Since most of the abundant organism are also shared by a large majority of samples, their presence and similar abundance across samples is not surprising.

Soil type and host species have been previously identified as important factors structuring microbial communities in soil (Garbeva *et al*. 2004, Berg *et al*. 2009). In our study when we look at the complete and rare microbiome tables, we found that soil factor plays a larger role in structuring these communities compared to species factor. One possible reason that lets Soil play such an integral role is the duration of the study. A single generation might not be sufficient for hosts to structure microbial communities, however over time this relationship may shift. Finally, we see a significant interaction effect but only for the rare microbiome. The rarest members of the community are more prone to being lost due to extinction events. Their presence in the microbiome depends on not only surviving in the soil metacommunity but also selection by the host. The abundant microbiome members on the other hand are less prone to being lost due to their sheer abundance. This result could indicate that different members of the microbiome need not all be selected under the same selection factors.

Neutral community analysis

Plants growing in their own soil were growing on a soil legacy that was conditioned by a member of the same species and thus would be exposed to microbial members whose abundances were structured by the ancestor. Whereas plants growing in a soil condition by a member of a different species would have to select upon a microbial community with members that will interact with it. This should lead to a change in member

abundance, specifically the abundant members of the microbiome. This is assuming that the reason these members are abundant is prior selection by the ancestral host. Further, we should also see an increase in selection for the rare members of the microbiome as these members might interact with the new species. Thus, we should not only detect higher number of species within away samples but also have more species that do not fit within the neutral distribution. Further, in the away samples we also see that a lot of the low frequency OTUs are selected for while the high frequency OTUs are selected against as seen by the few points that fall above and below predicted neutral model values.

It seems that most of the OTUs in home and away samples fit the neutral model pretty well. The interesting aspect is the selection for low frequency OTUs in the away samples. These results indicate that while neutral process play an important role in the assembly of *Trifolium* nodule microbiome, different hosts also play a role by selecting for and against microbial partners. A further study using multi-generation time series data will allow us to track the increase and decrease of specific OTUs, thus informing us of their importance with the host.

Broader Impacts

We live in a microbial world with microbes and their communities forming the foundation of biosphere and playing an integral role in the functioning of all trophic levels. For example, rhizobia are involved in increasing useable soil nitrogen resources by reduction of atmospheric nitrogen to ammonia through biological nitrogen fixation in exchange for photosynthetically fixed carbon from the host (Hayatt *et al*. 2010). Rice

fields associated with *Azolla* can fix upto 600kg N/ha/year during the growing periods (Fattah *et al* 2005). Thereby reducing dependence on commercial fertilizers and need for highly fertile land. Further, microbes play a crucial role in mediating phosphorus availability too. Further, plants can associate with different kinds of fungi. Plant-mycorrhizal associations are one of the better studied symbioses and about 80% of all land plants are capable of associating with mycorrhiza. Such associations lead to improved phosphate uptake by the host (Vance 2001). Apart from improving nutrient uptake, mycorrhizal associations also enhance root surface area by sending out their hyphae and creating secondary root systems. Understanding how plant-microbe interactions are affected under ecological conditions will help us introduce better land management practices. Further, reducing the dependence on commercial fertilizer not only provides monetary benefits to the farmer but also reduces leaching of nutrients into local water bodies and preventing large scale algal blooms.

Microbes are also shown to protect their hosts from biotic and abiotic stresses. For example, isolates cultured from wheat grown in saline environments gave rise to about 24 salt tolerant isolates, all of which were able to produce phytohormones like indole 3 acetic acid or gibberelins and improve plant productivity under salt stress (Upadhyay *et al.* 2009, Culligan *et al.* 2012). Further, rhizobacteria produce metabolites that can inhibit the growth of other taxa (Kim *et al.* 2006). Rhizospheric fungi are also well known producers of antibiotic metabolites that can inhibit the growth of other microbes or defend the host against predatory protozoa, improving protection against biotic stress (Hoffmeister 2007, Brakhage 2011). Thus, it is important to consider feedbacks from the local microbiome community when exploring factors that contribute to host growth,

development and succession.

Methods

Plant experiment

Seeds from *Trifolium barbigerum, T. bifidum, T. fucatum, T. macraei, T. microdon,* and *T. wormskioldii* were collected from Bodega Marine Reserve, California, in 2012. Soil was collected from below each species and placed in pots in a UC Davis greenhouse. Seeds were scarified by razor nicking and planted directly in field soil, then watered as needed. At 6 weeks of age, plants were harvested, soil shaken from the roots and individual nodules plucked from roots and flash-frozen.

Sequencing experiment

DNA was extracted from single nodules using Zymo's quick gDNA kit and stored at -20C. Two loci were targeted by PCR, nodC and 16S.

nodC

The symbiotic nodC locus was targeted with nodC_4192-4845 (GGCGAGACCCTKTTYTGCTA, GTGACKACCATYSCAAGGCT), with a PCR program 95C 1:00 followed by 35 cycles of 95C 0:30, 51C 0:30, 72C 1:00, and a final extension of 72C for 1:00. Amplicons were Sanger sequenced at UC Davis and traces were quality trimmed and aligned with CodonCode to identify polymorphic sites.

16S

The same DNA extractions were used as template in a 16S PCR targeting the 799-1115 region (AACMGGATTAGATACCCKG, KGGTYKCGCTCGTTRC), with a PCR program

95C 1:00 followed by 35 cycles of 95C 0:30, 60C 0:30, 72C 1:00, and a final extension of 72C for 1:00 PCR was done with combinatorially barcoded primers to enable a high degree of multiplexing; our combinatorial approach enabled this with only 4 barcodes per end, combined with 9 standard Illumina indexing adaptors (See supplemental methods). Briefly, first stage PCR was conducted with barcoded primers containing part of the Illumina adaptor, and then pools were made for the second stage PCR that completed the adaptor and added the Illumina barcodes. Data was deposited at SRA under the code **SRP070507 and bioproject accession code PRJNA297440.**

Microbiome community profile: Building OTU table

Above-mentioned 16s rRNA regions was sequenced on Illumina platform. The generated paired-end, non-overlapping reads were demultiplexed, trimmed to do away with bases that had a quality score of lower than 25. Qiime (v1.6.0) was used for all further downstream analysis. Demultiplexed forward and reverse reads are binned individually after their headers were renamed.

Each Operational Taxonomic Unit (OTU) is picked based on sequence similarity of the reads. Clustering is carried out with only forward reads using uclust_ref clustering algorithm. Against the Greengenes reference database (version: Greengenes 13_5) at 97% similarity. Reads that failed to hit the database were removed from further analysis. Taxonomic classification was assigned with the RTAX procedure in QIIME, using the Greengenes database. The RTAX method makes use of reads from both ends before assignment. The additional information from the second end allows for a more precise taxonomic assignment.

The OTU table was further filtered to remove poorly labeled OTUs. Further the OTU table was divided into complete, abundant and rare microbiome tables based on the most abundant order of bacteria; Rhizobiales.

For alpha and beta diversity analysis both tables were then rarefied to ensure they had equal number of reads. All samples in the complete microbiome table were rarefied to 6500 reads/ sample and the rare microbiome table after dropping the “Rhizobiales” order was rarefied to 163 reads/ sample. The abundant microbiome table was rarefied to 6250 reads/sample.

Doing this allowed us to study if the rare/ less abundant members of the microbiome community and the more abundant members were being affected differently.

Core microbiome

We computed the core microbiome within samples grown in home soil alone. This allowed us to look at the unique OTUs present in each species and those that are shared by all 6 species. We used 2 different definitions of “core” microbiome. The first was based on a presence absence matrix alone the second used an abundance threshold too.

Using presence absence matrix, an OTU was considered a part of the core if it was present in at least 2 or 3 samples. We also made use of an abundance threshold (0. 00001% of all reads) to select which OTUs would be a part of the presence absence matrix. Abundances were normalized to library sizes. Percentage of common OTUs was visualized by plotting a Euler grid. UpsetR, an R library was used to plot the grid. Taxonomies of common OTUs were saved as a table.

Diversity analysis

Alpha (within samples) and beta diversity (between samples) were used to estimate microbial community diversity. Alpha diversity was measured using Chao1, Observed species and Shannon metrics. All samples were rarefied to 160 and 500 reads/sample to keep sampling size the same. A permanently set seed was used to make results reproducible. We carried out an Anova to test if species richness of microbiome community depends on soil origin and host species. However, the rarified data lead to inconsistent diversity estimates. Also, results of Anova were depended heavily on the diversity metric chosen. We imported the OTU tables into QIIME to plot rarefaction curves. A minimum of 10 reads/sample with increments of 15 reads/sample was used as parameters.

Beta diversity analyses were carried out using Bray-Curtis distance on a count normalized OTU table and were visualized by different ordination techniques. R package “Phyloseq” (1.16.2) was used for principal coordinate analysis. Non parametric permutation test Adonis with 999 permutations was used on bray-curtis dissimilarity matrix to test between sample similarity and factors affecting it. As factors we included Soil (the host conditioning the soil), Species (the host currently planted), Host_PD (Pairwise distance matrix between host conditioning the soil and host currently planted), Nod_size (Nodule size of the hosts).

Ordination techniques

PCoA: Principal coordinate analysis

PCoA also called classical multidimensional scaling is a distance-based ordination

method that can be performed via the ordination function in Phyloseq. The major benefit of PCoA is the ability to choose a different distance measure.

CCA: Canonical Correspondence Analysis

CCA can be used to explore the relationship between two sets of variables. This is particularly useful as we make use of this ordination technique to explore how the Species and Soil variable affects microbial community structure. Further, by plotting out the eigenvalues of the constrained variables we can visualize how individual groups are clustering the samples.

NMDS: (Non metric MultiDimensional Scaling)

NMDS is a rank based approach, which substitutes the object distances with ranks. The NMDS algorithm introduces a parameter called “stress” which is used to measure the lack of fit between the object distances and calculated distance matrix. Then, all objects are iteratively repositioned to minimize the stress parameter or lack of fit. A total of 20 iterations were run to identify best fit. We made use of NMDS ordination to explore the relationships between microbial members themselves.

Sloan neutral community analysis

We wanted to test if our communities were neutrally distributed and if there were members within our communities that were selected for or against by the host. We made use of the Sloan neutral community assembly model to test our data. This model allows us to predict the relationship between the abundance of microbiome members

and their presence across samples. Our experimental design made use of growing plants in soil conditioned by the same species (home soil) and in a soil conditioned by another species (away soil). First the complete microbiome table was rarefied to 6500 reads/ sample and then samples were split into home and away. Samples where new species were the same as the ancestor were termed as home whereas samples where new species were different compared to ancestor were termed as away. The model was run to test the fit of the microbiome to the neutral model.

APPENDICES

APPENDIX 1: List of tables

Table 1: Techniques used to study microbial population structures

Technique	Description	Advantage	Disadvantage	Throughput
Culturing on plates	Microbes are cultured and isolated on selective media	Cheap, can identify basic abundance patterns	Labor intensive, Needs more downstream work to identify taxa of isolated members	Low
PFLA	Isolation of the phospholipids, conversion of the phospholipid fatty acids to their corresponding fatty acid methyl esters (known by the acronym FAME) and the separation, identification and quantification of the FAME by gas chromatography.	Can measure microbial biomass, no prior knowledge of sample required	Labor intensive, taxonomic assignments are not possible, don't get abundance data, need expertise with GC MS	Medium
qPCR	Known 16s regions are amplified.	Phylogenetic relationships can be explored, can get abundance data	PCR bias, can't be used for novel species	Low-Medium
DGGE/TGGE	16s fragments are separated by gradient of temperature or denaturant	Good for profiling and exploring community diversity without prior knowledge of samples	Hard to standadize gradients from one gel to another, don't get abundance data, taxonomic assignments can't be generated	Low-Medium

Table 1: (cont'd)

FISH	Oligonucleotide probes are designed for known 16s regions. Probes are fluorescently labelled.	Phylogenetic relationships can be explored, can get abundance data	Need to know which members are present to design probes or use general set of probes, Taxonomic assignments are hard	Medium-High

Table 2: Read distribution across samples.

SampleID	Soil	Species	Comb	TotalReads	Reads/sample
master.48.F4	Bar	Bar	Bar	189849	342329
master.48.E4	Bar	Bar	Bar	106421	
master.48.G4	Bar	Bar	Bar	6637	
master.48.H4	Bar	Bar	Bar	39422	
master.48.B6	Bar	Bif	Bar_Bif	98968	371526
master.48.A6	Bar	Bif	Bar_Bif	92645	
master.48.D6	Bar	Bif	Bar_Bif	92405	
master.48.C6	Bar	Bif	Bar_Bif	87508	
master.48.G6	Bar	Fuc	Bar_Fuc	128604	594061
master.48.H6	Bar	Fuc	Bar_Fuc	232217	
master.48.F6	Bar	Fuc	Bar_Fuc	180828	
master.48.E6	Bar	Fuc	Bar_Fuc	52412	
master.48.H5	Bar	Mac	Bar_Mac	117023	319847
master.48.E5	Bar	Mac	Bar_Mac	18793	
master.48.F5	Bar	Mac	Bar_Mac	162214	
master.48.G5	Bar	Mac	Bar_Mac	21817	
master.48.A5	Bar	Mic	Bar_Mic	19272	226774
master.48.C5	Bar	Mic	Bar_Mic	46511	
master.48.D5	Bar	Mic	Bar_Mic	98248	
master.48.B5	Bar	Mic	Bar_Mic	62743	
master.48.D4	Bar	Wor	Bar_Wor	154560	270937
master.48.C4	Bar	Wor	Bar_Wor	51693	
master.48.A4	Bar	Wor	Bar_Wor	27108	
master.48.B4	Bar	Wor	Bar_Wor	37576	
master.96.G7	Bif	Bar	Bif_Bar	126863	327944
master.96.G8	Bif	Bar	Bif_Bar	80669	
master.96.G6	Bif	Bar	Bif_Bar	63119	
master.96.G5	Bif	Bar	Bif_Bar	57293	
master.96.H7	Bif	Bif	Bif	186780	361000
master.96.H5	Bif	Bif	Bif	40991	
master.96.H8	Bif	Bif	Bif	93458	
master.96.H6	Bif	Bif	Bif	39771	
master.96.H11	Bif	Fuc	Bif_Fuc	131472	319272

Table 2: (cont'd)

master.96.H12	Bif	Fuc	Bif_Fuc	89609	
master.96.H10	Bif	Fuc	Bif_Fuc	61811	
master.96.H9	Bif	Fuc	Bif_Fuc	36380	
master.96.H1	Bif	Mac	Bif_Mac	111121	303176
master.96.H2	Bif	Mac	Bif_Mac	82313	
master.96.H3	Bif	Mac	Bif_Mac	76418	
master.96.H4	Bif	Mac	Bif_Mac	33324	
master.96.G11	Bif	Mic	Bif_Mic	123254	413155
master.96.G9	Bif	Mic	Bif_Mic	80630	
master.96.G12	Bif	Mic	Bif_Mic	54401	
master.96.G10	Bif	Mic	Bif_Mic	154870	
master.96.G3	Bif	Wor	Bif_Wor	73225	607602
master.96.G4	Bif	Wor	Bif_Wor	73347	
master.96.G1	Bif	Wor	Bif_Wor	213251	
master.96.G2	Bif	Wor	Bif_Wor	247779	
master.48.G1	Fuc	Bar	Fuc_Bar	20758	132620
master.48.E1	Fuc	Bar	Fuc_Bar	44726	
master.48.H1	Fuc	Bar	Fuc_Bar	25048	
master.48.F1	Fuc	Bar	Fuc_Bar	42088	
master.48.B3	Fuc	Bif	Fuc_Bif	33938	463251
master.48.C3	Fuc	Bif	Fuc_Bif	243008	
master.48.D3	Fuc	Bif	Fuc_Bif	98047	
master.48.A3	Fuc	Bif	Fuc_Bif	88258	
master.48.G3	Fuc	Fuc	Fuc	21970	382749
master.48.E3	Fuc	Fuc	Fuc	227933	
master.48.F3	Fuc	Fuc	Fuc	116516	
master.48.H3	Fuc	Fuc	Fuc	16330	
master.48.H2	Fuc	Mac	Fuc_Mac	38522	342113
master.48.G2	Fuc	Mac	Fuc_Mac	88391	
master.48.E2	Fuc	Mac	Fuc_Mac	172485	
master.48.F2	Fuc	Mac	Fuc_Mac	42715	
master.48.C2	Fuc	Mic	Fuc_Mic	438811	568420
master.48.D2	Fuc	Mic	Fuc_Mic	54160	
master.48.A2	Fuc	Mic	Fuc_Mic	61288	
master.48.B2	Fuc	Mic	Fuc_Mic	14161	
master.48.D1	Fuc	Wor	Fuc_Wor	198303	449714

Table 2: (cont'd)

master.48.C1	Fuc	Wor	Fuc_Wor	174940	
master.48.A1	Fuc	Wor	Fuc_Wor	16656	
master.48.B1	Fuc	Wor	Fuc_Wor	59815	
master.96.C8	Mac	Bar	Mac_Bar	37786	166469
master.96.C7	Mac	Bar	Mac_Bar	94665	
master.96.C6	Mac	Bar	Mac_Bar	27472	
master.96.C5	Mac	Bar	Mac_Bar	6546	
master.96.D8	Mac	Bif	Mac_Bif	60178	154224
master.96.D7	Mac	Bif	Mac_Bif	71358	
master.96.D6	Mac	Bif	Mac_Bif	15733	
master.96.D5	Mac	Bif	Mac_Bif	6955	
master.96.D11	Mac	Fuc	Mac_Fuc	79379	284968
master.96.D12	Mac	Fuc	Mac_Fuc	49608	
master.96.D9	Mac	Fuc	Mac_Fuc	66328	
master.96.D10	Mac	Fuc	Mac_Fuc	89653	
master.96.D1	Mac	Mac	Mac	38213	306450
master.96.D2	Mac	Mac	Mac	36878	
master.96.D4	Mac	Mac	Mac	112389	
master.96.D3	Mac	Mac	Mac	118970	
master.96.C10	Mac	Mic	Mac_Mic	130701	455614
master.96.C9	Mac	Mic	Mac_Mic	217853	
master.96.C12	Mac	Mic	Mac_Mic	37171	
master.96.C11	Mac	Mic	Mac_Mic	69889	
master.96.C3	Mac	Wor	Mac_Wor	111235	694542
master.96.C2	Mac	Wor	Mac_Wor	34134	
master.96.C4	Mac	Wor	Mac_Wor	462142	
master.96.C1	Mac	Wor	Mac_Wor	87031	
master.96.A7	Mic	Bar	Mic_Bar	53540	256638
master.96.A8	Mic	Bar	Mic_Bar	113895	
master.96.A5	Mic	Bar	Mic_Bar	73570	
master.96.A6	Mic	Bar	Mic_Bar	15633	
master.96.B7	Mic	Bif	Mic_Bif	77289	483865
master.96.B6	Mic	Bif	Mic_Bif	269418	
master.96.B8	Mic	Bif	Mic_Bif	62763	
master.96.B5	Mic	Bif	Mic_Bif	74395	
master.96.B9	Mic	Fuc	Mic_Fuc	146436	385375

Table 2: (cont'd)

master.96.B11	Mic	Fuc	Mic_Fuc	130256	
master.96.B10	Mic	Fuc	Mic_Fuc	47123	
master.96.B12	Mic	Fuc	Mic_Fuc	61560	
master.96.B2	Mic	Mac	Mic_Mac	41173	139607
master.96.B4	Mic	Mac	Mic_Mac	29519	
master.96.B1	Mic	Mac	Mic_Mac	53367	
master.96.B3	Mic	Mac	Mic_Mac	15548	
master.96.A9	Mic	Mic	Mic	48316	250180
master.96.A10	Mic	Mic	Mic	78005	
master.96.A12	Mic	Mic	Mic	18385	
master.96.A11	Mic	Mic	Mic	105474	
master.96.A1	Mic	Wor	Mic_Wor	92763	302158
master.96.A2	Mic	Wor	Mic_Wor	114872	
master.96.A4	Mic	Wor	Mic_Wor	43467	
master.96.A3	Mic	Wor	Mic_Wor	51056	
master.96.E8	Wor	Bar	Wor_Bar	172014	580237
master.96.E7	Wor	Bar	Wor_Bar	113958	
master.96.E5	Wor	Bar	Wor_Bar	272356	
master.96.E6	Wor	Bar	Wor_Bar	21909	
master.96.F7	Wor	Bif	Wor_Bif	130342	477186
master.96.F8	Wor	Bif	Wor_Bif	93743	
master.96.F6	Wor	Bif	Wor_Bif	182642	
master.96.F5	Wor	Bif	Wor_Bif	70459	
master.96.F9	Wor	Fuc	Wor_Fuc	50120	339721
master.96.F10	Wor	Fuc	Wor_Fuc	65384	
master.96.F12	Wor	Fuc	Wor_Fuc	80574	
master.96.F11	Wor	Fuc	Wor_Fuc	143643	
master.96.F2	Wor	Mac	Wor_Mac	179688	477332
master.96.F4	Wor	Mac	Wor_Mac	60320	
master.96.F3	Wor	Mac	Wor_Mac	88943	
master.96.F1	Wor	Mac	Wor_Mac	148381	
master.96.E10	Wor	Mic	Wor_Mic	73516	360102
master.96.E9	Wor	Mic	Wor_Mic	46202	
master.96.E11	Wor	Mic	Wor_Mic	201793	
master.96.E12	Wor	Mic	Wor_Mic	38591	

Table 2: (cont'd)

master.96.E4	Wor	Wor	Wor	117697	558645
master.96.E1	Wor	Wor	Wor	242639	
master.96.E2	Wor	Wor	Wor	122870	
master.96.E3	Wor	Wor	Wor	75439	

Table 3a: Alpha diversity modelling results using the complete microbiome table

Factor	Rarefaction depth	Metric	Df	Sum sq	Mean sq	F value	pvalue
Soil	160	Observed	5	63	12.594	2.147	0.0653
Species	160	Observed	5	42.2	8.444	1.44	0.2159
Soil:Species	160	Observed	25	247.9	9.918	1.691	0.0344
Residuals	160	Observed	108	633.5	5.866		
Soil	160	Shannon	5	1082	216.31	2.371	0.0439
Species	160	Shannon	5	375	74.94	0.822	0.5369
Soil:Species	160	Shannon	25	2693	107.7	1.181	0.2738
Residuals	160	Shannon	108	9852	91.22		
Soil	160	Chao1	5	0.1133	0.02265	1.648	0.1534
Species	160	Chao1	5	0.1052	0.02104	1.531	0.1863
Soil:Species	160	Chao1	25	0.5278	0.02111	1.536	0.0686
Residuals	160	Chao1	108	1.4842	0.01374		
Soil	500	Observed	5	138	27.59	1.175	0.326
Species	500	Observed	5	162.6	32.52	1.385	0.235
Soil:Species	500	Observed	25	731.8	29.27	1.247	0.217
Residuals	500	Observed	108	2535.2	23.47		
Soil	500	Shannon	5	1556	311.2	1.113	0.358
Species	500	Shannon	5	1616	323.3	1.156	0.336
Soil:Species	500	Shannon	25	7155	286.2	1.024	0.444
Residuals	500	Shannon	108	30196	279.6		
Soil	500	Chao1	5	0.1081	0.02161	1.404	0.228
Species	500	Chao1	5	0.1321	0.02642	1.717	0.137
Soil:Species	500	Chao1	25	0.3876	0.0155	1.007	0.465
Residuals	500	Chao1	108	1.6619	0.01539		
Soil	1000	Observed	5	211	42.22	0.986	0.43
Species	1000	Observed	5	263	52.56	1.227	0.301
Soil:Species	1000	Observed	25	1021	40.83	0.954	0.534
Residuals	1000	Observed	108	4624	42.82		
Soil	1000	Shannon	5	4802	960.4	0.994	0.425
Species	1000	Shannon	5	3747	749.4	0.776	0.569
Soil:Species	1000	Shannon	25	20789	831.6	0.861	0.656
Residuals	1000	Shannon	108	104316	965.9		
Soil	1000	Chao1	5	0.0996	0.01991	1.386	0.2353
Species	1000	Chao1	5	0.1387	0.02773	1.93	0.0951
Soil:Species	1000	Chao1	25	0.3407	0.01363	0.948	0.5404
Residuals	1000	Chao1	108	1.5517	0.01437		

Table 3b: Alpha diversity modelling results using the abundant microbiome table

Factor	Rarefaction depth	Metric	Df	Sum sq	Mean sq	F value	pvalue
Soil	160	Observed	5	8.83	1.7667	1.998	0.0847
Species	160	Observed	5	2.42	0.4833	0.547	0.7406
Soil:Species	160	Observed	25	43	1.72	1.945	0.0103
Residuals	160	Observed	108	95.5	0.8843		
Soil	160	Shannon	5	39.1	7.82	2.413	0.0408
Species	160	Shannon	5	6.8	1.366	0.421	0.833
Soil:Species	160	Shannon	25	155.6	6.225	1.92	0.0116
Residuals	160	Shannon	108	350.1	3.241		
Soil	160	Chao1	5	0.01434	0.002869	1.416	0.2243
Species	160	Chao1	5	0.00615	0.00123	0.607	0.6947
Soil:Species	160	Chao1	25	0.09245	0.003698	1.825	0.0183
Residuals	160	Chao1	108	0.2188	0.002026		
Soil	500	Observed	5	4.23	0.8458	0.448	0.814
Species	500	Observed	5	10.81	2.1625	1.146	0.341
Soil:Species	500	Observed	25	47.15	1.8858	1	0.474
Residuals	500	Observed	108	203.75	1.8866		
Soil	500	Shannon	5	26.9	5.39	0.487	0.786
Species	500	Shannon	5	46.6	9.317	0.841	0.523
Soil:Species	500	Shannon	25	226.1	9.043	0.817	0.713
Residuals	500	Shannon	108	1196	11.074		
Soil	500	Chao1	5	0.00173	0.000346	0.331	0.893
Species	500	Chao1	5	0.00663	0.001326	1.266	0.284
Soil:Species	500	Chao1	25	0.03719	0.001488	1.421	0.111

Table 3b: (cont'd)

Residuals	500	Chao1	108	0.11305	0.001047		
Soil	1000	Observed	5	6.23	1.246	0.48	0.79
Species	1000	Observed	5	23.23	4.646	1.79	0.121
Soil:Species	1000	Observed	25	68.73	2.749	1.059	0.401
Residuals	1000	Observed	108	280.25	2.595		
Soil	1000	Shannon	5	78.8	15.76	0.775	0.57
Species	1000	Shannon	5	73.2	14.65	0.72	0.61
Soil:Species	1000	Shannon	25	408.9	16.36	0.804	0.729
Residuals	1000	Shannon	108	2197.2	20.34		
Soil	1000	Chao1	5	0.00057	0.000114	0.193	0.965
Species	1000	Chao1	5	0.00381	0.000762	1.289	0.274
Soil:Species	1000	Chao1	25	0.01632	0.000653	1.105	0.35
Residuals	1000	Chao1	108	0.06382	0.000591		

Table 3c: Alpha diversity modelling results using the rare microbiome table

Factor	Rarefaction depth	Metric	Df	Sum sq	Mean sq	F value	pvalue
Soil	160	Observed	5	391	78.2	3.354	0.00743
Species	160	Observed	5	210.7	42.15	1.808	0.11731
Soil:Species	160	Observed	25	734.5	29.38	1.26	0.20715
Residuals	160	Observed	108	2517.8	23.31		
Soil	160	Shannon	5	1797	359.4	5.187	0.000264
Species	160	Shannon	5	464	92.9	1.341	0.252622
Soil:Species	160	Shannon	25	1942	77.7	1.121	0.33291
Residuals	160	Shannon	108	7482	69.3		
Soil	160	Chao1	5	0.424	0.0848	1.468	0.206
Species	160	Chao1	5	0.495	0.09901	1.714	0.138
Soil:Species	160	Chao1	25	1.984	0.07935	1.373	0.135
Residuals	160	Chao1	108	6.239	0.05777		

Table 4a: Distribution of core OTUs when only presence absence in 2 samples is used as a threshold.

OTUs	mac_a_freq	mic_a_freq	wor_a_freq	fuc_a_freq	bif_a_freq	bar_a_freq	Genus
102142	1	1	1	1	1	1	g__Rhizobium
1068486	1	1	1	1	1	1	g__
1119924	1	1	1	1	1	1	g__
1124017	1	1	1	1	1	1	g__Streptomyces
134359	1	1	1	1	1	1	g__Agrobacterium
136077	1	1	1	1	1	1	NA
144740	1	1	1	1	1	1	g__
150328	1	1	1	1	1	1	NA
1546879	1	1	1	1	1	1	g__
180756 1	1	1	1	1	1	1	g__
191698	1	1	1	1	1	1	g__Mycobacterium
204795	1	1	1	1	1	1	g__
210746	1	1	1	1	1	1	g__
212186	1	1	1	1	1	1	NA
217917	1	1	1	1	1	1	g__
220539	1	1	1	1	1	1	g__Agrobacterium
226906	1	1	1	1	1	1	g__

Table 4.a: (cont'd)

226964	1	1	1	1	1	1	g__Rhizobium
227191	1	1	1	1	1	1	g__Rhizobium
239819	1	1	1	1	1	1	NA
2675752	1	1	1	1	1	1	g__Rhizobium
275052	1	1	1	1	1	1	NA
279206	1	1	1	1	1	1	g__
322972	1	1	1	1	1	1	NA
3334369	1	1	1	1	1	1	g__
339053	1	1	1	1	1	1	g__
34879	1	1	1	1	1	1	g__
3555041	1	1	1	1	1	1	g__
360124	1	1	1	1	1	1	g__
360253	1	1	1	1	1	1	g__Sphingomonas
3611663	1	1	1	1	1	1	g__Uliginosibacterium
3746876	1	1	1	1	1	1	g__
40073	1	1	1	1	1	1	NA
4254478	1	1	1	1	1	1	NA
4261880	1	1	1	1	1	1	NA
4295043	1	1	1	1	1	1	g__
4321864	1	1	1	1	1	1	g__
4329093	1	1	1	1	1	1	g__Agrobacterium
4344241	1	1	1	1	1	1	g__Mesorhizobium
4346059	1	1	1	1	1	1	g__
4348172	1	1	1	1	1	1	g__Rhizobium
4371517	1	1	1	1	1	1	g__

Table 4.a: (cont'd)

4376885	1	1	1	1	1	1	g__Rhizobium
4376886	1	1	1	1	1	1	g__Rhizobium
4394922	1	1	1	1	1	1	g__
4396481	1	1	1	1	1	1	g__Rhizobium
4412141	1	1	1	1	1	1	g__
4424553	1	1	1	1	1	1	g__Kaistia
4426965	1	1	1	1	1	1	g__
4455570	1	1	1	1	1	1	g__
4476548	1	1	1	1	1	1	NA
57713	1	1	1	1	1	1	g__Fimbriimonas
589483	1	1	1	1	1	1	NA
628400	1	1	1	1	1	1	NA
654155	1	1	1	1	1	1	g__Phenylobacterium
792073	1	1	1	1	1	1	g__Rhizobium
800671	1	1	1	1	1	1	g__
993711	1	1	1	1	1	1	g__Pseudonocardia

Table 4b: Distribution of core OTUs when abundance along with presence absence in 2 samples is used as a threshold.

OTUs	mac_a_freq	mic_a_freq	wor_a_freq	fuc_a_freq	bif_a_freq	bar_a_freq	Genus
102142	1	1	1	1	1	1	g__Rhizobium
1068486	1	1	1	1	1	1	g__
1119924	1	1	1	1	1	1	g__
1124017	1	1	1	1	1	1	g__Streptomyces
134359	1	1	1	1	1	1	g__Agrobacterium
136077	1	1	1	1	1	1	NA
144740	1	1	1	1	1	1	g__
150328	1	1	1	1	1	1	NA
1546879	1	1	1	1	1	1	g__
180756 1	1	1	1	1	1	1	g__
191698	1	1	1	1	1	1	g__Mycobacterium
204795	1	1	1	1	1	1	g__
210746	1	1	1	1	1	1	g__
212186	1	1	1	1	1	1	NA
217917	1	1	1	1	1	1	g__
220539	1	1	1	1	1	1	g__Agrobacterium

Table 4b: (cont'd)

226906	1	1	1	1	1	1	g__
226964	1	1	1	1	1	1	g__Rhizobium
227191	1	1	1	1	1	1	g__Rhizobium
239819	1	1	1	1	1	1	NA
2675752	1	1	1	1	1	1	g__Rhizobium
275052	1	1	1	1	1	1	NA
279206	1	1	1	1	1	1	g__
322972	1	1	1	1	1	1	NA
3334369	1	1	1	1	1	1	g__
339053	1	1	1	1	1	1	g__
34879	1	1	1	1	1	1	g__
3555041	1	1	1	1	1	1	g__
360124	1	1	1	1	1	1	g__
360253	1	1	1	1	1	1	g__Sphingomonas
3611663	1	1	1	1	1	1	g__Uliginosibacterium
3746876	1	1	1	1	1	1	g__
40073	1	1	1	1	1	1	NA
4254478	1	1	1	1	1	1	NA
4261880	1	1	1	1	1	1	NA
4295043	1	1	1	1	1	1	g__
4321864	1	1	1	1	1	1	g__
4329093	1	1	1	1	1	1	g__Agrobacterium
4344241	1	1	1	1	1	1	g__Mesorhizobium
4346059	1	1	1	1	1	1	g__

Table 4b: (cont'd)

4348172	1	1	1	1	1	1	g__Rhizobium
4371517	1	1	1	1	1	1	g__
4376885	1	1	1	1	1	1	g__Rhizobium
4376886	1	1	1	1	1	1	g__Rhizobium
4394922	1	1	1	1	1	1	g__
4396481	1	1	1	1	1	1	g__Rhizobium
4412141	1	1	1	1	1	1	g__
4424553	1	1	1	1	1	1	g__Kaistia
4426965	1	1	1	1	1	1	g__
4455570	1	1	1	1	1	1	g__
4476548	1	1	1	1	1	1	NA
57713	1	1	1	1	1	1	g__Fimbriimonas
589483	1	1	1	1	1	1	NA
628400	1	1	1	1	1	1	NA
654155	1	1	1	1	1	1	g__Phenylobacterium
792073	1	1	1	1	1	1	g__Rhizobium
800671	1	1	1	1	1	1	g__
993711	1	1	1	1	1	1	g__Pseudonocardia

Table 5a: Adonis modelling results for complete microbiome

Factor	pvalue	F	pvalue	F	R2	pvalue
Home	0.0028	0.823				
Species	0.05961	0.043				
Soil	0.07711	0.006				
nod_size	0.00528	0.465				
	Soil		**Species**		**Soil:Species**	
Soil*Species	0.07711	0.003	0.05961	0.02	0.15806	0.522

Table 5b: Adonis modelling results for abundant microbiome.

Factor	R2	pvalue	R2	pvalue	R2	pvalue	R2	pvalue
Home	0.00164	0.953						
Species	0.07655	0.006						
Soil	0.05605	0.054						
nod_size	0.00561	0.496						
	Soil		**Species**					
Soil+Species	0.05605	0.040	0.07655	0.004				
	Soil		**Species**		**Soil:Species**			
Soil*Species	0.05605	0.025	0.07655	0.002	0.19574	0.079		

Table 5c: Adonis modelling results for rare microbiome.

	R2	pvalue	R2	pvalue	R2	pvalue
Home	0.011	0.047				
Soil	0.115	0.001				
Species	0.07676	0.001				
nod_size	0.01223	0.032				
	Soil		**Species**		**Soil:Species**	
Soil*Species	0.115	0.001	0.07676	0.001	0.183	0.001

Table 6: Model statistics for neutral model of away and home samples

	model fit	mci	rsquared	AIC	BIC	Samples	Richness
away	0.69	0.078	0.918	-1722.48	-1713.76	120	578
home	0.691	0.121	0.877	-597.97	-590.61	24	293

Table 7.a: OTUs that were selected for or against under the neutral model for all home samples

Selection	OTU.ID	p	freq	freq.pred	pred.lwr	Taxonomy
Negative	239819	0.00100122	0.416667	0.265074	0.130366	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales
Negative	279206	7.69E-05	0.25	0.158284	0.061886	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__Haliangiaceae; g__; s__
Negative	4365533	4.49E-05	0.25	0.086823	0.024713	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__Polyangiaceae; g__; s__
Negative	4426965	0.00100135	0.416667	0.296105	0.152338	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Methylophilales; f__Methylophilaceae; g__; s__
Negative	654155	0.00100135	0.583333	0.296105	0.152338	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__Phenylobacterium; s__
Positive	1068486	0.00200237	0.708333	0.532261	0.342418	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__; s__
Positive	111050	5.77E-05	0.166667	0.11469	0.03811	k__Bacteria; p__Actinobacteria; c__Thermoleophilia; o__Solirubrobacterales; f__; g__; s__
Positive	1124017	0.00200282	0.583333	0.621312	0.423621	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Streptomycetaceae; g__Streptomyces; s__
Positive	1360 77	5.77E-05	0.208333	0.11469	0.03811	k__Bacteria; p__TM7; c__TM7-3

Table 7.a: (cont'd)

Positive	1916 9898	7.05E-05	0.25	0.1435 4646	0.0535 1	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Mycobacteriaceae; g__Mycobacterium; s__
Positive	3601 2424	0.0003 3367	0.7916 6773	0.7088 7366	0.5088 66	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	3798 902	3.85E-05	0.1666 6767	0.0733 2477	0.0188 77	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Methylocystaceae; g__; s__
Positive	4007 73	0.0001 3567	0.2916 6705	0.2961 0538	0.1523 38	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae
Positive	4295 043	0.0001 5433	0.5833 3303	0.3425 6629	0.1866 29	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__; s__
Positive	4321 864	0.0001 2233	0.4583 3374	0.2650 6666	0.1303 66	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__; s__
Positive	4408 890	5.13E-05	0.25	0.1006 2	0.0311 35	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__; s__
Positive	4424 553	8.97E-05	0.375	0.1882 6229	0.0798 29	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Kaistia; s__
Positive	4473 178	0.0001 9967	0.2916 6743	0.4478 4343	0.2701 87	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Burkholderiaceae; g__; s__
Positive	7346	2.56E-05	0.0833 33	0.0473 1975	0.0091 75	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales
Positive	8151 02	8.33E-05	0.25	0.1731 9797	0.0706 69	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__; s__

Table 7.b: OTUs that were selected for or against under the neutral model for all away samples.

	OTU.ID	Abundance	freq	freq.pred	pred.lwr	pred.upr	Taxonomy
Negative	1119492	0.000463	0.65	0.8116	0.7690	0.9427	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
Negative	1124017	0.000333	0.5	0.7086	0.6222	0.7982	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Streptomycetaceae; g__Streptomyces; s__
Negative	134359	0.000658	0.8	0.9555	0.9002	0.9942	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Agrobacterium; s__
Negative	204795	0.000383	0.45	0.7006	0.6119	0.7880	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Negative	210746	0.000979	0.8	0.9597	0.9064	0.9942	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Oxalobacteraceae; g__; s__
Negative	212186	0.001363	0.9	0.9968	0.9964	1	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Bradyrhizobiaceae
Negative	226906	0.002021	0.9	1	1	1	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
Negative	361182	0.002103	0.9	1	1	1	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Rhodocyclales; f__Rhodocyclaceae; g__Uliginosibacterium; s__

Table 7.b: (cont'd)

	37	0.0			0.9	0.9	
Ne	46	01		0.9	68	99	k__Bacteria; p__Proteobacteria;
gat	87	20	0.9	99	40	99	c__Alphaproteobacteria; o__Rickettsiales;
ive	6	8	25	7	2	7	f__mitochondria; g__; s__
	42		0.9		0.9		k__Bacteria; p__Proteobacteria;
Ne	54	0.0	41		68		c__Betaproteobacteria;
gat	47	01	66		98		o__Burkholderiales;
ive	8	96	7	1	1	1	f__Comamonadaceae
	43	0.0	0.8	0.9		0.9	
Ne	46	00	16	81	0.9	94	k__Bacteria; p__Bacteroidetes;
gat	05	74	66	16	38	45	c__Cytophagia; o__Cytophagales;
ive	9	1	7	8	03	5	f__Cytophagaceae; g__; s__
	44	0.0		0.5	0.4	0.6	k__Bacteria; p__Proteobacteria;
Ne	26	00		86	97	70	c__Betaproteobacteria;
gat	96	26	0.4	93	47	99	o__Methylophilales; f__Methylophilaceae;
ive	5	4	5	3	4	8	g__; s__
	44			0.9	0.9	0.9	k__Bacteria; p__Proteobacteria;
Ne	60	0.0		96	62	99	c__Betaproteobacteria;
gat	87	00	0.4	71	89	71	o__Methylophilales; f__Methylophilaceae;
ive	1	95	5	6	8	8	g__; s__
		0.0	0.9	0.9	0.9	0.9	k__Bacteria; p__Proteobacteria;
Ne	58	00	41	97	63	99	c__Betaproteobacteria;
gat	94	97	66	26	87	79	o__Burkholderiales;
ive	83	1	7	1	4	9	f__Comamonadaceae
	10		0.0	0.0	0.0	0.0	
Po	13	1.7	91	32	12	81	k__Bacteria; p__Bacteroidetes;
siti	95	9E	66	42	52	33	c__[Saprospirae]; o__[Saprospirales];
ve	4	-05	7	5	6	2	f__Chitinophagaceae; g__; s__
		0.0	0.8	0.7	0.6	0.7	k__Bacteria; p__Proteobacteria;
Po	10	00	58	20	34	93	c__Alphaproteobacteria; o__Rhizobiales;
siti	21	34	33	55	38	03	f__Rhizobiaceae; g__Rhizobium;
ve	42	1	3	1	2	6	s__leguminosarum
	10	0.0	0.7	0.6	0.5	0.6	k__Bacteria; p__Bacteroidetes;
Po	68	00	16	11	22	94	c__Sphingobacteriia;
siti	48	27	66	60	23	05	o__Sphingobacteriales;
ve	6	7	7	7	6	4	f__Sphingobacteriaceae; g__; s__

Table 7.b: (cont'd)

Positive	1072352	8.9E-06	0.058333	0.01578	0.00420	0.057	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__; g__; s__
Positive	1076405	3.21E-05	0.125	0.06015	0.02973	0.171783	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	1078362	2.95E-05	0.133333	0.05497	0.00226	0.011234	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__; s__
Positive	1106411	3.97E-05	0.208333	0.07600	0.00406	0.137646	k__Bacteria; p__Armatimonadetes; c__Armatimonadia; o__FW68; f__; g__; s__
Positive	1110150	3.59E-05	0.175	0.06801	0.03509	0.274741	k__Bacteria; p__Actinobacteria; c__Thermoleophilia; o__Solirubrobacterales; f__; g__; s__
Positive	1118089	1.92E-05	0.1	0.03486	0.01391	0.008468	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__Devosia; s__
Positive	1119031	6.28E-05	0.225	0.12605	0.07811	0.097332	k__Bacteria; p__Actinobacteria; c__Thermoleophilia; o__Solirubrobacterales; f__Solirubrobacteraceae; g__; s__
Positive	1134053	4.36E-05	0.133333	0.08410	0.04610	0.047559	k__Bacteria; p__TM7; c__TM7-3; o__; f__; g__; s__
Positive	1137888	2.44E-05	0.1	0.04480	0.01984	0.009977	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
Positive	1162272	1.67E-05	0.083333	0.02999	0.01183	0.007797	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Frankiaceae; g__; s__

Table 7.b: (cont'd)

Positive	1293 4	1.79E-05	0.1	0.032 425	0.012 526	0.081 332	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Pseudonocardiaceae; g__Pseudonocardia; s__
Positive	1447 40	7.69E-05	0.3083 333	0.1581 281	0.1037 27	0.2340 34	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__; s__
Positive	1503 28	5.00E-05	0.2025	0.0978 41	0.0565 23	0.1641 08	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae
Positive	1546 879	7.05E-05	0.2754	0.1434 4	0.0918 66	0.2217 337	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Micromonosporaceae; g__; s__
Positive	1553 204	2.95E-05	0.1416 67	0.0549 78	0.0263 33	0.1110 234	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Sphingomonas
Positive	1807 561	7.56E-05	0.2833 333	0.1553 19	0.0132 98	0.2306 93	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	1916 98	8.21E-05	0.3416 67	0.1701 96	0.1134 53	0.2047 4	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Mycobacteriaceae; g__Mycobacterium; s__
Positive	2154 23	1.67E-05	0.1	0.0292 998	0.0118 83	0.0777 3	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__Rhodoplanes; s__
Positive	2271 91	0.000 565	0.975	0.9306 06	0.8705 42	0.9563 7	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum
Positive	2398 19	0.000 129	0.475	0.2836 64	0.2107	0.3704 09	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales

Table 7.b: (cont'd)

Positive	243118	1.15E-05	0.0666667	0.020453	0.006298	0.06435	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	2568663	1.15E-05	0.0666667	0.020453	0.006298	0.06435	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Rhodocyclales; f__Rhodocyclaceae; g__; s__
Positive	2675752	0.000046	0.95	0.860893	0.78773	0.911666	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum
Positive	273185	1.28E-05	0.0833333	0.022814	0.007439	0.06779 2	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__Devosia; s__
Positive	275052	0.00006 5	0.991667	0.962207	0.91135 2	0.98438 8	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae
Positive	279206	4.10E-05	0.175	0.07869	0.04256 9	0.14094 9	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__Haliangiaceae; g__; s__
Positive	287547	1.79E-05	0.0833333	0.032425	0.01252 6	0.08133 2	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Burkholderiaceae; g__Burkholderia; s__
Positive	2984014	2.69E-05	0.116667	0.04986	0.02302 5	0.10462 1	k__Bacteria; p__Chloroflexi; c__TK10; o__B07_WMSP1
Positive	324252	5.51E-05	0.2583333	0.109033	0.06494 3	0.17737 8	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria
Positive	3798902	2.44E-05	0.1333333	0.04802	0.01984 7	0.09799 7	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Methylocystaceae; g__; s__

Table 7.b: (cont'd)

Positive	4007 3	4.36E-05	0.208333	0.008410 3	0.004644 8	0.147559	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae
Positive	4178575	1.92E-05	0.009166 7	0.003486 9	0.001391 3	0.008468 2	k__Bacteria; p__Acidobacteria; c__Solibacteres; o__Solibacterales; f__Solibacteraceae; g__; s__
Positive	4261880	4.36E-05	0.225	0.008410 3	0.004644 8	0.147559	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae
Positive	4295043	0.000145	0.5	0.320861	0.201544 5	0.408282	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__; s__
Positive	4303161	1.15E-05	0.066667	0.002045 3	0.000629 8	0.006435 7	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__Propionicimonas; s__
Positive	4303530	2.31E-05	0.1	0.004296 6	0.001830 9	0.009467 8	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
Positive	4314391	2.82E-05	0.141667	0.005241 2	0.002466 2	0.010792 9	k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__EB1017; g__; s__
Positive	4321864	0.000019	0.533333	0.427193	0.301966	0.516129	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__; s__
Positive	4327907 6	2.05E-05	0.108333 3	0.003732 9	0.001534 2	0.008802 2	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__; s__
Positive	4329093	4.23E-05	0.175	0.008139 39	0.004498 8	0.147559 3	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Agrobacterium; s__

Table 7.b: (cont'd)

Positive	4344241	6.28E-05	0.225	0.1265	0.07811	0.19733	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Phyllobacteriaceae; g__Mesorhizobium; s__
Positive	4348172	5.64E-05	0.3	0.1185	0.06709	0.18807	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum
Positive	4365533	5.05E-05	0.225	0.0978	0.05623	0.16408	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__Polyangiaceae; g__; s__
Positive	4371517	0.00017	0.467	0.3916	0.3086	0.4068	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__; g__; s__
Positive	4376885	2.95E-05	0.175	0.0978	0.02633	0.12344	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum
Positive	4376886	5.05E-05	0.233	0.0978	0.05623	0.16408	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum
Positive	4390891	8.97E-06	0.058333	0.0158	0.00420	0.05754	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	4391318	5.64E-05	0.25	0.1185	0.06709	0.18807	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__Sphingobacteriaceae; g__; s__
Positive	4394922	5.13E-05	0.225	0.0100	0.00623	0.16742	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Ellin6067; f__; g__; s__
Positive	4396481	6.54E-05	0.333	0.1318	0.08263	0.20305	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum

Table 7.b: (cont'd)

Positive	4400328	2.4E-05	0.125	0.044802	0.019847	0.097599	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__; g__; s__
Positive	4408890	4.87E-05	0.216667	0.095407	0.054607	0.160795	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__; s__
Positive	4412141	0.000101	0.358333	0.215673	0.151746	0.246524	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__; s__
Positive	4417475	2.18E-05	0.108333	0.039805	0.016805	0.091353	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__; g__; s__
Positive	4418695	2.56E-05	0.166667	0.047321	0.021042	0.101311	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Sphingomonas
Positive	4424055	0.000100	0.466667	0.215673	0.151746	0.246524	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Kaistia; s__
Positive	4427861	2.05E-05	0.166667	0.037325	0.015902	0.088202	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__Rhodoplanes; s__elegans
Positive	4429884	1.54E-05	0.125	0.027095	0.009888	0.074595	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Bradyrhizobiaceae; g__; s__
Positive	4430568	1.92E-05	0.125	0.034869	0.013913	0.084232	k__Bacteria; p__Acidobacteria; c__Acidobacteriia; o__Acidobacteriales; f__Acidobacteriaceae; g__; s__
Positive	4434889	2.31E-05	0.166667	0.042916	0.018309	0.094678	k__Bacteria; p__Acidobacteria; c__Sva0725; o__Sva0725; f__; g__; s__

Table 7.b: (cont'd)

Positive	44 45 43 2	2.0 5E -05	0.1	0.0 37 32 9	0.0 15 34	0.0 88 02 2	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	44 56 11 3	2.4 4E -05	0.1	0.0 44 80 2	0.0 19 84 7	0.0 97 99 7	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Rhodoferax; s__
Positive	44 58 21 4	1.9 2E -05	0.0 91 66 7	0.0 34 86 9	0.0 13 91 3	0.0 84 68 2	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__; s__
Positive	44 63 76 7	1.7 9E -05	0.0 83 33 3	0.0 32 42 5	0.0 12 52 6	0.0 81 33 2	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__; s__
Positive	44 65 15 8	8.9 7E -06	0.0 58 33 3	0.0 15 78 2	0.0 04 20 5	0.0 57 4	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Kaistobacter; s__
Positive	44 65 43 1	8.9 7E -06	0.0 58 33 3	0.0 15 78 2	0.0 04 20 5	0.0 57 4	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Methylocystaceae; g__; s__
Positive	44 76 08 9	2.1 8E -05	0.1 08 33 3	0.0 39 80 5	0.0 16 80 6	0.0 91 35 3	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__; s__
Positive	44 76 54 8	0.0 00 11 8	0.4 5	0.2 55 75 5	0.1 86 11 9	0.3 40 54 3	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Phyllobacteriaceae
Positive	44 78 41 3	4.7 4E -05	0.1 91 66 7	0.0 92 31 1	0.0 52 43 1	0.1 57 48 4	k__Bacteria; p__Bacteroidetes; c__[Saprospirae]; o__[Saprospirales]; f__Chitinophagaceae; g__; s__
Positive	44 80 02 2	2.3 1E -05	0.1 08 33 3	0.0 42 29 6	0.0 18 30 9	0.0 94 67 8	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae

Table 7.b: (cont'd)

Positive	4943667	1.03E-05	0.066667	0.018109	0.005218	0.006089	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae
Positive	5434657	1.54E-05	0.091667	0.027586	0.009885	0.007459	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__; s__
Positive	5674886	1.92E-05	0.116667	0.034869	0.013913	0.008482	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Dokdonella; s__
Positive	5771 3	1.54E-05	0.091667	0.027586	0.009885	0.007459	k__Bacteria; p__Armatimonadetes; c__[Fimbriimonadia]; o__[Fimbriimonadales]; f__[Fimbriimonadaceae]; g__Fimbriimonas; s__
Positive	6104 18	8.97E-06	0.058333	0.015782	0.004205	0.005744	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Hyphomicrobiaceae; g__Rhodoplanes; s__
Positive	6284 00	6.03E-05	0.291667	0.120389	0.073654	0.190674	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae
Positive	6541 55	0.000228	0.675	0.513094	0.424655	0.600719	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__Phenylobacterium; s__
Positive	6568 9	7.56E-05	0.266667	0.155313	0.101328	0.230693	k__Bacteria; p__Chloroflexi; c__Chloroflexi; o__[Roseiflexales]; f__[Kouleothrixaceae]; g__; s__
Positive	6668 83	3.72E-05	0.141667	0.070666	0.036926	0.131042	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Bdellovibrionales; f__Bdellovibrionaceae; g__Bdellovibrio; s__

Table 7.b: (cont'd)

Positive	76 71 42	0.000 10 4	0.3 25	0.2 21 81	0.1 56 9	0.304 11 8	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae
Positive	79 20 73	0.000 66 39	0.891 7	0.787 17 8	0.705 63 4	0.850 90 6	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Rhizobium; s__leguminosarum
Positive	80 06 71	5.38E-05	0.258 33 3	0.106 21 9	0.062 81	0.174 05 8	k__Bacteria; p__Actinobacteria; c__Acidimicrobiia; o__Acidimicrobiales; f__C111; g__; s__
Positive	81 51 02	4.36E-05	0.1 75	0.084 10 3	0.046 44 8	0.147 55 9	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Nocardioidaceae; g__; s__
Positive	81 77 14	3.72E-05	0.1 5	0.070 66 6	0.036 92 6	0.131 04 2	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Myxococcales; f__; g__; s__
Positive	86 95 13	1.54E-05	0.083 33 3	0.027 58 6	0.009 88 5	0.074 59 5	k__Bacteria; p__Bacteroidetes; c__Sphingobacteriia; o__Sphingobacteriales; f__; g__; s__
Positive	99 37 11	8.72E-05	0.341 66 7	0.182 21 2	0.123 37 2	0.260 76 7	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Pseudonocardiaceae; g__Pseudonocardia; s__

APPENDIX 2: List of figures

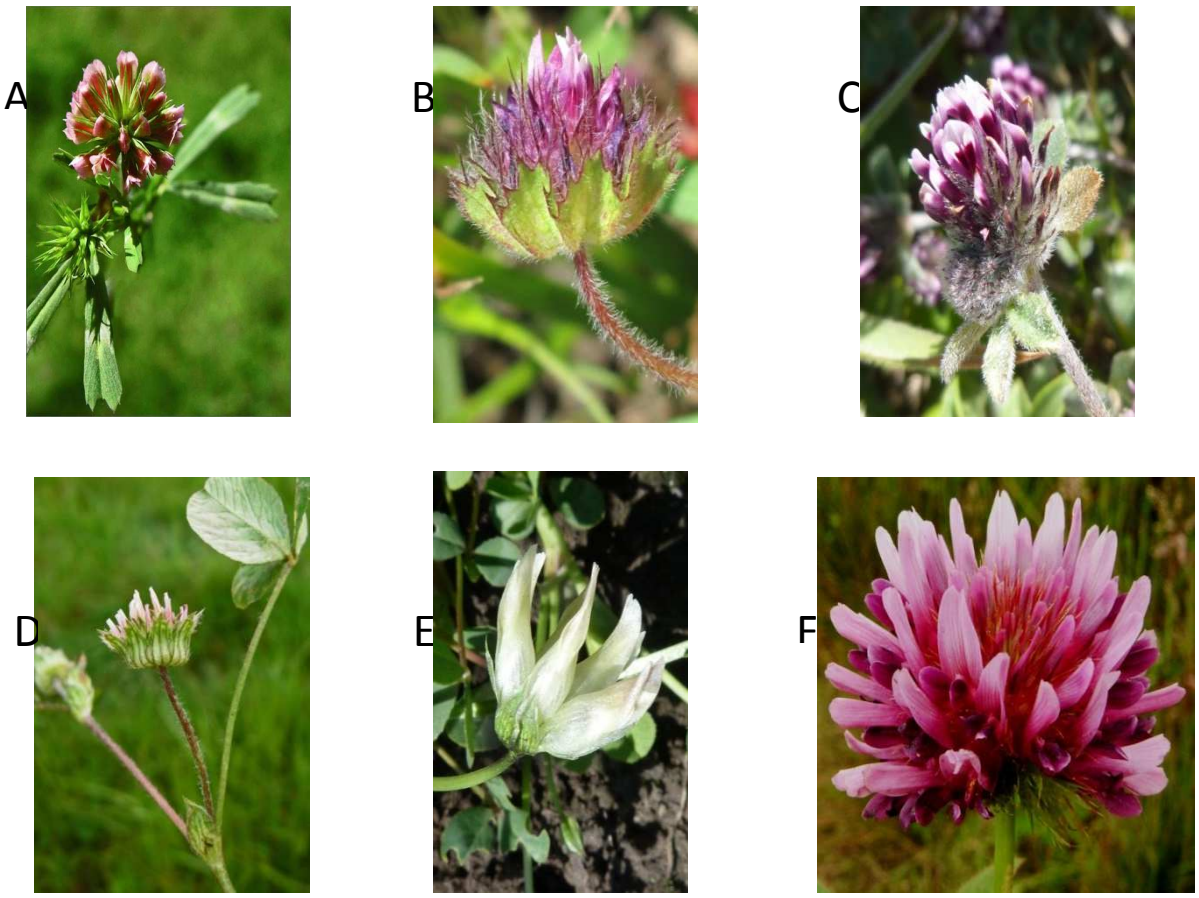


Figure 1: Trifolium species used in this study.
A) *Trifolium bifidum*, B)*Trifolium barbigerum*, C)*Trifolium macraei*, D) *Trifolium microdon*, E)*Trifolium fucatum*, F) *Trifolium wormskioldii*

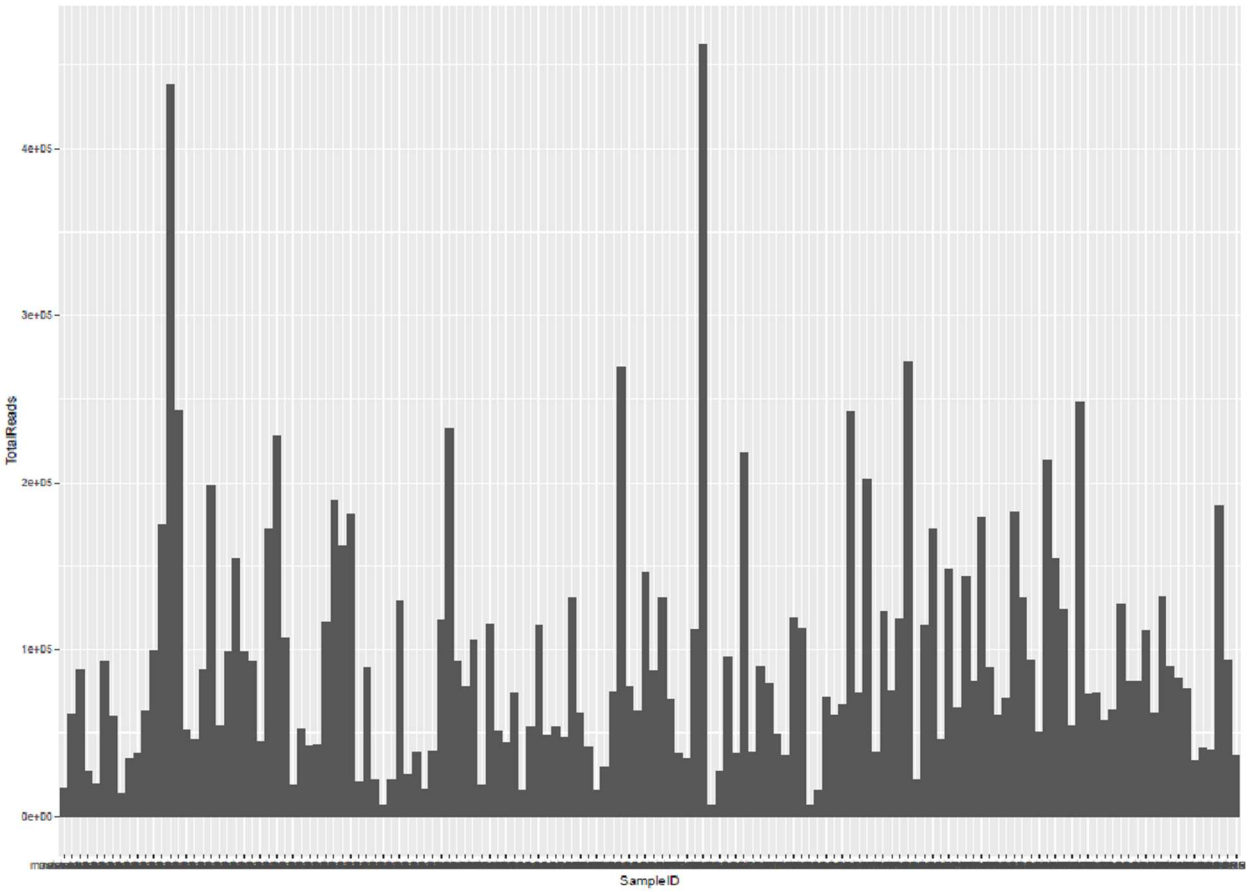


Figure 2a: Distribution of reads across samples – Basic.

This graph represents the distribution of reads across all 144 samples. X axis has the sample names and Y axis show the total number of reads.

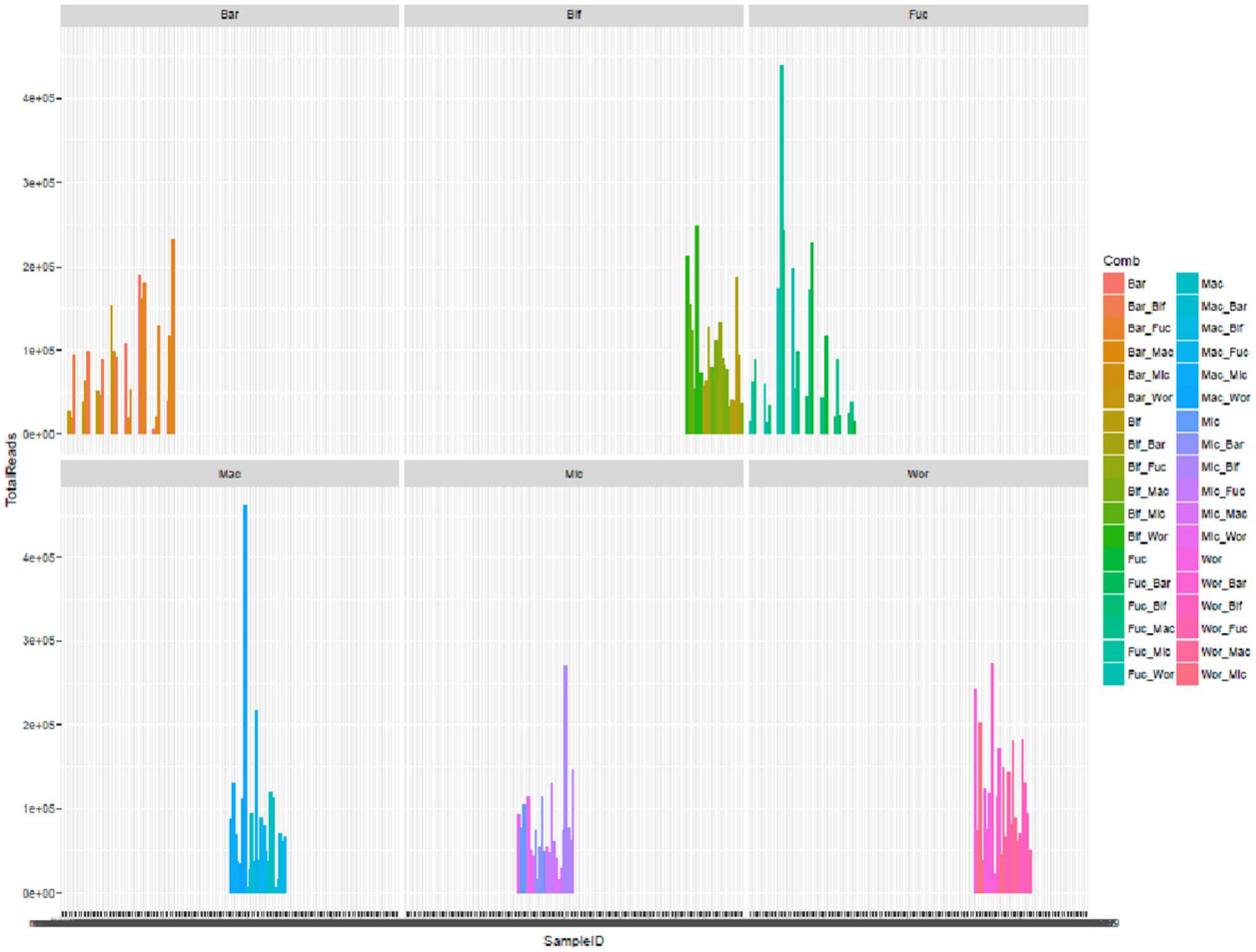
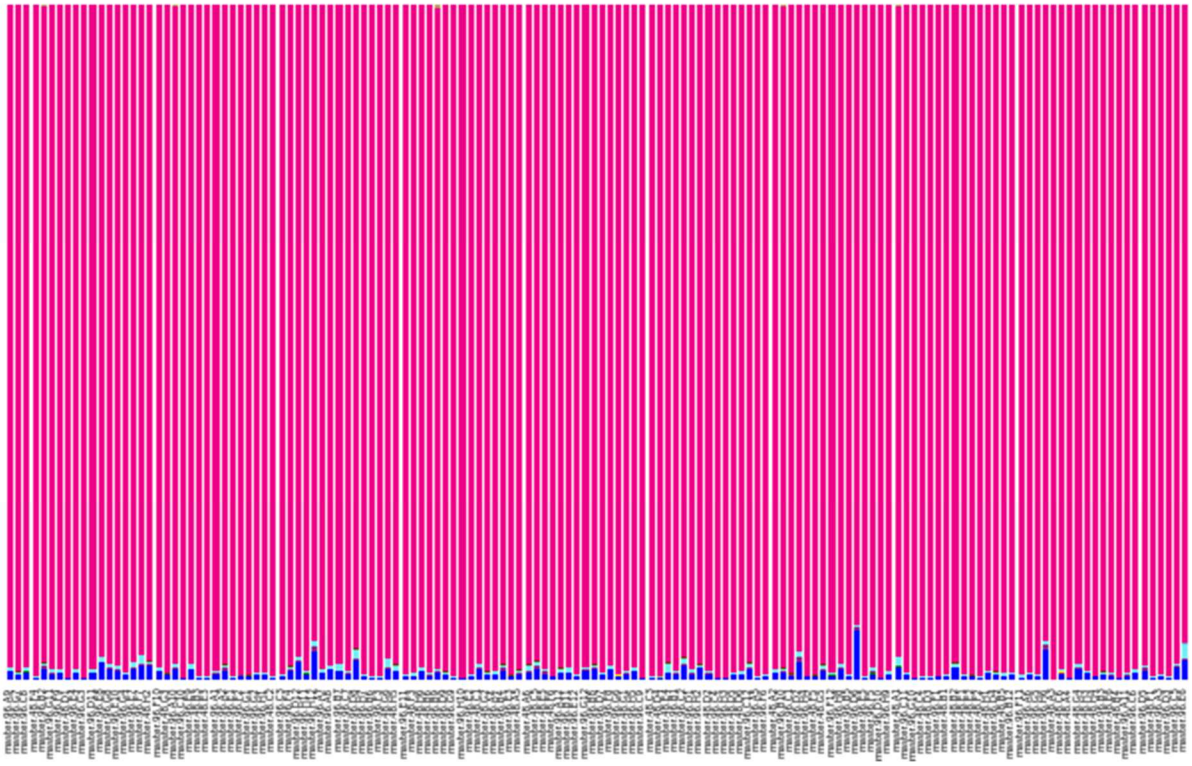


Figure 2b: Distribution of reads: Coloured.

This graph represents the distribution of reads across all 144 samples, coloured by the different growing combinations. X axis has the sample names and Y axis show the total number of reads.

a)



b)

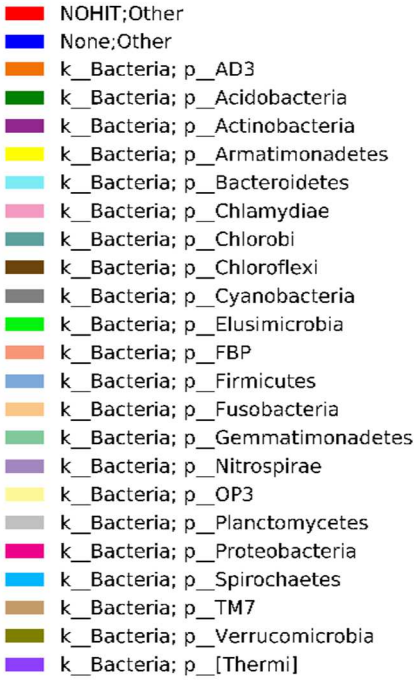
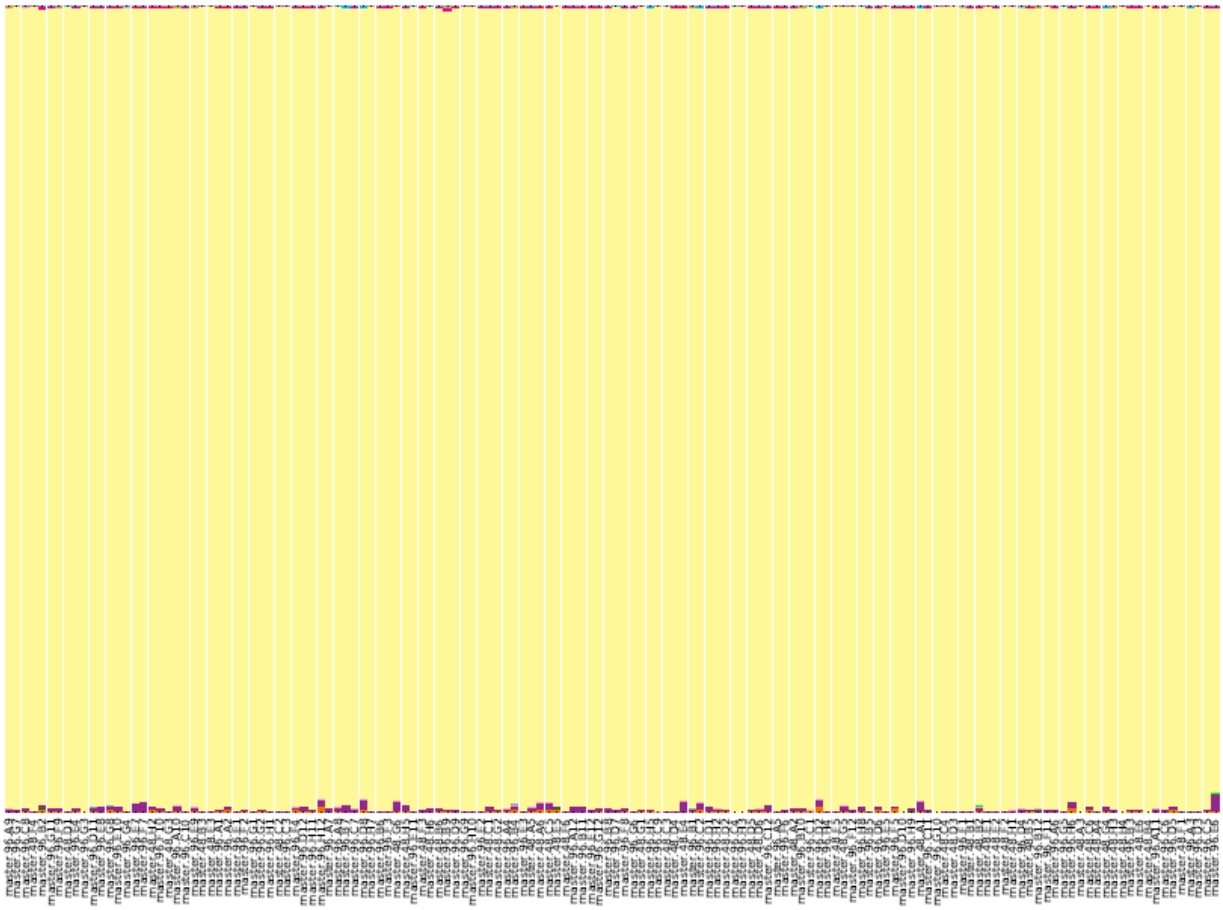


Figure 3a: Individual phyla contributions: Unfiltered OTU table

This graph represents the distribution of different phyla across the 144 samples.

a)



b)

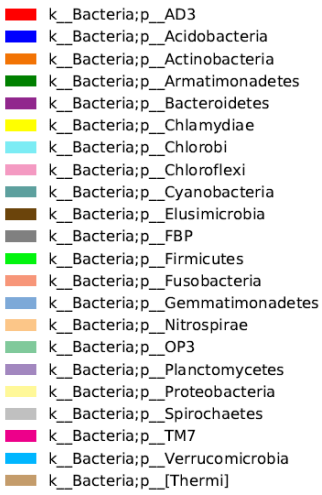
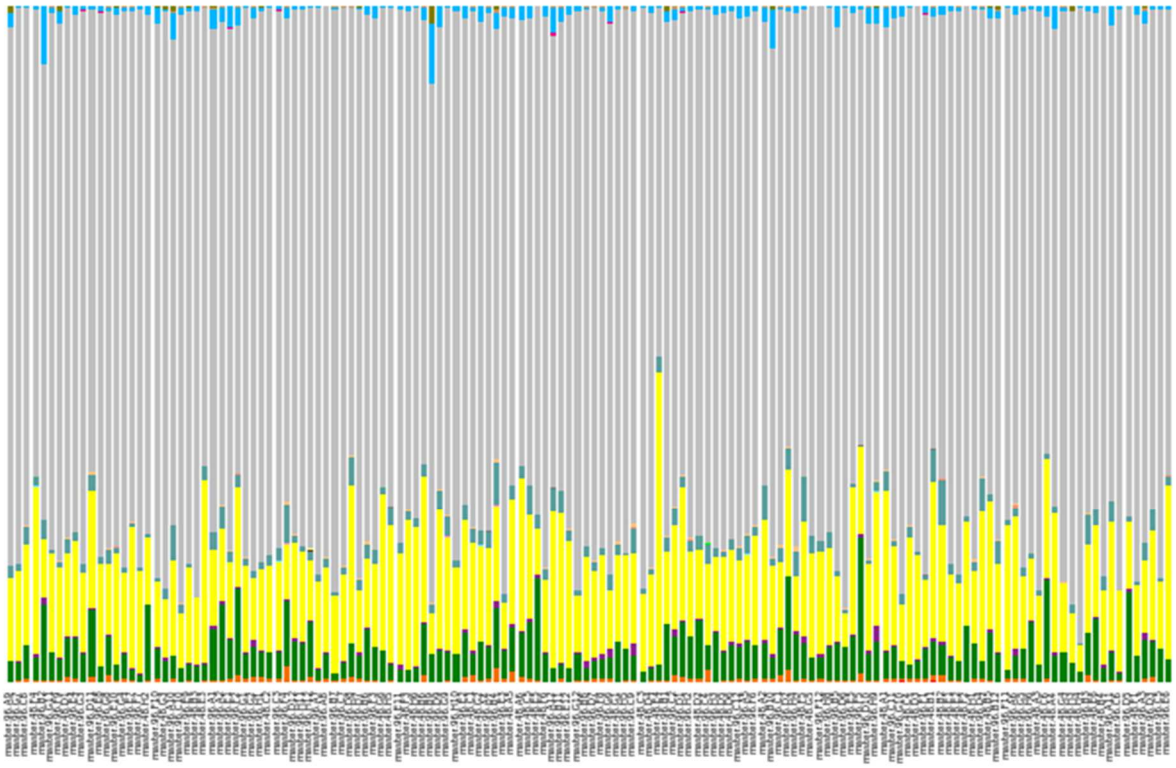


Figure 3b: Individual phyla contributions: Complete filtered OTU table.

This graph represents the distribution of different phyla across the 144 samples. Table is filtered to remove bad taxonomic assignments.

a)



b)

- k__Bacteria; p__AD3
- k__Bacteria; p__Acidobacteria
- k__Bacteria; p__Actinobacteria
- k__Bacteria; p__Armatimonadetes
- k__Bacteria; p__Bacteroidetes
- k__Bacteria; p__Chlamydiae
- k__Bacteria; p__Chlorobi
- k__Bacteria; p__Chloroflexi
- k__Bacteria; p__Cyanobacteria
- k__Bacteria; p__Elusimicrobia
- k__Bacteria; p__FBP
- k__Bacteria; p__Firmicutes
- k__Bacteria; p__Fusobacteria
- k__Bacteria; p__Gemmatimonadetes
- k__Bacteria; p__Nitrospirae
- k__Bacteria; p__OP3
- k__Bacteria; p__Planctomycetes
- k__Bacteria; p__Proteobacteria
- k__Bacteria; p__Spirochaetes
- k__Bacteria; p__TM7
- k__Bacteria; p__Verrucomicrobia
- k__Bacteria; p__[Thermi]

Figure 3c: Individual phyla contributions: Rare OTU table.

This graph represents the distribution of different phyla across the 144 samples. Table is filtered to remove bad taxonomic assignments and OTUs belonging to the “Rhizobiales” group.

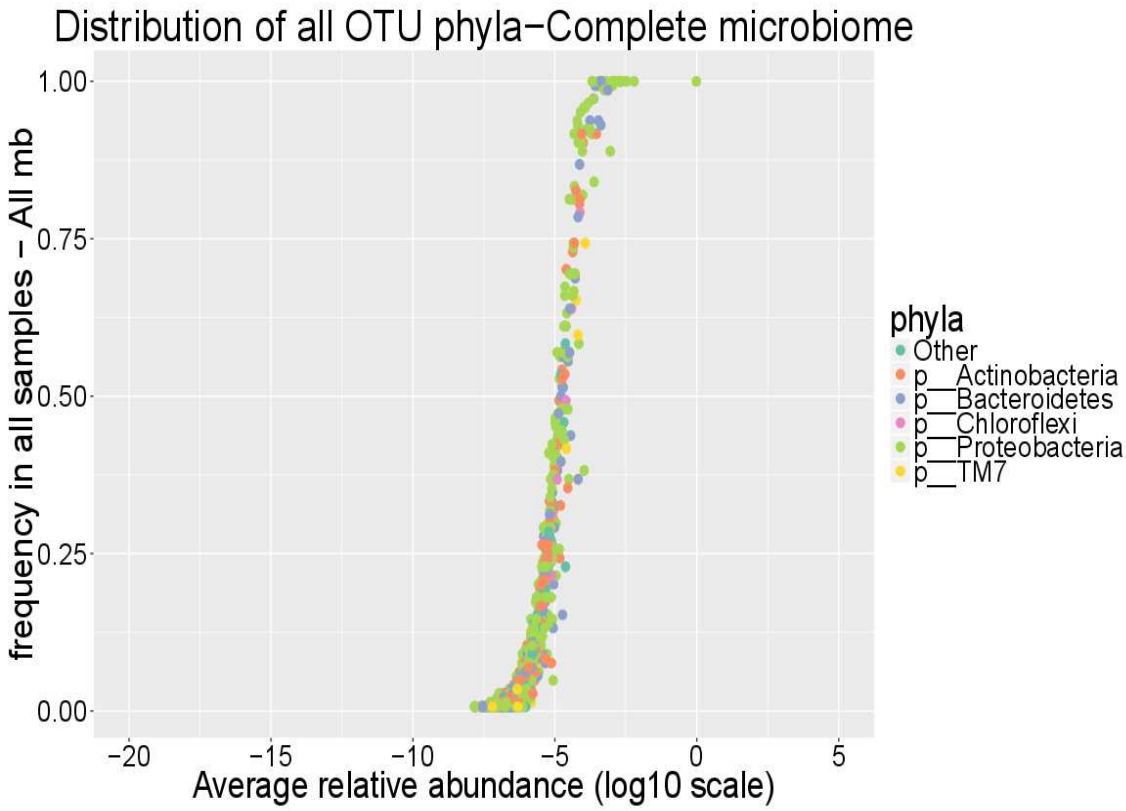


Figure 4a: Top five phyla contributions: Complete OTU table.

This graph represents how many OTUs are present in high frequency across samples. Here the image shows that the order Proteobacteria is not only the dominant order present in high abundance but also in high frequency across samples. Log scale along x axis is log10

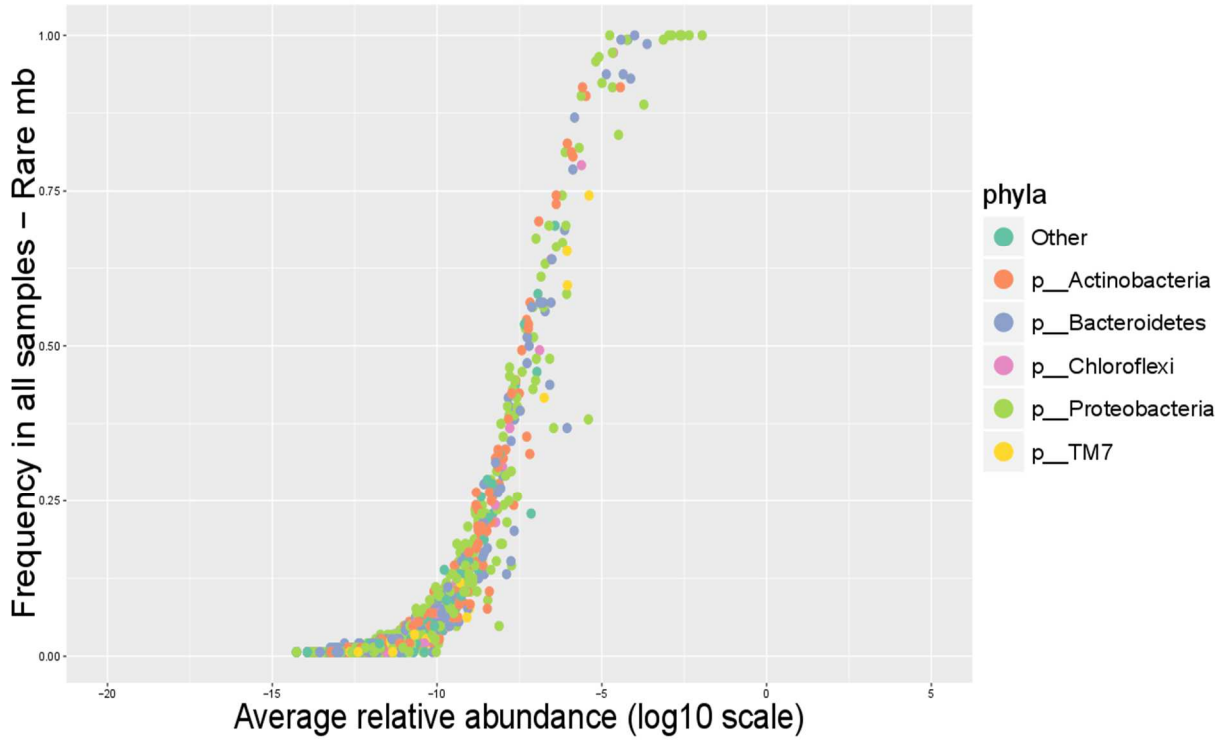


Figure 4b: Top five phyla contributions: Rare OTU table.

This graph represents how many OTUs are present in high frequency across samples. Here the image shows that no single phylum is present in all samples. Thus the abundance within the rare microbiome is more variable. The order Proteobacteria and Bacteriodetes is not only the dominant order present in high abundance but also in high frequency across samples. Log scale along x axis is log10

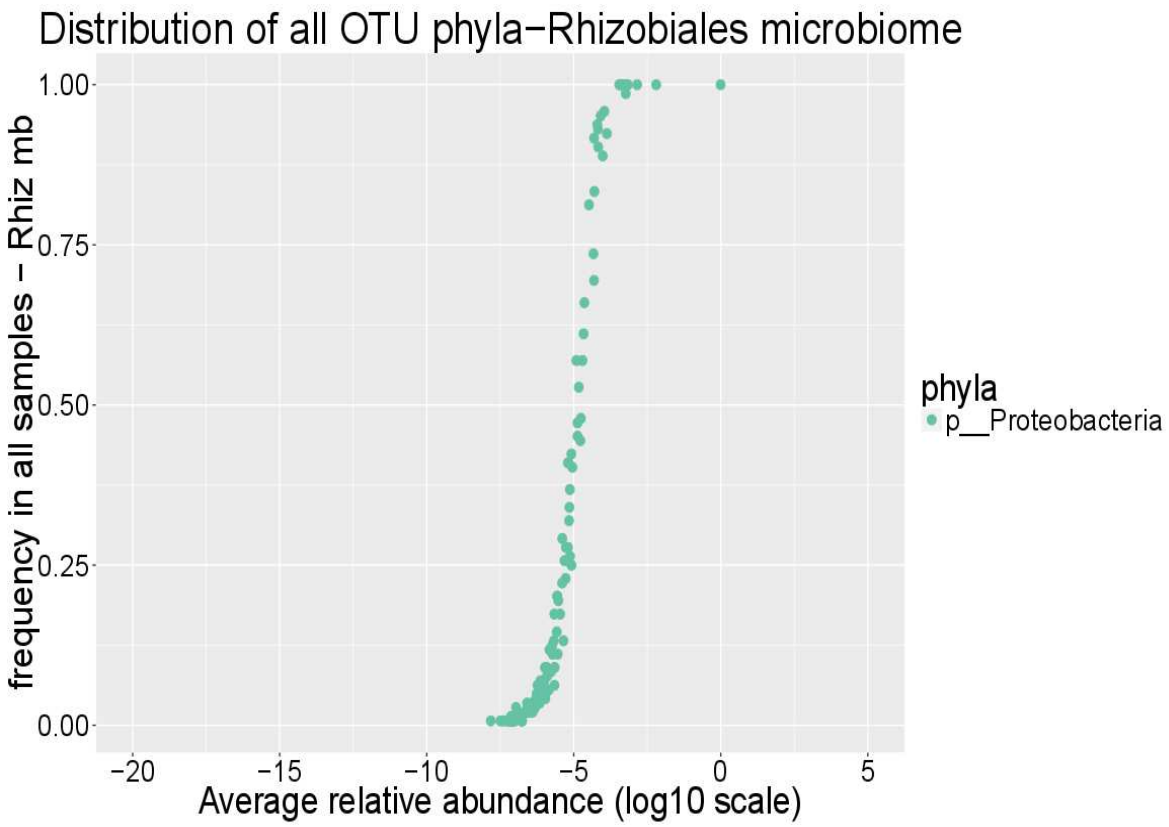


Figure 4c: Top five phyla contributions: Abundant OTU table.

This graph represents how many OTUs are present in high frequency across samples for the abundant (Rhizobiales) OTU table. Log scale along x axis is log10

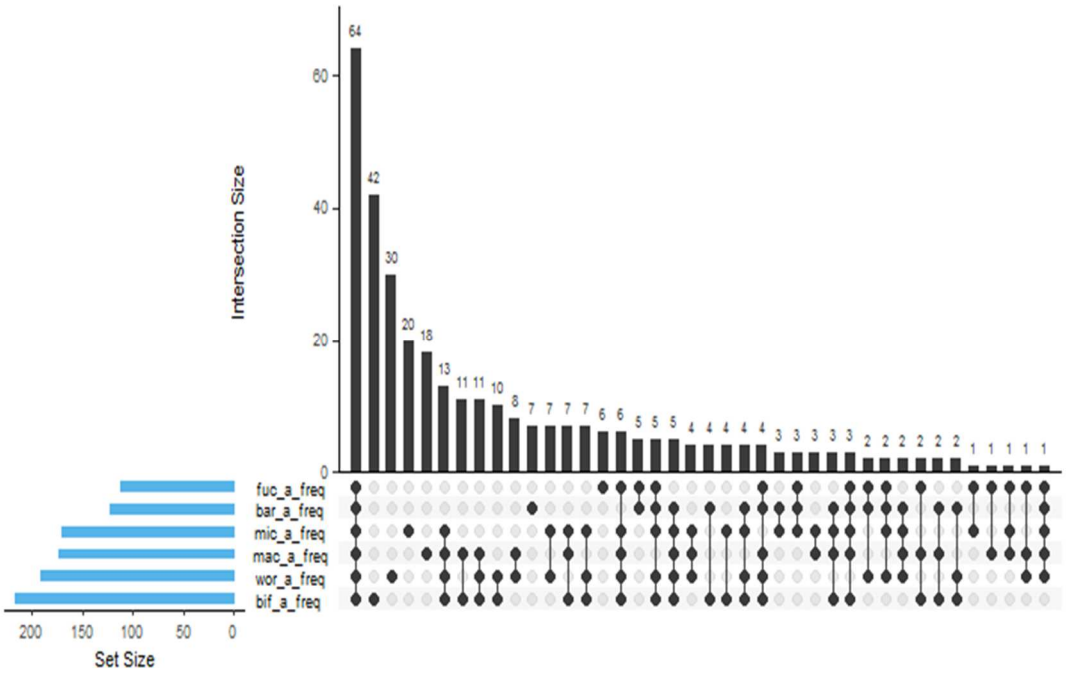


Figure 5a: Number of OTUs identified as core within at least 2 home samples.

The plot is built using presence absence data from OTUs present in at least 2 of the 4 native samples. Blue bars on the left indicate number of OTUs identified as core in the corresponding sample while the black histogram bars indicate number of intersecting OTUs.

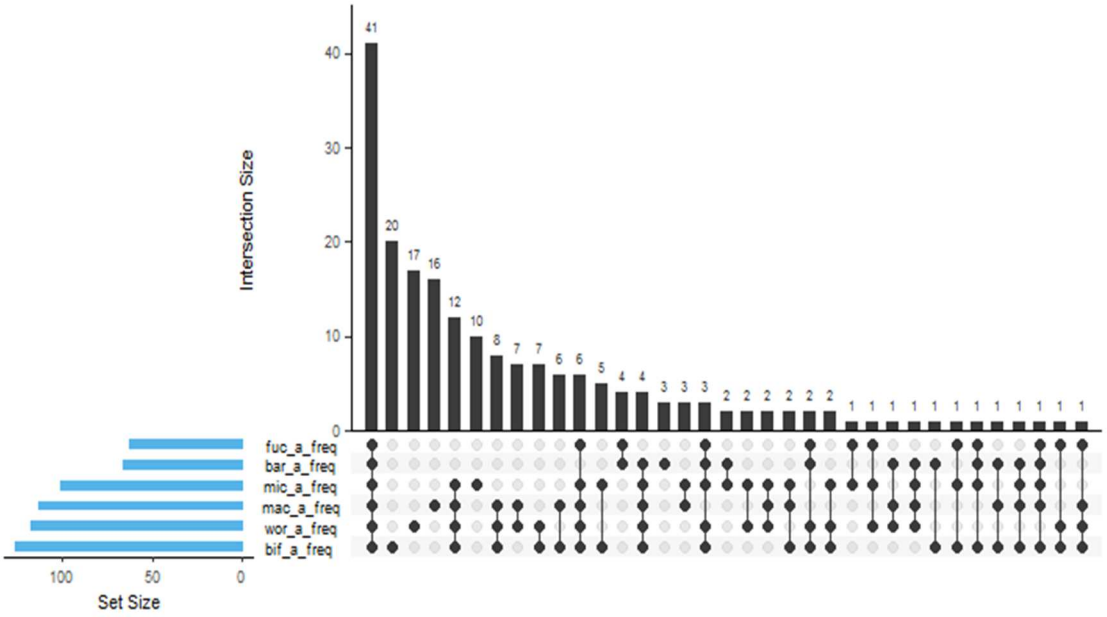


Figure 5b: Number of OTUs identified as core within at least 3 home samples.

The plot is built using presence absence data from OTUs present in at least 3 of the 4 native samples. Blue bars on the left indicate number of OTUs identified as core in the corresponding sample while the black histogram bars indicate number of intersecting OTUs.

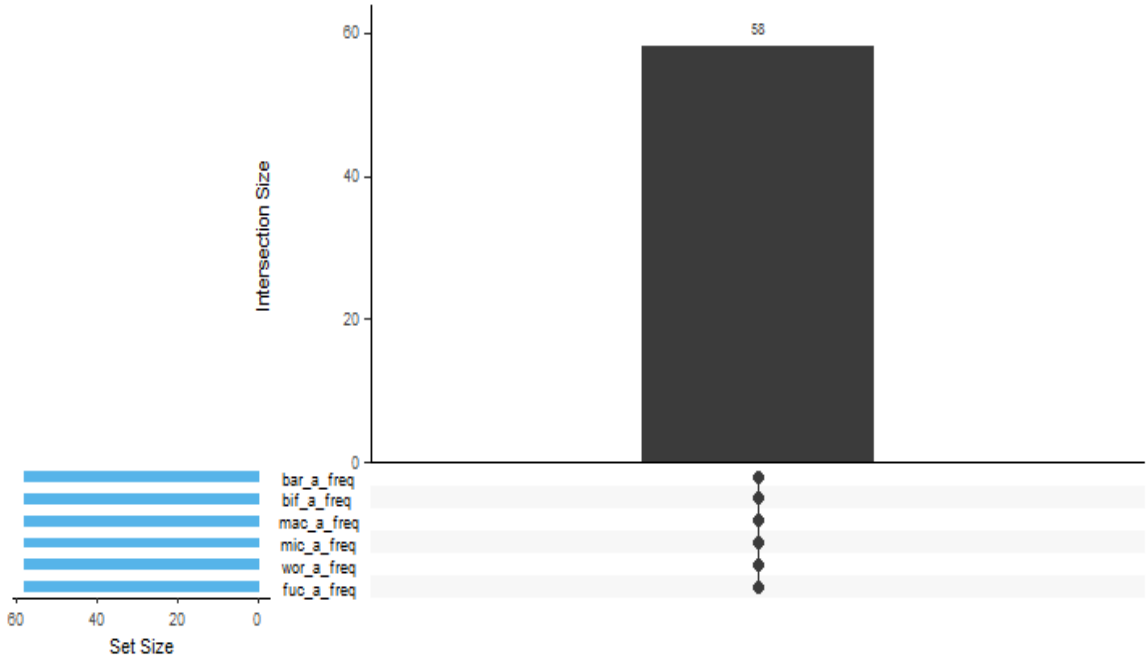


Figure 5c: Number of OTUs identified as core within at least 2 home samples and an abundance threshold of 0.00001%.

Blue bars on the left indicate size of the library while the black histogram bars indicate number of intersecting OTUs. Abundance threshold: the above mentioned threshold was used as this seemed to be the faint cut off between abundance vs presence. Any higher abundance and I select only OTUs belonging to Rhizobiales and there isn't a comparison.

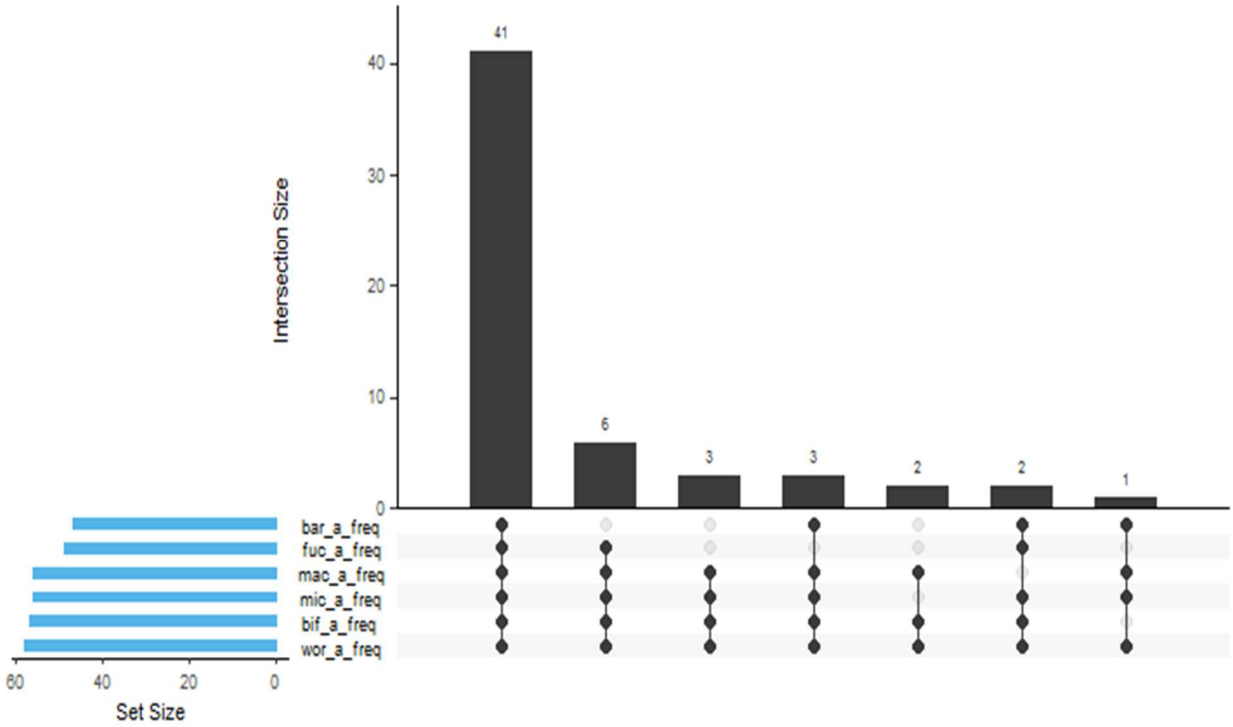
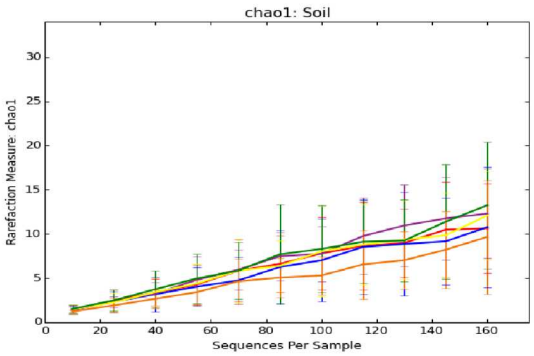


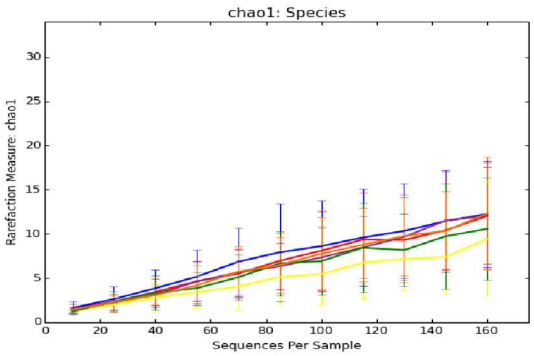
Figure 5d: Number of OTUs identified as core within at least 3 home samples and an abundance threshold of 0.00001%.

Blue bars on the left indicate size of the library while the black histogram bars indicate number of intersecting OTUs.

a)



b)

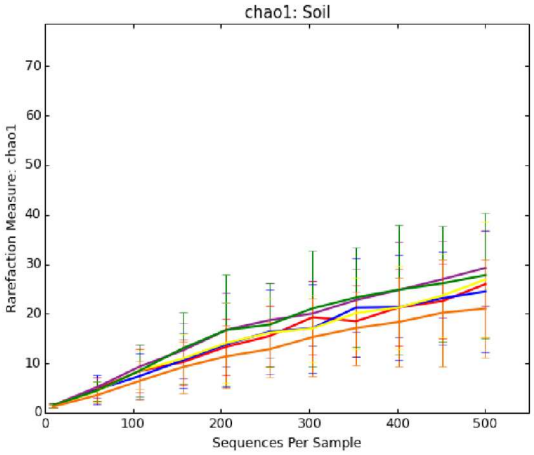


c)

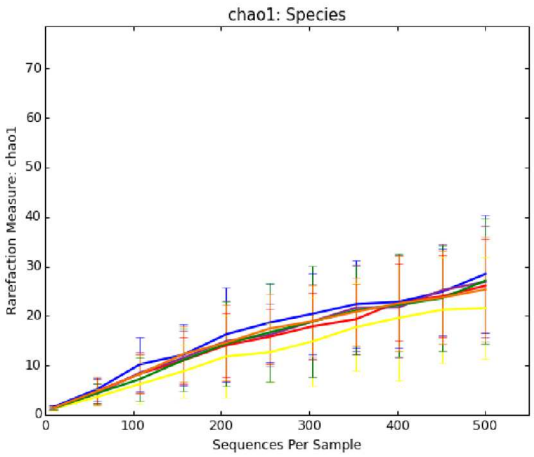
Figure 6a: Rarefaction at 160reads/ sample using the complete OTU table.

This figure shows rarefaction curves for species richness when all home samples are subsampled to 160 sequences each. Metric for species richness is Observed species and rarefaction curves are colored by a)“Soil” followed by b)“Species” c) legend for the colored rarefaction curves.

a)



b)

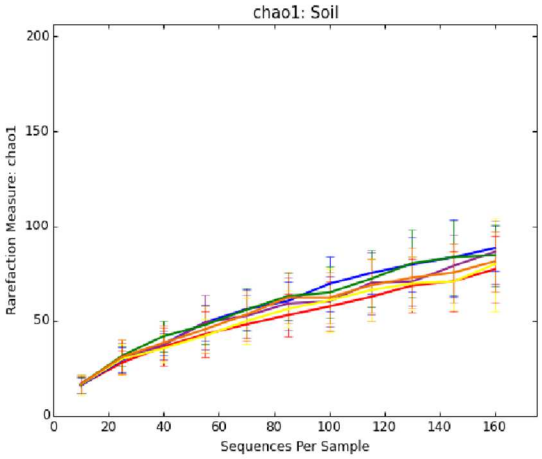


c)

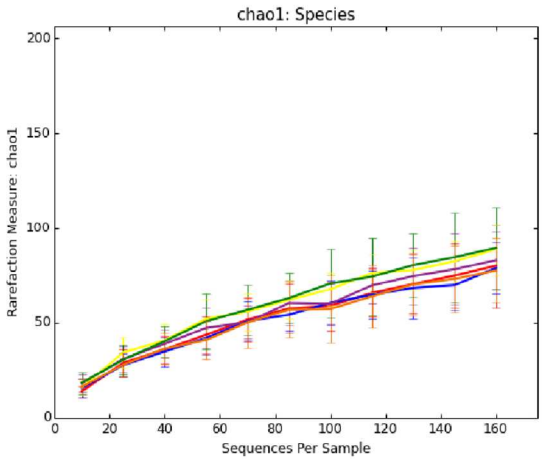
Figure 6b: Rarefaction at 500 reads/ sample using the complete OTU table.

This figure shows rarefaction curves for species richness when all home samples are subsampled to 500 sequences each. Metric for species richness is Observed species and rarefaction curves are colored by a) “Soil” followed by b) “Species” c) legend for the colored rarefaction curves.

a)



b)

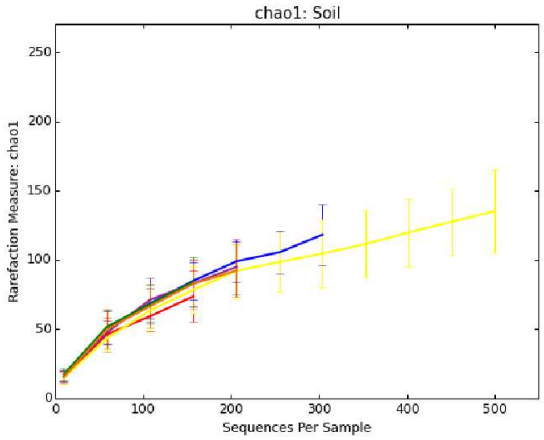


c)

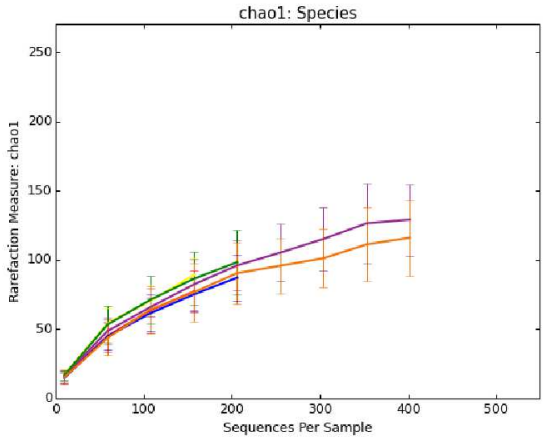
Figure 6c: Rarefaction at 160 reads/ sample using the rare OTU table.

This figure shows rarefaction curves for species richness when all home samples are subsampled to 160 sequences each. Metric for species richness is Observed species and rarefaction curves are colored by a) “Soil” followed by b) “Species” c) legend for the colored rarefaction curves.

a)



b)



c)

Figure 6d: Rarefaction at 500 reads/ sample using the rare OTU table.

This figure shows rarefaction curves for species richness when all home samples are subsampled to 500 sequences each. Metric for species richness is Observed species and rarefaction curves are colored by a) “Soil” followed by b) “Species” c) legend for the colored rarefaction curves. Both graphs show incomplete rarefaction curves indicating that these samples did not have sufficient read depth to rarefy to 500 reads/ sample.

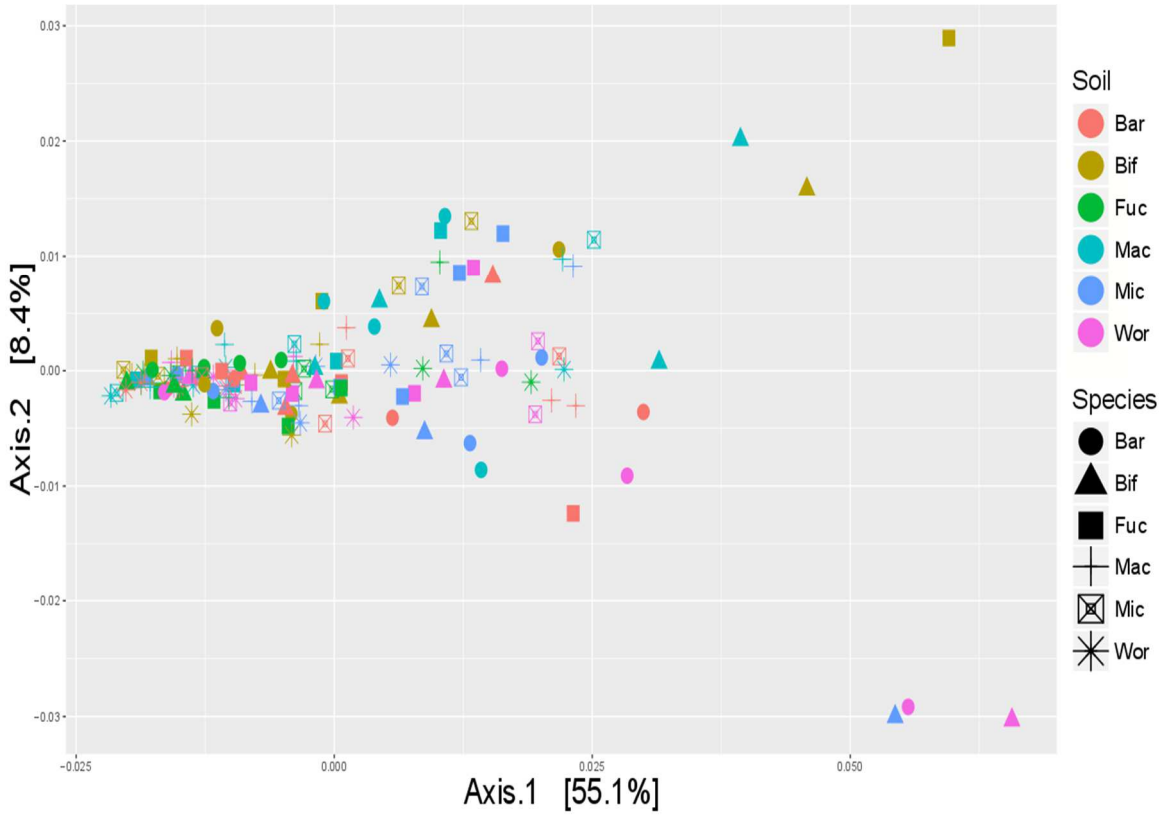


Figure 7a.i: PCoA ordination using the complete filtered OTU table

This graph a principal coordinate analysis performed using the complete OTU table rarefied and normalized to 6500 reads/ sample (586 OTUs detected). Samples are coloured by Soil factor and shaped by Species factor to represent all the combinations used in the study. Samples group along a single, dominant axis. One possible reason this could happen in the abundance of the OTUs belonging to the Rhizobiales category

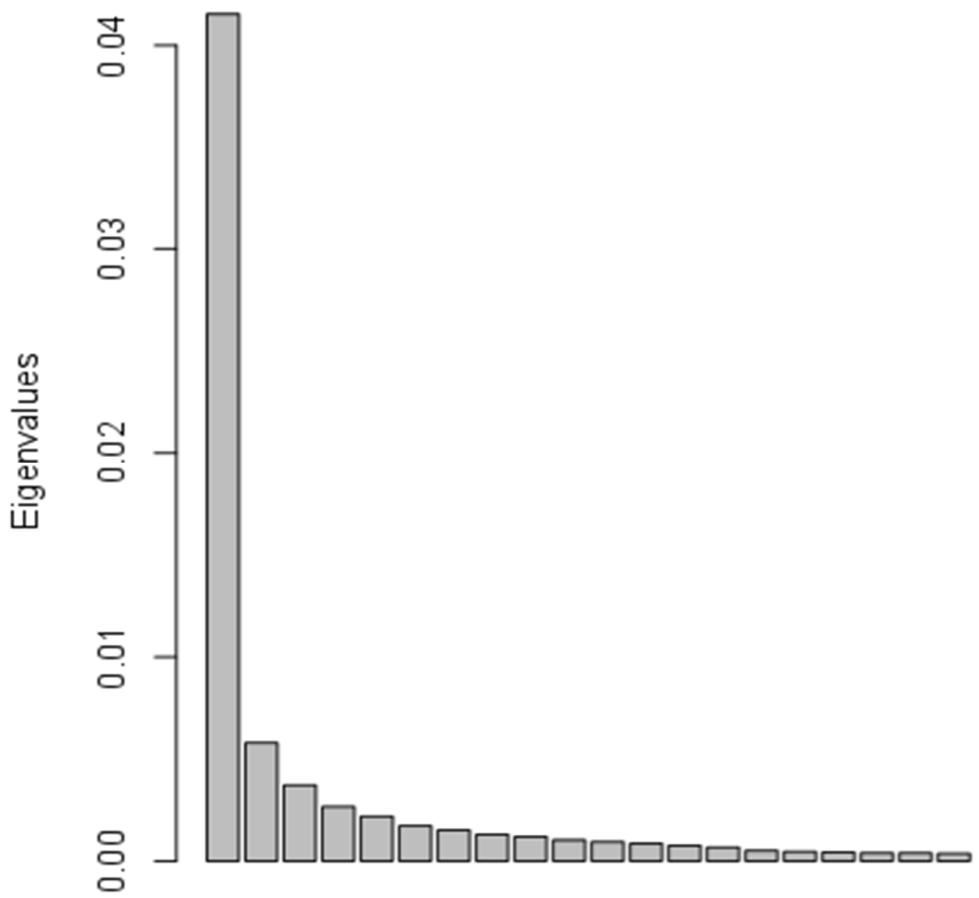


Figure 7a.ii) Scree plot for the PCoA ordination using the complete OTU table.

Notice large eigenvalue first axis is also seen in the scree plot, indicating that one dominant factor is driving the relationship.

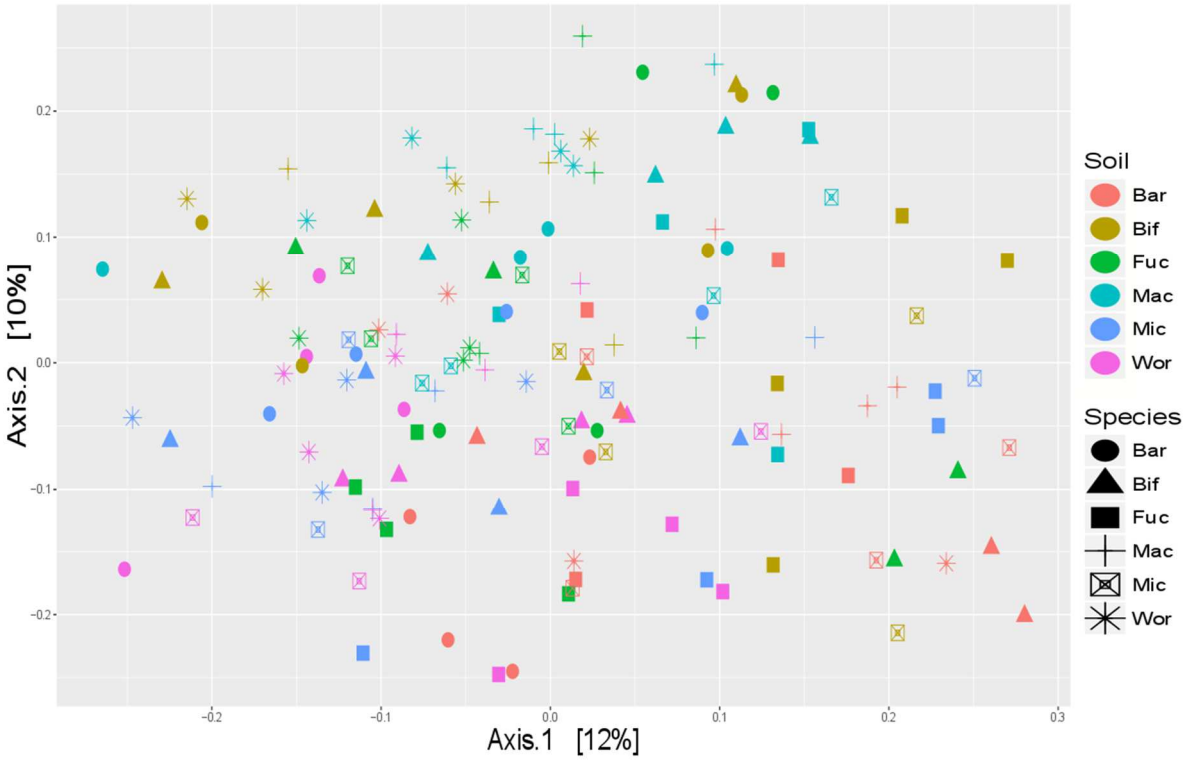


Figure 7b.i: PCoA ordination using the rare OTU table

Rare table was rarefied and then normalized to 163 reads/ sample (513 OTUs identified). Removing the abundant taxa also removes the grouping along a single axis. However, there still seems to be an overall clustering by the Soil factor. Samples are coloured by Soil factor and shaped by Species factor to represent all the combinations used in the study.

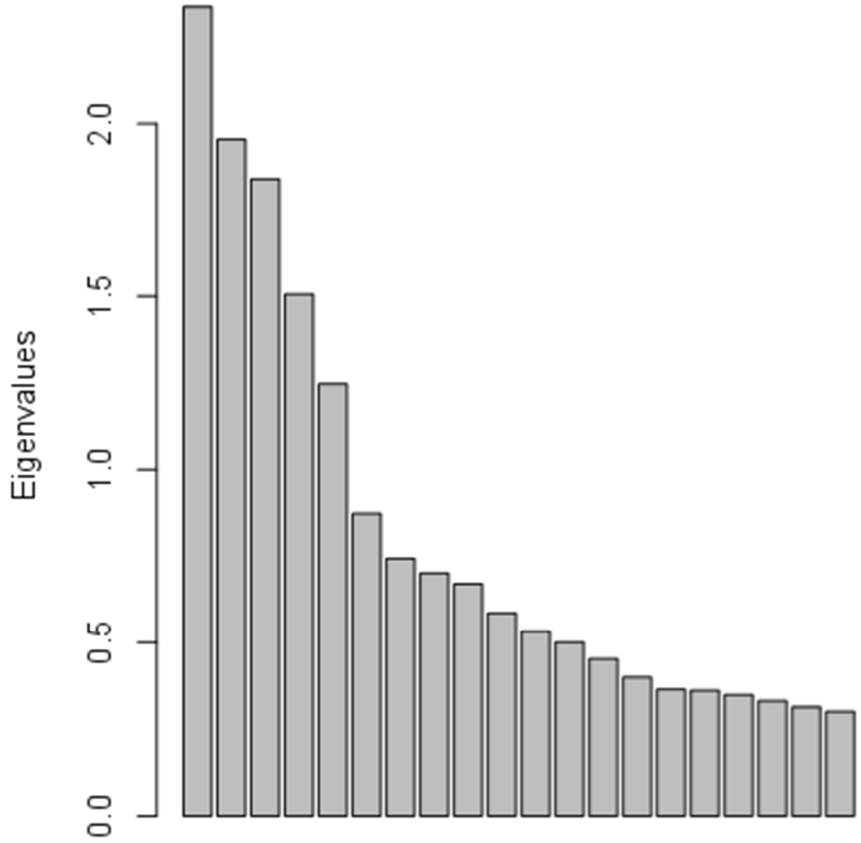


Figure 7b.ii:) Scree plot for the PCoA ordination using the rare OTU table

Removing the dominant taxa also removes the single dominant eigen value we see in the complete OTU PCoA plot. This could indicate that a single factor is pushing the grouping that we see earlier.

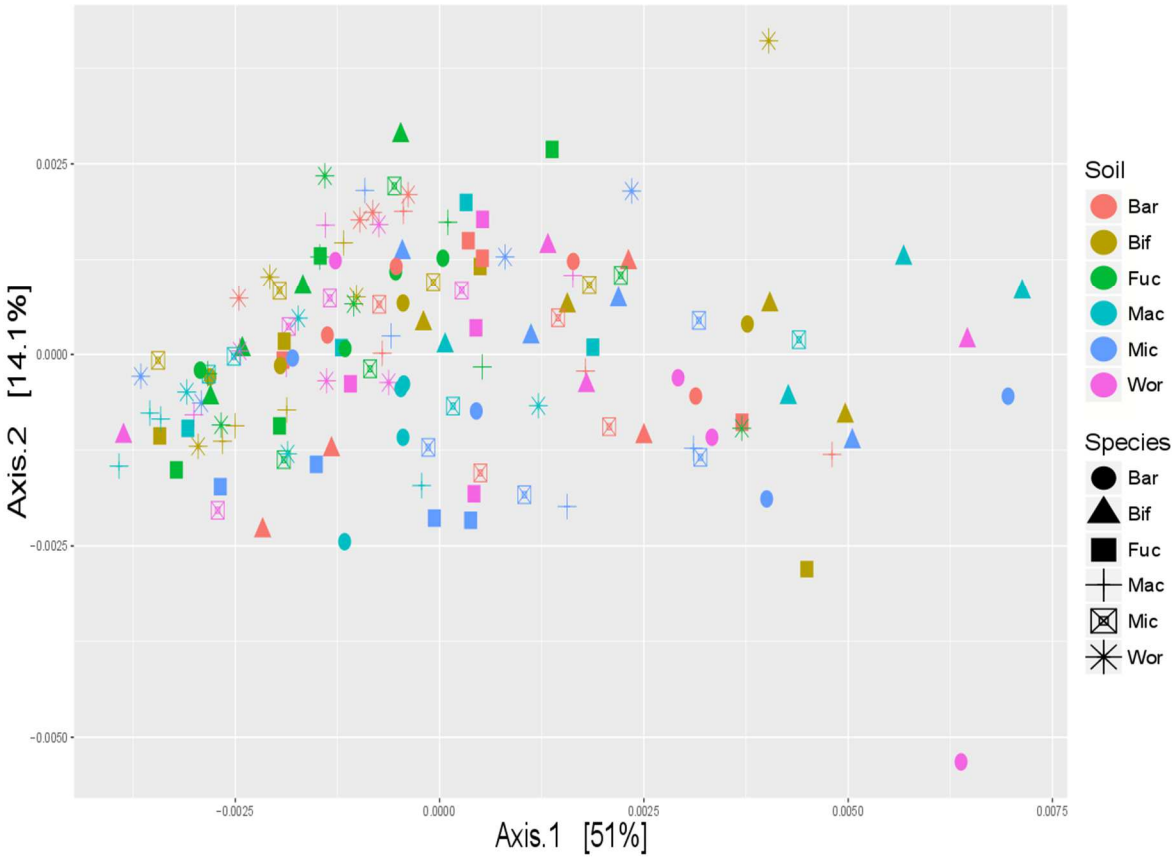


Figure 7c.i PCoA ordination using the abundant OTU table (comprising of 75 OTUs)

Rhizobiales OTU table was rarefied and then normalized to 6250 reads/sample. Samples tend to group towards the left half go the ordination, similar to the complete OUT.table.

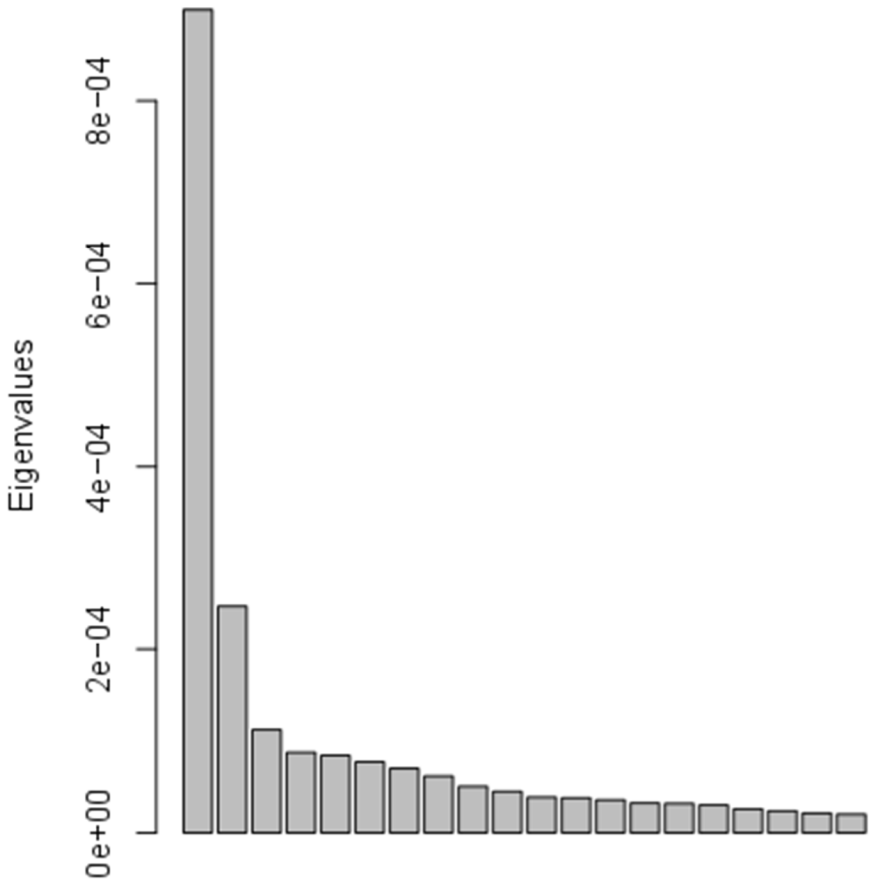


Figure7c.ii:) Scree plot for the PCoA ordination using the abundant OTUtable.

Samples again have the large single dominant axis also seen when looking at the scree plot for complete OTU table, which is absent in the rare microbiome table. This indicates that samples grouping along a single axis are driven by the dominant taxa.

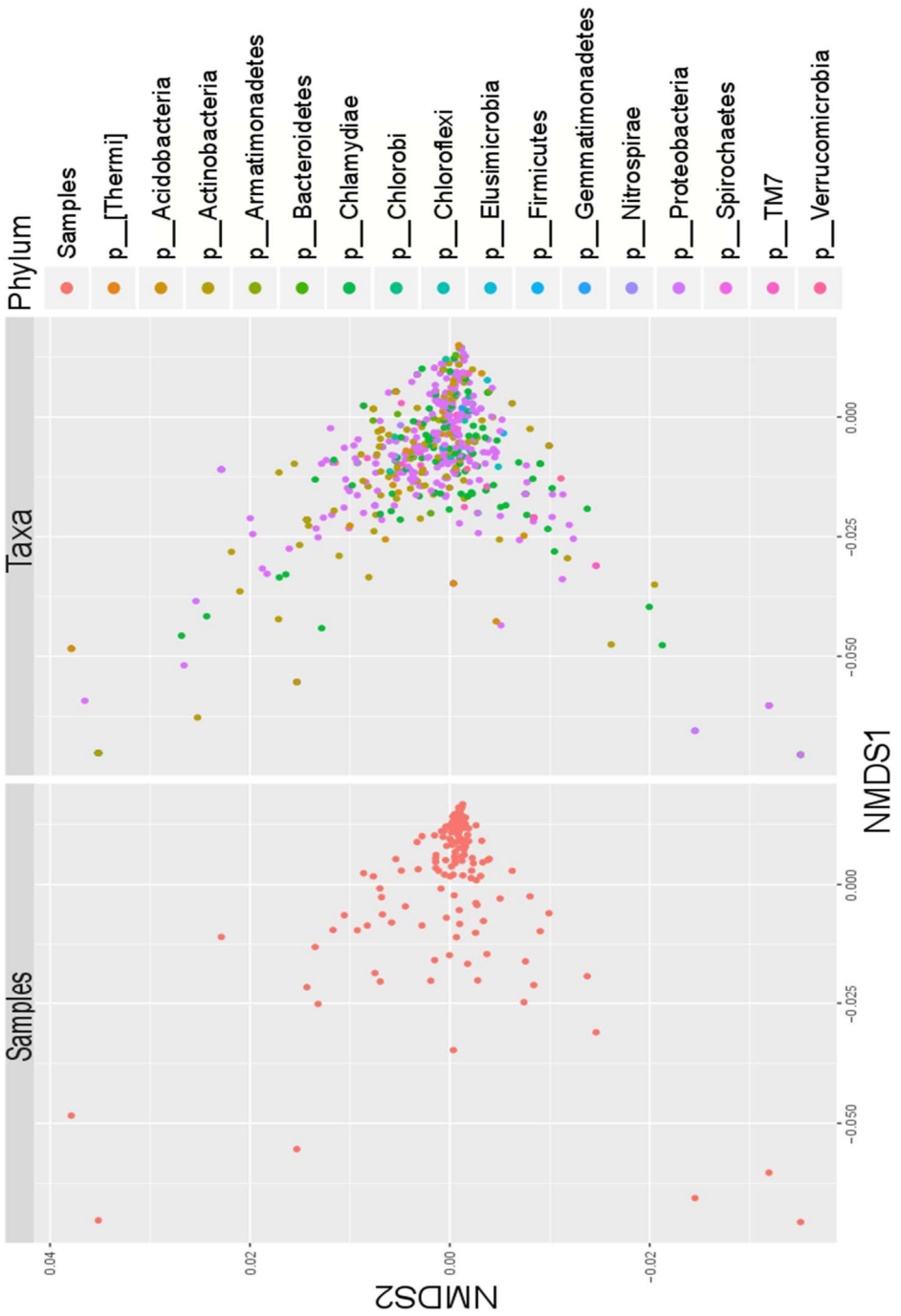


Figure 8a: NMDS plot performed on the complete OTU table.

Samples are represented on the left and their corresponding microbial communities on the right. This plot explores how clustering of samples relates to clustering of microbial community members. Samples tend to cluster along a single axis.

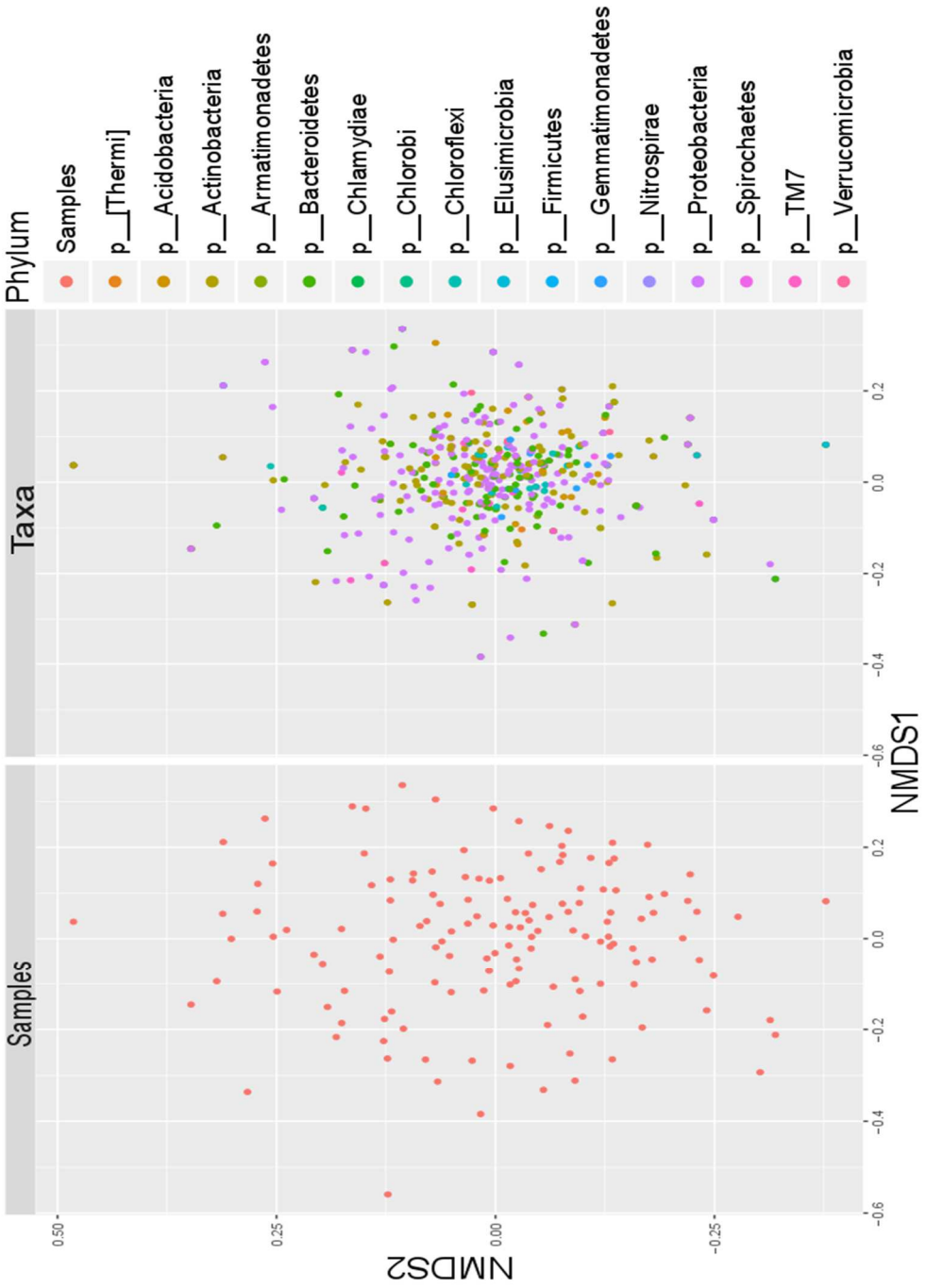


Figure 8b: NMDS plot performed on the rare OTU table

Samples are represented on the left and their corresponding microbial communities on the right. This plot explores how clustering of samples relates to clustering of microbial community members. Here we do not see the close grouping in the center that we see for both the samples and microbial communities of the complete table. The close the points are, the most similar the samples and microbiome. This graph gives further evidence that the rare OTUs are more variable than the complete OTUs.

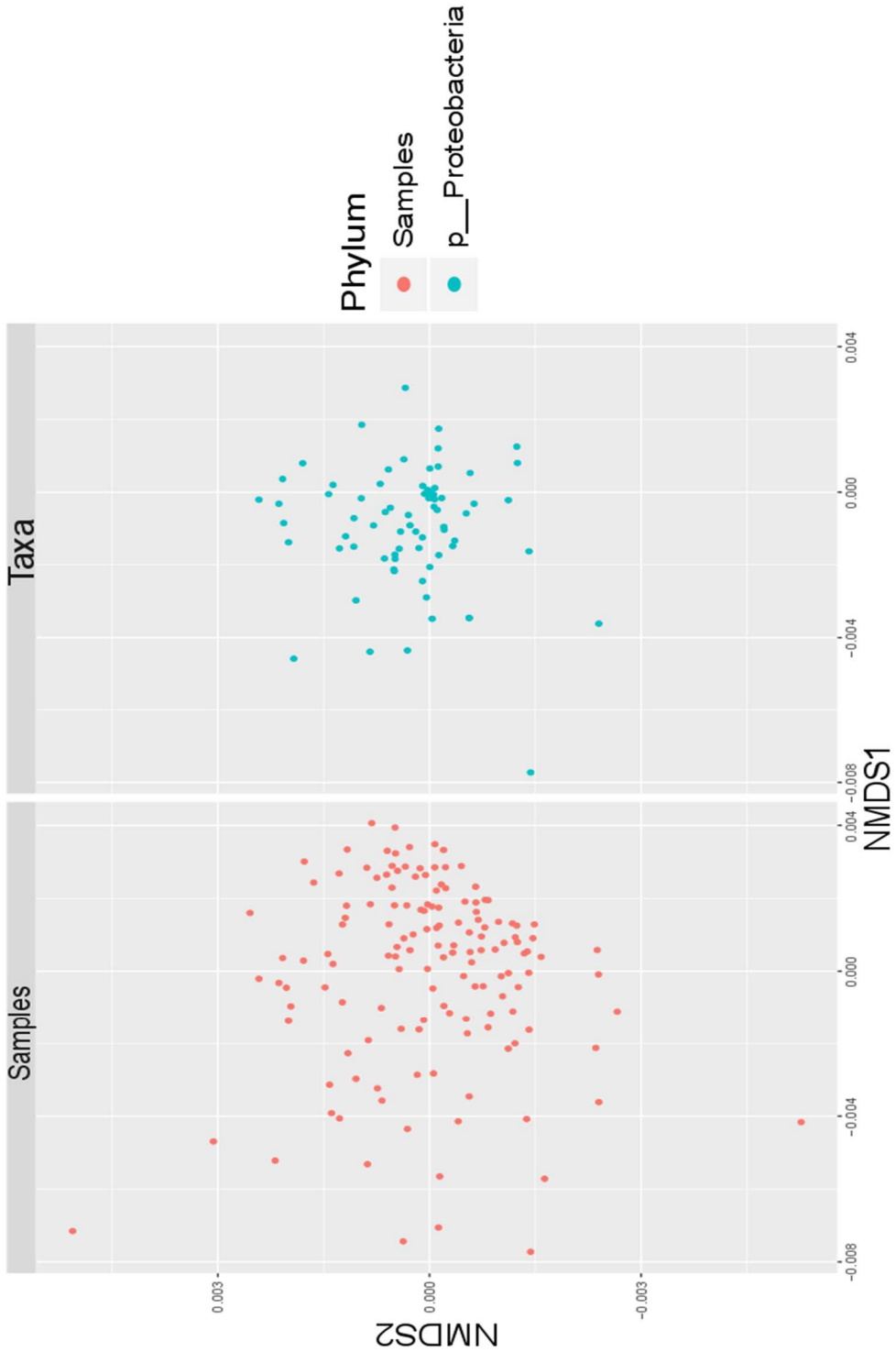


Figure 8c: NMDS plot performed on the abundant OTU table.

Samples are represented on the left and their corresponding microbial communities on the right. Sample grouping overlaps with abundant taxa of microbial community specifically with the abundant members: Agrobacterum and leguminosarum.

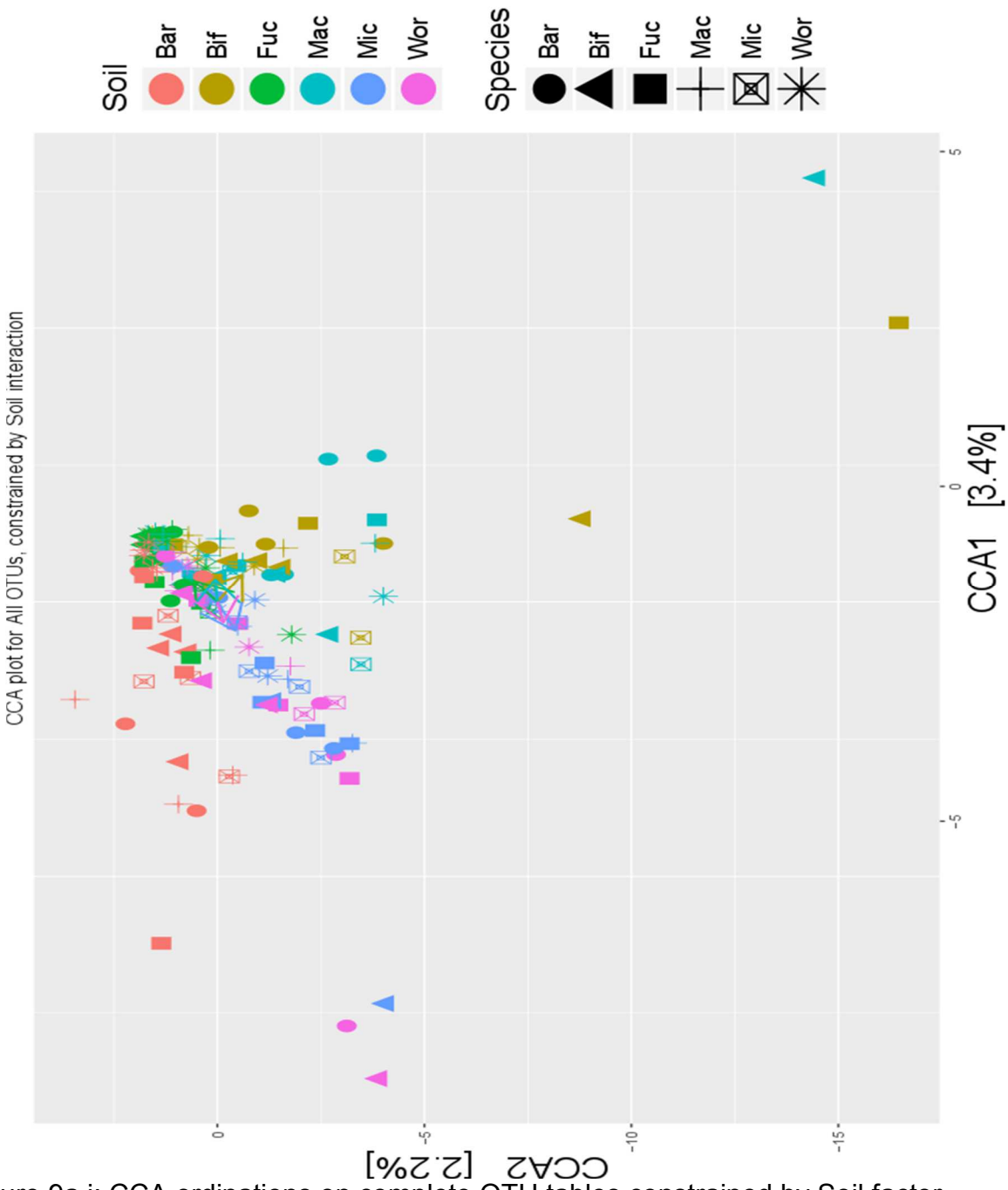


Figure 9a.i: CCA ordinations on complete OTU tables constrained by Soil factor.

This plot shows that all samples are highly similar to each other when Soil factor is constrained. This means that the variation explained by the other factors is pretty small. This could indicate that either the relationships are driven by the dominant taxa or communities are highly similar when removing the soil factor. The arrows are coloured by the corresponding Soil factor.

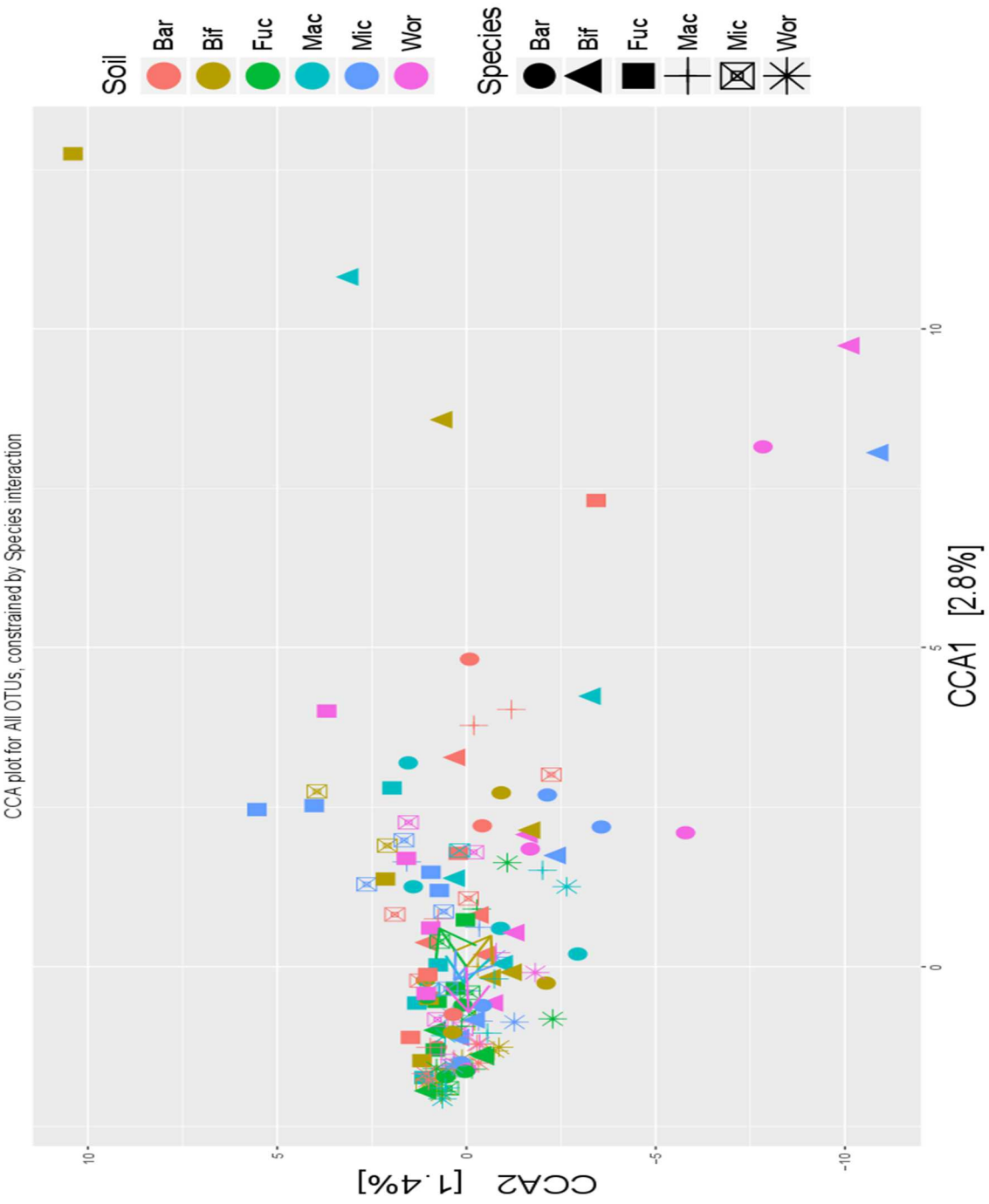


Figure 9a.ii: CCA ordinations on complete OTU tables constrained by Species factor.

This plot too shows that all samples are highly similar to each other when Species factor is constrained. However, we see larger eigenvalues for the CCA axis when Soil factor (CCA1: 3.4%) is constrained compared to species factor (2.8%). This means that the variation explained by the other factors is pretty small and relationships are driven by the dominant taxa. The arrows are coloured by the corresponding Species factor.

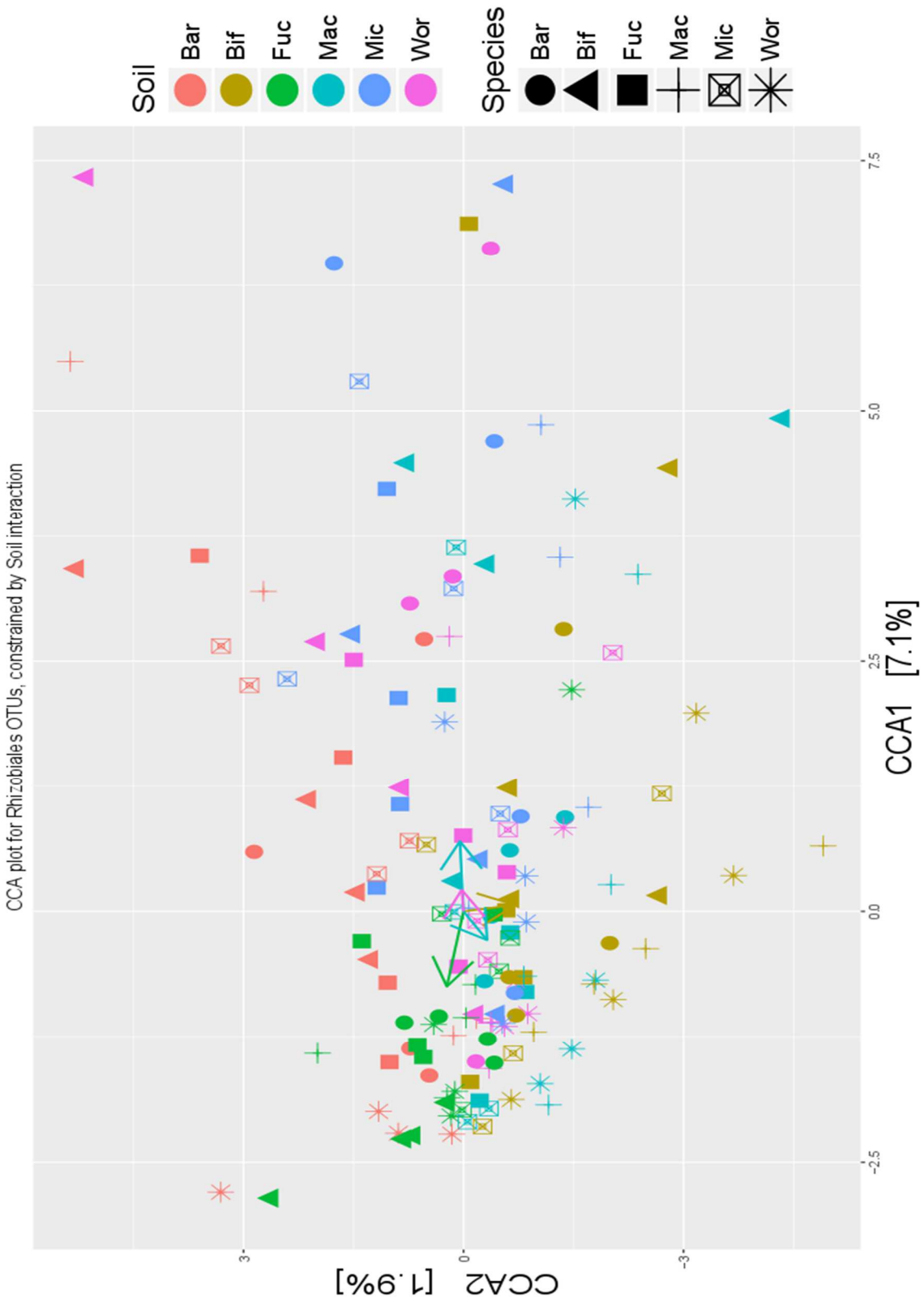


Figure 9b.i: CCA ordinations on abundant OTU tables constrained by Soil factor

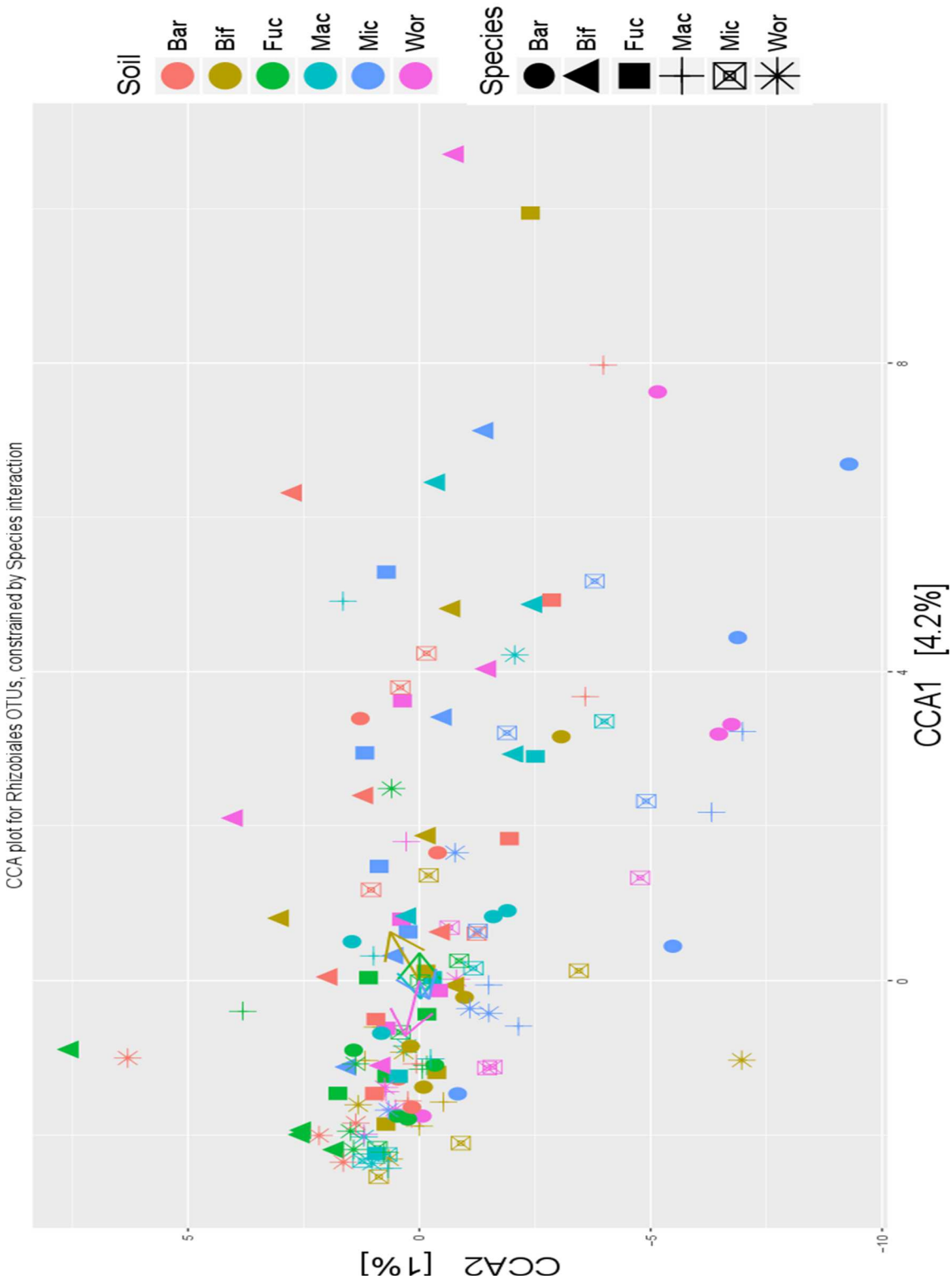


Figure 9b.ii: CCA ordinations on abundant OTU tables constrained by Species factor

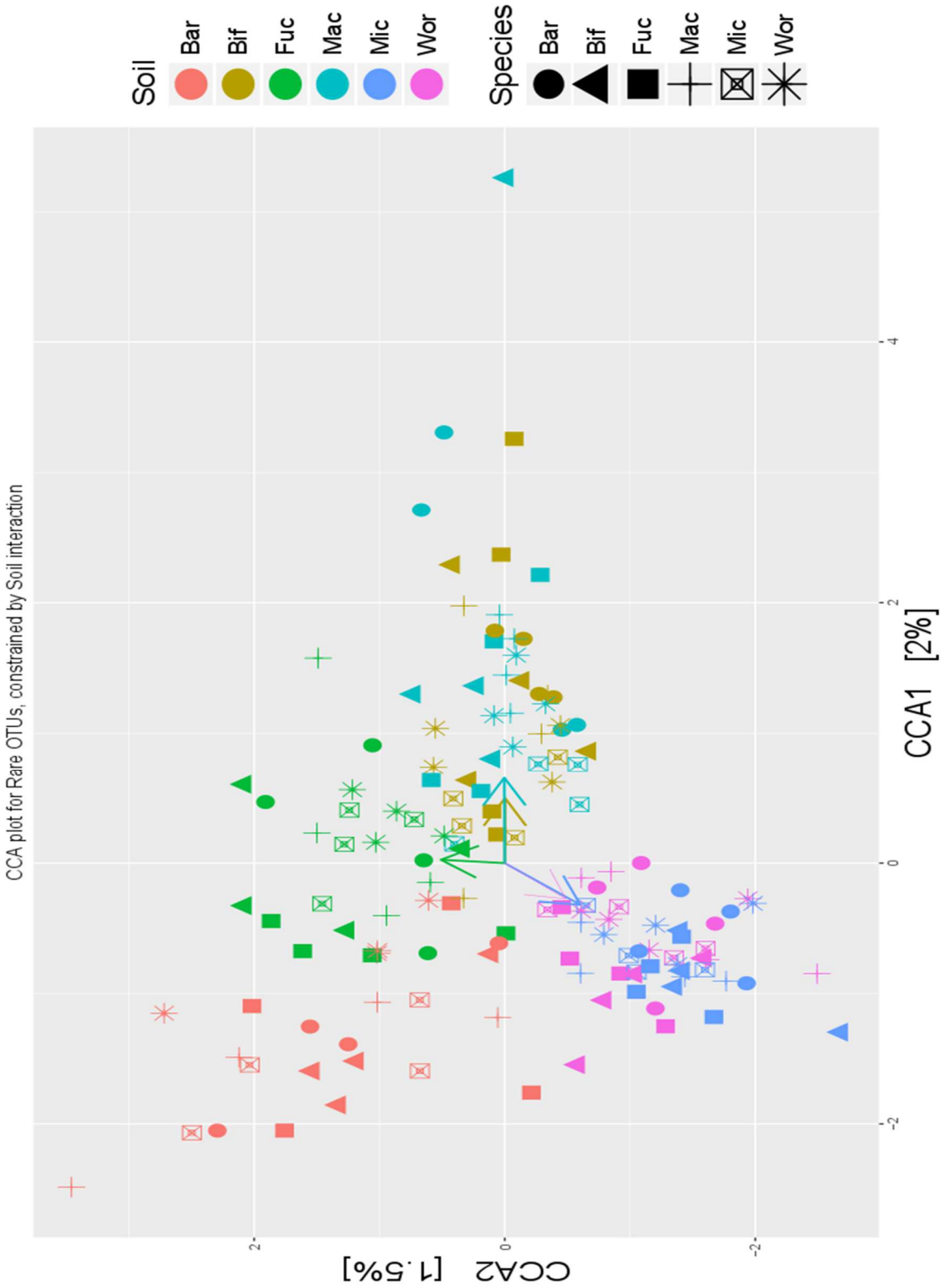


Figure 9c.i: CCA ordinations on rare OTU tables constrained by Soil factor

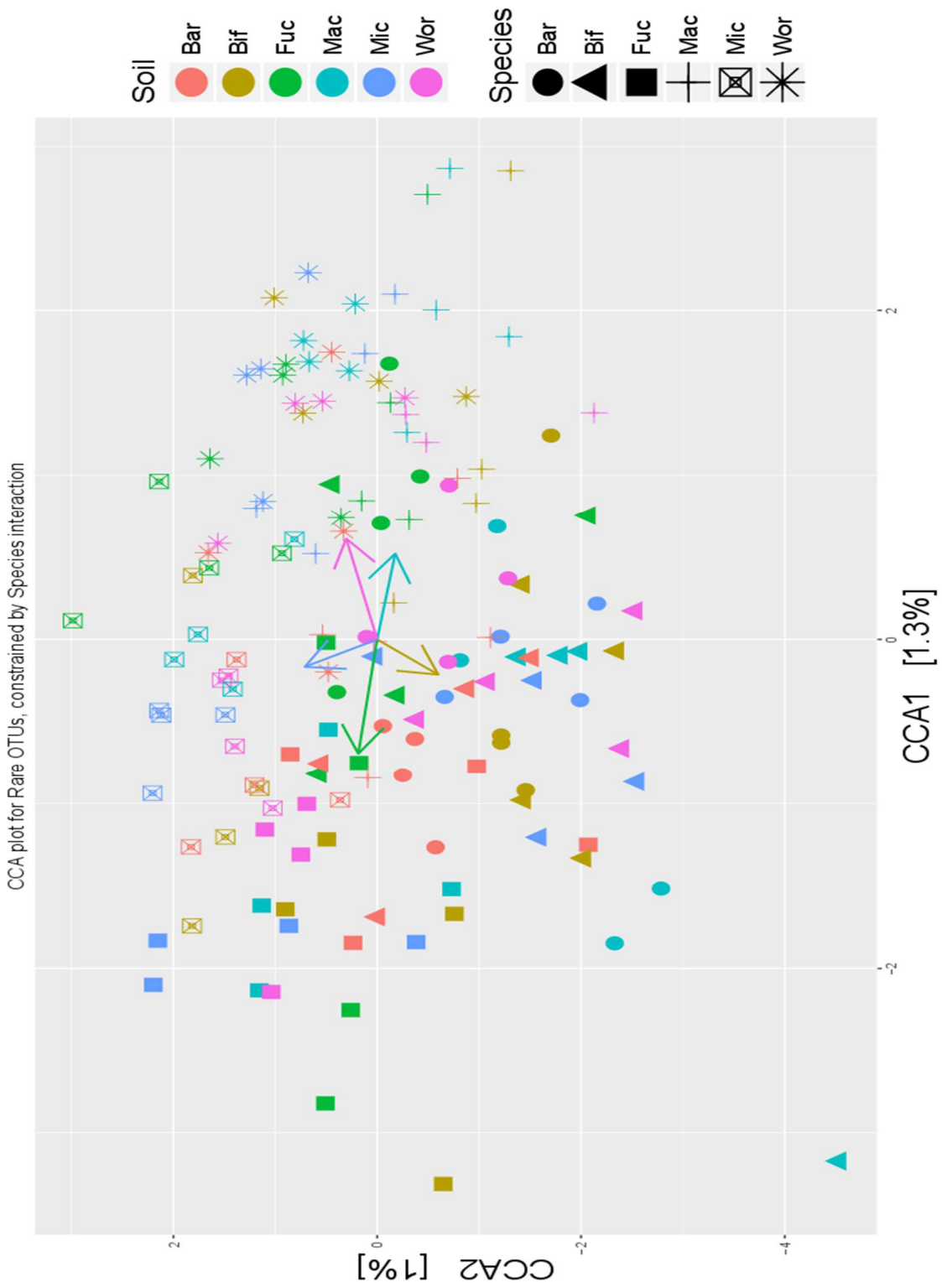


Figure 9c.ii: CCA ordinations on rare OTU tables constrained by Species factor

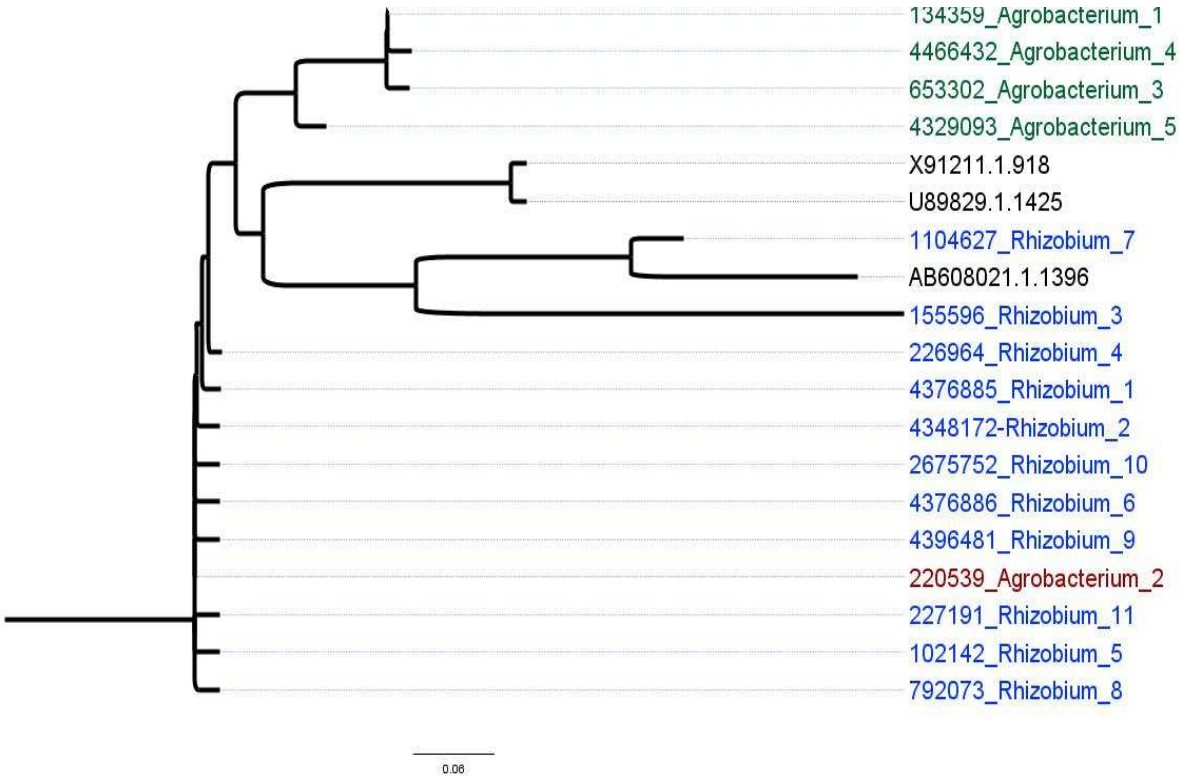


Figure 10: Phylogenetic tree of Agrobacterium and Rhizobium.

The tree represents a phylogeny between agrobacterium OTUs and rhizobium OTUs from our samples. The agrobacterium coloured in red is the one that is present in high abundance within our samples (~85-95% of OTUs). Labels in black are reference rhizobium sequences.

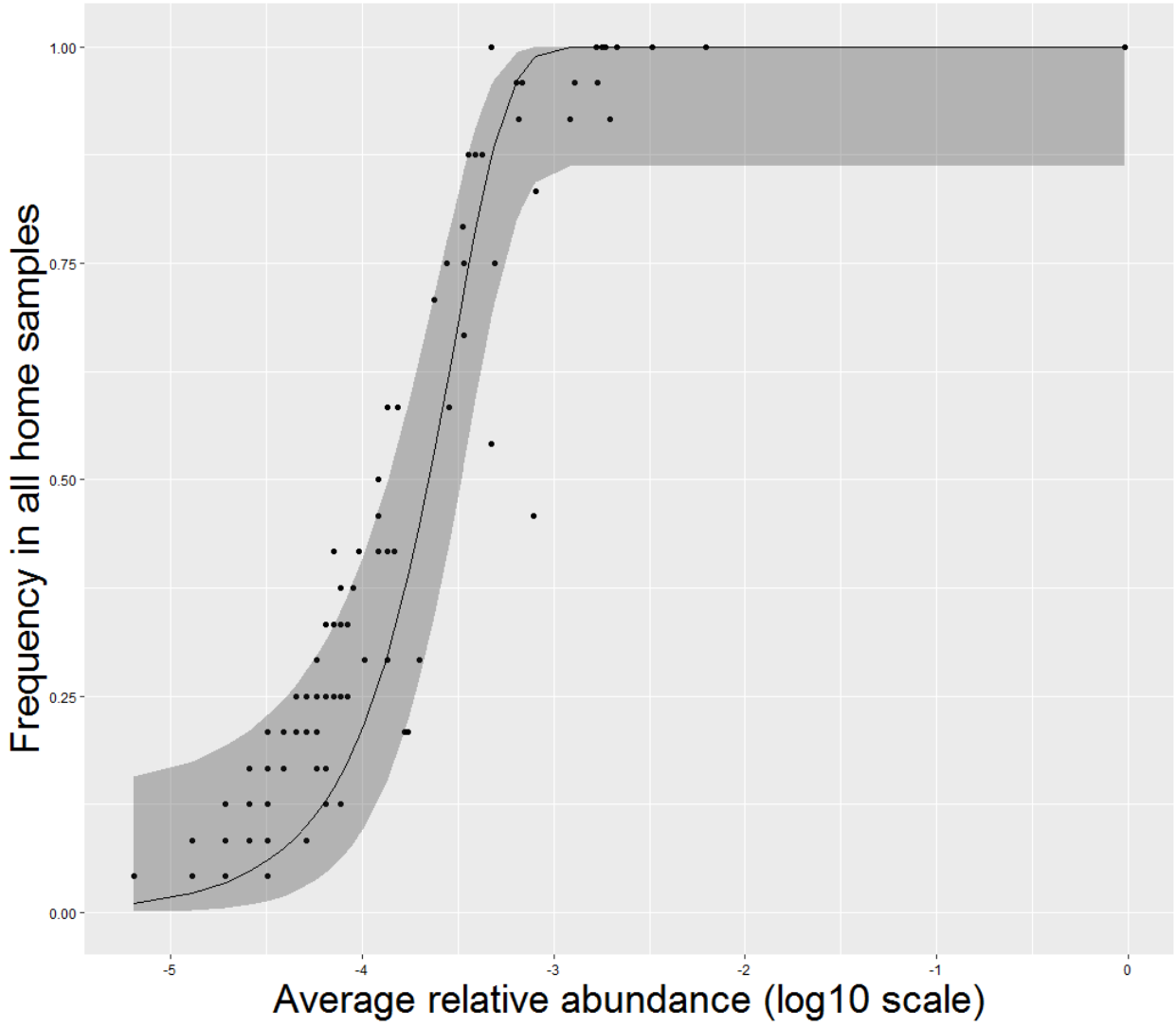


Figure 11.a: Fit of neutral model for home samples

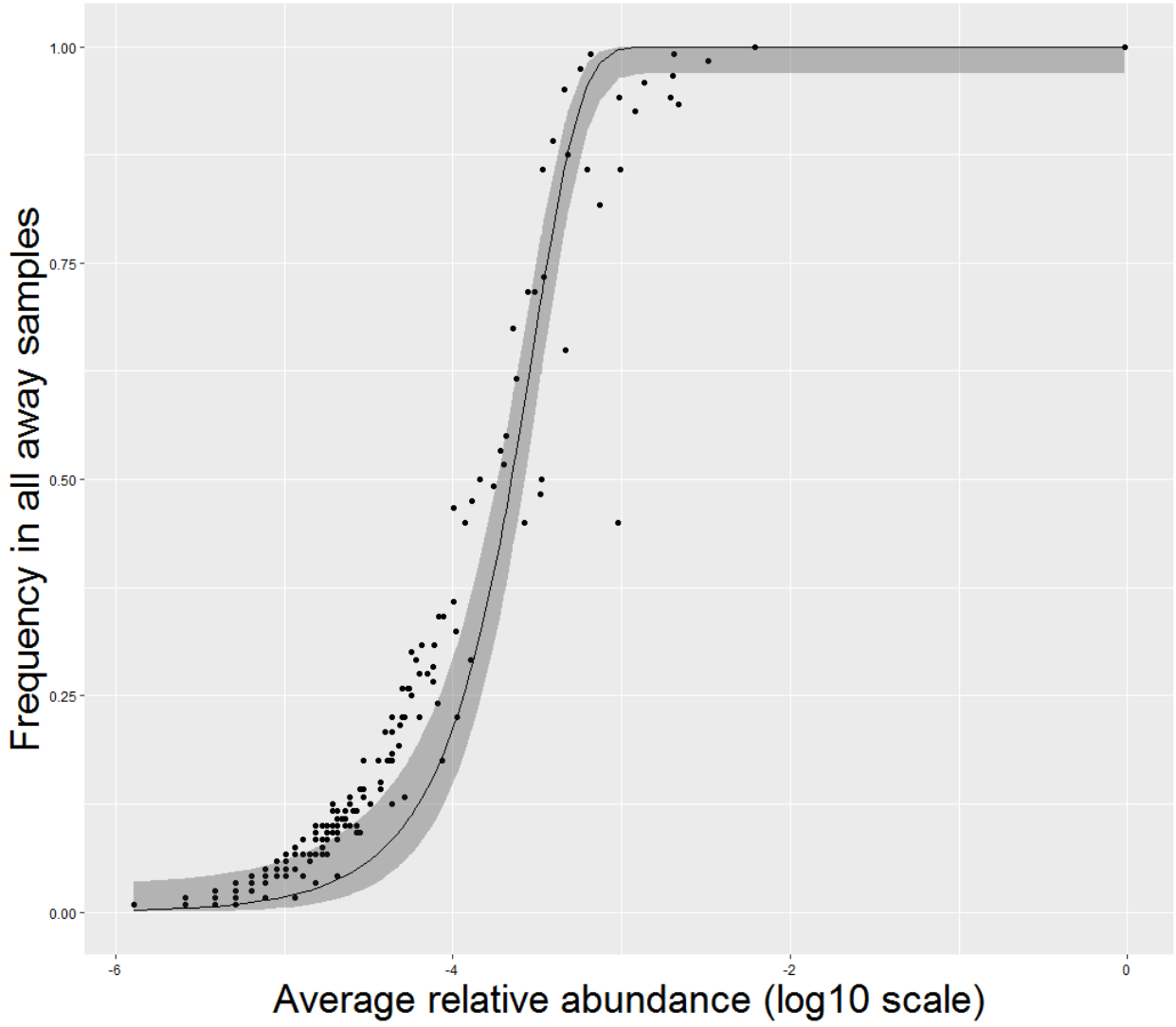


Figure 11.b: Fit of neutral model for away data

APPENDIX 3: Main script

This script is used to process the OTU table received from QIIME. The script runs a filtering step, diversity analysis, ordinations and Adonis models.

```
#required libraries
library("RColorBrewer")
library("ggplot2")
library("plyr")
library("vegan")
library("reshape2")
library("ape")
library("phyloseq")
library("data.table")
library("biome")

#otu files
otu_all = "../Desktop/nod_mb/R_files/otus.biom"
m=read.csv("../Desktop/nod_mb/R_files/nodule_map_new.txt", row.names=1)
file<-import_biom(otu_all)
map = sample_data(m)
comp <-merge_phyloseq(file,map)
head(sample_data(comp))
colnames(tax_table(comp)) <- c(k = "Kingdom", p = "Phylum", c = "Class", o = "Order", f
= "Family", g = "Genus", s = "Species")
taxa_sums(comp)
#complete otu mb
comp

#removing low quality taxa matches and normalize
mb_all= subset_taxa(comp, Kingdom!="None" & Kingdom!="NOHIT" )
norm_all = transform_sample_counts(mb_all, function(x) x/sum(x))
taxa_sums(mb_all)
mb_all

df_all = as(sample_data(norm_all), "data.frame")
d_all = phyloseq::distance(norm_all, "bray")
norm_all

#only rhiz and normalize
mb_rhiz= subset_taxa(mb_all, Order=="o__Rhizobiales")
norm_rhiz = transform_sample_counts(mb_rhiz, function(x) x/sum(x))
df_rhiz = as(sample_data(norm_rhiz), "data.frame")
d_rhiz = phyloseq::distance(norm_rhiz, "bray")
```

```
#take rare as a percentage of the reads => OTUS <0.01% in abundance is "rare"

#only rare and normalize
#get OTUs
x = names(sort(taxa_sums(mb_rhiz), decreasing = TRUE))
y = names(sort(taxa_sums(mb_all), decreasing = TRUE))
select_rare <- setdiff(y, x)

#subsample and normalize
mb_rare= prune_taxa(select_rare, mb_all)
norm_rare = transform_sample_counts(mb_rare, function(x) x/sum(x))
df_rare = as(sample_data(norm_rare), "data.frame")
d_rare = phyloseq::distance(norm_rare, "bray")

taxa <- tax_table(mb_all)
otus <- otu_table(mb_all)
write.table(otus, file='otus-genus.txt')
write.table(taxa, file='taxa.txt')

taxa_rhiz <- tax_table(mb_rhiz)
otus_rhiz <- otu_table(mb_rhiz)
write.table(otus_rhiz, file='otus-rhiz.txt')
write.table(taxa_rhiz, file='taxa-rhiz.txt')

taxa_rare <- tax_table(mb_rare)
otus_rare <- otu_table(mb_rare)
write.table(otus_rare, file='otus-rare.txt')
write.table(taxa_rare, file='taxa-rare.txt')

#rarefy tables to same depth
even_depth <- function(ss_no, otu_table){
#set.seed(3336)
rarefied <- rarefy_even_depth(otu_table, sample.size = ss_no)
return(rarefied)
}


set.seed(123)
mb_rare <- even_depth(163, mb_rare)
mb_rhiz <- even_depth(6200, mb_rhiz)
mb_all <- even_depth(6500, mb_all)

#read distribution
readsumsdf_all = data.frame(nreads = sort(sample_sums(mb_all),TRUE), sorted =
1:nsamples(mb_all), type = "Samples")
range(readsumsdf_all$nreads)
```

```
readsumsdf_rhiz = data.frame(nreads = sort(sample_sums(mb_rhiz),TRUE), sorted =
1:nsamples(mb_all), type = "Samples")
range(readsumsdf_rhiz$nreads)
readsumsdf_rare = data.frame(nreads = sort(sample_sums(mb_rare),TRUE), sorted =
1:nsamples(mb_all), type = "Samples")
range(readsumsdf_rare$nreads)

#plot read_counts
sdt = data.table(as(sample_data(mb_all), "data.frame"), TotalReads =
sample_sums(mb_all), keep.rownames = TRUE)
setnames(sdt, "rn", "SampleID")
ggplot(sdt, aes(x=SampleID,y=TotalReads,fill=Comb)) + geom_bar(stat="identity",
position=position_dodge()) +facet_wrap(facets = ~Soil)
ggplot(sdt, aes(x=Soil,y=TotalReads,fill=Comb)) + geom_bar(stat="identity")
+facet_wrap(facets = ~Soil)
ggplot(sdt, aes(x=SampleID,y=TotalReads)) + geom_bar(stat="identity")

#save file
read_dist <- subset(sdt, select=c("SampleID", "Soil", "Species","Comb", "TotalReads"))
write.table(read_dist,file="../Desktop/read_dist.txt")

#otu abundances and frequencies
abund_val <- function(normalized){
  otu.abun = apply(otu_table(normalized),1,mean)
  # Calculate the frequency of each OTU across all samples
  otu.freq = rowSums(otu_table(normalized) != 0)/144
  # Reassign names of phyla so we only color by the top 5 phyla and mark all others as
"other"
  phyla = as.vector(data.frame(tax_table(normalized))$Phylum)
  levels(phyla) = c(levels(phyla),"other")
  keephyla = c("p__Bacteroidetes","p__Proteobacteria","p__Actinobacteria",
"p__Chloroflexi", "p__TM7")
  phyla[!(phyla %in% keephyla)] = "Other"
  phyla = as.vector(phyla)
  phyla=as.factor(phyla)
  otuabun = cbind.data.frame(abundance=log(otu.abun),frequency=otu.freq,phyla)
  return(otuabun)
}


#get values to plot chart
abun_all <-  abund_val(norm_all)
abun_rhiz <- abund_val(norm_rhiz)
abun_rare <- abund_val(norm_rare)

# Use color brewer to pick a color scheme for the phyla
```

```
brew = brewer.pal(6, "Set1")
# Create a scatterplot of OTUs showing their average relative abundance and frequency
ggplot(abun_all, aes(x=abundance,y=frequency,color=phyla)) + geom_point(size=3) +
xlab("Average relative abundance (log scale)") + ylab("frequency in all samples - All
mb") + scale_colour_brewer(palette="Set2")+labs(title="Distribution of all OTU phyla-
Complete microbiome")+ xlim(-20, 5)
ggplot(abun_rare, aes(x=abundance,y=frequency,color=phyla)) + geom_point(size=3) +
xlab("Average relative abundance (log scale)") + ylab("frequency in all samples - Rare
mb") + scale_colour_brewer(palette="Set2")+labs(title="Distribution of all OTU phyla-
Rare microbiome")+ xlim(-20, 5)
ggplot(abun_rhiz, aes(x=abundance,y=frequency,color=phyla)) + geom_point(size=3) +
xlab("Average relative abundance (log scale)") + ylab("frequency in all samples - Rhiz
mb") + scale_colour_brewer(palette="Set2")+labs(title="Distribution of all OTU phyla-
Rhizobiales microbiome")+ xlim(-20, 5)
#bar plots needed for rare and rhizobiales taxa

#distribution of phylums
abudance_cal <- function(normalized_otu, Taxa_sel){
  bar_table <- plot_bar(normalized_otu, fill="Phylum")
  bar_table <- bar_table$data[bar_table$data$Abundance>0,]
  bar_taxa <- bar_table[bar_table$Phylum==Taxa_sel,]
  bar_taxa <- bar_taxa[c(2,3,17)]
  sum_prot <- ddply(bar_taxa, .(Sample), summarize, Sum=sum(Abundance))
  return(sum_prot)
}

prot_all <- abudance_cal(norm_all, "p__Proteobacteria")
bact_all <- abudance_cal(norm_all, "p__Bacteroidetes")
act_all <- abudance_cal(norm_all, "p__Actinobacteria")

prot_rare <- abudance_cal(norm_rare, "p__Proteobacteria")
bact_rare <- abudance_cal(norm_rare, "p__Bacteroidetes")
act_rare <- abudance_cal(norm_rare, "p__Actinobacteria")


#Richness
adiv_anova <- function(ss_no, otu_table, outf){
  mlist <- factor(c("Observed", "Shannon", "Chao1"))
  set.seed(3336)
  rarefied <- rarefy_even_depth(otu_table, sample.size = ss_no)
  richness <- estimate_richness(rarefied, measures=mlist)
  richness <- richness[-3]
  sdata <- sample_data(rarefied)
  for ( i in 1:3){
    aov_facs <- aov(richness[[i]]~ sdata$Soil*sdata$Species)
    capture.output(summary(aov_facs), file=outf, append=T)
```

```
  }
}

adiv_anova(160, mb_all, "ALL_160.txt")
adiv_anova(500, mb_all, "ALL_500.txt")
adiv_anova(1000, mb_all, "ALL_1000.txt")

adiv_anova(160, mb_rhiz, "Rhiz_160.txt")
adiv_anova(500, mb_rhiz, "Rhiz_500.txt")
adiv_anova(1000, mb_rhiz, "Rhiz_1000.txt")

adiv_anova(160, mb_rare, "Rare_160.txt")
adiv_anova(500, mb_rare, "Rare_500.txt")
adiv_anova(1000, mb_rare, "Rare_1000.txt")


#One more for nod_size
adiv_anova_ns <- function(ss_no, otu_table, outf){
  mlist <- factor(c("Observed", "Shannon", "Chao1"))
  set.seed(3336)
  rarefied <- rarefy_even_depth(otu_table, sample.size = ss_no)
  richness <- estimate_richness(rarefied, measures=mlist)
  richness <- richness[-3]
  sdata <- sample_data(rarefied)
  for ( i in 1:3){
    aov_facs <- aov(richness[[i]]~ sdata$nod_size)
    capture.output(summary(aov_facs), file=outf, append=T)
  }
}

adiv_anova_ns(160, mb_all, "ALL_160_ns.txt")
adiv_anova_ns(500, mb_all, "ALL_500_ns.txt")

adiv_anova_ns(160, mb_rhiz, "Rhiz_160_ns.txt")
adiv_anova_ns(500, mb_rhiz, "Rhiz_500_ns.txt")

adiv_anova_ns(160, mb_rare, "Rare_160_ns.txt")
adiv_anova_ns(500, mb_rare, "Rare_500_ns.txt")


#Ordinations: PCoA principal
mb.pcoa.rare <- ordinate(norm_rare, d_rare, method="PCoA")
pcoa.rare<-plot_ordination(norm_rare, mb.pcoa.rare, color="Soil",
shape="Species")+geom_point(size=5)+labs(title="PCoA plot for rare OTUs")
pcoa.rare
plot_scree(mb.pcoa.rare, "Scree plot for PCoA with bray-curtis distance")
```

```
mb.pcoa.rhiz <- ordinate(norm_rhiz, d_rhiz, method="PCoA")
pcoa.rhiz<-plot_ordination(norm_rhiz, mb.pcoa.rhiz, color="Soil",
shape="Species")+geom_point(size=5)+labs(title="PCoA plot for Rhizobales OTUs")
pcoa.rhiz
plot_scree(mb.pcoa.rhiz, "Scree plot for PCoA with bray-curtis distance")

mb.pcoa.all <- ordinate(norm_all, d_all, method="PCoA")
pcoa.all<-plot_ordination(norm_all, mb.pcoa.all, color="Soil",
shape="Species")+geom_point(size=5)+labs(title="PCoA plot for all OTUs")
pcoa.all
plot_scree(mb.pcoa.all, "Scree plot for PCoA with bray-curtis distance")


#Ordination: CCA constrained by species soil interaction
#seems like most are along a single axis now.. not sure how to communicate this
result?
mball.cca <- ordinate(mb_all, method= "CCA", formula=norm_all~Soil)
all_cca <- plot_ordination(norm_all, mball.cca, color="Soil",
shape="Species")+geom_point(size=5)+labs(title="CCA plot for All OTUs, constrained
by Species interaction")
all_cca

arrowmat = vegan::scores(mball.cca, display = "bp")
arrowdf <- data.frame(labels = rownames(arrowmat), arrowmat)
# Define the arrow aesthetic mapping
arrow_map = aes(xend = CCA1, yend = CCA2, x = 0, y = 0, shape = NULL, color =
NULL,
              label = labels)
label_map = aes(x = 1.2 * CCA1, y = 1.2 * CCA2, shape = NULL, color = NULL,
              label = labels)
# Make a new graphic
arrowhead = arrow(length = unit(0.05, "npc"))
p1 = all_cca+ geom_segment(arrow_map, size = 0.2, data = arrowdf, color = "black",
                       arrow = arrowhead) + geom_text(label_map, size = 2, data = arrowdf)
p1

############CCA with rare otu table
mbrare.cca <- ordinate(mb_rare, method= "CCA", formula=norm_rare~Species)
rare_cca <- plot_ordination(norm_rare, mbrare.cca, color="Soil",
shape="Species")+geom_point(size=5)+labs(title="CCA plot for Rare OTUs,
constrained by Species interaction")
rare_cca

arrowmat = vegan::scores(mbrare.cca, display = "bp")
arrowdf <- data.frame(labels = rownames(arrowmat), arrowmat)
```

```
# Define the arrow aesthetic mapping
arrow_map = aes(xend = CCA1, yend = CCA2, x = 0, y = 0, shape = NULL, color =
NULL,
              label = labels)
label_map = aes(x = 1.2 * CCA1, y = 1.2 * CCA2, shape = NULL, color = NULL,
              label = labels)
# Make a new graphic
arrowhead = arrow(length = unit(0.05, "npc"))
p1 = rare_cca+ geom_segment(arrow_map, size = 0.2, data = arrowdf, color = "black",
                        arrow = arrowhead) + geom_text(label_map, size = 2, data = arrowdf)
p1

##############CCA with rhizobiaes otu table
mbrhiz.cca <- ordinate(mb_rhiz, method= "CCA", formula=norm_rhiz~Soil)
rhiz_cca <- plot_ordination(norm_rhiz, mbrhiz.cca, color="Soil",
shape="Species")+geom_point(size=5)+labs(title="CCA plot for Rhizobiales OTUs,
constrained by Soil interaction")
rhiz_cca

arrowmat = vegan::scores(mbrhiz.cca, display = "bp")
arrowdf <- data.frame(labels = rownames(arrowmat), arrowmat)
# Define the arrow aesthetic mapping
arrow_map = aes(xend = CCA1, yend = CCA2, x = 0, y = 0, shape = NULL, color =
NULL,
              label = labels)
label_map = aes(x = 1.2 * CCA1, y = 1.2 * CCA2, shape = NULL, color = NULL,
              label = labels)
# Make a new graphic
arrowhead = arrow(length = unit(0.05, "npc"))
p1 = rhiz_cca+ geom_segment(arrow_map, size = 0.2, data = arrowdf, color = "black",
                        arrow = arrowhead) + geom_text(label_map, size = 2, data = arrowdf)
p1

#NMDS visualize the distance between the points themselves.
nmds_ordinations <- function(mb_otu, norm_otu, d_otu, title){
mb.nmds<- ordinate(norm_otu, d_otu, method="NMDS", "bray")
  plot_ordination(norm_otu, mb.nmds, type = "taxa", color = "Phylum", title = "taxa")
}

nmds_ordinations(mb_all,norm_all, d_all, "ALL")
nmds_ordinations(mb_rhiz,norm_rhiz, d_rhiz, "Rhizobiales")
nmds_ordinations(mb_rare,norm_rare, d_rare, "Rare")

#adonis
#beta div -- adonis rare using distances identified above
#rare
```

```
set.seed(123)
ads1<-adonis(d_rare ~ Native, df_rare, permutations = 999) # not imp
ads2<-adonis(d_rare ~ Species , df_rare, permutations = 999) # imp 8
ads3<-adonis(d_rare ~ Soil , df_rare, permutations = 999) #
ads4<-adonis(d_rare ~ nod_size , df_rare, permutations = 999) #
ads5<-adonis(d_rare ~ Soil+Species , df_rare, permutations = 999)
ads6<-adonis(d_rare ~ Soil:Species , df_rare, permutations = 999)
ads7<-adonis(d_rare ~ Soil*Species , df_rare, permutations = 999)
ads8<-adonis(d_rare ~ X18s_PD , df_rare, permutations = 999) #
ads9<-adonis(d_rare ~ trna_PD , df_rare, permutations = 999) #
ads10<-adonis(d_rare ~ elison_PD , df_rare, permutations = 999) #
ads11<-adonis(d_rare ~ Soil * Species * nod_size , df_rare, permutations = 999) #

#rhiz
 set.seed(123)
 ads1<-adonis(d_rhiz ~ Native, df_rhiz, permutations = 999) # not imp
 ads2<-adonis(d_rhiz ~ Species , df_rhiz, permutations = 999) # imp 8
ads3<-adonis(d_rhiz ~ Soil , df_rhiz, permutations = 999) #
 ads4<-adonis(d_rhiz ~ nod_size , df_rhiz, permutations = 999) #
 ads5<-adonis(d_rhiz ~ Soil+Species , df_rhiz, permutations = 999)
 ads6<-adonis(d_rhiz ~ Soil:Species , df_rhiz, permutationads1s = 999)
ads7<-adonis(d_rhiz ~ Soil*Species , df_rhiz, permutations = 999)
 ads8<-adonis(d_rhiz ~ Native*nod_size , df_rhiz, permutations = 999)
 ads9<-adonis(d_rhiz ~ nod_size*Soil*Species , df_rhiz, permutations = 999)
  ads10<-adonis(d_rhiz ~ X18s_PD , df_rhiz, permutations = 999) # ads11<-
adonis(d_rhiz ~ trna_PD , df_rhiz, permutations = 999) #
 ads12<-adonis(d_rhiz ~ elison_PD , df_rhiz, permutations = 999) #

ads13<-adonis(d_rhiz ~ Soil+Species+Soil:Species+nod_size , df_rhiz, permutations =
999) #


#all
set.seed(123)
ads1<-adonis(d_all ~ Native, df_all, permutations = 999) # not imp
ads2<-adonis(d_all ~ Species , df_all, permutations = 999) # imp 8
ads3<-adonis(d_all ~ Soil , df_all, permutations = 999) #
ads4<-adonis(d_all ~ nod_size , df_all, permutations = 999) #
ads5<-adonis(d_all ~ Soil*Species , df_all, permutations = 999)

ads6<-adonis(d_all ~ Soil:Species , df_all, permutations = 999)
ads7<-adonis(d_all ~ Soil*Species , df_all, permutations = 999)
ads8<-adonis(d_all ~ Soil*Species*nod_size , df_all, permutations = 999)
ads10<-adonis(d_all ~ X18s_PD , df_all, permutations = 999) #
ads11<-adonis(d_all ~ trna_PD , df_all, permutations = 999) #
ads12<-adonis(d_all ~ elison_PD , df_all, permutations = 999) #
```

APPENDIX 4: Core microbiome calculations

```
library(UpSetR)

bar= subset_samples(mb_all, Comb=="Bar")
bif= subset_samples(mb_all, Comb=="Bif")
mac= subset_samples(mb_all, Comb=="Mac")
mic= subset_samples(mb_all, Comb=="Mic")
wor= subset_samples(mb_all, Comb=="Wor")
fuc= subset_samples(mb_all, Comb=="Fuc")

abun_freq <- function(otu_table){
  otu_table = filter_taxa(otu_table, function(x) mean(x) > 0, TRUE)
  norm_cat <-  transform_sample_counts(otu_table, function(x) x/sum(x))
  otu.abun = apply(otu_table(norm_cat),1,mean)
  otu.freq = rowSums(otu_table(otu_table) != 0)/nsamples(otu_table)
  otuabun = cbind.data.frame(abundance=otu.abun,frequency=otu.freq)
  return(otuabun)
}

bar_a<- abun_freq(bar)
bif_a<- abun_freq(bif)
fuc_a<- abun_freq(fuc)
wor_a<- abun_freq(wor)
mac_a<- abun_freq(mac)
mic_a<- abun_freq(mic)


trial <- merge(otu_table(mb_all),mac_a,  by="row.names", all=T)
rownames(trial) <- trial$Row.names
trial <- trial[,c(146:147)]

trial <- merge(trial,mic_a,  by="row.names", all=T)
rownames(trial) <- trial$Row.names
trial <- trial[,-c(1)]

trial <- merge(trial, wor_a,by="row.names", all = T)
rownames(trial) <- trial$Row.names
trial <- trial[,-c(1)]

trial <- merge(trial, fuc_a,by="row.names", all = T)
rownames(trial) <- trial$Row.names
trial <- trial[,-c(1)]

trial <- merge(trial, bif_a,by="row.names", all = T)
```

```
rownames(trial) <- trial$Row.names
trial <- trial[,-c(1)]

trial <- merge(trial, bar_a,by="row.names", all = T)
rownames(trial) <- trial$Row.names
trial <- trial[,-c(1)]

colnames(trial) <- c("mac_a_abun", "mac_a_freq", "mic_a_abun", "mic_a_freq",
"wor_a_abun", "wor_a_freq","fuc_a_abun", "fuc_a_freq","bif_a_abun",
"bif_a_freq","bar_a_abun", "bar_a_freq")

trial[is.na(trial)] <- 0
mat_freq <- trial[,c(2,4,6,8,10,12)]

mat_comp <- trial


#presence/absence core otus
half_freq <- mat_freq
half_freq[half_freq<0.5] <- 0
half_freq[half_freq>=0.5] <- 1


third_freq <- mat_freq
third_freq[third_freq<0.75] <- 0
third_freq[third_freq>=0.75] <- 1

full_freq <- mat_freq
full_freq[full_freq<1] <- 0
full_freq[full_freq>=1] <- 1

write.table(half_freq, file="half_freq.txt")
write.table(third_freq, file="third_freq.txt")
write.table(full_freq, file="full_freq.txt")

#Upset plots
upset(half_freq,  sets=c("fuc_a_freq", "wor_a_freq" ,"mic_a_freq" ,"mac_a_freq"
,"bif_a_freq" ,"bar_a_freq"), sets.bar.color = "#56B4E9",order.by = "freq")
upset(third_freq,  sets=c("fuc_a_freq", "wor_a_freq" ,"mic_a_freq" ,"mac_a_freq"
,"bif_a_freq" ,"bar_a_freq"), sets.bar.color = "#56B4E9",order.by = "freq")
upset(full_freq,  sets=c("fuc_a_freq", "wor_a_freq" ,"mic_a_freq" ,"mac_a_freq"
,"bif_a_freq" ,"bar_a_freq"), sets.bar.color = "#56B4E9",order.by = "freq")


#complete the taxonomy table
get_ids <- function(frequency_tab, out_file){
```

```
  row_sub = apply(frequency_tab, 1, function(row) all(row !=0 ))
  taxa_tab <- merge(frequency_tab[row_sub,], tax_table(mb_all), by="row.names")
  return(table(taxa_tab$Phylum))
#  write.table(taxa_tab,file=out_file)
}

get_ids(half_freq, "half_taxa.txt")
get_ids(third_freq, "third_taxa.txt")
get_ids(full_freq, "full_taxa.txt")


###abundance
head(mat_comp)
#add in abundace threshold.

abun.00001 <- mat_comp[mat_comp["mac_a_abun"]>=0.00001 &
mat_comp["mic_a_abun"]>=0.00001& mat_comp["fuc_a_abun"]>=0.00001&
mat_comp["wor_a_abun"]>=0.00001& mat_comp["bif_a_abun"]>=0.00001&
mat_comp["bar_a_abun"]>=0.00001,]
#half <- abun.00001[abun.00001["mac_a_freq"]>=0.5 &
abun.00001["mic_a_freq"]>=0.5& abun.00001["fuc_a_freq"]>=0.5&
abun.00001["wor_a_freq"]>=0.5& abun.00001["bif_a_freq"]>=0.5&
abun.00001["bar_a_freq"]>=0.5,]
abund_freq <- abun.00001[,c(2,4,6,8,10,12)]

half_abun <- abund_freq
half_abun[half_abun<0.5] <- 0
half_abun[half_abun>=0.5] <- 1
upset(half_abun,  sets=c("fuc_a_freq", "wor_a_freq" ,"mic_a_freq" ,"mac_a_freq"
,"bif_a_freq" ,"bar_a_freq"), sets.bar.color = "#56B4E9",order.by = "freq")

third_abun <- abund_freq
third_abun[third_abun<0.75] <- 0
third_abun[third_abun>=0.75] <- 1
upset(third_abun,  sets=c("fuc_a_freq", "wor_a_freq" ,"mic_a_freq" ,"mac_a_freq"
,"bif_a_freq" ,"bar_a_freq"), sets.bar.color = "#56B4E9",order.by = "freq")

full_abun <- abund_freq
full_abun[full_abun<1] <- 0
full_abun[full_abun>=1] <- 1
upset(full_abun,  sets=c("fuc_a_freq", "wor_a_freq" ,"mic_a_freq" ,"mac_a_freq"
,"bif_a_freq" ,"bar_a_freq"), sets.bar.color = "#56B4E9",order.by = "freq")

write.table(half_abun, file="half_abun.txt")
write.table(third_abun, file="third_abun.txt")
write.table(full_abun, file="full_abun.txt")
```

```
get_ids(half_abun, "half_abun_taxa.txt")
get_ids(third_abun, "third_abun_taxa.txt")
get_ids(full_abun, "full_abun_taxa.txt")
```

APPENDIX 5: Neutral model analysis

```
#Adam Burns - 2/10/2015
#aburns2@uoregon.edu
#Fits the neutral model from Sloan et al. 2006 to an OTU table and returns several
fitting statistics. Alternatively, will return predicted occurrence frequencies for each OTU
based on their abundance in the metacommunity when stats=FALSE.

sncm.fit <- function(spp, pool=NULL, stats=TRUE, taxon=NULL){
  require(minpack.lm)
  require(Hmisc)
  require(stats4)

  options(warn=-1)

  #Calculate the number of individuals per community
  N <- mean(apply(spp, 1, sum))

  #Calculate the average relative abundance of each taxa across communities
  if(is.null(pool)){
    p.m <- apply(spp, 2, mean)
    p.m <- p.m[p.m != 0]
    p <- p.m/N
  } else {
    p.m <- apply(pool, 2, mean)
    p.m <- p.m[p.m != 0]
    p <- p.m/N
  }

  #Calculate the occurrence frequency of each taxa across communities
  spp.bi <- 1*(spp>0)
  freq <- apply(spp.bi, 2, mean)
  freq <- freq[freq != 0]

  #Combine
  C <- merge(p, freq, by=0)
  C <- C[order(C[,2]),]
  C <- as.data.frame(C)
  C.0 <- C[!(apply(C, 1, function(y) any(y == 0))),] #Removes rows with any zero (absent
in either source pool or local communities)
  p <- C.0[,2]
  freq <- C.0[,3]
  names(p) <- C.0[,1]
  names(freq) <- C.0[,1]
```

```
  #Calculate the limit of detection
  d = 1/N

  ##Fit model parameter m (or Nm) using Non-linear least squares (NLS)
  m.fit <- nlsLM(freq ~ pbeta(d, N*m*p, N*m*(1-p), lower.tail=FALSE), start=list(m=0.1))
  m.ci <- confint(m.fit, 'm', level=0.95)

  ##Fit neutral model parameter m (or Nm) using Maximum likelihood estimation (MLE)
  sncm.LL <- function(m, sigma){
    R = freq - pbeta(d, N*m*p, N*m*(1-p), lower.tail=FALSE)
    R = dnorm(R, 0, sigma)
    -sum(log(R))
  }
  m.mle <- mle(sncm.LL, start=list(m=0.1, sigma=0.1), nobs=length(p))

  ##Calculate Akaike's Information Criterion (AIC)
  aic.fit <- AIC(m.mle, k=2)
  bic.fit <- BIC(m.mle)

  ##Calculate goodness-of-fit (R-squared and Root Mean Squared Error)
  freq.pred <- pbeta(d, N*coef(m.fit)*p, N*coef(m.fit)*(1-p), lower.tail=FALSE)
  Rsqr <- 1 - (sum((freq - freq.pred)^2))/(sum((freq - mean(freq))^2))
  RMSE <- sqrt(sum((freq-freq.pred)^2)/(length(freq)-1))

  pred.ci <- binconf(freq.pred*nrow(spp), nrow(spp), alpha=0.05, method="wilson",
return.df=TRUE)

  ##Calculate AIC for binomial model
  bino.LL <- function(mu, sigma){
    R = freq - pbinom(d, N, p, lower.tail=FALSE)
    R = dnorm(R, mu, sigma)
    -sum(log(R))
  }
  bino.mle <- mle(bino.LL, start=list(mu=0, sigma=0.1), nobs=length(p))

  aic.bino <- AIC(bino.mle, k=2)
  bic.bino <- BIC(bino.mle)

  ##Goodness of fit for binomial model
  bino.pred <- pbinom(d, N, p, lower.tail=FALSE)
  Rsqr.bino <- 1 - (sum((freq - bino.pred)^2))/(sum((freq - mean(freq))^2))
  RMSE.bino <- sqrt(sum((freq - bino.pred)^2)/(length(freq) - 1))

  bino.pred.ci <- binconf(bino.pred*nrow(spp), nrow(spp), alpha=0.05, method="wilson",
return.df=TRUE)
```

```
  ##Calculate AIC for Poisson model
  pois.LL <- function(mu, sigma){
    R = freq - ppois(d, N*p, lower.tail=FALSE)
    R = dnorm(R, mu, sigma)
    -sum(log(R))
  }
  pois.mle <- mle(pois.LL, start=list(mu=0, sigma=0.1), nobs=length(p))

  aic.pois <- AIC(pois.mle, k=2)
  bic.pois <- BIC(pois.mle)

  ##Goodness of fit for Poisson model
  pois.pred <- ppois(d, N*p, lower.tail=FALSE)
  Rsqr.pois <- 1 - (sum((freq - pois.pred)^2))/(sum((freq - mean(freq))^2))
  RMSE.pois <- sqrt(sum((freq - pois.pred)^2)/(length(freq) - 1))

  pois.pred.ci <- binconf(pois.pred*nrow(spp), nrow(spp), alpha=0.05, method="wilson",
return.df=TRUE)

  ##Results
  if(stats==TRUE){
    fitstats <- data.frame(m=numeric(), m.ci=numeric(), m.mle=numeric(),
maxLL=numeric(), binoLL=numeric(), poisLL=numeric(), Rsqr=numeric(),
Rsqr.bino=numeric(), Rsqr.pois=numeric(), RMSE=numeric(), RMSE.bino=numeric(),
RMSE.pois=numeric(), AIC=numeric(), BIC=numeric(), AIC.bino=numeric(),
BIC.bino=numeric(), AIC.pois=numeric(), BIC.pois=numeric(), N=numeric(),
Samples=numeric(), Richness=numeric(), Detect=numeric())
    fitstats[1,] <- c(coef(m.fit), coef(m.fit)-m.ci[1], m.mle@coef['m'], m.mle@details$value,
bino.mle@details$value, pois.mle@details$value, Rsqr, Rsqr.bino, Rsqr.pois, RMSE,
RMSE.bino, RMSE.pois, aic.fit, bic.fit, aic.bino, bic.bino, aic.pois, bic.pois, N, nrow(spp),
length(p), d)
    return(fitstats)
  } else {
    A <- cbind(p, freq, freq.pred, pred.ci[,2:3], bino.pred, bino.pred.ci[,2:3])
    A <- as.data.frame(A)
    colnames(A) <- c('p', 'freq', 'freq.pred', 'pred.lwr', 'pred.upr', 'bino.pred', 'bino.lwr',
'bino.upr')
    if(is.null(taxon)){
      B <- A[order(A[,1]),]
    } else {
      B <- merge(A, taxon, by=0, all=TRUE)
      row.names(B) <- B[,1]
      B <- B[,-1]
      B <- B[order(B[,1]),]
    }
    return(B)
```

```
  }
}

#required libraries
library("RColorBrewer")
library("ggplot2")
library("plyr")
library("vegan")
library("reshape2")
library("ape")
library("phyloseq")
library("data.table")
library("biome")
library("metagenomeSeq")

#otu files
otu_all = "../Desktop/Thesis_files/R_files/otus.biom"
m=read.csv("../Desktop//Thesis_files/R_files/nodule_map_new.txt", row.names=1)
file<-import_biom(otu_all)
map = sample_data(m)
comp <-merge_phyloseq(file,map)
head(sample_data(comp))
colnames(tax_table(comp)) <- c(k = "Kingdom", p = "Phylum", c = "Class", o = "Order", f
= "Family", g = "Genus", s = "Species")
taxa_sums(comp)
#complete otu mb
comp

#removing low quality taxa matches and normalize
mb_all= subset_taxa(comp, Kingdom!="None" & Kingdom!="NOHIT" )
norm_all = transform_sample_counts(mb_all, function(x) x/sum(x))
taxa_sums(mb_all)
mb_all

#rarefy tables to same depth
even_depth <- function(ss_no, otu_table){
  #set.seed(3336)
  rarefied <- rarefy_even_depth(otu_table, sample.size = ss_no)
  return(rarefied)
}


home= subset_samples(mb_all, Native=="yes")
away= subset_samples(mb_all, Native=="no")

set.seed(24)
```

```
home_mb <- even_depth(6500, home)
away_mb <- even_depth(6500, away)

tab_home <- t(otu_table(home_mb))
tab_away <- t(otu_table(away_mb))

tax <- read.table("../Desktop//Thesis_files/tax_tab.txt", head=T, sep="\t")
tax <- tax[c(2,1)]

mod_stats_home <-sncm.fit(spp=tab_home, stats=TRUE)
mb_table_home <- sncm.fit(spp=tab_home,  stats=FALSE)
otu_home <- cbind(row.names(mb_table_home), mb_table_home$freq)

mod_stats_away <-sncm.fit(spp=tab_away, stats=TRUE)
mb_table_away <- sncm.fit(spp=tab_away, stats=FALSE)
#otu_away <- cbind(row.names(mb_table_away),mb_table_away$freq)



ggplot(mb_table_home, aes(log10(p), freq)) +
  geom_point() +
  geom_ribbon(data=mb_table_home,aes(ymin=pred.lwr,ymax=pred.upr),alpha=0.3)+
  geom_line(data=mb_table_home,aes(y = freq.pred))+
  xlab("Average relative abundance (log10 scale)") + ylab("Frequency in all home
samples") +
  theme(axis.title = element_text(size = 26),axis.text = element_text(colour = "black"))

ggplot(mb_table_away, aes(log10(p), freq)) +
  geom_point() +
  geom_ribbon(data=mb_table_away,aes(ymin=pred.lwr,ymax=pred.upr),alpha=0.3)+
  geom_line(data=mb_table_away,aes(y = freq.pred))+
  xlab("Average relative abundance (log10 scale)") + ylab("Frequency in all away
samples") +
  theme(axis.title = element_text(size = 26),axis.text = element_text(colour = "black"))

home_lwr <- ifelse(mb_table_home$freq<mb_table_home$pred.lwr,
row.names(mb_table_home),0 )
home_upr <- ifelse(mb_table_home$freq>mb_table_home$pred.upr,
row.names(mb_table_home),0 )
table(home_lwr)
table(home_upr)


as <- away_upr %in% row.names(mb_table_away)
away_u <- mb_table_away[as,]
away_u$OTU.ID <- row.names(away_u)
```

```
as <- away_lwr %in% row.names(mb_table_away)
away_l <- mb_table_away[as,]
away_l$OTU.ID <- row.names(away_l)

as <- home_lwr %in% row.names(mb_table_home)
home_l <- mb_table_home[as,]
home_l$OTU.ID <- row.names(home_l)

as <- home_upr %in% row.names(mb_table_home)
home_u <- mb_table_home[as,]
home_u$OTU.ID <- row.names(home_u)

au <- merge(away_u, tax, by="OTU.ID")
al <- merge(away_l, tax, by="OTU.ID")
hu <- merge(home_u, tax, by="OTU.ID")
hl <- merge(home_l, tax, by="OTU.ID")

away_lwr <- ifelse(mb_table_away$freq<mb_table_away$pred.lwr,
row.names(mb_table_home),0 )
away_upr <- ifelse(mb_table_away$freq>mb_table_away$pred.upr,
row.names(mb_table_home),0 )
table(away_lwr)
table(away_upr)

## fishers exact test for OTU presence between home and away samples
otu_fish <- matrix(c(5,14,15,88,251,459),2,3)
fisher.test(otu_fish)

##################
otu_test <- merge(otu_home, otu_away, by=1, all=T)


write.table(au, "away_upr.txt")
write.table(al, "away_lwr.txt")
write.table(hu, "home_upr.txt")
write.table(hl, "home_lwr.txt")

h_ot <- read.table("../Desktop/Thesis_files/home_otus.txt", head=T)
a_ot <- read.table("../Desktop/Thesis_files/away_otus.txt", head=T)
h <- merge(h_ot, tax, by="OTU.ID")
a <- merge(a_ot, tax, by="OTU.ID")
```

REFERENCES

REFERENCES

1. **Ai, C., Liang, G., Sun, J., Wang, X., He, P., Zhou, W., & He, X. (2015).** Reduced dependence of rhizosphere microbiome on plant-derived carbon in 32-year long-term inorganic and organic fertilized soils. Soil Biology and Biochemistry, 80, 70–78. doi:10.1016/j.soilbio.2014.09.028

2. **Amann, R., & Fuchs, B. M. (2008).** Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. Nature Reviews. Microbiology, 6(5), 339–48. doi:10.1038/nrmicro1888

3. **Berg, G., & Smalla, K. (2009).** Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiology Ecology, 68(1), 1–13. doi:10.1111/j.1574-6941.2009.00654.x

4. **Bever, J. D., Platt, T. G., & Morton, E. R. (2012).** Microbial population and community dynamics on plant roots and their feedbacks in plant communities. Annual Review of Microbiology, 66(131), 265–283. doi:10.1146/annurev-micro-092611-150107.Microbial

5. **Bever, J. D., Westover, K. M., & Antonovics, J. (1997).** Incorporating the Soil Community into Plant Population Dynamics : The Utility of the Feedback" British Ecological Society: http://www.jstor.org/sta, 85(5), 561–573.

6. **Bell G (2001)**. Neutral Macroecology. Science, 293(5539) 10.1126/science.293.5539.2413

7. **Brakhage, A. A., & Schroeckh, V. (2011).** Fungal secondary metabolites - Strategies to activate silent gene clusters. Fungal Genetics and Biology, 48(1), 15–22. doi:10.1016/j.fgb.2010.04.004

8. **Broeckling, C. D., Broz, A. K., Bergelson, J., Manter, D. K., & Vivanco, J. M. (2008).** Root exudates regulate soil fungal community composition and diversity. Applied and Environmental Microbiology, 74(3), 738–744. doi:10.1128/AEM.02188-07

9. **Bulgarelli, D., Schlaeppi, K., Spaepen, S., Ver Loren van Themaat, E., & Schulze-Lefert, P. (2013).** Structure and functions of the bacterial microbiota of plants. Annual Review of Plant Biology, 64, 807–38. doi:10.1146/annurev-arplant-050312-120106

10. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, *et al.* (2010).** QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–336.

11. **Chaparro, J. M., Badri, D. V., Bakker, M. G., Sugiyama, A., Manter, D. K., & Vivanco, J. M. (2013).** Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions. PLoS ONE, 8(2), 1–10. doi:10.1371/journal.pone.0055731

12. **Chase J.M. and Leibold M.A. (2003).** Linking Classical and Contemporary Approaches. University of Chicago Press: Chicago, IL,USA.

13. **Chesson, P. (2016).** Mechanisms of Maintenance of Species Diversity.31(2000), 343–358.

14. **Checcucci, A., Azzarello, E, Bazzicalupo M., Galardini M., Lagomarsino A., Mancuso S., Marti L., Marzano M. C., Mocali S., Squartini A., Zanardo M., Mengoni A.(2016).** Mixed Nodule Infection in Sinorhizobium meliloti–Medicago sativa symbiosis suggest the presence of cheating behavior. Frontiers in Plant Science. doi:10.3389/fpls.2016.00835

15. **Cho, I., & Blaser, M. J. (2012).** The human microbiome: at the interface of health and disease. Nature Reviews Genetics, 13(4), 260–270. doi:10.1038/nrg3182

16. **Cipollini, D., Rigsby, C. M., & Barto, E. K. (2012).** Microbes as Targets and Mediators of Allelopathy in Plants. Journal of Chemical Ecology, 38(6), 714–727. doi:10.1007/s10886-012-0133-7

17. **Coleman-Derr D, Desgarennes D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T, North G, Visel A, Partida-Martinez LP& Tringe S (2016)**. Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. New Phytologist 209(2), 798-811. doi: 10.1111/nph.13697

18. **Cummings, S. (2009).** The application of plant growth promoting rhizobacteria (PGPR) in low input and organic cultivation of graminaceous crops; potential and problems. Environmental Biotechnology, 5(2), 43–50.

19. **Denison R.F., Kiers E.T. (2004).** "Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis." FEMS Microbiol. Lett. 237 187–193. 10.1016/j.femsle.2004.07.013

20. **Denison, R. F. (2000).** Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *American Naturalist* 156:567–576.

21. **Diagne, N., Arumugam, K., Ngom, M., Nambiar-Veetil, M., Franche, C., Narayanan, K. K., & Laplaze, L. (2013).** Use of frankia and actinorhizal plants for degraded lands reclamation. BioMed Research International, 2013. doi:10.1155/2013/948258

22. **Ding, J., Jiang, X., Ma, M., Zhou, B., Guan, D., Zhao, B., Li L., Cao F, Li, J. (2016).** Effect of 35 years inorganic fertilizer and manure amendment on structure of bacterial and archaeal communities in black soil of northeast China. Applied Soil Ecology, 105, 187–195. doi:10.1016/j.apsoil.2016.04.010

23. **Dohrmann, A. B., Küting, M., Jünemann, S., Jaenicke, S., Schlüter, A., & Tebbe, C. C. (2013).** Importance of rare taxa for bacterial diversity in the rhizosphere of Bt- and conventional maize varieties. The ISME Journal, 7(1), 37–49. doi:10.1038/ismej.2012.77

24. **Doty, S. L., Sher, A. W., Fleck, N. D., Khorasani, M., Bumgarner, R. E., Khan, Z., DeLuca, T. H. (2016).** Variable Nitrogen Fixation in Wild Populus. Plos One, 11(5), e0155979. doi:10.1371/journal.pone.0155979

25. **Dybzinski, R., & Tilman, D. (2007).** Resource use patterns predict long-term outcomes of plant competition for nutrients and light. The American Naturalist, 170(3), 305–318. doi:10.1086/519857

26. **Edgar, R. C. (2010).** Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460–2461. doi:10.1093/bioinformatics/btq461

27. **Elton, C. (1927).** Animal Ecology. Sidgwick and Jackson, London, UK.

28. **Faoro H, Alves AC, Souza EM, Rigo LU, Cruz LM, Al-Janabi SM, et al. (2010)** Influence of soil characteristics on the diversity of bacteria in the Southern Brazilian Atlantic Forest. Appl Environ Microbiol. 2010;76(14):4744–9. Epub 2010/05/25. AEM.03025-09 [pii] doi: 10.1128/AEM.03025-09 .

29. **Fattah, Q. A. (2005)**"Plant resources for human development." Third international botanical conference.

30. **Favier, C. F., Vaughan, E. E., De Vos, W. M., & Akkermans, A. D. L. (2002).** Molecular monitoring of succession of bacterial communities in human neonates. Applied and Environmental Microbiology, 68(1), 219–226. doi:10.1128/AEM.68.1.219-226.2002

31. **Fierer, N., & Jackson, R. B. (2006).** The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America, 103(3), 626–631. doi:10.1073/pnas.0507535103

32. **Finney, J. C., Pettay, D. T., Sampayo, E. M., Warner, M. E., Oxenford, H. A., & Lajeunesse, T. C. (2010).** The relative of host-habitat, depth, significance and speciation and geography on the ecology, in the genus Symbiodinium of coral endosymbionts. Microbial Ecology, 60(1), 250–263. doi:10.1007/s00248-0

33. **Fitzsimons, M. S., & Miller, R. M. (2010).** The importance of soil

microorganisms for maintaining diverse plant communities in tallgrass prairie. American Journal of Botany, 97(12), 1937–1943. doi:10.3732/ajb.0900237

34. **Frey, S. D., Knorr, M., Parrent, J. L., & Simpson, R. T. (2004).** Chronic nitrogen enrichment affects the structure and function of the soil microbial community in temperate hardwood and pine forests. Forest Ecology and Management, 196(1), 159–171. doi:10.1016/j.foreco.2004.03.018

35. ***Friesen ML, Porter SS, Stark SC, von Wettberg EJ, Sachs JL, Martinez-Romero E (2011).*** *Microbially mediated plant functional traits. Annu Rev Ecol Evol Syst. 2011;42(1):23–46.*

36. **Garbeva, P., van Veen, J. A. a, & van Elsas, J. D. D. (2004).** Microbial diversity in soil: Selection of Microbial Populations by Plant and Soil Type and Implications for Disease Suppressiveness. Annual Review of Phytopathology, 42(1), 243–270. doi:10.1146/annurev.phyto.42.012604.135455

37. **Gause, G. F. (1934).** Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence. Science (New York, N.Y.), 79(2036), 16–17. doi:10.1126/science.79.2036.16-a

38. **Hardin, G. (1960).** The competitive exclusion principle. Science. doi:10.1126/science.131.3409.1292

39. **Hardy, W. F., & Burns, R. C. (1968).** Biological nitrogen fixation. doi: 10.1146/annurev.pp.17.060166.001103

40. **Hayat, R., Ali, S., Amara, U., Khalid, R., & Ahmed, I. (2010).** Soil beneficial bacteria and their role in plant growth promotion: A review. Annals of Microbiology, 60(4), 579–598. doi:10.1007/s13213-010-0117-1

41. **Heath, K. D., & Tiffin, P. (2007).** Context dependence in the coevolution of plant and rhizobial mutualists. Proceedings. Biological Sciences / The Royal Society, 274(1620), 1905–12. doi:10.1098/rspb.2007.0495

42. **Hirsch, P. R., & Mauchline, T. H. (2012).** Who's who in the plant root microbiome? Nature Biotechnology, 30(10), 961–962. doi:10.1038/nbt.2387

43. **Hoffmeister, D., & Keller, N. P. (2007).** Natural products of filamentous fungi: enzymes, genes, and their regulation. Natural Product Reports, 24(2), 393–416. doi:10.1039/b603084j

44. **Inceoğlu, Ö., Al-Soud, W. A., Salles, J. F., Semenov, A. V., & van Elsas, J. D. (2011).** Comparative analysis of bacterial communities in a potato field as determined by pyrosequencing. PLoS ONE, 6(8). doi:10.1371/journal.pone.0023321

45. **Karadeniz, A., Topcuoğlu, Ş. & İnan, S.(2006).** Auxin, Gibberellin, Cytokinin and Abscisic Acid Production in Some Bacteria. World J Microbiol Biotechnol 22: 1061. doi:10.1007/s11274-005-4561-1

46. **Kim, J., & Rees, D. C. (1994).** Nitrogenase and biological nitrogen fixation. Biochemistry, 33(2), 389–397. doi:10.1021/bi00168a001

47. **Kim, J.-G., Park, B. K., Kim, S.-U., Choi, D., Nahm, B. H., Moon, J. S., Hwang, I. (2006).** Bases of biocontrol: sequence predicts synthesis and mode of action of agrocin 84, the Trojan horse antibiotic that controls crown gall. Proceedings of the National Academy of Sciences of the United States of America, 103(23), 8846–8851. doi:10.1073/pnas.0602965103

48. **Klironomos, J. N. (2002).** Feedback with soil biota contributes to plant rarity and invasiveness in communities. Nature, 417(6884), 67–70. doi:10.1038/417067a

49. **Koeppel A.F., Wu, M. (2016).** Species matter: the role of competition in the assembly of congeneric bacteria. ISME J., 8 (2014), pp. 531–540.

50. **Kucey, R. M. N. (1988).** Alteration of Size of Wheat Root Systems and Nitrogen-Fixation by Associative Nitrogen-Fixing Bacteria Measured under Field Conditions. Canadian Journal of Microbiology, 34(6), 735–739.

51. **Lagier, J. C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, Raoult, D. (2012).** Microbial culturomics: Paradigm shift in the human gut microbiome study. Clinical Microbiology and Infection, 18(12), 1185–1193. doi:10.1111/1469-0691.12023

52. **Lauber, C. L., Strickland, M. S., Bradford, M. A., & Fierer, N. (2008).** The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biology and Biochemistry, 40(9), 2407–2415. doi:10.1016/j.soilbio.2008.05.021

53. **Length, F. (2013).** Antimicrobial activity of soil actinomycetes Isolated from Alkharj , KSA. International Research Journal of Microbiology, 4(1), 12–20.

54. **Lou, Y., Clay, S. A., Davis, A. S., Dille, A., Felix, J., Ramirez, A. H. M., Yannarell, A. C. (2014).** An Affinity-Effect Relationship for Microbial Communities in Plant-Soil Feedback Loops. Microbial Ecology, 67(4), 866–876. doi:10.1007/s00248-013-0349-2

55. **Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Dangl, J. L. (2012).** Defining the core Arabidopsis thaliana root microbiome. Nature, 488(7409), 86–90. doi:10.1038/nature11237

56. **Lynch, M. D. J., & Neufeld, J. D. (2015).** Ecology and exploration of the rare biosphere. Nature Reviews. Microbiology, 13(4), 217–229. doi:10.1038/nrmicro3400

57. **Marcobal, A., Kashyap, P. C., Nelson, T. A., Aronov, P. A., Donia, M. S., Spormann, A., Sonnenburg, J. L. (2013).** A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. The ISME Journal, 7(10), 1933–43. doi:10.1038/ismej.2013.89

58. **Marilley, L., Vogt, G., Blanc, M., & Aragno, M. (1998).** Bacterial diversity in the bulk soil and rhizosphere fractions of Lolium perenne and Trifolium repens as revealed by PCR restriction analysis of 16S rDNA. Plant and Soil, 198(2), 219–224. doi:10.1023/A:1004309008799

59. **McMurdie, P. J., & Holmes, S. (2014).** Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Computational Biology, 10(4). doi:10.1371/journal.pcbi.1003531

60. **Mitchell, S. D., Bernard, E., Longo, A., Garcia, A., Hutchinson, M., & Causey, R. (2016).** The use of PLFA analysis to detect differences in microbial activity in compost from horses treated with and without antibiotics. Journal of Equine Veterinary Science, 39, S5. doi:10.1016/j.jevs.2016.02.010

61. **Moora, M., & Zobel, M. (1996).** Effect of arbuscular mycorrhiza on inter- and intraspecific competition of two grassland species. Oecologia, 108(1), 79–84. doi:10.1007/bf00333217

62. **Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Green, S., & Dangl, J. L. (2013).** Diversity and heritability of the maize rhizosphere microbiome under field conditions. doi:10.1073/pnas.1302837110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1302837110

63. **Pester, M., Bittner, N., Deevong, P., & Wagner, M. (2015).** Europe PMC Funders Group Europe PMC Funders Author Manuscripts A " rare biosphere " microorganism contributes to sulfate reduction in a peatland, 4(12), 1591–1602. doi:10.1038/ismej.2010.75.A

64. **Preston, F. W. (1948)** The Commonness, And Rarity, of Species Ecological Society of America http://dx.doi.org.proxy1.cl.msu.edu/10.2307/1930989

65. **Reinhold-Hurek, B., Bünger, W., Burbano, C. S., Sabale, M., & Hurek, T. (2015).** Roots Shaping Their Microbiome: Global Hotspots for Microbial Activity. Annual Review of Phytopathology, 53(1), 403–424. doi:10.1146/annurev-phyto-082712-102342

66. **Rettedal, E. A., Gumpert, H., & Sommer, M. O. A. (2014).** Cultivation-based

multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. Nature Communications, 5, 4714. doi:10.1038/ncomms5714

67. **Rodrigo Mendes, Marco Kruijt, Irene de Bruijn, Ester Dekkers, Menno van der Voort, Johannes H. M. Schneider, Yvette M. Piceno, Todd Z. DeSantis, Gary L. Andersen, Peter A. H. M. Bakker, Jos M. Raaijmakers. (2011).** Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. Science, 332(December 2011), 1097–1100. doi:10.1126/science.1203980

68. **Rolli, E., Marasco, R., Vigani, G., Ettoumi, B., Mapelli, F., Deangelis, M. L., Gandolfi, C., Casati, E., Previtali, F., Gerbino, R., Pierotti Cei, F., Borin, S., Sorlini, C., Zocchi, G. and Daffonchio, D. (2015).** Improved plant resistance to drought is promoted by the root-associated microbiome as a water stress-dependent trait. Environmental Microbiology, 17(2), 316–331. doi:10.1111/1462-2920.12439

69. **Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Fierer, N. (2010).** Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME Journal, 4(10), 1340–1351. doi:10.1038/ismej.2010.58

70. **Saikkonen, K., Faeth, S., Helander, M., & Sullivan, T. (1998).** Fungal Endophytes: A Continuum of Interactions with Host Plants. Annual Review of Ecology and Systematics, 29, 319-343. Retrieved from http://www.jstor.org/stable/221711

71. **Schloss, P. D.,*et al.* (2009).** Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology, 75(23), 7537–7541. doi:10.1128/AEM.01541-09

72. **Seringe, N. C. (1825).** "Trifolium." de Candolle AP, Prodromus Systematis Naturalis Regni Vegetabilis

73. **Shendure, J., & Ji, H. (2008).** Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135–1145. doi:10.1038/nbt1486

74. **Society, T. A., & Press, C. (2016).** Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? The American Naturalist, Vol. 93, No. 870 (May - Jun., 1959), pp. 145-159.

75. **Sørensen J., Sessitsch A.(2006).** Plant-associated bacteria lifestyle and molecular interactions J.D. van Elsas (Ed.), *et al*., Modern Soil Microbiology (2nd edn), CRC Press , pp. 211–236

76. **Sprent, J. I. (2001**). Nodulation in legumes. Kew, UK: Royal Botanic Gardens.

77. **Stachowicz J.J. (2001).** Mutualism, facilitation and structure of ecological communities. BioScience (2001) 51 (3): 235-246. doi: 10.1641/0006-3568(2001)051[0235:MFATSO]2.0.CO;2

78. **Steenhoudt, O., & Vandereyden, J. (2000).** Azospirillum, fee-living nitrogen fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. FEMS Microbiology Reviews, 24(4), 487–506. doi:10.1111/j.1574-6976.2000.tb00552.x

79. **Tilman D.(2016).** Constraints and Tradeoffs : Toward a Predictive Theory of Competition and Succession. (2016), 58(1), 3–15.

80. **Tilman, D. (1982).** Resource competition and community structure. Monographs in Population Biology, Princeton University Press, Princeton, New Jersey, USA

81. **Thompson J.N. (2005).** The Geographic Mosaic of Coevolution. University of Chicago Press: Chicago, IL, USA.

82. **Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-liggett, C., Knight, R., & Gordon, J. I. (2007).** The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature, 449(7164), 804–810. doi:10.1038/nature06244.The

83. **Turner, T. R., James, E. K., & Poole, P. S. (2013).** The plant microbiome. Genome Biology, 14(6), 209. doi:10.1186/gb-2013-14-6-209

84. **Turner, T. R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Poole, P. S. (2013).** Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants, 7(12), 2248–2258. doi:10.1038/ismej.2013.119

85. **Union, A. O., & Auk, T. (2016).** The Niche-Relationships of the California. American Ornithologists 34(4), 427–433.

86. **Upadhyay, S. K., Singh, D. P., & Saikia, R. (2009**). Genetic Diversity of Plant Growth Promoting Rhizobacteria Isolated from Rhizospheric Soil of Wheat under Saline Condition. Current Microbiology, 59(5), 489–496. doi:10.1007/s00284-009-9464-1

87. **Uroz, S., Buée, M., Murat, C., Frey-Klett, P., & Martin, F. (2010).** Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environmental Microbiology Reports, 2(2), 281–288. doi:10.1111/j.1758-2229.2009.00117.x

88. **Van der Putten, W. H., Klironomos, J. N., & Wardle, D. A. (2007).** Microbial ecology of biological invasions. The ISME Journal, 1(1), 28–37. doi:10.1038/ismej.2007.9

89. **Vance, C. P. (2016).** Update on the State of Nitrogen and Phosphorus Nutrition Symbiotic Nitrogen Fixation and Phosphorus Acquisition . Plant Nutrition in a World of Declining Renewable Resources, 127(October), 390–397. doi:10.1104/pp.010331.390

90. **Weinert, N., Piceno, Y., Ding, G. C., Meincke, R., Heuer, H., Berg, G.,Smalla, K. (2011).** PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: Many common and few cultivar-dependent taxa. FEMS Microbiology Ecology, 75(3), 497–506. doi:10.1111/j.1574-6941.2010.01025.x

91. **White, D. C., *et al* (1997).** "Chemical and molecular approaches for rapid assessment of the biological status of soils." Biological indicators of soil health. (1997): 371-396.

92. **Wieland, G., Neumann, R., & Backhaus, H. (2001).** Variation of Microbial Communities in Soil , Rhizosphere , and Rhizoplane in Response to Crop Species , Soil Type , and Crop Development Variation of Microbial Communities in Soil , Rhizosphere , and Rhizoplane in Response to Crop Species , Soil Type , an. Applied and Environmental Microbiolgy, 67(12), 5849–5854. doi:10.1128/AEM.67.12.5849

93. **Zelles, L. (1999).** Fatty acid patterns of phospholipids and lipopolysaccharides in the characterisation of microbial communities in soil: A review. Biology and Fertility of Soils, 29(2), 111–129. doi:10.1007/s003740050533