

Incubation Results

Jared

February 13, 2019

Contents

Setup	1
Results	1
Alpha Diversity	1
Beta Diversity	3
Aliens	8
Alien Heatmaps	11

Setup

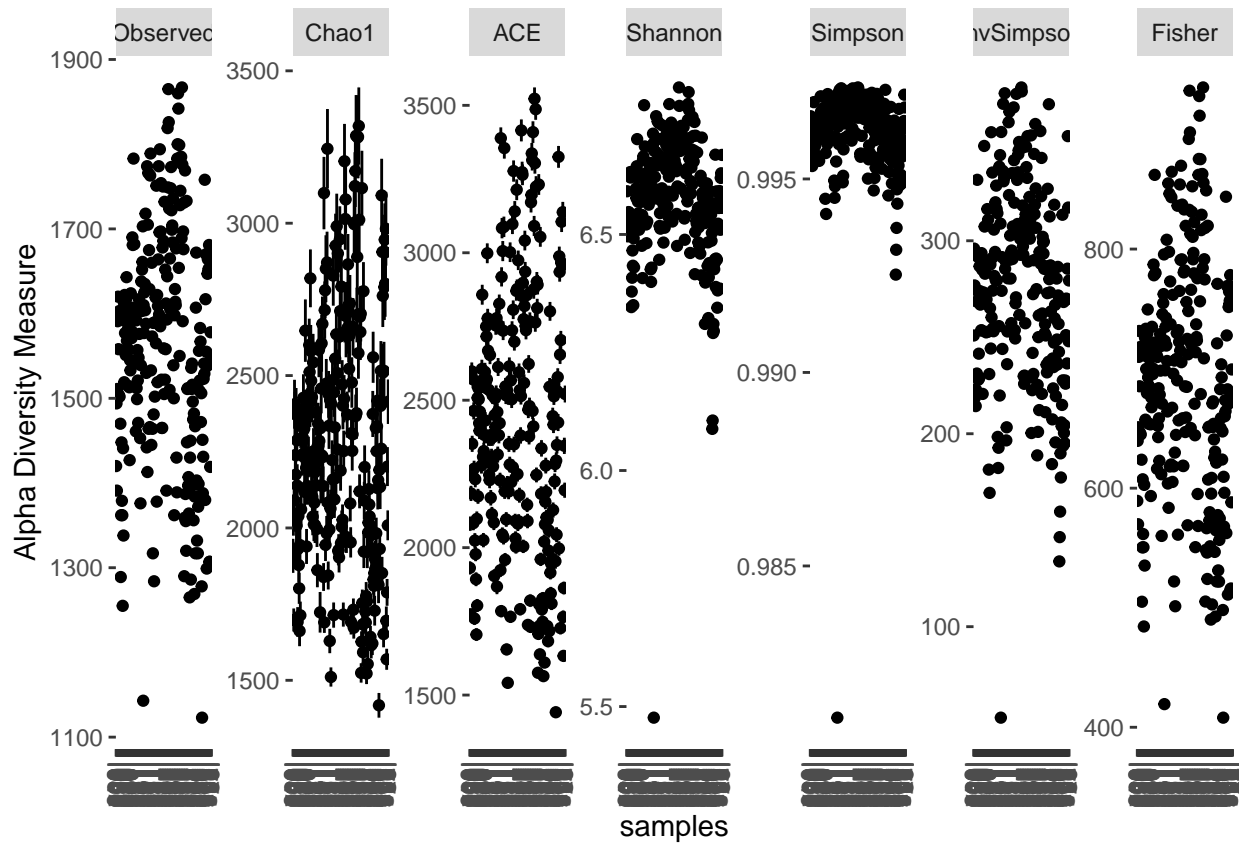
Load needed libraries

```
library(tidyverse)
library(phyloseq)
library(vegan)
library(ggplot2)
library(ggtree)
library(gplots)
```

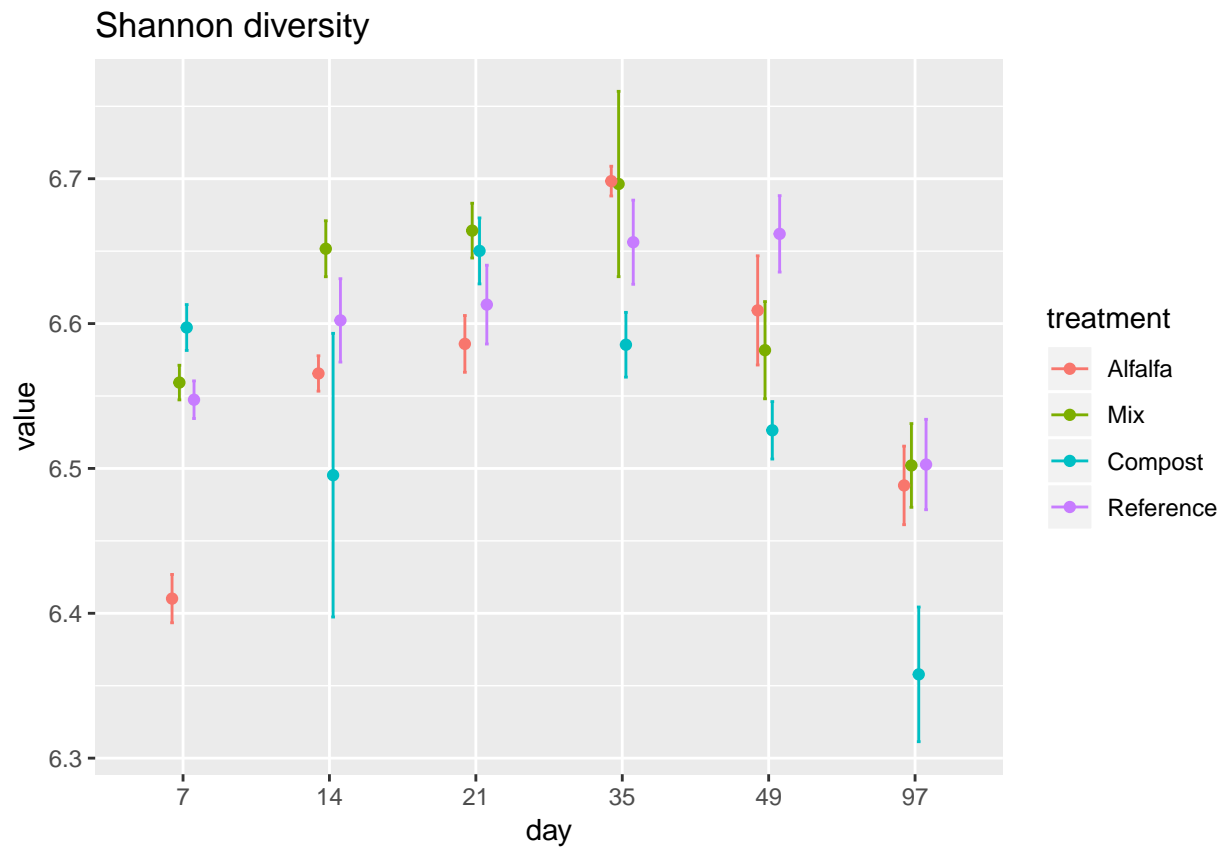
Results

Alpha Diversity

```
# This object has day 0 and amendment samples removed and is rarefied to 6000
physeq <- readRDS("data/IncPhyseqRareClusteredTree")
plot_richness(physeq)
```



```
shannon <- plot_richness(physeq, measures = "Shannon")
shannon.df <- shannon$data
# load summary function
source("Functions/summarySE.R")
summary.shannon.df <- summarySE(shannon.df, measurevar = "value", groupvars = c("treatment", "day"))
pd <- position_dodge(0.2) # move them .05 to the left and right
ggplot(summary.shannon.df, aes(x=day, y=value, colour=treatment)) +
  geom_errorbar(aes(ymin=value-se, ymax=value+se), width=.1, position=pd) +
  geom_point(position=pd) +
  ggtitle("Shannon diversity")
```

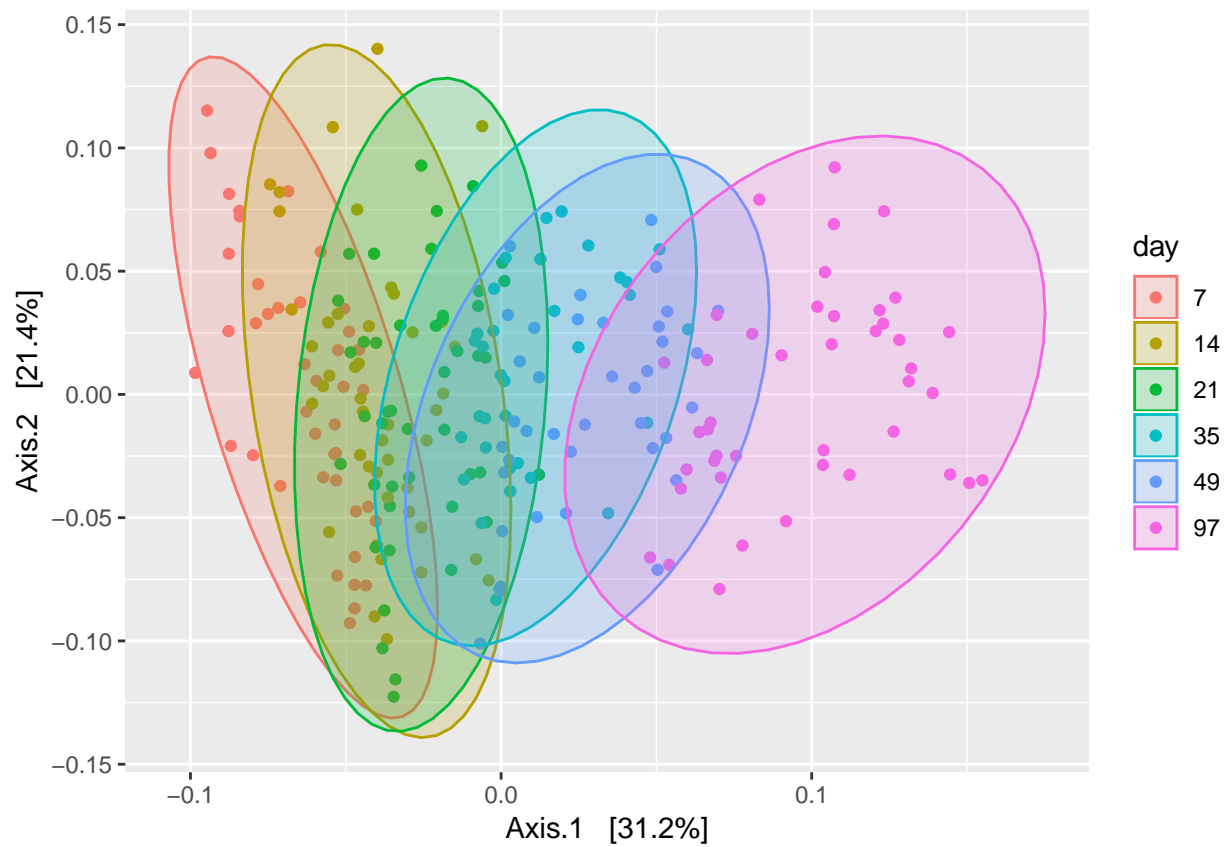


Beta Diversity

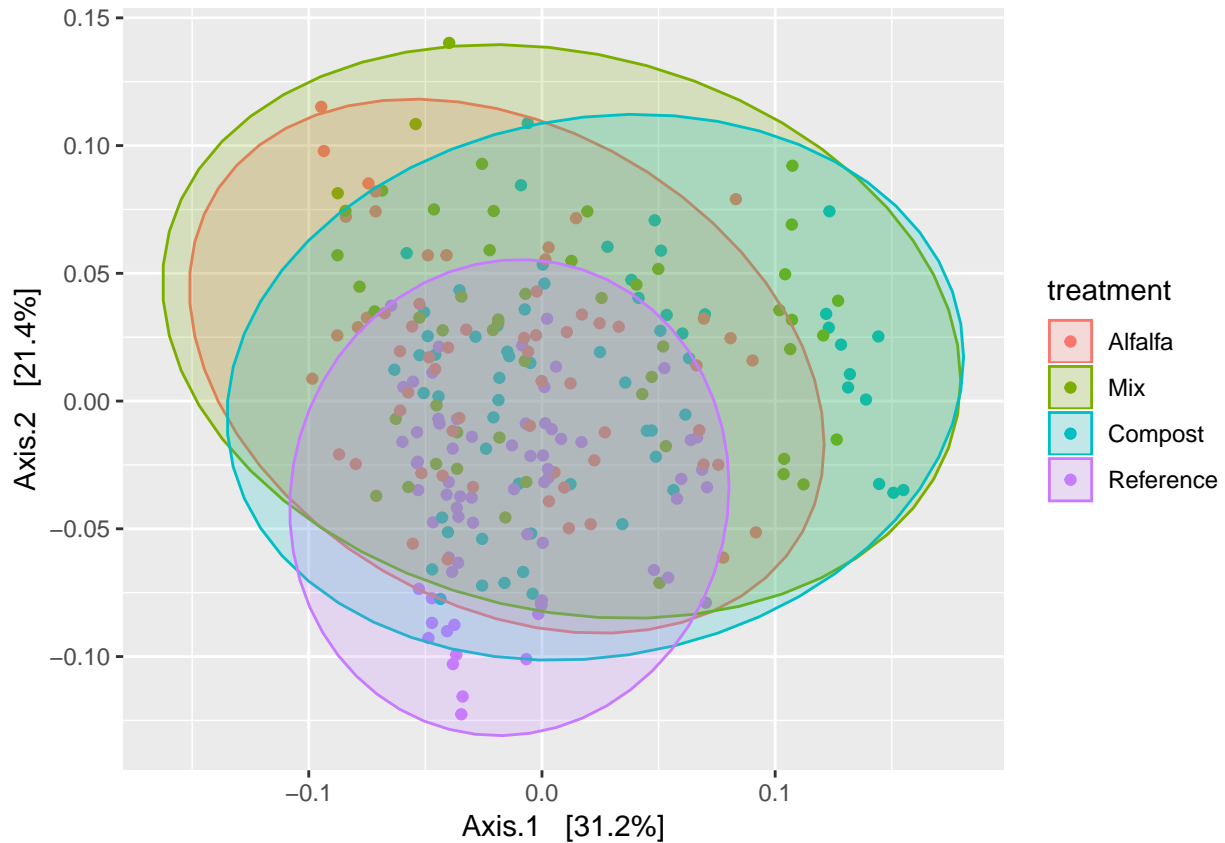
Now let's plot an PCA and NMDS ordination of the weighted unifracs distances, which uses the phylogenetic tree and considers the phylogenetic relationship between OTUs when calculating the distance matrix

PCA

```
PCoA <- ordinate(physeq, "PCoA", "wunifrac")
plot_ordination(physeq, PCoA, color = "day") + stat_ellipse(geom = "polygon", type = "norm", alpha = 0.1)
```



```
plot_ordination(physeq, PCoA, color = "treatment") + stat_ellipse(geom = "polygon", type = "norm", alpha = 0.5)
```



NMDS

```
NMDS <- ordinate(physeq, "NMDS", "wunifrac")

## Run 0 stress 0.1151871
## Run 1 stress 0.1151881
## ... Procrustes: rmse 0.0001294299 max resid 0.001279911
## ... Similar to previous best
## Run 2 stress 0.1151873
## ... Procrustes: rmse 2.764585e-05 max resid 0.0004068377
## ... Similar to previous best
## Run 3 stress 0.1151883
## ... Procrustes: rmse 0.0002414934 max resid 0.002234498
## ... Similar to previous best
## Run 4 stress 0.1151868
## ... New best solution
## ... Procrustes: rmse 0.0001812778 max resid 0.002123894
## ... Similar to previous best
## Run 5 stress 0.1151866
## ... New best solution
## ... Procrustes: rmse 5.209263e-05 max resid 0.0007885718
## ... Similar to previous best
## Run 6 stress 0.1245046
## Run 7 stress 0.1151874
## ... Procrustes: rmse 0.0001627992 max resid 0.002102698
```

```

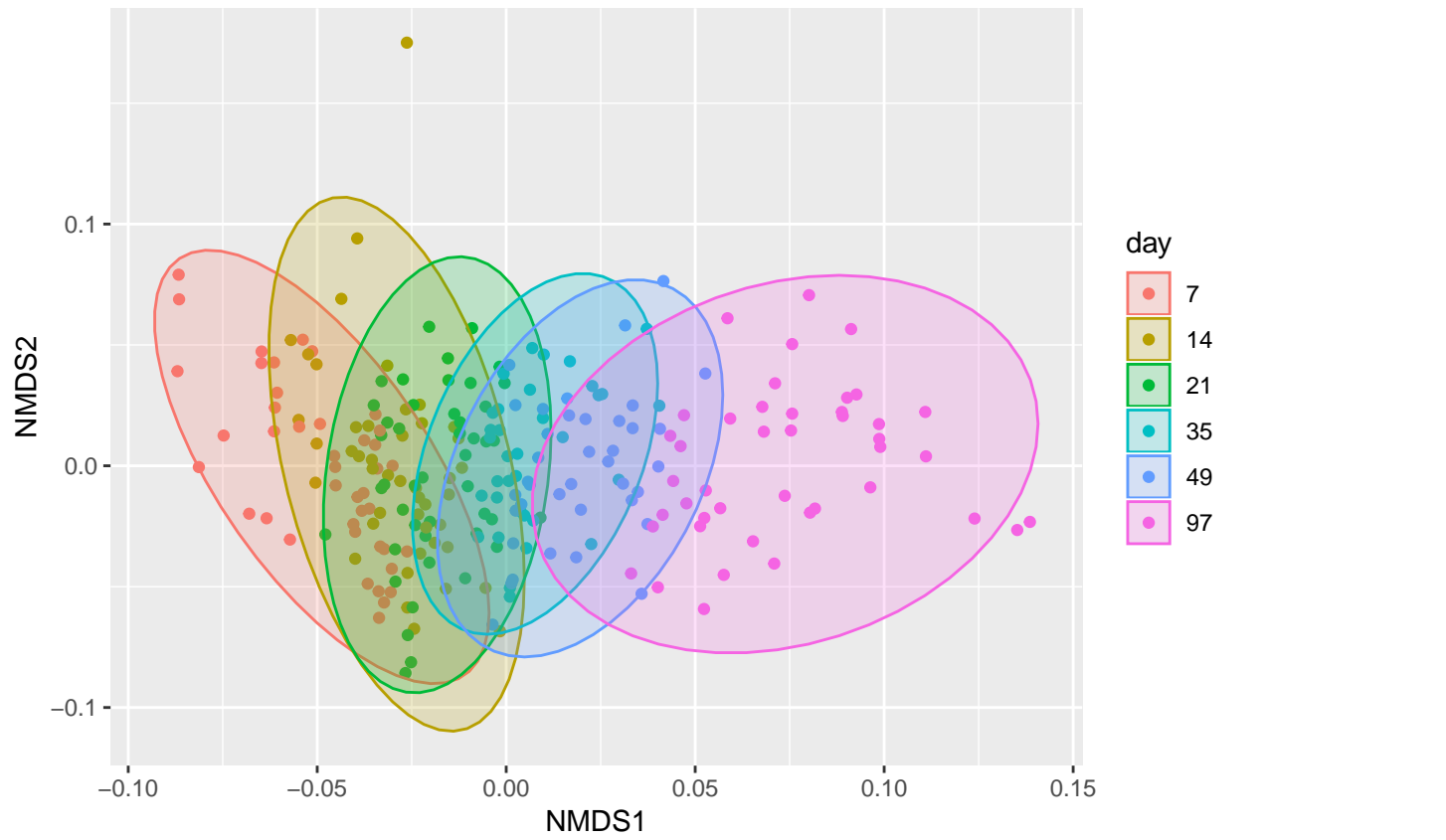
## ... Similar to previous best
## Run 8 stress 0.1151866
## ... New best solution
## ... Procrustes: rmse 1.570857e-05  max resid 0.0002149405
## ... Similar to previous best
## Run 9 stress 0.1151866
## ... Procrustes: rmse 3.04161e-05  max resid 0.0004297543
## ... Similar to previous best
## Run 10 stress 0.1245082
## Run 11 stress 0.1151872
## ... Procrustes: rmse 0.000159528  max resid 0.002089685
## ... Similar to previous best
## Run 12 stress 0.1151867
## ... Procrustes: rmse 4.776284e-05  max resid 0.0007108816
## ... Similar to previous best
## Run 13 stress 0.1151869
## ... Procrustes: rmse 7.714202e-05  max resid 0.001179512
## ... Similar to previous best
## Run 14 stress 0.1151871
## ... Procrustes: rmse 0.0001572979  max resid 0.002108385
## ... Similar to previous best
## Run 15 stress 0.1151866
## ... Procrustes: rmse 2.890565e-05  max resid 0.0004087085
## ... Similar to previous best
## Run 16 stress 0.1245049
## Run 17 stress 0.1151882
## ... Procrustes: rmse 0.0001943049  max resid 0.002111913
## ... Similar to previous best
## Run 18 stress 0.124504
## Run 19 stress 0.1151878
## ... Procrustes: rmse 0.000127666  max resid 0.001280578
## ... Similar to previous best
## Run 20 stress 0.1151881
## ... Procrustes: rmse 0.0001926802  max resid 0.002104732
## ... Similar to previous best
## *** Solution reached

```

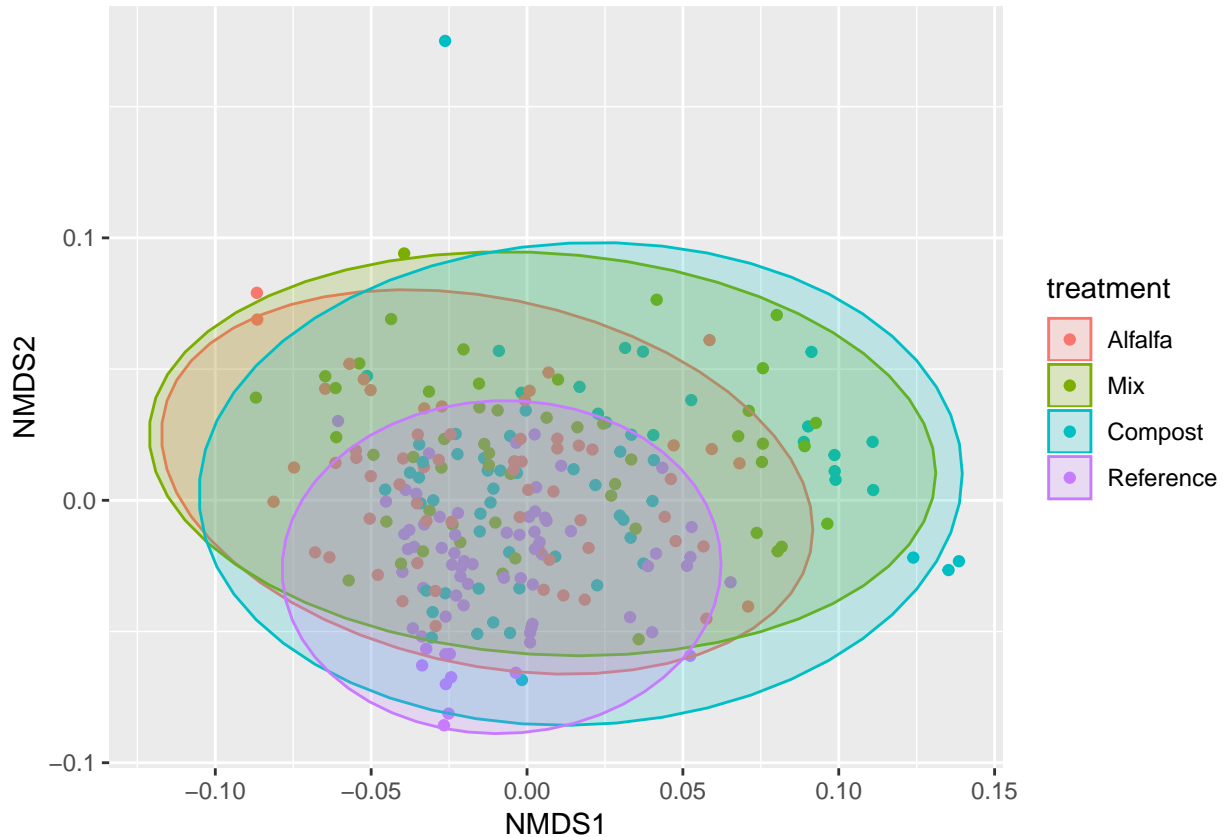
```

plot_ordination(physeq, NMDS, color = "day") + stat_ellipse(geom = "polygon", type = "norm", alpha = 0.1)

```



```
plot_ordination(physeq, NMDS, color = "treatment") + stat_ellipse(geom = "polygon", type = "norm", alpha
```



Aliens

Table with # of aliens for each treatment and what period of the incubation they were detected, early/late or throughout.

The lists of alien OTUs detected in amended microcosms was generated with , we can load those lists along with the phyloseq object with clustering information from :

```
alf.aliens <- readRDS("data/alf.aliens.rds")
comp.aliens <- readRDS("data/comp.aliens.rds")
mix.aliens <- readRDS("data/mix.aliens.rds")
# Read in the phyloseq object from the clustering script

alf <- prune_samples(sample_data(physeq)$treatment %in% c("Alfalfa"), physeq) %>%
  filter_taxa(function(x) sum(x) > 1, T)
early.alf <- prune_samples(sample_data(alf)$response.group %in% c("early"), alf) %>%
  filter_taxa(function(x) sum(x) > 1, T)
late.alf <- prune_samples(sample_data(alf)$response.group %in% c("late"), alf) %>%
  filter_taxa(function(x) sum(x) > 1, T)

number.aliens.early.alf <- prune_taxa(as.character(alf.aliens), early.alf) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

number.aliens.late.alf <- prune_taxa(as.character(alf.aliens), late.alf) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
```



```

otu_table() %>%
nrow()

number.aliens.total.alf <- prune_taxa(as.character(alf.aliens), alf) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

# print("Number of alfalfa alien OTUs detected in all alfalfa")
# number.aliens.total.alf
# print("Total number of OTUs detected in all alfalfa")
# nrow(otu_table(alf))
# print("Number of alfalfa alien OTUs detected in early alfalfa")
# number.aliens.early.alf
# print("Total number of OTUs detected in early alfalfa")
# nrow(otu_table(early.alf))
# print("Number of alfalfa alien OTUs detected in late alfalfa")
# number.aliens.late.alf
# print("Total number of OTUs detected in late alfalfa")
# nrow(otu_table(late.alf))

comp <- prune_samples(sample_data(physeq)$treatment %in% c("Compost"), physeq) %>%
  filter_taxa(function(x) sum(x) > 1, T)
early.comp <- prune_samples(sample_data(comp)$response.group %in% c("early"), comp) %>%
  filter_taxa(function(x) sum(x) > 1, T)
late.comp <- prune_samples(sample_data(comp)$response.group %in% c("late"), comp) %>%
  filter_taxa(function(x) sum(x) > 1, T)

number.aliens.early.comp <- prune_taxa(as.character(comp.aliens), early.comp) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

number.aliens.late.comp <- prune_taxa(as.character(comp.aliens), late.comp) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

number.aliens.total.comp <- prune_taxa(as.character(comp.aliens), comp) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

# print("Number of Compost alien OTUs detected in all Compost")
# number.aliens.total.comp
# print("Total number of OTUs detected in all Compost")
# nrow(otu_table(comp))
# print("Number of Compost alien OTUs detected in early Compost")
# number.aliens.early.comp
# print("Total number of OTUs detected in early Compost")
# nrow(otu_table(early.comp))
# print("Number of Compost alien OTUs detected in late Compost")
# number.aliens.late.comp
# print("Total number of OTUs detected in late Compost")

```

```

# nrow(otu_table(late.comp))

mix <- prune_samples(sample_data(physeq)$treatment %in% c("Mix"), physeq) %>%
  filter_taxa(function(x) sum(x) > 1, T)
early.mix <- prune_samples(sample_data(mix)$response.group %in% c("early"), mix) %>%
  filter_taxa(function(x) sum(x) > 1, T)
late.mix <- prune_samples(sample_data(mix)$response.group %in% c("late"), mix) %>%
  filter_taxa(function(x) sum(x) > 1, T)

number.aliens.early.mix <- prune_taxa(as.character(mix.aliens), early.mix) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

number.aliens.late.mix <- prune_taxa(as.character(mix.aliens), late.mix) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

number.aliens.total.mix <- prune_taxa(as.character(mix.aliens), mix) %>%
  filter_taxa(function(x) sum(x) > 1, T) %>%
  otu_table() %>%
  nrow()

# print("Number of Mix alien OTUs detected in all Mix")
# number.aliens.total.mix
# print("Total number of OTUs detected in all Mix")
# nrow(otu_table(mix))
# print("Number of Mix alien OTUs detected in early Mix")
# number.aliens.early.mix
# print("Total number of OTUs detected in early Mix")
# nrow(otu_table(early.mix))
# print("Number of Mix alien OTUs detected in late Mix")
# number.aliens.late.mix
# print("Total number of OTUs detected in late Mix")
# nrow(otu_table(late.mix))

library(knitr)
a <- c("Alfalfa", number.aliens.total.alf, number.aliens.early.alf, number.aliens.late.alf)
b <- c("Compost", number.aliens.total.comp, number.aliens.early.comp, number.aliens.late.comp)
c <- c("Mix", number.aliens.total.mix, number.aliens.early.mix, number.aliens.late.mix)
x <- as.data.frame(rbind(a,b,c))
colnames(x) <- c("Treatment", "Throughout", "Early", "Late")
rownames(x) <- NULL
kable(x, caption = "Number of OTUs considered aliens detected in each treatment")

```

Table 1: Number of OTUs considered aliens detected in each treatment

Treatment	Throughout	Early	Late
Alfalfa	26	25	4
Compost	74	60	28
Mix	58	38	20

Alien Heatmaps

Function to get a list of OTUs from a sample type

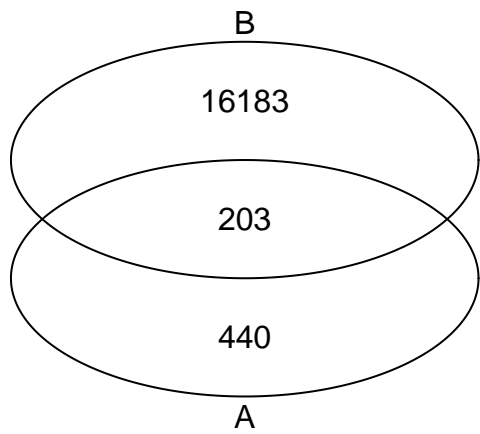
```
GetOTUs <- function(physeq, samples) {  
  prune_samples(sample_data(physeq)$treatment %in% c(samples), physeq) %>%  
  filter_taxa(function(x) sum(x) > 1, T) %>%  
  tax_table() %>%  
  row.names()  
}
```

Using function to generate list of OTUs from all starting soil, incubated soils and amendments

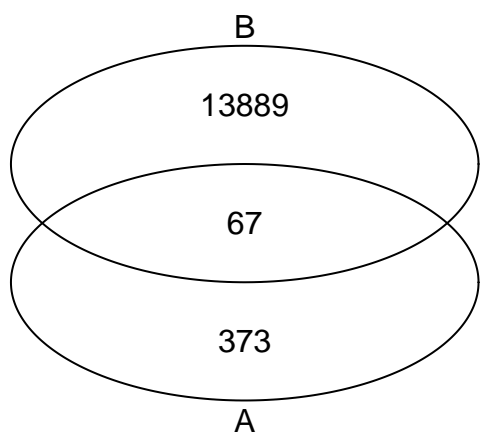
```
physeq.raw <- readRDS("data/RDS/incubation_physeq_Aug18.RDS")  
# Incubated samples:  
Alfalfa.otus <- GetOTUs(physeq.raw, c("Alfalfa"))  
Compost.otus <- GetOTUs(physeq.raw, c("Compost"))  
CompAlfa.otus <- GetOTUs(physeq.raw, c("CompAlfa"))  
Control.otus <- GetOTUs(physeq.raw, c("Control"))  
# Amendment samples:  
AlfalfaAmend.otus <- GetOTUs(physeq.raw, c("AlfalfaAmend"))  
CompostAmend.otus <- GetOTUs(physeq.raw, c("CompostAmend"))  
# Soil sample:  
AlfalfaSoil.otus <- GetOTUs(physeq.raw, c("AlfalfaSoil"))
```

```
GetAlienHeatMap <- function(physeq, control_otus, alien_otus, recieving_otus, samples){  
  otus <- list(alien_otus, control_otus)  
  print("Looking for aliens between amendment and control soil")  
  venn <- venn(otus)  
  alf.aliens <- attr(venn, "intersections")$A  
  aliens <- list(alf.aliens, recieving_otus)  
  print("Detecting aliens in amended soil")  
  aliens.venn <- venn(aliens)  
  aliens.detected <- attr(aliens.venn, "intersections")$`A:B`  
  incubated <- prune_samples(sample_data(physeq)$treatment %in% c(samples), physeq) %>%  
    filter_taxa(function(x) sum(x) > 5, T) %>%  
    transform_sample_counts(function(x) x / sum(x))  
  incubated.aliens <- prune_taxa(aliens.detected, incubated)  
  sample.order <- as.data.frame(sample_data(incubated.aliens)) %>%  
    arrange(day, replication) %>%  
    select(i_id) %>%  
    remove_rownames()  
  alien.heatmap <- plot_heatmap(incubated.aliens, sample.label = "day", taxa.order= "Phylum", taxa.labels =  
    sample.order = as.character(sample.order$i_id),  
    low = "#66CCFF", high = "#000033", na.value = "white")  
  alien.heatmap  
}
```

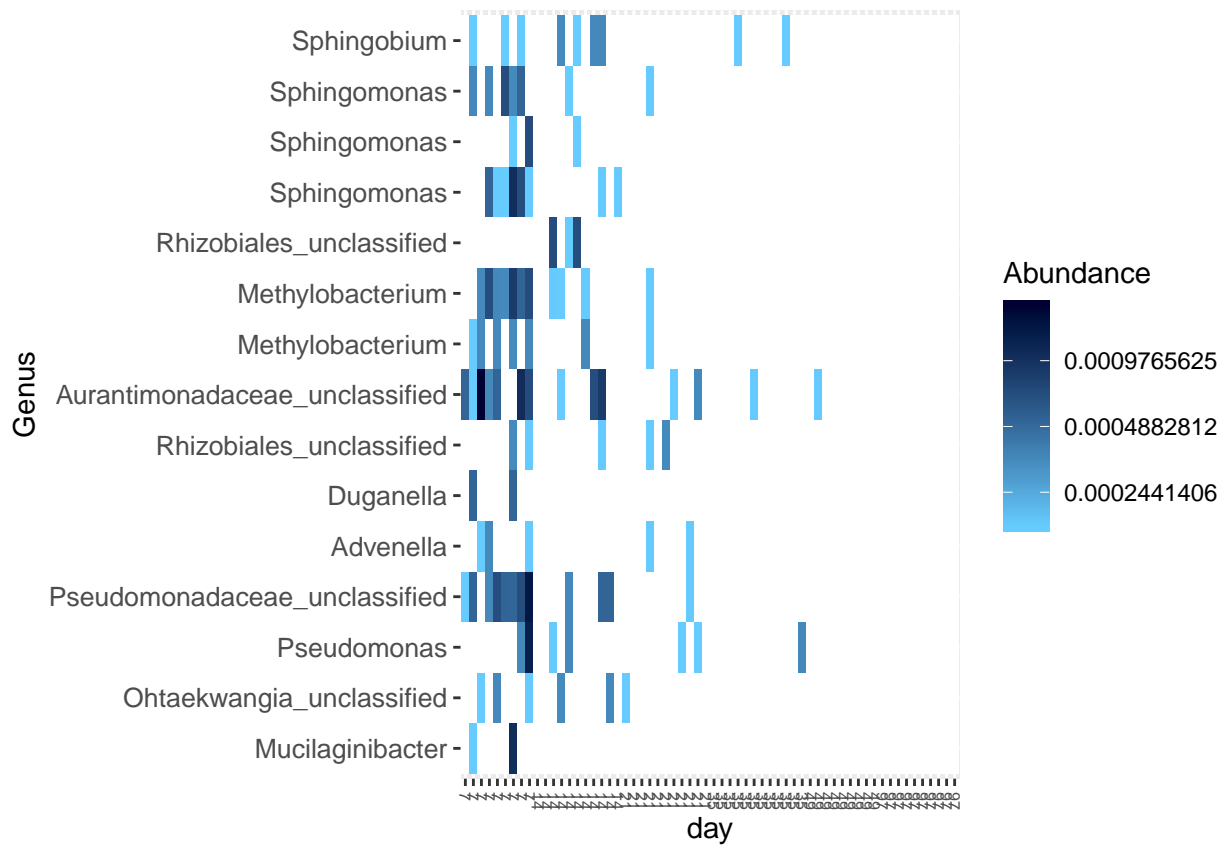
```
alf.heatmap <- GetAlienHeatMap(physeq, Control.otus, AlfalfaAmend.otus, Alfalfa.otus, c("Alfalfa"))  
  
## [1] "Looking for aliens between amendment and control soil"
```



```
## [1] "Detecting aliens in amended soil"
```

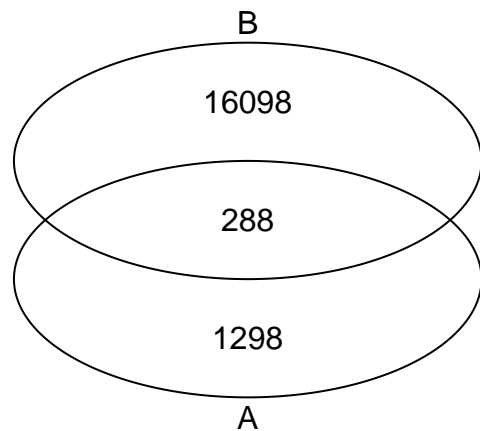


```
alf.heatmap
```

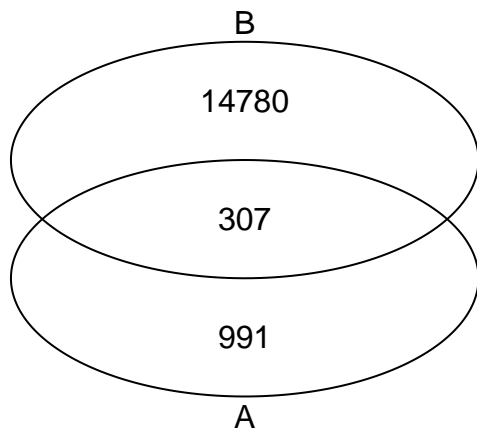


```
comp.heatmap <- GetAlienHeatMap(physeq, Control.otus, CompostAmend.otus, Compost.otus, c("Compost"))
```

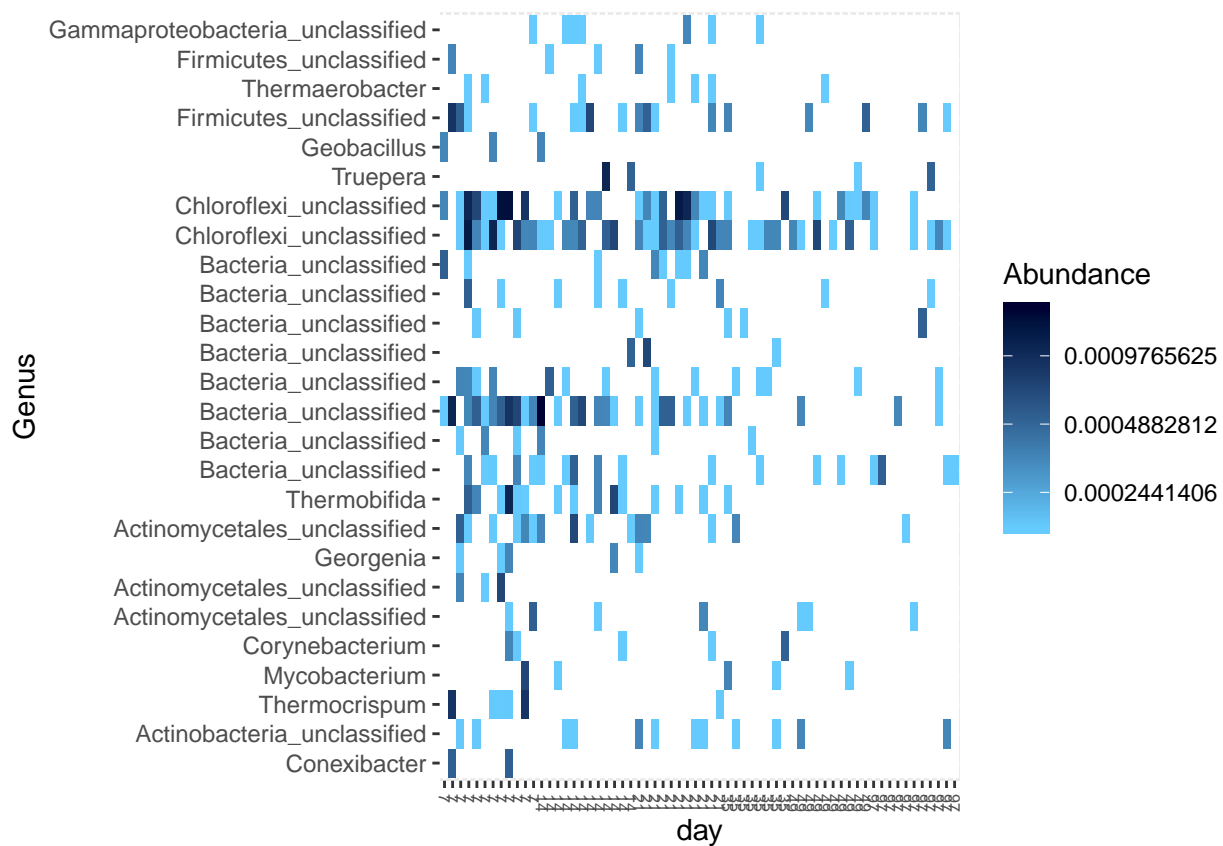
```
## [1] "Looking for aliens between amendment and control soil"
```



```
## [1] "Detecting aliens in amended soil"
```



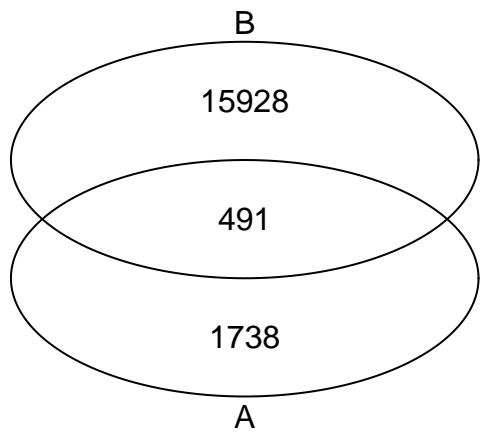
comp.heatmap



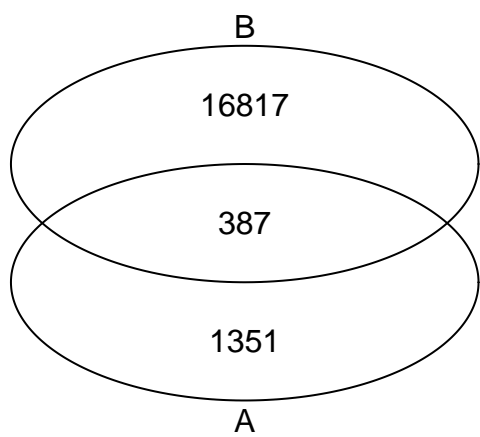
I didn't extract DNA from a mixed sample of compost and alfalfa, to get the list of potential aliens, we will combine the two lists from alfalfa and compost.

```
mix <- c(AlfalfaAmend.otus, CompostAmend.otus)
mix.heatmap <- GetAlienHeatMap(physeq, Control.otus, mix, CompAlfa.otus, c("Mix"))
```

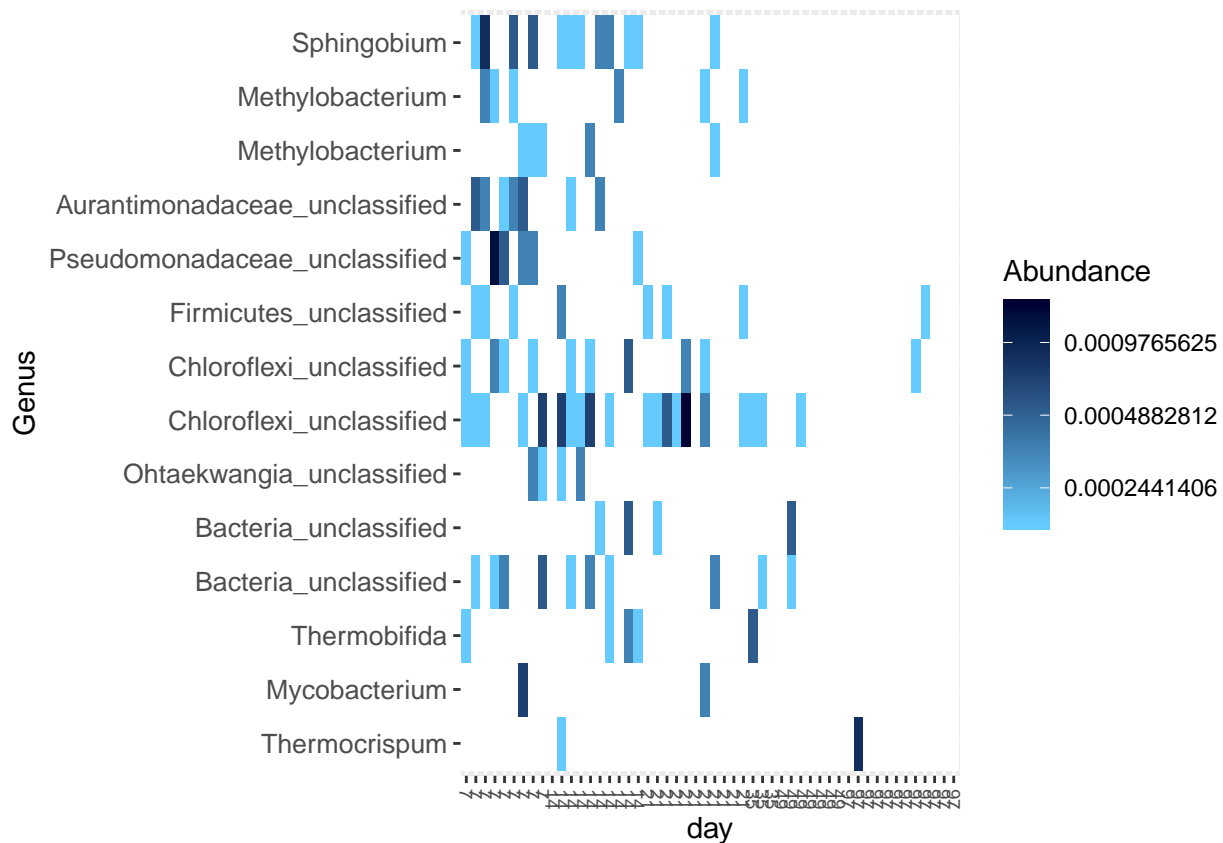
```
## [1] "Looking for aliens between amendment and control soil"
```



```
## [1] "Detecting aliens in amended soil"
```



```
mix.heatmap
```



The relative abundance of the alien OTUs through the incubation is relatively low, but different between treatments. Could low abundance OTUs be drivers in this situation? We should compare the relative abundance of the dominant OTUs from each treatment.

Let's make another heatmap, this one will have a list of OTUs from a treatment that is composed of the top ten OTUs by relative abundance from each day.

```
Day_top10 <- function(physeq, trt, days){
  trt <- prune_samples(sample_data(physeq)$treatment %in% c(trt), physeq)

  d0 <- subset_samples(trt, day == days[1])
  10 <- names(sort(taxa_sums(d0), TRUE)[1:10])

  d7 <- subset_samples(trt, day == days[2])
  17 <- names(sort(taxa_sums(d7), TRUE)[1:10])

  d14 <- subset_samples(trt, day == days[3])
  114 <- names(sort(taxa_sums(d14), TRUE)[1:10])

  d21 <- subset_samples(trt, day == days[4])
  121 <- names(sort(taxa_sums(d21), TRUE)[1:10])

  d35 <- subset_samples(trt, day == days[5])
  135 <- names(sort(taxa_sums(d35), TRUE)[1:10])

  d49 <- subset_samples(trt, day == days[6])
  149 <- names(sort(taxa_sums(d49), TRUE)[1:10])
}
```



```

d97 <- subset_samples(trt, day == days[7])
197 <- names(sort(taxa_sums(d97), TRUE)[1:10])
list <- c(10, 17, 114, 121, 135, 149, 197)
list
phy <- prune_taxa(list, trt) %>%
  filter_taxa(function(x) sum(x) > 5, T) %>%
  transform_sample_counts(function(x) x / sum(x))
sample.order <- as.data.frame(sample_data(phy)) %>%
  arrange(day, replication) %>%
  select(i_id) %>%
  remove_rownames()
heatmap <- plot_heatmap(phy, sample.label = "day", taxa.order= "Phylum", taxa.label = "Genus",
  sample.order = as.character(sample.order$i_id),
  low = "#66CCFF", high = "#000033", na.value = "white")

heatmap
}

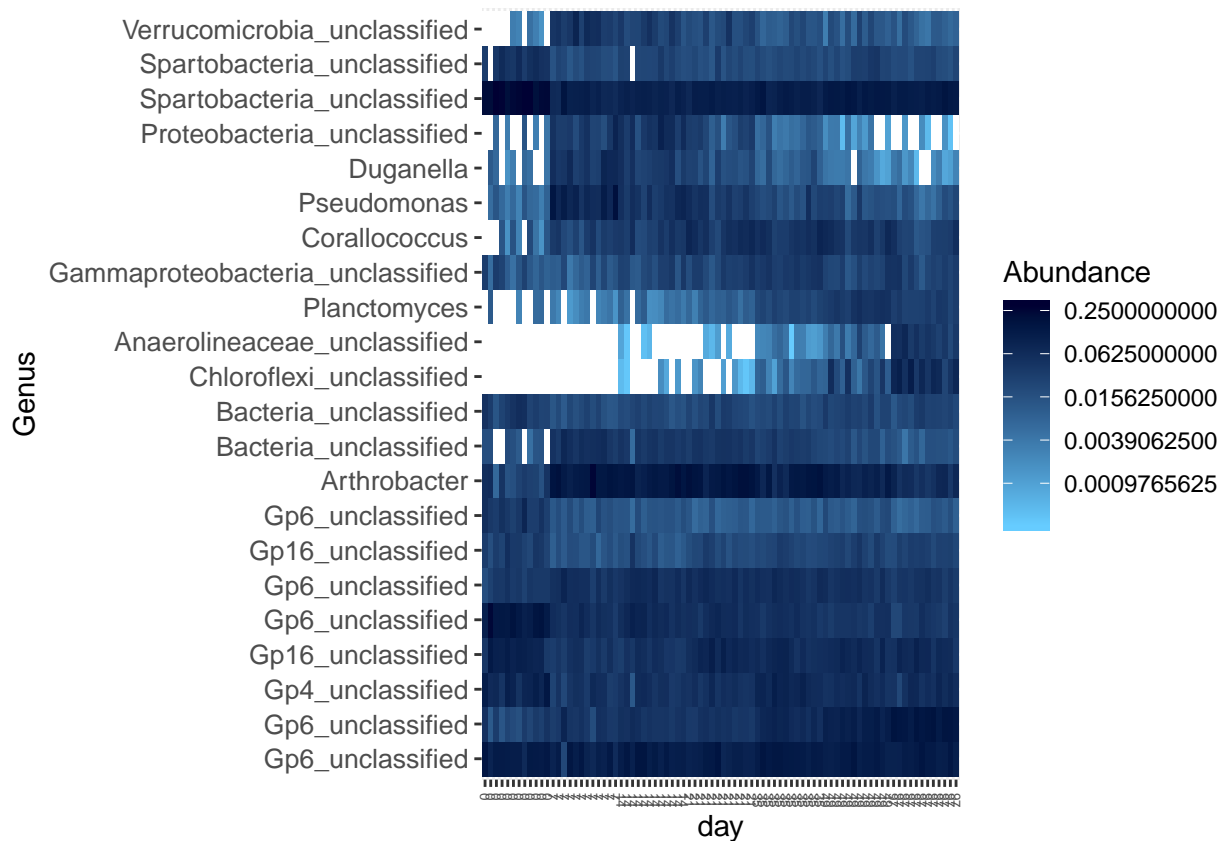
```

Not rarefied for these heatmaps:

```

days <- c("0", "7", "14", "21", "35", "49", "97")
alf <- Day_top10(physeq.raw, c("Alfalfa"), days)
alf

```

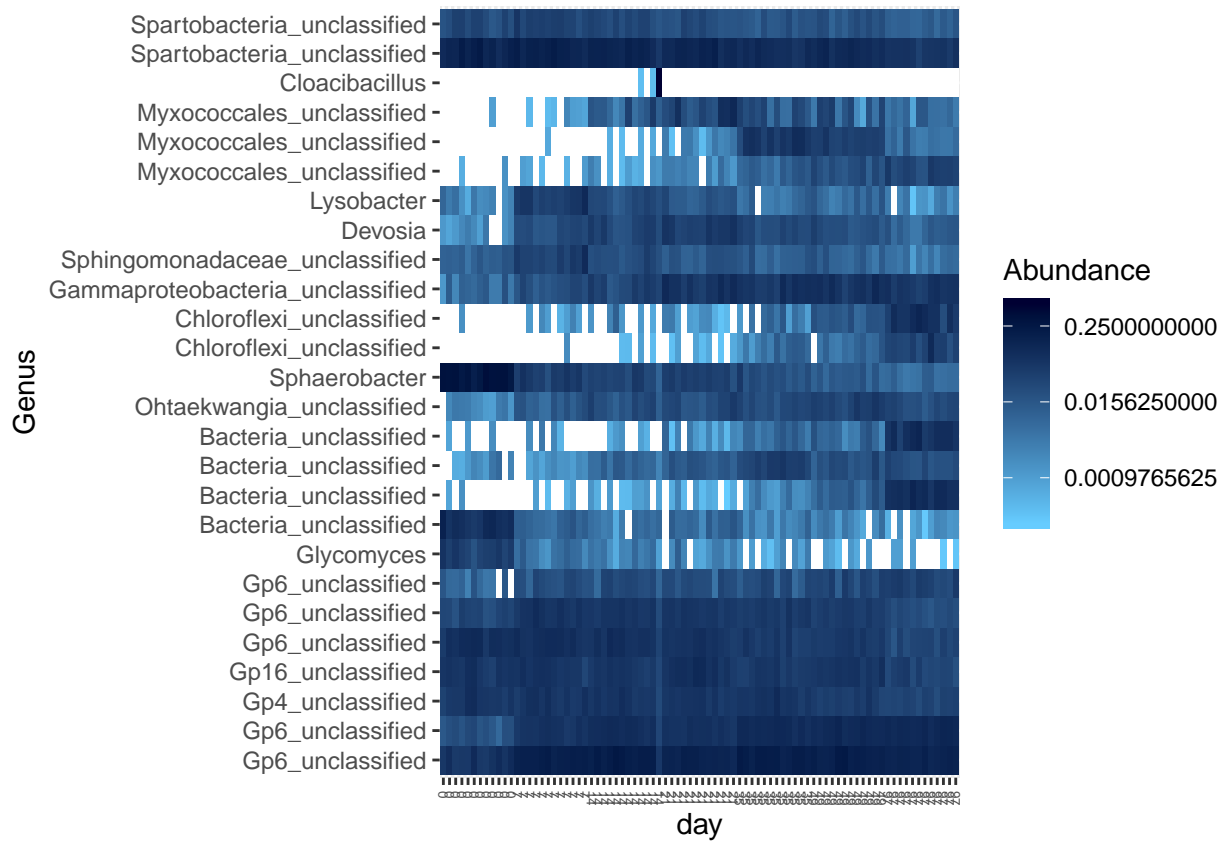


Compost:

```

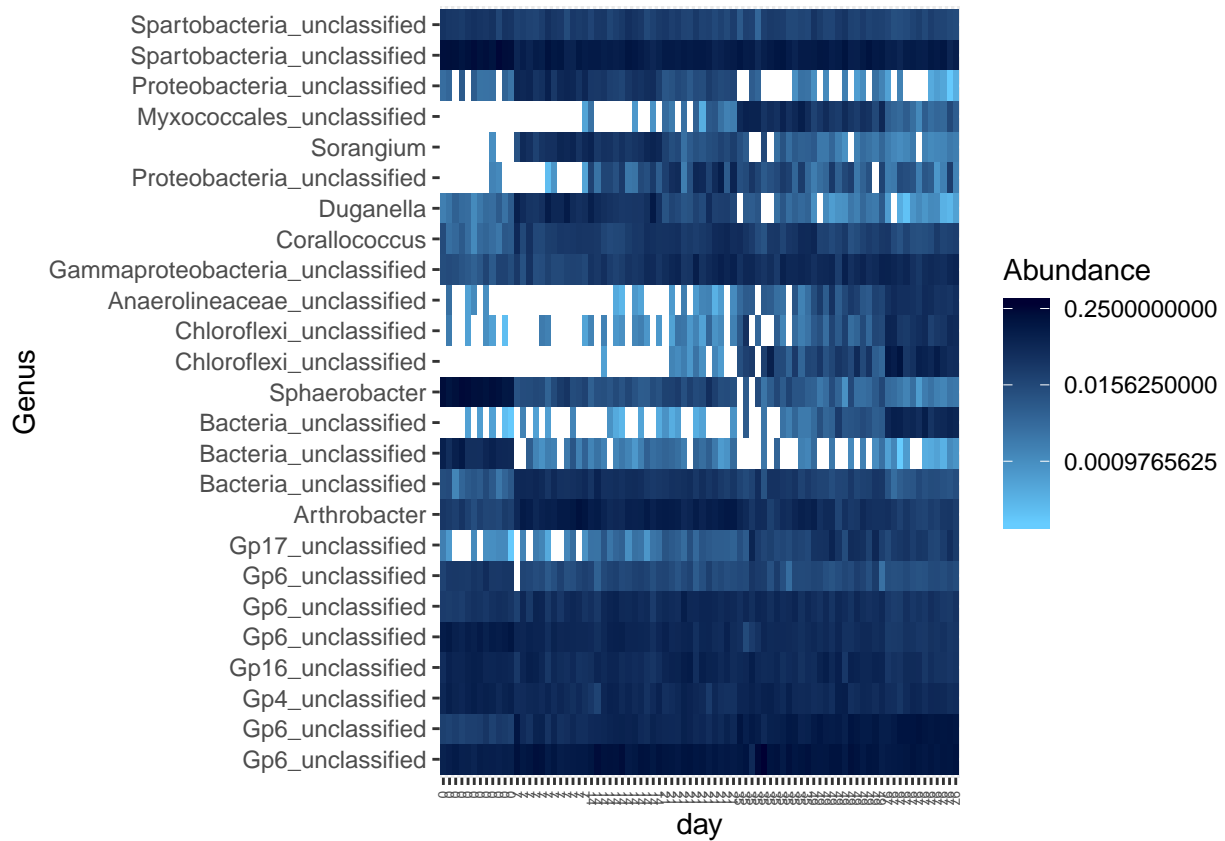
comp <- Day_top10(physeq.raw, c("Compost"), days)
comp

```



Mix:

```
mix <- Day_top10(physeq.raw, c("CompAlfa"), days)
mix
```



What are the common top 10 OTUs from the three treatments?

```
venn(list("Alfalfa" = levels(alf$data$OTU), "Compost" = levels(comp$data$OTU), "Mix" = levels(mix$data$OTU)))
```

