

Group Project Report - Up_All_Night-Rs

Tianyi Zuo, Roshael Chellappah, Tram Duong, Malcolm Stewart

April 4th, 2022

Introduction

Soil microbial communities can play a crucial role in building plant communities. Also, individual plants frequently cause changes in soil microbial community composition. This reciprocal change then affects the growth of neighbouring or individual plants that subsequently colonize the soil. As this effect increases, it can cause interspecific differences and thus affect plant community composition. In addition, the intensity of this interaction varies over time, so we focus on how the intensity of the interaction between plant and microbial populations varies over time.

In our data, the experimenters mainly collected changes in the soil microbial community associated with individuals of four plant species for more than 20 years. And the effect of soil biota on the performance of conspecific and heterospecific seedlings was quantified in a greenhouse experiment that preserved the soil characteristics of these monocultures. We used this dataset to investigate the associated effects of individual plants and soil microbes and whether they change in intensity over time. We propose relevant questions about this topic:

Question 1. What effect do plant species have on the bacterial community composition?

- a. Does the plant species affect the bacterial community composition?
- b. Does the plant age affect the bacterial community?
- c. How do the bacterial communities' composition change across time as the various plant species age?

Question 2. How does the Bacterial community affect the plant species?

- a. Do bacteria communities have an effect on plant root mass (below-ground dry mass)?
- b. Which bacteria has the greatest effect on plant biomass when the plant is 4-12 years old?

Package loading

Input the data

```
MBData<-read.csv("./Data/BacteriaCommunityMatrix.csv")
SBData<-read.csv("./Data/BacteriaSampleData.csv")
GData<-read.csv("./Data/GreenhouseData.csv")
```

Question 1

What effect do plant species have on the bacterial community composition?

Question 1a)

Does the plant species affect the bacterial community composition?

In order to test the similarity between the bacterial community in different plant species areas, we will use a Non-Metric Multidimensional Scaling (NMDS) to visualize differences among our samples within different plant species. For this question, we mainly used the Bacteria Community Matrix to compare the composition of bacterial communities among different plant species. The matrix data collected soil microbial communities associated with individuals of the four dominant plants species, which is a community matrix of bacterial OTU characterized from each soil sample.

Transform the Data

According the data we have, we transformed the column and row first, and removed errors which resulted in the sum of the row and column to be zero.

```
rownames(MBData) <- MBData[,1]
MBData<- MBData[,-1 ]
Matrix<-t(MBData)
Matrix <- Matrix[!(rowSums(Matrix) == 0), !(colSums(Matrix) == 0)]
```

Distance Matrix

```
#Create the Bray_Curtis Dissimilarity metric
Dist<- vegdist(Matrix,method="bray",binary=F)
DistMat<-as.matrix(Dist)
PDat<-melt(DistMat)
```

Non-Metric Multidimensional Scaling

Using Non-metric Dimensional Scaling(NMDS), we can visualize the cluster of samplings from different plant species. We can find out whether different plant species area have a stronger effect on bacterial communities.

```
set.seed(12138)
NMDSdat<-metaMDS(Dist,k=2 ,trymax = 200)
```

```

## Run 0 stress 0.1524127
## Run 1 stress 0.1524131
## ... Procrustes: rmse 0.0001786159 max resid 0.001660839
## ... Similar to previous best
## Run 2 stress 0.1524127
## ... Procrustes: rmse 1.805165e-05 max resid 0.0001266726
## ... Similar to previous best
## Run 3 stress 0.1538359
## Run 4 stress 0.1543444
## Run 5 stress 0.1538358
## Run 6 stress 0.1524127
## ... Procrustes: rmse 2.130521e-05 max resid 0.0002153781
## ... Similar to previous best
## Run 7 stress 0.1524127
## ... Procrustes: rmse 1.208819e-05 max resid 7.751305e-05
## ... Similar to previous best
## Run 8 stress 0.1543443
## Run 9 stress 0.1538359
## Run 10 stress 0.1524127
## ... Procrustes: rmse 1.463694e-05 max resid 9.318012e-05
## ... Similar to previous best
## Run 11 stress 0.1538359
## Run 12 stress 0.1524127
## ... Procrustes: rmse 7.263723e-06 max resid 4.023348e-05
## ... Similar to previous best
## Run 13 stress 0.1524127
## ... Procrustes: rmse 6.87498e-06 max resid 4.722308e-05
## ... Similar to previous best
## Run 14 stress 0.1524127
## ... Procrustes: rmse 1.684329e-05 max resid 0.0001480371
## ... Similar to previous best
## Run 15 stress 0.1549423
## Run 16 stress 0.1538359
## Run 17 stress 0.1543444
## Run 18 stress 0.1549423
## Run 19 stress 0.1524127
## ... Procrustes: rmse 1.388171e-05 max resid 0.0001094046
## ... Similar to previous best
## Run 20 stress 0.1524127
## ... New best solution
## ... Procrustes: rmse 7.132119e-06 max resid 5.354022e-05
## ... Similar to previous best
## *** Solution reached

```

```

PDat<-data.frame(NMDS1=NMDSdat$points[,1],
                   NMDS2=NMDSdat$points[,2],
                   X=row.names(Matrix))
PDat<-merge(PDat,SBData,by="X",all.x=T,all.y=F)
qplot(x=NMDS1,y=NMDS2,colour=Species,alpha=I(0.6),data=PDat)+theme_bw()

```

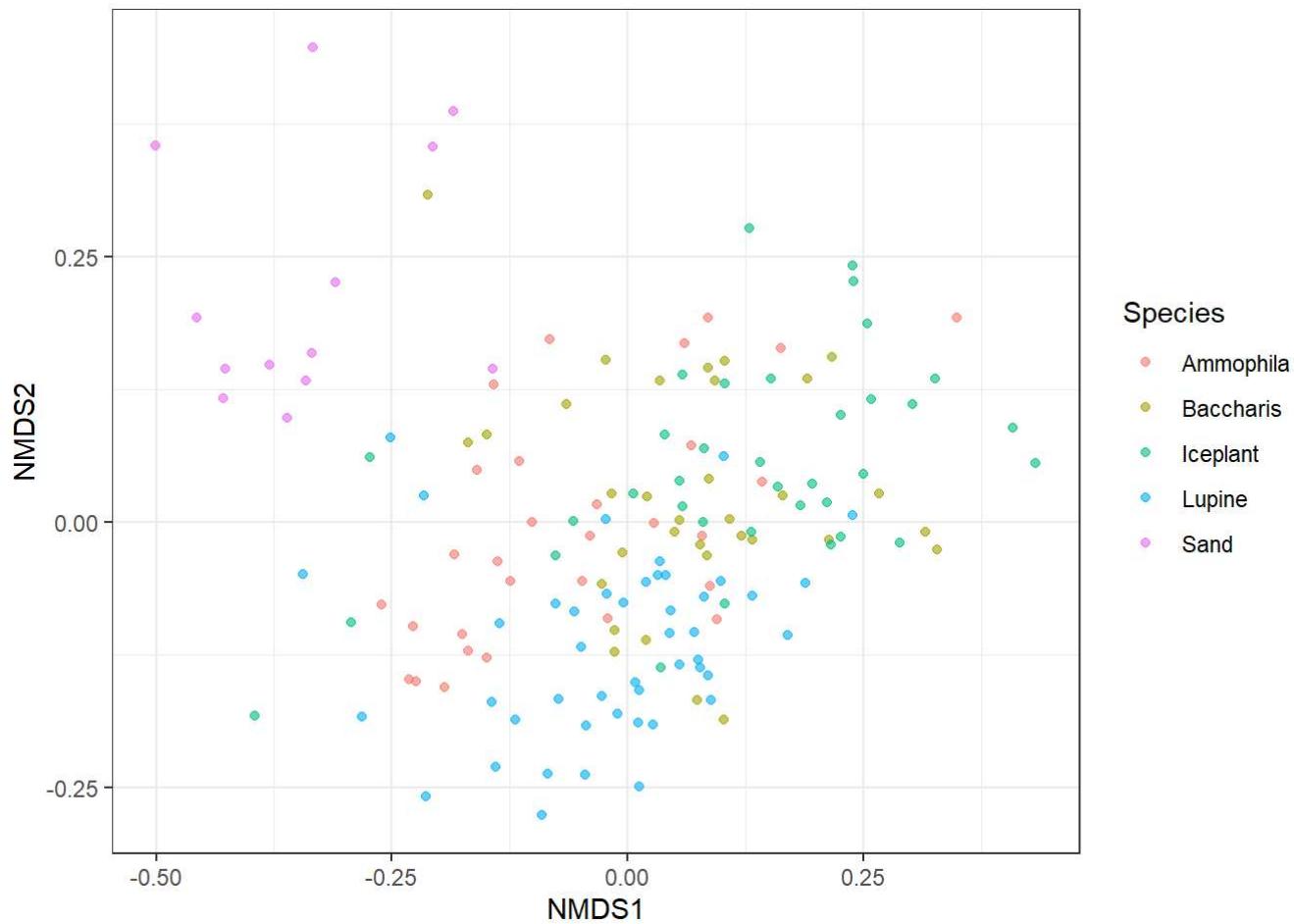


Figure 1. NMDS plot show the overall variation in bacterial community composition among the different specific plant species. Points are color-coded based on the identity of the plant species: Ammophila (red); Baccharis(green); IcePlant (orange); Lupine (blue); and Sand (pink).

Discussion

According to Figure 1, the bacterial community composition differed between plant species, but the results were not significant. Based on the graph, the control group(sand) has less similarity with the other four plant species. Samplings from same plant species area are more likely to cluster together, but there is still some overlap between differing plant species. From Figure 1, we can conclude that the bacterial community composition differs between plant species.

Question 1b)

Does the plant age affect the bacterial community composition?

We next wanted to investigate in more detail whether there is a temporal pattern in bacterial communities based on our available data and whether the composition of their communities changes depending on the time of plant growth.

Transform Matrix

In order to test whether plant age will effect the bacterial community, we separated the bacterial community matrix into four different matrices, filtering by plant species.

```
MatrixA <- Matrix[grep("Ammophila", row.names(Matrix)), ]
MatrixB <- Matrix[grep("Baccharis", row.names(Matrix)), ]
MatrixI <- Matrix[grep("Iceplant", row.names(Matrix)), ]
MatrixL <- Matrix[grep("Lupine", row.names(Matrix)), ]
```

Distance Matrix

```
#Create the Bray_Curtis Dissimilarity metric for each plant species
DistA<- vegdist(MatrixA,method="bray",binary=F)
DistMatA<-as.matrix(DistA)
PDatA<-melt(DistMatA)

DistB<- vegdist(MatrixB,method="bray",binary=F)
DistMatB<-as.matrix(DistB)
PDatB<-melt(DistMatB)

DistI<- vegdist(MatrixI,method="bray",binary=F)
DistMatI<-as.matrix(DistI)
PDatI<-melt(DistMatI)

DistL<- vegdist(MatrixL,method="bray",binary=F)
DistMatL<-as.matrix(DistL)
PDatL<-melt(DistMatL)
```

Non-Metric Multidimensional Scaling

Using Non-metric Dimensional Scaling(NMDS), we can visualize the cluster of samplings with plant age in four different plant species. We can find out whether the plant growth age has a stronger effect on bacterial communities.

```
#Non-Metric Multidimensional Scaling for Ammophila species.
set.seed(12138)
NMDSdatA<-metaMDS(DistA,k=2 ,trymax = 200)
```

```

## Run 0 stress 0.1027607
## Run 1 stress 0.1027607
## ... New best solution
## ... Procrustes: rmse 2.98912e-06 max resid 1.202099e-05
## ... Similar to previous best
## Run 2 stress 0.1409243
## Run 3 stress 0.1399827
## Run 4 stress 0.1027607
## ... Procrustes: rmse 3.063112e-05 max resid 9.637482e-05
## ... Similar to previous best
## Run 5 stress 0.1429074
## Run 6 stress 0.1039323
## Run 7 stress 0.1027607
## ... Procrustes: rmse 1.841972e-05 max resid 5.480584e-05
## ... Similar to previous best
## Run 8 stress 0.1027607
## ... Procrustes: rmse 3.647275e-05 max resid 0.0001322759
## ... Similar to previous best
## Run 9 stress 0.1039323
## Run 10 stress 0.1426368
## Run 11 stress 0.1027607
## ... Procrustes: rmse 1.354018e-05 max resid 5.041662e-05
## ... Similar to previous best
## Run 12 stress 0.1034352
## Run 13 stress 0.1409243
## Run 14 stress 0.1426369
## Run 15 stress 0.1034352
## Run 16 stress 0.1038899
## Run 17 stress 0.1027607
## ... Procrustes: rmse 5.67144e-06 max resid 1.713808e-05
## ... Similar to previous best
## Run 18 stress 0.1027607
## ... Procrustes: rmse 2.246673e-05 max resid 7.091051e-05
## ... Similar to previous best
## Run 19 stress 0.1027607
## ... Procrustes: rmse 1.210752e-05 max resid 3.99117e-05
## ... Similar to previous best
## Run 20 stress 0.1027607
## ... Procrustes: rmse 8.40356e-06 max resid 2.696386e-05
## ... Similar to previous best
## *** Solution reached

```

```

PData<-data.frame(NMDS1=NMDSDatA$points[,1],
                    NMDS2=NMDSDatA$points[,2],
                    X=row.names(MatrixA))
PData<-merge(PData,SBData,by="X",all.x=T,all.y=F)
NMDSA<-qplot(x=NMDS1,y=NMDS2,colour=Age,alpha=I(0.6),data=PData) + theme_bw() + ggtitle("Ammophila")

```

```
#Non-Metric Multidimensional Scaling for Baccharis species.
```

```
set.seed(12138)
```

```
NMDSdatB<-metaMDS(DistB,k=2 ,trymax = 200)
```

```
## Run 0 stress 0.1663259
## Run 1 stress 0.2047861
## Run 2 stress 0.182558
## Run 3 stress 0.1915501
## Run 4 stress 0.1879523
## Run 5 stress 0.1825581
## Run 6 stress 0.1663259
## ... New best solution
## ... Procrustes: rmse 2.33032e-05 max resid 0.0001170446
## ... Similar to previous best
## Run 7 stress 0.1914446
## Run 8 stress 0.1663259
## ... Procrustes: rmse 4.166328e-05 max resid 0.000158607
## ... Similar to previous best
## Run 9 stress 0.1953224
## Run 10 stress 0.2015502
## Run 11 stress 0.1914446
## Run 12 stress 0.1663259
## ... Procrustes: rmse 2.501495e-05 max resid 0.0001061822
## ... Similar to previous best
## Run 13 stress 0.1663789
## ... Procrustes: rmse 0.008680836 max resid 0.04391118
## Run 14 stress 0.1663259
## ... Procrustes: rmse 1.838879e-05 max resid 7.281169e-05
## ... Similar to previous best
## Run 15 stress 0.1831176
## Run 16 stress 0.2066152
## Run 17 stress 0.1879523
## Run 18 stress 0.2004666
## Run 19 stress 0.1988136
## Run 20 stress 0.1828698
## *** Solution reached
```

```
PDatB<-data.frame(NMDS1=NMDSdatB$points[,1],
```

```
                    NMDS2=NMDSdatB$points[,2],
```

```
                    X=row.names(MatrixB))
```

```
PDatB<-merge(PDatB,SBData,by="X",all.x=T,all.y=F)
```

```
NMDSB<-qplot(x=NMDS1,y=NMDS2,colour=Age,alpha=I(0.6),data=PDatB)+theme_bw() + ggtitle("Baccharis")
)
```

```
#Non-Metric Multidimensional Scaling for Iceplant species.
```

```
set.seed(12138)
```

```
NMDSdatI<-metaMDS(DistI,k=2 ,trymax = 200)
```

```

## Run 0 stress 0.1240002
## Run 1 stress 0.1236591
## ... New best solution
## ... Procrustes: rmse 0.02090319 max resid 0.1103319
## Run 2 stress 0.1237757
## ... Procrustes: rmse 0.007462105 max resid 0.03557219
## Run 3 stress 0.1237757
## ... Procrustes: rmse 0.007461848 max resid 0.03557372
## Run 4 stress 0.1595197
## Run 5 stress 0.1743154
## Run 6 stress 0.1240002
## ... Procrustes: rmse 0.02138832 max resid 0.1123285
## Run 7 stress 0.1595197
## Run 8 stress 0.1649935
## Run 9 stress 0.1236591
## ... New best solution
## ... Procrustes: rmse 3.963904e-05 max resid 0.0001625783
## ... Similar to previous best
## Run 10 stress 0.1237757
## ... Procrustes: rmse 0.007464133 max resid 0.03550868
## Run 11 stress 0.1749454
## Run 12 stress 0.124
## ... Procrustes: rmse 0.02129476 max resid 0.111973
## Run 13 stress 0.1711729
## Run 14 stress 0.1237757
## ... Procrustes: rmse 0.007462836 max resid 0.03556195
## Run 15 stress 0.1402901
## Run 16 stress 0.1239998
## ... Procrustes: rmse 0.02115019 max resid 0.1114109
## Run 17 stress 0.1245378
## Run 18 stress 0.1699085
## Run 19 stress 0.178428
## Run 20 stress 0.1721498
## *** Solution reached

```

```

PDatI<-data.frame(NMDS1=NMDSDatI$points[,1],
                    NMDS2=NMDSDatI$points[,2],
                    X=row.names(MatrixI))
PDatI<-merge(PDatI,SBData,by="X",all.x=T,all.y=F)
NMDSI<-qplot(x=NMDS1,y=NMDS2,colour=Age,alpha=I(0.6),data=PDatI)+theme_bw() + ggtitle("Iceplant")
)
```

```

#Non-Metric Multidimensional Scaling for Lupine species.
set.seed(12138)
NMDSDatL<-metaMDS(DistL,k=2 ,trymax = 200)

```

```

## Run 0 stress 0.1434957
## Run 1 stress 0.1510527
## Run 2 stress 0.1434957
## ... Procrustes: rmse 6.191069e-05 max resid 0.0003274777
## ... Similar to previous best
## Run 3 stress 0.1510527
## Run 4 stress 0.1434706
## ... New best solution
## ... Procrustes: rmse 0.004583149 max resid 0.02566037
## Run 5 stress 0.1510527
## Run 6 stress 0.1434925
## ... Procrustes: rmse 0.004946751 max resid 0.02597845
## Run 7 stress 0.1510527
## Run 8 stress 0.1560766
## Run 9 stress 0.1434925
## ... Procrustes: rmse 0.004946838 max resid 0.02597613
## Run 10 stress 0.1434705
## ... New best solution
## ... Procrustes: rmse 3.645327e-05 max resid 0.0001710421
## ... Similar to previous best
## Run 11 stress 0.1516846
## Run 12 stress 0.1559927
## Run 13 stress 0.1434707
## ... Procrustes: rmse 0.0001238281 max resid 0.0006621348
## ... Similar to previous best
## Run 14 stress 0.1510527
## Run 15 stress 0.1778942
## Run 16 stress 0.1881897
## Run 17 stress 0.1510527
## Run 18 stress 0.1510527
## Run 19 stress 0.1782528
## Run 20 stress 0.1559928
## *** Solution reached

```

```

PDatL<-data.frame(NMDS1=NMDSDatL$points[,1],
                     NMDS2=NMDSDatL$points[,2],
                     X=row.names(MatrixL))
PDatL<-merge(PDatL,SBData,by="X",all.x=T,all.y=F)
NMDSL<-qplot(x=NMDS1,y=NMDS2,colour=Age,alpha=I(0.6),data=PDatL)+
  theme_bw()+
  ggtitle("Lupine")

```

```

#Plotting the NMDS result
plot_grid(NMDSA, NMDSB, NMDSI,NMDSL, ncol = 2)

```

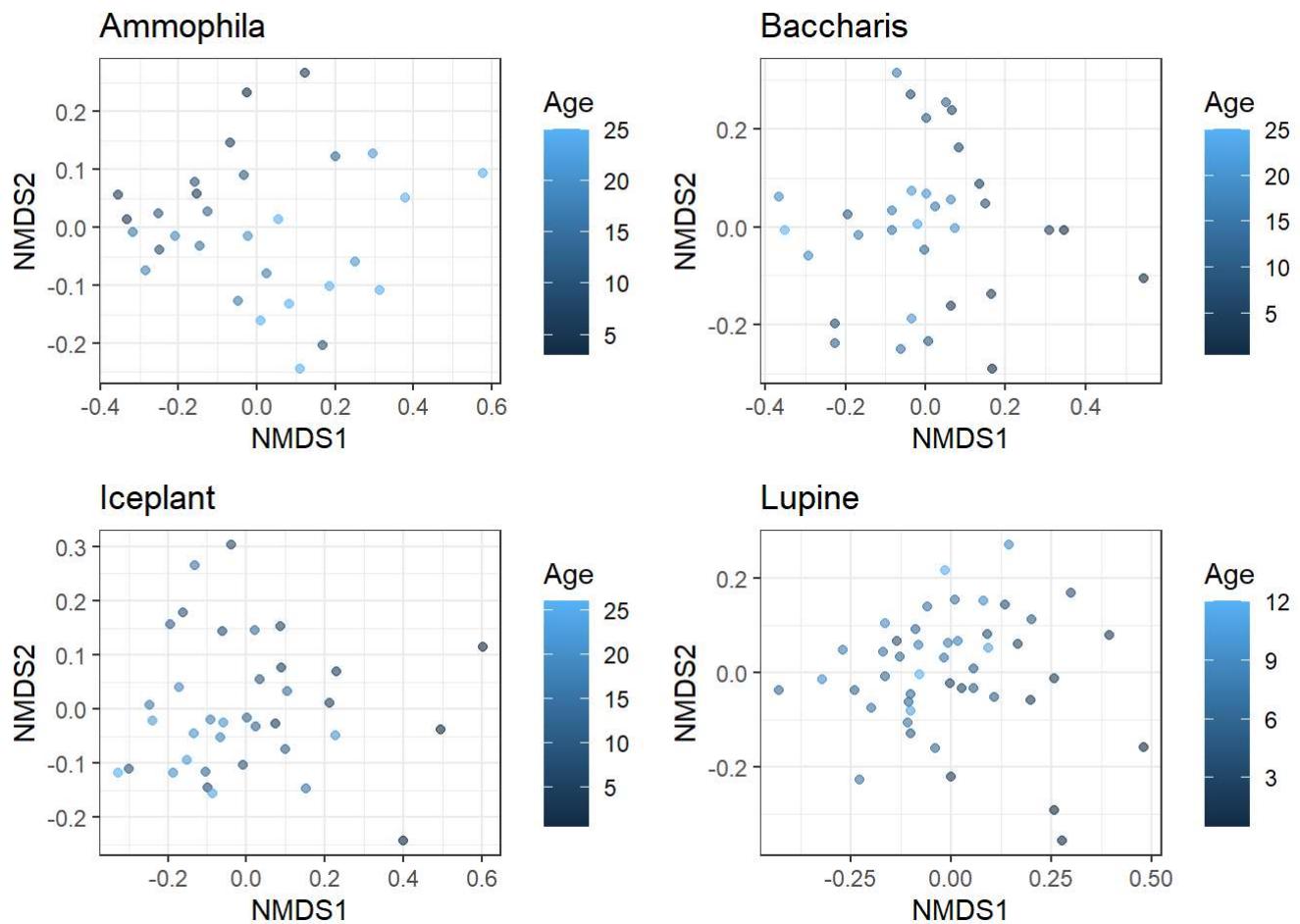


Figure 2. The NMDS ordination plot shows the microbial community composition as a factor of plant species and the age of plant individuals. Bacterial communities associated with the specific species are color-coded by individual plant age; the light to dark colour of data points represent the age gradient from old to young, with species-specific minimum and maximum age.

Discussion

According to Figure 2, in each plant, the bacterial composition changes with the age of the plant. Among Ammophila species, the largest changes in bacterial community composition were observed. Bacterial abundance also changes with increasing plant age in the other three species but this change is not significant. Overall, we can find that the plant age is an important factor which has some effect on changes within the bacterial community, but we still need further research.

Question 1c)

How do the bacterial communities' composition change across time as the various plant species age?

Transform Data

Subset Data - remove unwanted columns

```
MDSsubset <- GData %>% select(c("PlantSpecies", "SoilHostSpecies", "Age")) # creates new dataset
from the GreenhouseData.csv with only these 4 columns
```

Data Preparation

The purpose of this section is to identify whether the bacterial soil host species composition is in any way correlated with or impacted by the age of the examined plant species. From answering this question and if some significant correlation is identified, we can further look into why bacterial soil host species compositions may differ across time.

First, we will filter the data by plant species:

```
AaSub <- MDSubset %>% filter(PlantSpecies == "Aa") # creates subset of MDSubset with only Aa plant data  
BpSub <- MDSubset %>% filter(PlantSpecies == "Bp") # does the same as above with Bp plant data  
CeSub <- MDSubset %>% filter(PlantSpecies == "Ce")  
LaSub <- MDSubset %>% filter(PlantSpecies == "La")
```

Since both age and bacterial soil host species are categorical variables, it is difficult to statistically identify whether the relationship between these two variables is significant or not. Therefore, we believe that the best way to identify any difference in bacterial soil host species composition across time would be to visualize the data. Therefore, we will create a stacked bar graphs for each plant species, identifying the approximate bacterial compositions within plant species at every age that samples were collected.

```

# Graph for Aa
AaPlot <- ggplot(data = AaSub, aes(x = Age, fill = SoilHostSpecies)) +
  geom_bar() +
  scale_x_continuous(breaks = seq(0, 28, by = 2)) +
  scale_y_continuous(breaks = seq(0, 20, by = 2)) +
  ggtitle("Ammophila arenaria") +
  xlab("Age (years)") + ylab("Number of Samples") +
  guides(fill=guide_legend(title="Soil Host Species")) +
  theme_bw()

# Graph for Bp
BpPlot <- ggplot(data = BpSub, aes(x = Age, fill = SoilHostSpecies)) +
  geom_bar() +
  scale_x_continuous(breaks = seq(0, 28, by = 2)) +
  scale_y_continuous(breaks = seq(0, 20, by = 2)) +
  ggtitle("Baccharis pilularis") +
  xlab("Age (years)") + ylab("Number of Samples") +
  guides(fill=guide_legend(title="Soil Host Species")) +
  theme_bw()

# Graph for Ce
CePlot <- ggplot(data = CeSub, aes(x = Age, fill = SoilHostSpecies)) +
  geom_bar() +
  scale_x_continuous(breaks = seq(0, 28, by = 2)) +
  scale_y_continuous(breaks = seq(0, 20, by = 2)) +
  ggtitle("Carpobrotus edulis") +
  xlab("Age (years)") + ylab("Number of Samples") +
  guides(fill=guide_legend(title="Soil Host Species")) +
  theme_bw()

# Graph for La
LaPlot <- ggplot(data = LaSub, aes(x = Age, fill = SoilHostSpecies)) +
  geom_bar() +
  scale_x_continuous(breaks = seq(0, 28, by = 2)) +
  scale_y_continuous(breaks = seq(0, 20, by = 2)) +
  ggtitle("Lupine arboreus") +
  xlab("Age (years)") + ylab("Number of Samples") +
  guides(fill=guide_legend(title="Soil Host Species")) +
  theme_bw()

# Arrange all these graphs together on one page
Plot <- ggarrange(AaPlot, BpPlot + rremove("ylab"), CePlot, LaPlot + rremove("ylab"), # helps id
  entify which plots to use within the multiplot figure, and rremove("ylab") removes the y-axis La
  bel for 2 of the 4 graphs
  labels = c("A", "B", "C", "D"), # sets labels for every plot
  ncol = 2, nrow = 2, # sets the dimensions for the multiplot
  common.legend = TRUE, legend = "right") # sets a common plot and places it on the left
  of the page

annotate_figure(Plot,
  top = text_grob("Bacterial Soil Host Species Compositions Per Plant Species", co
  lor = "black", face = "bold", size = 14)) # creates a common title for the multiplot

```

Bacterial Soil Host Species Compositions Per Plant Species

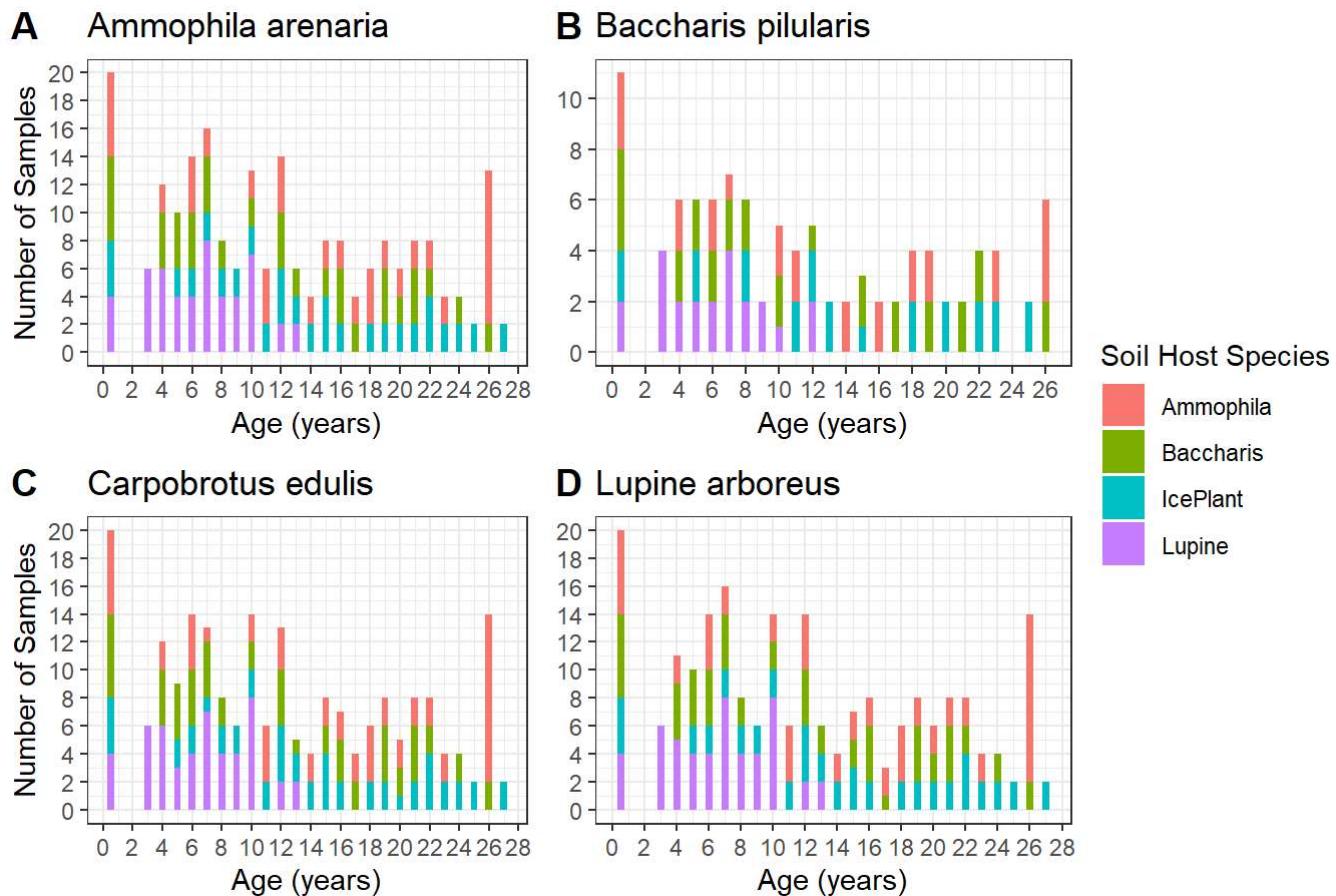


Figure 3. Multi-plot figure of stacked bar graphs of bacterial soil host species for every plant species sampled. Y-axis identifies the numbers of samples collected at each age group and bacterial composition can be identified by the various colours within every bar (each colour represents a different bacterial species, as outlined in the legend). Greater amount of one particular colour within any given bar indicates a greater abundance of that bacteria at that specific age.

Discussion

Unlike what we had predicted, it seems that the studied plants and their related bacterial species have no great correlation when examining bacterial compositions. There are however some trends present within these graphs. One conclusion that can be made from these stacked bar graphs is that the bacterial species Lupine disappears from the soil environments of all 4 plant species after 13 years of age. Although there were less total *Baccharis pilularis* samples taken in the Greenhouse experiment, all plant species began with a similar composition of all studied bacterial soil host species. Furthermore, it also seems that *Ammophila arenaria* is always in greatest abundance when the plants are 26 years old. All plant species also seem to follow similar trends in how the bacterial community composition fluctuates over time. Further research could identify why Lupine suddenly disappears from the soil microbiome after 13 years of age, and if there is some other factor that results in this disappearance. It would also be useful to identify whether the relationship between age and bacterial soil host species is significant; however, since both of these variables are categorical and since the data is not normally distributed, it is difficult to identify a particular statistical test that could help us identify this significance.

Question 2a)

Do bacterial species have an effect on plant root mass (below-ground dry mass)?

In order to account for different root masses among species of plants, the relationship between soil host species and root mass will be looked at per plant species. In order to account for difference in root masses among age, an older age group and younger sampled age group will be chosen among all plant species (10 years and 0.5 years, respectively) for visual analysis. In order to account for differences in root mass by age and for statistical confirmation, the ANCOVA test will be used.

Transform the Data

```
#Subset the data to get only the columns of interest  
root <- GData[c("PlantSpecies", "SoilHostSpecies", "BG_Dry_mg", "Age")]
```

```
#Age needs to be changed into a categorical variable  
class(root$Age)
```

```
## [1] "numeric"
```

```
root$Age <- as.factor(root$Age)  
class(root$Age)
```

```
## [1] "factor"
```

Visualizing the data:

```
#Subset data again to get only the ages of interest  
oldplants <- root[grep('10', root$Age),]  
youngplants <- root[grep('0.5', root$Age),]
```

```
oldplot <- ggplot(oldplants, aes(x=PlantSpecies, y=BG_Dry_mg, fill=SoilHostSpecies)) +  
  geom_boxplot() +  
  facet_wrap(~PlantSpecies, scale="free") +  
  labs(x= "Plant species (15 years)", y = "Below-ground dry mass (mg)", fill= "Soil host species") +  
  theme_classic()  
oldplot
```

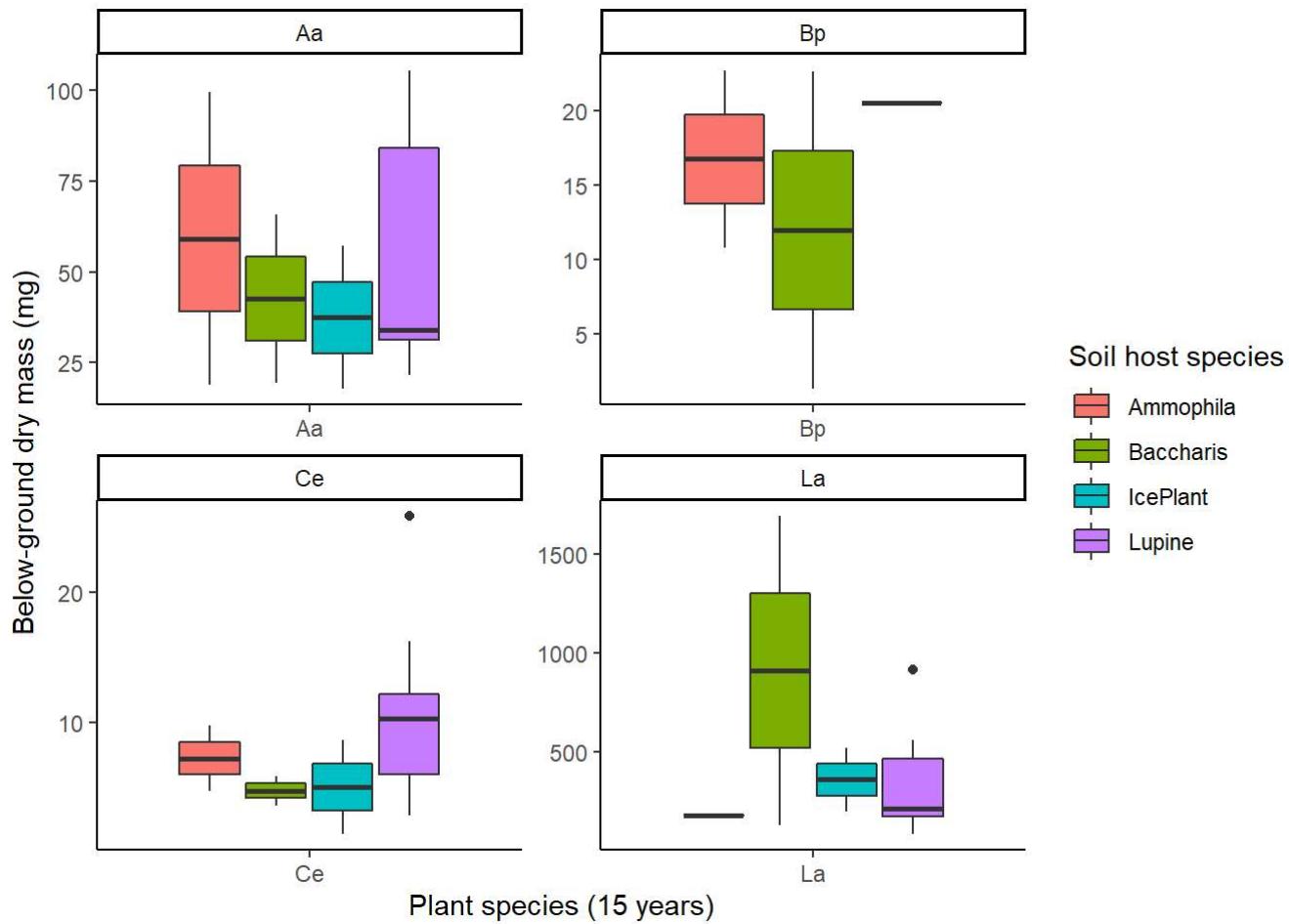


Figure 4. Box plot of the relationship between soil host species and below-ground dry mass in milligrams of each plant species at age 10 years (*A. arenaria*, *B. pilularis*, *C. edulis*, *L. arboreus*, respectively). Bp at age 15 was not treated with Lupine for an unknown reason (possibly due to lack on samples).

```
youngplot <- ggplot(youngplants, aes(x=PlantSpecies, y=BG_Dry_mg, fill=SoilHostSpecies)) +
  geom_boxplot() +
  facet_wrap(~PlantSpecies, scale="free") +
  labs(x= "Plant species (0.5 years)", y = "Below-ground dry mass (mg)", fill= "Soil host species") +
  theme_classic()
youngplot
```

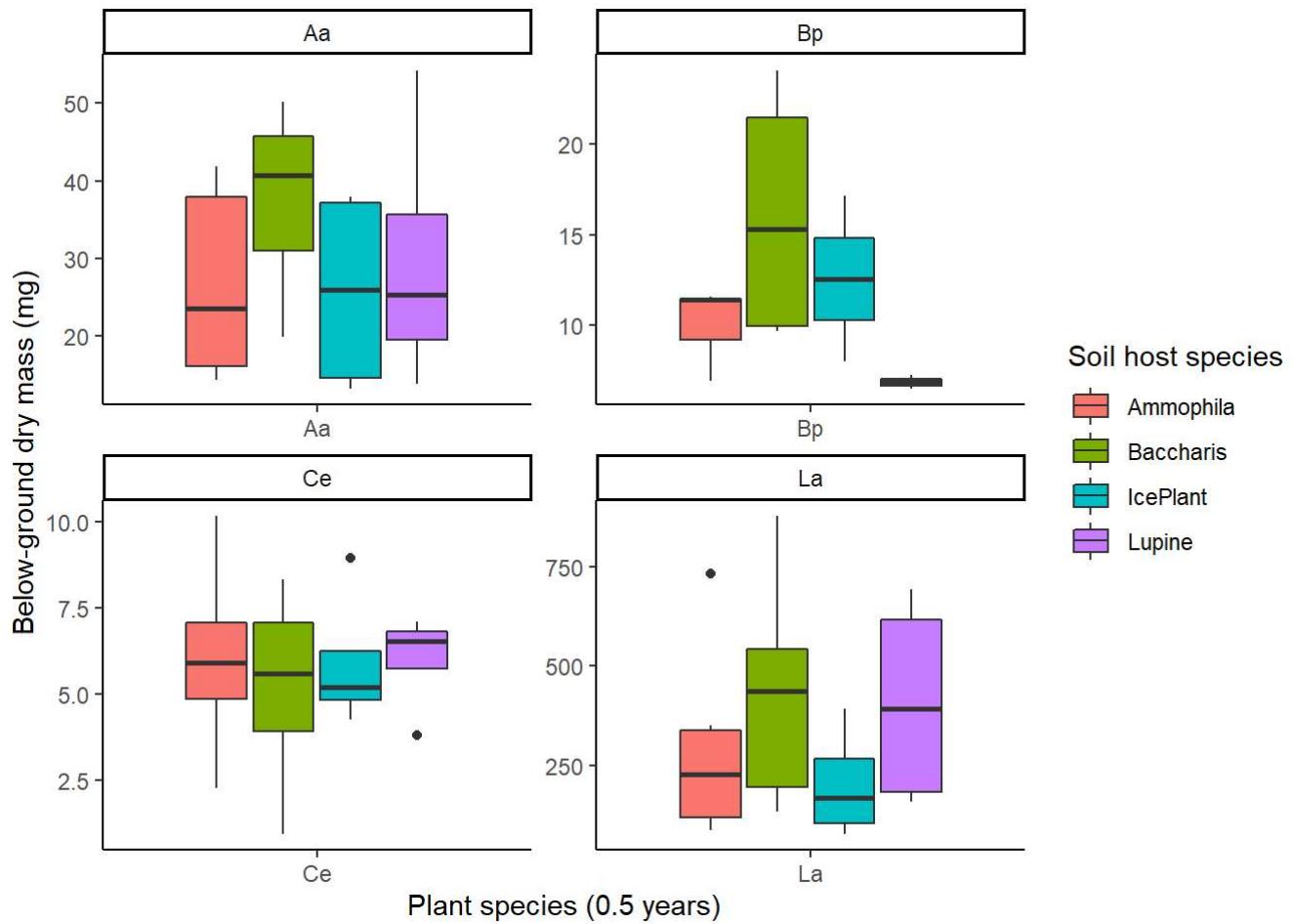


Figure 5. Box plot of the relationship between soil host species and below-ground dry mass in milligrams of each plant species at age 0.5 years (*A. arenaria*, *B. pilularis*, *C. edulis*, *L. arboreus*, respectively).

Statistical analysis using ANCOVA:

```
#Checking data to meet assumptions
```

```
#Assumption 1: the covariate and treatment should be independent of each other.
```

```
root$Age <- as.numeric(root$Age)
rootmodel <- aov(Age ~ SoilHostSpecies, data = root)
summary(rootmodel)
```

```
##                               Df Sum Sq Mean Sq F value Pr(>F)
## SoilHostSpecies      3   8075    2692   58.57 <2e-16 ***
## Residuals            738  33914     46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Assumption 2: There is homogeneity of variance.
```

```
leveneTest(BG_Dry_mg~SoilHostSpecies, data = root)
```

```

## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group     3  0.2262 0.8782
##          738

```

The covariance model ANCOVA will be used to statistically analyse for significance in soil host species.

```

#Take age into consideration
ancova_model <- aov(BG_Dry_mg ~ SoilHostSpecies + PlantSpecies + Age, data = root)
Anova(ancova_model, type="III")

```

```

## Anova Table (Type III tests)
##
## Response: BG_Dry_mg
##             Sum Sq  Df  F value Pr(>F)
## (Intercept) 16514   1   0.4677 0.4943
## SoilHostSpecies 48626   3   0.4591 0.7110
## PlantSpecies 20813562   3 196.5027 <2e-16 ***
## Age          33572   1   0.9509 0.3298
## Residuals    25915091 734
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Discussion

In Figures 4 and 5, box plots mostly overlap with each other, showing weak correlation between soil host species and below-ground dry mass. Across species, there doesn't appear to be a prominent soil host species contributing to below-ground dry mass. Across ages 0.5 to 10 years for each species, there doesn't appear to be a prominent soil host species contributing to below-ground dry mass. According to the ANCOVA analysis, the p-value for soil host species is more than 0.05. This indicates that soil host species are not significantly correlated to below-ground dry mass, even when taking age into account. Therefore, it is assumed that the bacteria community formed by the soil host species is not significant to plant root growth.

Question 2b)

Which bacteria has the greatest effect on plant biomass when the plant is 4-12 years old?

The goal of this section is to determine which bacteria has the greatest effect on plant biomass when the plant is between 4 and 12 years old. This was chosen as this is the average age of the plants in the data set.

Transform the Data

```

# Removing all but wanted data points
GData = GData %>%
  select(BG_Dry_mg, PlantSpecies, SoilHostSpecies, Age)

```

Setting up Data

```
#create a new data set with only the plants that are this old
subdata = GData %>%
  subset(Age == 4:12)
```

```
## Warning in Age == 4:12: longer object length is not a multiple of shorter object
## length
```

```
head(subdata)
```

```
##   BG_Dry_mg PlantSpecies SoilHostSpecies Age
## 31    380.26          La      Baccharis    7
## 37    216.89          La      Baccharis    4
## 60     40.21          Aa       Lupine     9
## 63    12.42          Bp      Baccharis   12
## 66    229.14          La     Ammophila    6
## 70   1696.73          La      Baccharis   10
```

Now with this data, we can begin our analysis

Analysis

We will use a box plot to see the impact that the bacteria have on the plants that are this age

```
Plot1 = ggplot(subdata, aes(x = SoilHostSpecies, y = BG_Dry_mg, fill = SoilHostSpecies)) +
  geom_boxplot() +
  theme_classic() +
  scale_fill_discrete(name = "Soil Bacteria") +
  scale_x_discrete(name = "", labels = c("", "", "", "")) +
  scale_y_continuous(name = "Root Dry Mass (mg)")

Plot1
```

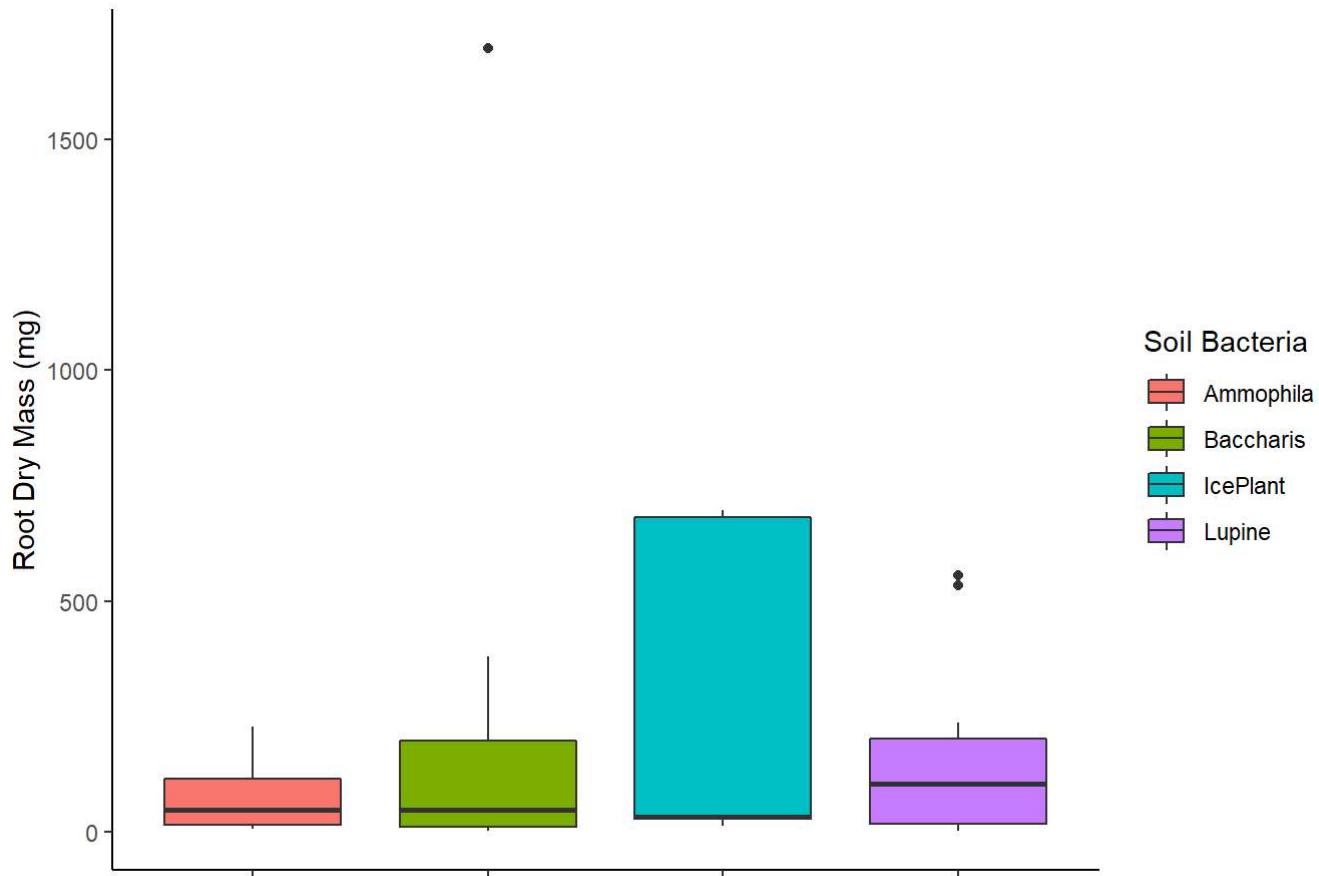


Figure 6: Shows the relationship between the different species of Soil Bacteria and Root Dry mass in mg. Lupine has the highest average value while IcePlant has the greatest variation

```
new_subdata = subdata
for (i in 1:length(subdata$PlantSpecies)) {
  if (subdata$SoilHostSpecies[i] != "Lupine"){
    new_subdata$SoilHostSpecies[i] = "Other"
  }
}
```

Now we will recreate the plot with the new data set

```
Plot2 = ggplot(new_subdata, aes(x = SoilHostSpecies, y = BG_Dry_mg, fill = SoilHostSpecies)) +
  geom_boxplot() +
  theme_classic() +
  scale_fill_discrete(name = "Soil Bacteria") +
  scale_x_discrete(name = "", labels = c("", "")) +
  scale_y_continuous(name = "Root Dry Mass (mg)")
Plot2
```

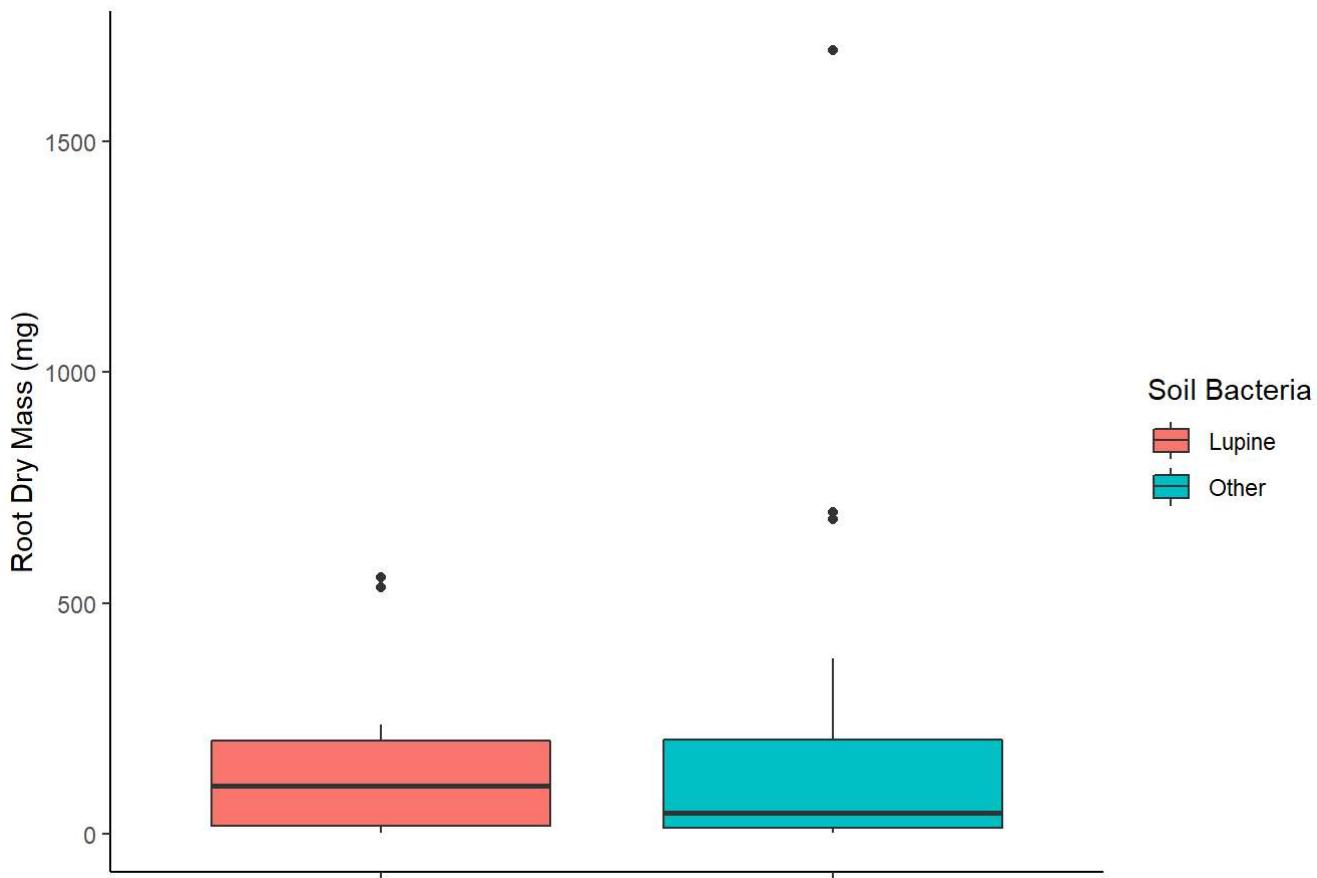


Figure 7: Shows the difference in Root Dry Mass between Lupine and all the other different species of soil bacteria.

Now we will conduct an unpaired two-sample t-test to determine if this difference between the two groups is significant or not. We will use a two-sample t-test as the two groups are unrelated.

```
t.test(subset(new_subdata$BG_Dry_mg, new_subdata$SoilHostSpecies == "Other"),
       subset(new_subdata$BG_Dry_mg, new_subdata$SoilHostSpecies == "Lupine"),
       alternative = "two.sided", var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: subset(new_subdata$BG_Dry_mg, new_subdata$SoilHostSpecies == "Other") and subset(new_subdata$BG_Dry_mg, new_subdata$SoilHostSpecies == "Lupine")
## t = 0.41412, df = 31.912, p-value = 0.6816
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -163.3390 246.6905
## sample estimates:
## mean of x mean of y
## 207.5341 165.8583
```

Table 1: Shows the results of a Welch Two Sample t-test of below ground dry mass of plants with Lupine against plants with any of the other different species. With a p-value of 0.6816, the results are not statistically significant

Discussion

According to Figure 6, we can see that Lupine has an impact on the dry weight of the roots between 4 and 12 years. To further examine this relationship, we will create a new set of data that groups together all the other species of soil bacteria, then create a boxplot using that set. Based on Figure 7, we can see that there is a difference between samples with Lupine and all the other species of soil bacteria at this point in time. Besides, based on the results of this two-sample t-test, we can see that while there is a difference between the two groups, this difference is not enough for it to be considered significant as the p-value is not less than 0.05.

Conclusion

Our data suggest that plant-soil microbial interactions do exist, but we were not able to demonstrate whether the strength of this interaction is affected by the age of the individual plant. In our data, the different ages of plant individuals can result in fluctuating bacterial community compositions, though this interaction is not significant. We came to a similar conclusion in analyzing soil microorganisms affecting plant root weight. At present, it seems that the age of the individual plant does not affect the interaction between the plant and the soil microorganisms greatly. That is different from the results in the research article from which we obtained our data. It is suggested in the article that the strength of this interaction correlates with the length of time that individual plants condition the soil. That may be because other influencing factors, such as soil conditions, were not included in our analysis, resulting in differing conclusions. However, since changes to the bacterial soil host community composition were identified as plants aged, it could be helpful to focus subsequent studies on finding some way to quantify whether this interaction is significant. Furthermore, looking into the soil conditioning effects of plants and its resulting impact on bacterial community compositions could help researchers further understand the plant-soil microbe interaction and relationship.