Taxonomic_Analysis

Before you begin:

These scripts were tailored for the analyses performed in:

Seibert et al, 2021, *Mild and severe SARS-CoV-2 infection induces respiratory and intestinal microbiome changes in the K18-hACE2 transgenic mouse model*

Purpose:

The purpose of this script is to analyze taxonomic abundances of the different groups within the ceca and the lungs. This will correspond to Figures 3A-D, 4, 7A-D, and 8

Load packages

```
library(phyloseq)
library(dplyr)
library(ggplot2)
library(RColorBrewer)
library(ggpubr)
library(forcats)
library(DESeq2)
library(microbiome)
library(Hmisc)
library(plyr)
```

Import the PhyloSeq object needed for this analysis from the alpha diversity analysis

```
# Load the file RarifiedASVs.rds
Phylo.samples.final <- readRDS(file =
"/Users/baseibert/Perez_Lab/Projects/Microbiome/Projects/K18_SARS/DataAnalysis/R_Files/Phylo.samples.fi
nal.rds")
```

Agglomerate at the phylum level

```
Phylum <- Phylo.samples.final %>%
  tax_glom(taxrank = "Phylum") %>%                    # agglomerate at the Phylum level
  transform_sample_counts(function(x) {x/sum(x)} ) %>%  # Transform to relative abundance
  psmelt() %>%                                        # Melt to long format (this is for
ggplot), create dataframe
  arrange(Phylum)                                     # Sort data frame alphabetically by class
```

Agglomerate at the family level

```
Family <- Phylo.samples.final %>%
  tax_glom(taxrank = "Family") %>%                    # agglomerate at the Family level
  transform_sample_counts(function(x) {x/sum(x)} ) %>%  # Transform to relative abundance
  psmelt() %>%                                        # Melt to long format (this is for
ggplot), create dataframe
  arrange(Family)                                     # Sort data frame alphabetically by class
```

Figure 3A: Taxonomic boxplots of the cecum at the Phylum level

Filter the Phylum into the cecum with the 3 groups that we will analyze

```
# Filter the samples so that you only have cecum samples and only PBS, 3 and 5 infected vehicele
Phylum.cecum <- Phylum %>%
  filter(SampleType == "Cecum") %>%
  filter(Group == "PBS" | Group == "Infected-3-Vehicle" | Group == "Infected-5-Vehicle")
```

First, lets plot all of the Phyla on a single bar graph.

```
Phylum.cecum.box <- Phylum.cecum

# Change phylum to a factor so I can change the order
```

```
Phylum.cecum.box$Phylum <- as.factor(Phylum.cecum.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.cecum.box$Abundance <- Phylum.cecum.box$Abundance*100

# Reorder the groups
Phylum.cecum.box$Group<- factor(Phylum.cecum.box$Group, levels = c('PBS', 'Infected-3-Vehicle',
'Infected-5-Vehicle'))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot using ggplot using boxplots
ggplot(data=Phylum.cecum.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Since there are many phyla that have extremely low abundance, we are going to group all of the taxa that have less than 0.1% relative abundance since these taxa are very rare and we are more interested in the more prominant taxa to see trends

```
# Filter for only the most abundant phylum. Adding too many phyla will cause the graph to look muddled
Phylum.cecum.box <- Phylum.cecum

# Group taxa that have overall abundance of less than 0.1%
Phylum.cecum.box$Phylum[Phylum.cecum.box$Abundance < 0.01] <- "Taxa < 1% abund."

# Change phylum to a factor so I can change the order
Phylum.cecum.box$Phylum <- as.factor(Phylum.cecum.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.cecum.box$Abundance <- Phylum.cecum.box$Abundance*100

# Reorder the groups
Phylum.cecum.box$Group<- factor(Phylum.cecum.box$Group, levels = c('PBS', 'Infected-3-Vehicle',
'Infected-5-Vehicle'))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot using ggplot using boxplots
ggplot(data=Phylum.cecum.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Since it is harder to see the less abundant taxa when all phyla are plotted in the same graph, lets split the phyla into 2 graphs.

First, lets calculate the abundances, plot the boxplots and perform statistical analysis for the more abundant phyla

```
# Filter for only the most abundant phylum. Adding too many phyla will cause the graph to look muddled
Phylum.cecum.box <- Phylum.cecum %>%
  filter(Phylum == "Bacteroidota" | Phylum == "Firmicutes" | Phylum == "Verrucomicrobiota")

# Change phylum to a factor so I can change the order
Phylum.cecum.box$Phylum <- as.factor(Phylum.cecum.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.cecum.box$Abundance <- Phylum.cecum.box$Abundance*100

# Reorder the groups
Phylum.cecum.box$Group<- factor(Phylum.cecum.box$Group, levels = c('PBS', 'Infected-3-Vehicle',
'Infected-5-Vehicle'))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot using ggplot using boxplots
```

```
ggplot(data=Phylum.cecum.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
Phylum.cecum.box.Special <- Phylum.cecum.box %>%
  filter(Phylum == "Bacteroidota")
compare_means(Abundance ~ Group, data = Phylum.cecum.box.Special, method = "wilcox.test")

Phylum.cecum.box.Special <- Phylum.cecum.box %>%
  filter(Phylum == "Firmicutes")
compare_means(Abundance ~ Group, data = Phylum.cecum.box.Special, method = "wilcox.test")

Phylum.cecum.box.Special <- Phylum.cecum.box %>%
  filter(Phylum == "Verrucomicrobiota")
compare_means(Abundance ~ Group, data = Phylum.cecum.box.Special, method = "wilcox.test")
```

Next, lets calculate the abundances, plot the boxplots and perform statistical analysis for the less abundant phyla

```
# Filter for only the most abundant phylum. Adding too many phyla will cause the graph to look muddled
Phylum.cecum.box <- Phylum.cecum %>%
  filter(Phylum == "Actinobacteriota" | Phylum == "Proteobacteria")

# Change phylum to a factor so I can change the order
Phylum.cecum.box$Phylum <- as.factor(Phylum.cecum.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.cecum.box$Abundance <- Phylum.cecum.box$Abundance*100

# Reorder the groups
Phylum.cecum.box$Group<- factor(Phylum.cecum.box$Group, levels = c('PBS', 'Infected-3-Vehicle',
'Infected-5-Vehicle'))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot using ggplot using boxplots
ggplot(data=Phylum.cecum.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 3, by=1), limits=c(0, 3))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
Phylum.cecum.box.Special <- Phylum.cecum.box %>%
  filter(Phylum == "Actinobacteriota")
compare_means(Abundance ~ Group, data = Phylum.cecum.box.Special, method = "wilcox.test")

Phylum.cecum.box.Special <- Phylum.cecum.box %>%
  filter(Phylum == "Proteobacteria")
compare_means(Abundance ~ Group, data = Phylum.cecum.box.Special, method = "wilcox.test")
```

Figure 3B: Firmicutes/Bacteroidetes ratio in the cecum

Lets analyze the Firmicutes/Bacteroidetes ratio as this can be use as an indicator of dysbiosis.

I tested calculating the ratio with raw counts and with relative abundances and the results were very similar. This script is for calculating the ration using relative abundances

```
# Create a data frame that only has the relative abundances of those classified as Firmicutes
Phylum.cecum_RA_Firm <- Phylum.cecum %>%
  filter(Phylum == "Firmicutes")

# Remove access columns that are not needed
# I will keep Abundance, SampleID, MouseID, Group and dpi
Phylum.cecum_RA_Firm.edit <- Phylum.cecum_RA_Firm[,c(3,4,8,9,10)]
```

```
# Change the first column header aka Abundance to Firmicutes
colnames(Phylum.cecum_RA_Firm.edit)[1] <- "Firmicutes"

# Export the data frame as a csv file
write.csv(Phylum.cecum_RA_Firm.edit, "Phylum.cecum_RA_Firm.csv")

# Create a data frame that only has the relative abundances of those classified as Bacteroidota
Phylum.cecum_RA_Bact <- Phylum.cecum %>%
  filter(Phylum == "Bacteroidota")

# Remove access columns that are not needed
# I will keep Abundance, SampleID, MouseID, Group and dpi
Phylum.cecum_RA_Bact.edit <- Phylum.cecum_RA_Bact[,c(3,4,8,9,10)]

# Change the first column header aka Abundance to Bacteroidota
colnames(Phylum.cecum_RA_Bact.edit)[1] <- "Bacteroidota"

# Export the data frame as a csv file
write.csv(Phylum.cecum_RA_Bact.edit, "Phylum.cecum_RA_Bact.csv")

# I calculated the ratios of the Firmicutes/Bacteroidota in excel and then imported it back into R
Phylum.cecum_RA_FBTOTAL <-
read.csv("/Users/baseibert/Perez_Lab/Projects/Microbiome/Projects/K18_SARS/DataAnalysis/R_Files/Phylum.
cecum_RA_FBRatio.csv")

# Make Group a factor and relevel
Phylum.cecum_RA_FBTOTAL$Group <- factor(Phylum.cecum_RA_FBTOTAL$Group, levels = c("PBS", "Infected-3-
Vehicle", "Infected-5-Vehicle"))

# Set the colors for the different groups
colorgroups = c("black","darkorange2", "blue")

# Plot the data with ggplot
ggplot(data=Phylum.cecum_RA_FBTOTAL, aes(x=FB_Ratio, y=Group, color=Group)) +
  geom_boxplot()+
   theme_bw()+
  scale_color_manual(values=colorgroups)+
  scale_x_continuous(breaks = seq(0, 4, by=2), limits=c(0, 4))+
  coord_flip()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
compare_means(FB_Ratio ~ Group, data = Phylum.cecum_RA_FBTOTAL, method = "wilcox.test")
```

Figure 3C: Taxonomic boxplots of the cecum at the Family level

Filter the Family into the cecum with the 3 groups that we will analyze. We will group the taxa that have less than 1% relative abundance since we are interested in overall trends and those sequences that have extremely small abundances could be sequencing artifacts.

```
# Filter the samples so that you only have cecum samples and only PBS, 3 and 5 infected vehicele
Family.cecum <- Family %>%
  filter(SampleType == "Cecum") %>%
 filter(Group == "PBS" | Group == "Infected-3-Vehicle" | Group == "Infected-5-Vehicle")
```

Lets look at all of the families in the cecum

```
# Filter for the most abundant families
Family.cecum.box <- Family.cecum

# Change Family to a factor so i can change the order
Family.cecum.box$Family <- as.factor(Family.cecum.box$Family)

# Multiply the Abundance by 100 to show an easier percentage
Family.cecum.box$Abundance <- Family.cecum.box$Abundance*100

# Make Group a factor and relevel
Family.cecum.box$Group <- factor(Family.cecum.box$Group, levels = c("PBS", "Infected-3-Vehicle",
"Infected-5-Vehicle"))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot the graph using ggplot
ggplot(data=Family.cecum.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
```

```
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Lets make a table of the prevalences by family
FamilyAverage <- Family.cecum.box %>%
    group_by(Family, Group) %>%
    summarise(num = n(),
              total_Abundance = mean(Abundance))
```

Since the abundances of all of the very rare families are making the trends of the bacterial families that are prominant hard to see, we are going to group the families that have less than 1% relative abundance in a single sample.

```
# I only want the families that have at least 1% relative abundance in a sample
Family.cecum_greater_1_abundance <- Family.cecum.box %>%
  filter(Family.cecum.box[,3] >= 1.0) #this is 1 since we already multiplied the abundances by 100 to
convert into %

unique(Family.cecum_greater_1_abundance$Family)

# Change phylum to a factor so I can change the order
Family.cecum_greater_1_abundance$Family <- as.factor(Family.cecum_greater_1_abundance$Family)

# Reorder the groups
Family.cecum_greater_1_abundance$Group<- factor(Family.cecum_greater_1_abundance$Group, levels =
c('PBS', 'Infected-3-Vehicle', 'Infected-5-Vehicle'))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot using ggplot using boxplots
ggplot(data=Family.cecum_greater_1_abundance, aes(x=fct_rev(Family), y=Abundance,
color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

All families appeared in more than 1 sample so they were not filtered out.

Next, lets graph the most abundant families separately so that we can have different y axis for the less abundant and see the relationships among rare taxa more clear.

```
# Filter for the most abundant families
Family.cecum.box <- Family.cecum %>%
  filter(Family == "Lachnospiraceae" | Family == "Muribaculaceae" | Family == "Akkermansiaceae" |
Family == "Bacteroidaceae" | Family == "Erysipelotrichaceae" | Family == "Oscillospiraceae")

# Relevel the families to be alphabetical
Family.cecum.box$Family <- factor(Family.cecum.box$Family, levels = c("Bacteroidaceae",
"Muribaculaceae", "Erysipelotrichaceae", "Lachnospiraceae", "Oscillospiraceae", "Akkermansiaceae"))

# Change Family to a factor so i can change the order
Family.cecum.box$Family <- as.factor(Family.cecum.box$Family)

# Multiply the Abundance by 100 to show an easier percentage
Family.cecum.box$Abundance <- Family.cecum.box$Abundance*100

# Make Group a factor and relevel
Family.cecum.box$Group <- factor(Family.cecum.box$Group, levels = c("PBS", "Infected-3-Vehicle",
"Infected-5-Vehicle"))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot the graph using ggplot
ggplot(data=Family.cecum.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
```

```
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all abundant families
Family.cecum.box.Special <- Family.cecum.box %>%
  filter(Family == "Bacteroidaceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")

Family.cecum.box.Special <- Family.cecum.box %>%
  filter(Family == "Muribaculaceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")

Family.cecum.box.Special <- Family.cecum.box %>%
  filter(Family == "Erysipelotrichaceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")

Family.cecum.box.Special <- Family.cecum.box %>%
  filter(Family == "Lachnospiraceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")

Family.cecum.box.Special <- Family.cecum.box %>%
  filter(Family == "Oscillospiraceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")

Family.cecum.box.Special <- Family.cecum.box %>%
  filter(Family == "Akkermansiaceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")
```

After, lets graph the less abundant families in their own graph so the trends are easier to view in the graph

```
# Filter for the less abundant families
Family.cecum.less.box <- Family.cecum %>%
  filter(Family == "Acholeplasmataceae" | Family == "Clostridiaceae" | Family == "Lactobacillaceae" |
Family == "Monoglobaceae" | Family == "Ruminococcaceae" | Family == "Sutterellaceae")

# Relevel the families so that they are grouped by phylum
Family.cecum.less.box$Family <- factor(Family.cecum.less.box$Family, levels = c("Acholeplasmataceae",
"Clostridiaceae","Lactobacillaceae", "Monoglobaceae", "Ruminococcaceae","Sutterellaceae"))

# Change family to a factor so I can change the order
Family.cecum.less.box$Family <- as.factor(Family.cecum.less.box$Family)

# Multiply the Abundance by 100 to show an easier percentage
Family.cecum.less.box$Abundance <- as.numeric(Family.cecum.less.box$Abundance)
Family.cecum.less.box$Abundance <- Family.cecum.less.box$Abundance*100

# Make Group a factor and relevel
Family.cecum.less.box$Group <- factor(Family.cecum.less.box$Group, levels = c("PBS", "Infected-3-
Vehicle", "Infected-5-Vehicle"))

# Set the colors for the different groups
colorgroups = c("blue","darkorange2", "black")

# Plot the graph using ggplot
ggplot(data=Family.cecum.less.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 12, by=2), limits=c(0, 12))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all less abundant families
Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Acholeplasmataceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")

Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Clostridiaceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")

Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Lactobacillaceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")
```

```
Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Monoglobaceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")

Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Ruminococcaceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")

Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Sutterellaceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")
```

Figure 3D: Correlation analysis of the cecum at the Family level

We will now perform a correlation analysis when analyzing the factors of infection to bacterial families

First, lets create a function to flatten a matrix as this will be needed for the correlation matrix

```
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor  =(cormat)[ut],
    p = pmat[ut]
    )
}
```

Next, we will perform the correlation analysis using rcorr from the package Hmisc

```
# Filter the samples so that you only have cecum samples and only PBS, 3 and 5 infected vehicle
cecum_added_factors <- read.csv("/Users/baseibert/Desktop/Added_factors_cecum.csv")
Cecum_total_factors <- merge(Family.cecum, cecum_added_factors, by = "SampleID")

Cecum_total_factors_edited <- Cecum_total_factors[,-
c(1,2,3,5,6,7,8,9,11,12,13,14,15,16,17,18,19,21,23)]

#Acholeplasmataceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Acholeplasmataceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter$row <- revalue(resultsfilter$row, c("Abundance"="Acholeplasmataceae"))

#Akkermansiaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Akkermansiaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Akkermansiaceae"))
resultsfilter_final <- rbind(resultsfilter, resultsfilter2)

#Bacteroidaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Bacteroidaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Bacteroidaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Clostridiaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Clostridiaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
```

```
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Clostridiaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Erysipelotrichaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Erysipelotrichaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Erysipelotrichaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Lachnospiraceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Lachnospiraceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Lachnospiraceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Lactobacillaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Lactobacillaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Lactobacillaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Monoglobaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Monoglobaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Monoglobaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Muribaculaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Muribaculaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Muribaculaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Oscillospiraceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Oscillospiraceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Oscillospiraceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Ruminococcaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Ruminococcaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
```

```
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Ruminococcaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

#Sutterellaceae
Cecum_total_factors_edited_filter <- Cecum_total_factors_edited %>%
  filter(Cecum_total_factors_edited[,3] == "Sutterellaceae")
Cecum_total_factors_edited_filter <- as.matrix(Cecum_total_factors_edited_filter[,-3])
res <- rcorr(Cecum_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter2$row <- revalue(resultsfilter2$row, c("Abundance"="Sutterellaceae"))
resultsfilter_final <- rbind(resultsfilter_final, resultsfilter2)

# Heatmap
resultsfilter_final$row <- factor(resultsfilter_final$row, levels = c("Bacteroidaceae",
"Muribaculaceae", "Acholeplasmataceae", "Clostridiaceae", "Erysipelotrichaceae", "Lachnospiraceae",
"Lactobacillaceae", "Monoglobaceae", "Oscillospiraceae", "Ruminococcaceae", "Sutterellaceae",
"Akkermansiaceae"))

resultsfilter_final$column <- factor(resultsfilter_final$column, levels = c("Inoculum", "dpi.x",
"Acitivity", "Viral.NT.Titer", "Viral.Lung.Titer", "Viral.Brain.Titer"))

ggplot(resultsfilter_final, aes(column, fct_rev(row))) +
  geom_point(aes(colour = cor),
                  size =8)+
  scale_color_gradient2(limits = c(-0.7, 0.7), breaks = c(-0.6,-0.3,0,0.3,0.6),
                        low = "royalblue1",
                        mid = "brown3",
                        high = "yellow")+
  theme_bw()

# Produce a table with those correlation results that have are significant or have a p value < 0.05
resultsfilter_final_significant <- resultsfilter_final %>%
  filter(resultsfilter_final[,4] < 0.05)
```

The figure was further edited in Adobe Illustrator to indicate the relationships that were considered statistically significant.

Figure 4: Taxonomic barplot of the cecum at the Family level

Now lets produce a graph for the taxonomic barplot

```
# Combine the Taxa that are less than 1% relative abundance together
Family.cecum.barplot <- Family.cecum
Family.cecum.barplot$Family[Family.cecum.barplot$Abundance < 0.01] <- "Taxa < 1% abund."

ColorTest = c("gray31", "yellow","forestgreen", "midnightblue", "blue", "steelblue3", "lightblue2",
"royalblue1", "cyan", "mediumblue",  "deepskyblue1", "darkorchid4", "orchid1")

Family.cecum.barplot$MouseID <- factor(Family.cecum.barplot$MouseID, levels = c("5", "6", "23", "24",
"25", "34", "35", "36", "3", "4", "20", "21", "22", "31", "32", "33", "1", "2", "15", "17", "19"))

Family.cecum.barplot$Family <- factor(Family.cecum.barplot$Family, levels = c("Muribaculaceae",
"Bacteroidaceae", "Lachnospiraceae", "Oscillospiraceae", "Erysipelotrichaceae","Acholeplasmataceae",
"Lactobacillaceae", "Ruminococcaceae", "Clostridiaceae", "Monoglobaceae", "Sutterellaceae",
"Akkermansiaceae", "Taxa < 1% abund."))

#plot using ggplot
ggplot(data=Family.cecum.barplot, aes(x=MouseID, y=Abundance, fill=fct_rev(Family))) +
  geom_bar(aes(), colour = "black", stat="identity", position="stack", width = 0.8) +
  scale_fill_manual(values=ColorTest)+
  facet_grid(~dpi, scales = "free", space = "free")+
  theme(legend.position="bottom") +
  guides(fill=guide_legend(nrow=5))+
  theme(legend.position="right") + guides(fill=guide_legend(ncol =1))+
  theme_bw()+
  theme(panel.grid = element_blank())+
  theme(strip.text = element_text(size=11, face="bold"))
```

The figure is then combined with activity score and viral titers in Adobe Illustrator

Figure 7A: Taxonomic boxplots of the lung at the Phylum level

Filter the Phylum into the cecum with the 3 groups that we will analyze

```
# Filter the samples so that you only have cecum samples and only PBS, 3 and 5 infected vehicele
Phylum.lung.GC376 <- Phylum %>%
  filter(SampleType == "Lung") %>%
  filter(Group == "Mock-GC376" | Group == "Infected-3-GC376" | Group == "Infected-5-GC376")
```

First, lets plot all of the Phyla on a single bar graph.

```
Phylum.lung.box <- Phylum.lung.GC376

# Change phylum to a factor so I can change the order
Phylum.lung.box$Phylum <- as.factor(Phylum.lung.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.lung.box$Abundance <- Phylum.lung.box$Abundance*100

# Reorder the groups
Phylum.lung.box$Group<- factor(Phylum.lung.box$Group, levels = c("Mock-GC376", "Infected-3-GC376",
"Infected-5-GC376"))

# Set the colors for the different groups
colorgroups = c("maroon1","forestgreen","chocolate4")

# Plot using ggplot using boxplots
ggplot(data=Phylum.lung.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Like the cecum, there are many phyla that have extremely low abundance, we are going to group all of the taxa that have less than 0.1% relative abundance since these taxa are very rare and we are more interested in the more prominant taxa to see trends

```
# Filter for only the most abundant phylum as seen in the bargraph above. Adding too many phyla will
cause the graph to look muddled
Phylum.lung.GC376.box <- Phylum.lung.GC376

# Group taxa that have overall abundance of less than 1%
Phylum.lung.GC376.box$Phylum[Phylum.lung.GC376.box$Abundance < 0.01] <- "Taxa < 1% abund."

# Change phylum to a factor so i can change the order
Phylum.lung.GC376.box$Phylum <- as.factor(Phylum.lung.GC376.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.lung.GC376.box$Abundance <- Phylum.lung.GC376.box$Abundance*100

# Relevel the groups
Phylum.lung.GC376.box$Group <- factor(Phylum.lung.GC376.box$Group, levels = c("Mock-GC376", "Infected-
3-GC376", "Infected-5-GC376"))

# Set the colors for the different groups
colorgroups = c("maroon1","forestgreen","chocolate4")

# Graph the plot using ggplot
ggplot(data=Phylum.lung.GC376.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Since it is harder to see the less abundant taxa when all phyla are plotted in the same graph, lets split the phyla into 2 graphs.

First, lets calculate the abundances, plot the boxplots and perform statistical analysis for the more abundant phyla

```
# Filter for only the most abundant phylum as seen in the bargraph above. Adding too many phyla will
cause the graph to look muddled
Phylum.lung.GC376.box <- Phylum.lung.GC376 %>%
  filter(Phylum == "Firmicutes" | Phylum == "Bacteroidota" | Phylum == "Proteobacteria")

# Change phylum to a factor so i can change the order
Phylum.lung.GC376.box$Phylum <- as.factor(Phylum.lung.GC376.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.lung.GC376.box$Abundance <- Phylum.lung.GC376.box$Abundance*100

# Relevel the groups
Phylum.lung.GC376.box$Group <- factor(Phylum.lung.GC376.box$Group, levels = c("Mock-GC376", "Infected-
3-GC376", "Infected-5-GC376"))

# Set the colors for the different groups
colorgroups = c("maroon1","forestgreen","chocolate4")

# Graph the plot using ggplot
ggplot(data=Phylum.lung.GC376.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
Phylum.lung.GC376.box.Special <- Phylum.lung.GC376.box %>%
  filter(Phylum == "Firmicutes")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")

Phylum.lung.GC376.box.Special <- Phylum.lung.GC376.box %>%
  filter(Phylum == "Bacteroidota")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")

Phylum.lung.GC376.box.Special <- Phylum.lung.GC376.box %>%
  filter(Phylum == "Proteobacteria")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")
```

Next, lets calculate the abundances, plot the boxplots and perform statistical analysis for the less abundant phyla

```
# Filter for only the most abundant phylum as seen in the bargraph above. Adding too many phyla will
cause the graph to look muddled
Phylum.lung.GC376.box <- Phylum.lung.GC376 %>%
  filter(Phylum == "Actinobacteriota" | Phylum == "Deinococcota"| Phylum == "Verrucomicrobiota")

# Change phylum to a factor so i can change the order
Phylum.lung.GC376.box$Phylum <- as.factor(Phylum.lung.GC376.box$Phylum)

# Multiply the Abundance by 100 to show an easier percentage
Phylum.lung.GC376.box$Abundance <- Phylum.lung.GC376.box$Abundance*100

# Relevel the groups
Phylum.lung.GC376.box$Group <- factor(Phylum.lung.GC376.box$Group, levels = c("Mock-GC376", "Infected-
3-GC376", "Infected-5-GC376"))

# Set the colors for the different groups
colorgroups = c("maroon1","forestgreen","chocolate4")

# Graph the plot using ggplot
ggplot(data=Phylum.lung.GC376.box, aes(x=fct_rev(Phylum), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 15, by=3), limits=c(0, 15))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
Phylum.lung.GC376.box.Special <- Phylum.lung.GC376.box %>%
  filter(Phylum == "Actinobacteriota")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")
```

```
Phylum.lung.GC376.box.Special <- Phylum.lung.GC376.box %>%
  filter(Phylum == "Deinococcota")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")

Phylum.lung.GC376.box.Special <- Phylum.lung.GC376.box %>%
  filter(Phylum == "Verrucomicrobiota")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")
```

Figure 7B: Firmicutes/Bacteroidetes ration in the lung

Lets analyze the Firmicutes/Bacteroidetes ratio as this can be use as an indicator of dysbiosis.

I tested calculating the ratio with raw counts and with relative abundances and the results were very similar. This script is for calculating the ration using relative abundances

```
# Create a data frame that only has the relative abundances of those classified as Firmicutes
Phylum.lung_RA_Firm <- Phylum.lung.GC376 %>%
  filter(Phylum == "Firmicutes")

# Remove access columns that are not needed
# I will keep Abundance, SampleID, MouseID, Group and dpi
Phylum.lung_RA_Firm.edit <- Phylum.lung_RA_Firm[,c(3,4,8,9,10)]

# Change the first column header aka Abundance to Firmicutes
colnames(Phylum.lung_RA_Firm.edit)[1] <- "Firmicutes"

# Export the data frame as a csv file
write.csv(Phylum.lung_RA_Firm.edit, "Phylum.lung_RA_Firm.csv")

# Create a data frame that only has the relative abundances of those classified as Bacteroidota
Phylum.lung_RA_Bact <- Phylum.lung.GC376 %>%
  filter(Phylum == "Bacteroidota")

# Remove access columns that are not needed
# I will keep Abundance, SampleID, MouseID, Group and dpi
Phylum.lung_RA_Bact.edit <- Phylum.lung_RA_Bact[,c(3,4,8,9,10)]

# Change the first column header aka Abundance to Bacteroidota
colnames(Phylum.lung_RA_Bact.edit)[1] <- "Bacteroidota"

# Export the data frame as a csv file
write.csv(Phylum.lung_RA_Bact.edit, "Phylum.lung_RA_Bact.csv")

# I calculated the ratios of the Firmicutes/Bacteroidota in excel and then imported it back into R
Phylum.lung_RA_FBTOTAL <-
read.csv("/Users/baseibert/Perez_Lab/Projects/Microbiome/Projects/K18_SARS/DataAnalysis/R_Files/Phylum.
lung.RA_FBRatio.csv")

# Make Group a factor
Phylum.lung_RA_FBTOTAL$Group <- factor(Phylum.lung_RA_FBTOTAL$Group, levels = c("Mock-GC376",
"Infected-3-GC376", "Infected-5-GC376"))

# Set the colors for the different groups
colorgroups = c("chocolate4","forestgreen","maroon1" )

# Graph the plot using ggplot
ggplot(data=Phylum.lung_RA_FBTOTAL, aes(x=FB_Ratio, y=Group, color=Group)) +
  geom_boxplot()+
   theme_bw()+
  scale_color_manual(values=colorgroups)+
  scale_x_continuous(breaks = seq(0, 6, by=2), limits=c(0, 6))+
  coord_flip()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
compare_means(FB_Ratio ~ Group, data = Phylum.lung_RA_FBTOTAL, method = "wilcox.test")
```

There is one outlier that is not shown in the Infected-5-GC376 group that is not shown (FB Ratio = 18)

Fig 7C: Taxonomic boxplots of the lung at the Family level

Filter the Family into the lung with the 3 groups that we will analyze. We will group the taxa that have less than 2% relative abundance since we are interested in overall trends and there were more families with lower relative abundances in the lungs compared to the ceca

```
# Filter the samples so that you only have lung samples and only GC376
Family.lung.GC376 <- Family %>%
  filter(SampleType == "Lung") %>%
  filter(Group == "Mock-GC376" | Group == "Infected-3-GC376" | Group == "Infected-5-GC376")

write.csv(Family.lung.GC376,  "Family_lung_GC376.csv")
```

Lets look at all of the families in the lungs

```
# Filter the most abundant bacteria
Family.lung.GC376.box <- Family.lung.GC376

# Multipy abundances by 100
Family.lung.GC376.box$Abundance <- as.numeric(Family.lung.GC376.box$Abundance)
Family.lung.GC376.box$Abundance <- Family.lung.GC376.box$Abundance*100

# Assign the colors for the graphs
colorgroups = c("maroon1","forestgreen","chocolate4")

# Relevel the groups
Family.lung.GC376.box$Group <- factor(Family.lung.GC376.box$Group, levels = c("Mock-GC376", "Infected-
3-GC376", "Infected-5-GC376"))

# Graph plot using ggplot
ggplot(data=Family.lung.GC376.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Since the abundances of all of the very rare families are making the trends of the bacterial families that are prominant hard to see, we are going to group the families that have less than 1% relative abundance in a single sample.

```
# I only want the families that have at least 1% relative abundance in a sample
Family.lung_greater_1_abundance <- Family.lung.GC376.box %>%
  filter(Family.lung.GC376.box[,3] >= 1.0) #this is 1 since we already multiplied the abundances by 100
to convert into %

unique(Family.lung_greater_1_abundance$Family)

# Change phylum to a factor so I can change the order
Family.lung_greater_1_abundance$Family <- as.factor(Family.lung_greater_1_abundance$Family)

# Reorder the groups
Family.lung_greater_1_abundance$Group<- factor(Family.lung_greater_1_abundance$Group, levels = c("Mock-
GC376", "Infected-3-GC376", "Infected-5-GC376"))

# Set the colors for the different groups
colorgroups = c("maroon1","forestgreen","chocolate4")

# Plot using ggplot using boxplots
ggplot(data=Family.lung_greater_1_abundance, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group)))
+
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Then, I filtered out families that only appeared in a single sample (most likely contamination) Blattabacteriaceae, Caulobacteraceae, Halomonadaceae, Hydrogenophilaceae, Microbacteriaceae, Pasteurellaceae

Next, I will graph the most abundant families separately so that we can have different y axis for the less abundant and see the relationships among rare taxa more clear.

```
# Filter the most abundant bacteria
Family.lung.GC376.box <- Family.lung.GC376 %>%
  filter(Family == "Muribaculaceae" | Family == "Lachnospiraceae")

# Order the families
Family.lung.GC376.box$Family <- factor(Family.lung.GC376.box$Family, levels = c("Muribaculaceae",
"Lachnospiraceae"))

# Multipy abundances by 100
Family.lung.GC376.box$Abundance <- as.numeric(Family.lung.GC376.box$Abundance)
Family.lung.GC376.box$Abundance <- Family.lung.GC376.box$Abundance*100

# Assign the colors for the graphs
colorgroups = c("maroon1","forestgreen","chocolate4")

# Relevel the groups
Family.lung.GC376.box$Group <- factor(Family.lung.GC376.box$Group, levels = c("Mock-GC376", "Infected-
3-GC376", "Infected-5-GC376"))

# Graph plot using ggplot
ggplot(data=Family.lung.GC376.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 100, by=25), limits=c(0, 100))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Perform statistical
Family.lung.GC376.box.Special <- Family.lung.GC376.box %>%
  filter(Family == "Muribaculaceae")
compare_means(Abundance ~ Group, data = Family.lung.GC376.box.Special, method = "wilcox.test")

Family.lung.GC376.box.Special <- Family.lung.GC376.box %>%
  filter(Family == "Lachnospiraceae")
compare_means(Abundance ~ Group, data = Family.lung.GC376.box.Special, method = "wilcox.test")
```

Now, I will graph the less abundant families

```
# Filter the families that have lower abundance
Family.lung.GC376.less.box <- Family.lung.GC376 %>%
  filter(Family == "Acholeplasmataceae" | Family == "Aerococcaceae" | Family == "Akkermansiaceae" |
Family == "Bacteroidaceae" | Family == "Beijerinckiaceae" | Family == "Burkholderiaceae"| Family ==
"Butyricicoccaceae" | Family == "Caulobacteraceae" | Family == "Clostridiaceae" | Family ==
"Corynebacteriaceae" | Family == "Eggerthellaceae" | Family == "Enterobacteriaceae" | Family ==
"Enterococcaceae" | Family == "Erysipelatoclostridiaceae" | Family == "Erysipelotrichaceae" | Family ==
"Lactobacillaceae" | Family == "Leuconostocaceae" |  Family =="Microbacteriaceae" |  Family ==
"Moraxellaceae" | Family == "Morganellaceae" | Family == "Oscillospiraceae" | Family ==
"Peptococcaceae" | Family == "Peptostreptococcaceae" | Family == "Pseudomonadaceae" | Family ==
"Rhodanobacteraceae" | Family == "Ruminococcaceae" | Family == "Sphingomonadaceae" | Family ==
"Staphylococcaceae" | Family == "Sutterellaceae" | Family == "Thermaceae" | Family == "Weeksellaceae" |
Family == "Xanthomonadaceae")

# Relevel the families so that they are grouped by phylum
Family.lung.GC376.less.box$Family <- factor(Family.lung.GC376.less.box$Family, levels =
c("Corynebacteriaceae","Eggerthellaceae", "Microbacteriaceae",
"Bacteroidaceae", "Weeksellaceae",
"Thermaceae",
"Acholeplasmataceae", "Aerococcaceae", "Butyricicoccaceae","Clostridiaceae", "Enterococcaceae",
"Erysipelatoclostridiaceae","Erysipelotrichaceae", "Lactobacillaceae",
"Leuconostocaceae","Oscillospiraceae", "Peptococcaceae", "Peptostreptococcaceae", "Ruminococcaceae",
"Staphylococcaceae",

"Akkermansiaceae",

"Beijerinckiaceae", "Burkholderiaceae", "Caulobacteraceae","Enterobacteriaceae", "Moraxellaceae",
"Morganellaceae","Pseudomonadaceae","Rhodanobacteraceae", "Sphingomonadaceae", "Sutterellaceae",
"Xanthomonadaceae"))

# Multipy abundances by 100
Family.lung.GC376.less.box$Abundance <- as.numeric(Family.lung.GC376.less.box$Abundance)
Family.lung.GC376.less.box$Abundance <- Family.lung.GC376.less.box$Abundance*100

# Assign the colors for the graphs
colorgroups = c("maroon1","forestgreen","chocolate4")

# Relevel the groups
```

```
Family.lung.GC376.less.box$Group <- factor(Family.lung.GC376.less.box$Group, levels = c("Mock-GC376",
"Infected-3-GC376", "Infected-5-GC376"))

# Graph the plot using ggplot
ggplot(data=Family.lung.GC376.less.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 22, by=5), limits=c(0, 22))+
  theme_bw()+
  theme(panel.grid = element_blank())
```

Since there are still many families to place in the graph and it is hard to see. I will split the graphs into 2 separate graphs.

```
# Filter the families that have lower abundance
Family.lung.GC376.less.box <- Family.lung.GC376 %>%
  filter(Family == "Acholeplasmataceae" | Family == "Aerococcaceae" | Family == "Bacteroidaceae" |
Family == "Butyricicoccaceae" | Family == "Clostridiaceae" | Family == "Enterococcaceae" | Family ==
"Erysipelatoclostridiaceae" | Family == "Erysipelotrichaceae" | Family == "Lactobacillaceae" | Family
== "Leuconostocaceae" | Family == "Oscillospiraceae" | Family == "Peptococcaceae" | Family ==
"Peptostreptococcaceae" | Family == "Ruminococcaceae" | Family == "Staphylococcaceae" | Family ==
"Weeksellaceae")

# Relevel the families so that they are grouped by phylum
Family.lung.GC376.less.box$Family <- factor(Family.lung.GC376.less.box$Family, levels =
c("Bacteroidaceae", "Weeksellaceae",
"Acholeplasmataceae", "Aerococcaceae", "Butyricicoccaceae","Clostridiaceae", "Enterococcaceae",
"Erysipelatoclostridiaceae","Erysipelotrichaceae", "Lactobacillaceae",
"Leuconostocaceae","Oscillospiraceae", "Peptococcaceae", "Peptostreptococcaceae", "Ruminococcaceae",
"Staphylococcaceae"))

# Multipy abundances by 100
Family.lung.GC376.less.box$Abundance <- as.numeric(Family.lung.GC376.less.box$Abundance)
Family.lung.GC376.less.box$Abundance <- Family.lung.GC376.less.box$Abundance*100

# Assign the colors for the graphs
colorgroups = c("maroon1","forestgreen","chocolate4")

# Relevel the groups
Family.lung.GC376.less.box$Group <- factor(Family.lung.GC376.less.box$Group, levels = c("Mock-GC376",
"Infected-3-GC376", "Infected-5-GC376"))

# Graph the plot using ggplot
ggplot(data=Family.lung.GC376.less.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 22, by=5), limits=c(0, 22))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
Family.lung.GC376.less.box.Special <- Family.lung.GC376.less.box %>%
  filter(Family == "Staphylococcaceae")
compare_means(Abundance ~ Group, data = Family.lung.GC376.less.box.Special, method = "wilcox.test")

#-------------------------------------------------------------------------------------------------------

# Filter the families that have lower abundance
Family.lung.GC376.less.box <- Family.lung.GC376 %>%
  filter(Family =="Akkermansiaceae" | Family == "Beijerinckiaceae" | Family == "Burkholderiaceae" |
Family == "Caulobacteraceae" | Family == "Corynebacteriaceae" | Family == "Enterobacteriaceae" | Family
== "Eggerthellaceae" | Family =="Microbacteriaceae" | Family == "Moraxellaceae" | Family ==
"Morganellaceae" | Family == "Pseudomonadaceae" | Family == "Rhodanobacteraceae" | Family ==
"Sphingomonadaceae" | Family == "Sutterellaceae" | Family == "Thermaceae" | Family ==
"Xanthomonadaceae")

# Relevel the families so that they are grouped by phylum
Family.lung.GC376.less.box$Family <- factor(Family.lung.GC376.less.box$Family, levels =
c("Corynebacteriaceae","Eggerthellaceae", "Microbacteriaceae",
"Thermaceae",
"Akkermansiaceae",
"Beijerinckiaceae", "Burkholderiaceae", "Caulobacteraceae","Enterobacteriaceae", "Moraxellaceae",
"Morganellaceae","Pseudomonadaceae","Rhodanobacteraceae", "Sphingomonadaceae", "Sutterellaceae",
"Xanthomonadaceae"))
```

```
# Multipy abundances by 100
Family.lung.GC376.less.box$Abundance <- as.numeric(Family.lung.GC376.less.box$Abundance)
Family.lung.GC376.less.box$Abundance <- Family.lung.GC376.less.box$Abundance*100

# Assign the colors for the graphs
colorgroups = c("maroon1","forestgreen","chocolate4")

# Relevel the groups
Family.lung.GC376.less.box$Group <- factor(Family.lung.GC376.less.box$Group, levels = c("Mock-GC376",
"Infected-3-GC376", "Infected-5-GC376"))

# Graph the plot using ggplot
ggplot(data=Family.lung.GC376.less.box, aes(x=fct_rev(Family), y=Abundance, color=fct_rev(Group))) +
  geom_boxplot()+
  scale_color_manual(values=colorgroups)+
  coord_flip()+
  scale_y_continuous(breaks = seq(0, 22, by=5), limits=c(0, 22))+
  theme_bw()+
  theme(panel.grid = element_blank())

# Statistical Comparison using wilcox test using specific bacteria. I need to manually enter the
bacteria name. It gets too confusing when i add multiple lines with all phyla
Family.lung.GC376.less.box.Special <- Family.lung.GC376.less.box %>%
  filter(Family == "Xanthomonadaceae")
compare_means(Abundance ~ Group, data = Family.lung.GC376.less.box.Special, method = "wilcox.test")
```

One outlier in low group for Lactobacillaceae (~35%) and High dose Staphyloccoccacae (40%)

Figure 7D: Correlation analysis of the lungs at the Family level

We will now perform a correlation analysis when analyzing the factors of infection to bacterial families

First, lets create a function to flatten a matrix as this will be needed for the correlation matrix

```
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor  =(cormat)[ut],
    p = pmat[ut]
    )
}
```

Next, we will perform the correlation analysis using rcorr from the package Hmisc

```
# Filter the samples so that you only have cecum samples and only PBS, 3 and 5 infected vehicele
lung_added_factors <- read.csv("/Users/baseibert/Desktop/Added_factors_lung.csv")
lung_total_factors <- merge(Family.lung.GC376, lung_added_factors, by = "SampleID")

lung_total_factors_edited <- lung_total_factors[,-c(1,2,3,5,6,7,8,9,11,12,13,14,15,16,17,18,19,21,23)]

#Corynebacteriaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Corynebacteriaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung$row <- revalue(resultsfilter_Lung$row, c("Abundance"="Corynebacteriaceae"))

#Eggerthellaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Eggerthellaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Eggerthellaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_Lung)
```

```r
#Microbacteriaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Microbacteriaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Microbacteriaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Bacteroidaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Bacteroidaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Bacteroidaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Muribaculaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Muribaculaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Muribaculaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Weeksellaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Weeksellaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Weeksellaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Thermaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Thermaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Thermaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Acholeplasmataceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Acholeplasmataceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Acholeplasmataceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Aerococcaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Aerococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Aerococcaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Butyricicoccaceae
```

```
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Butyricicoccaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Butyricicoccaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Clostridiaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Clostridiaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Clostridiaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Enterococcaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Enterococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Enterococcaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Erysipelatoclostridiaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Erysipelatoclostridiaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Erysipelatoclostridiaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Erysipelotrichaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Erysipelotrichaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Erysipelotrichaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Lachnospiraceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Lachnospiraceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Lachnospiraceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Lactobacillaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Lactobacillaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Lactobacillaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Leuconostocaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
```

```
  filter(lung_total_factors_edited[,3] == "Leuconostocaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Leuconostocaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Oscillospiraceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Oscillospiraceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Oscillospiraceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Peptococcaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Peptococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Peptococcaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Peptostreptococcaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Peptostreptococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Peptostreptococcaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Ruminococcaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Ruminococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Ruminococcaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Staphylococcaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Staphylococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Staphylococcaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Beijerinckiaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Beijerinckiaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Beijerinckiaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Burkholderiaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Burkholderiaceae")
```

```
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Burkholderiaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Caulobacteraceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Caulobacteraceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Caulobacteraceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Enterobacteriaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Enterobacteriaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Enterobacteriaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Moraxellaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Enterococcaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Moraxellaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Morganellaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Morganellaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Morganellaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Pseudomonadaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Pseudomonadaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Pseudomonadaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Rhodanobacteraceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Rhodanobacteraceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Rhodanobacteraceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Sphingomonadaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Sphingomonadaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
```

```
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Sphingomonadaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Sutterellaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Sutterellaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Sutterellaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Xanthobacteraceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Xanthobacteraceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Xanthobacteraceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

#Akkermansiaceae
Lung_total_factors_edited_filter <- lung_total_factors_edited %>%
  filter(lung_total_factors_edited[,3] == "Akkermansiaceae")
Lung_total_factors_edited_filter <- as.matrix(Lung_total_factors_edited_filter[,-3])
res <- rcorr(Lung_total_factors_edited_filter, type = "spearman")
results <- flattenCorrMatrix(res$r, res$P)
resultsfilter_Lung2 <- results %>%
  filter(results[,1] == "Abundance")
resultsfilter_Lung2$row <- revalue(resultsfilter_Lung2$row, c("Abundance"="Akkermansiaceae"))
resultsfilter_final <- rbind(resultsfilter_Lung2, resultsfilter_final)

# Heatmap
resultsfilter_final$row <- factor(resultsfilter_final$row, levels = c("Corynebacteriaceae",
"Eggerthellaceae", "Microbacteriaceae", "Bacteroidaceae", "Muribaculaceae", "Weeksellaceae",
"Thermaceae", "Acholeplasmataceae", "Aerococcaceae", "Butyricicoccaceae", "Clostridiaceae",
"Enterococcaceae", "Erysipelatoclostridiaceae", "Erysipelotrichaceae", "Lachnospiraceae",
"Lactobacillaceae", "Leuconostocaceae", "Oscillospiraceae", "Peptococcaceae", "Peptostreptococcaceae",
"Ruminococcaceae", "Staphylococcaceae", "Beijerinckiaceae", "Burkholderiaceae", "Caulobacteraceae",
"Enterobacteriaceae", "Moraxellaceae", "Morganellaceae", "Pseudomonadaceae", "Rhodanobacteraceae",
"Sphingomonadaceae", "Sutterellaceae", "Xanthobacteraceae", "Akkermansiaceae"))

resultsfilter_final$column <- factor(resultsfilter_final$column, levels = c("Inoculum", "dpi.x",
"Acitivity", "Viral.NT.Titer", "Viral.Lung.Titer", "Viral.Brain.Titer"))

ggplot(resultsfilter_final, aes(row, fct_rev(column))) +
  #geom_tile(aes(fill = p))+
  #scale_fill_gradient(low = "snow",
  #                    high = "snow")+
  geom_point(aes(colour = cor),
                     size =8)+
  scale_color_gradient2(limits = c(-0.7, 0.7), breaks = c(-0.6,-0.3,0,0.3,0.6),
                        low = "royalblue1",
                        mid = "brown3",
                        high = "yellow")+
  theme_bw()

resultsfilter_final_significant <- resultsfilter_final %>%
  filter(resultsfilter_final[,4] < 0.05)
```

The figure was further edited in Adobe Illustrator to indicate the relationships that were considered statistically significant.

Figure 8: Taxonomic barplot of the lungs at the Family level

Now lets produce a graph for the taxonomic barplot

```
# Combine the Taxa that are less than 1% relative abundance together
Family.lung.barplot <- Family.lung.GC376
```

```
Family.lung.barplot$Family[Family.lung.barplot$Abundance < 0.02] <- "Taxa < 2% abund."

# Set the MouseID to a factor and assign the order
Family.lung.barplot$MouseID <- factor(Family.lung.barplot$MouseID, levels = c("13", "14", "45", "46",
"47", "56", "57", "58", "11", "12", "42", "43", "44", "53", "54", "55", "7", "8", "9", "37", "38",
"39", "51"))

# Set the level of the families
Family.lung.barplot$Family <- factor(Family.lung.barplot$Family, levels = c("Eggerthellaceae",
"Corynebacteriaceae","Muribaculaceae", "Bacteroidaceae", "Weeksellaceae","Lachnospiraceae",
"Staphylococcaceae", "Erysipelatoclostridiaceae", "Peptostreptococcaceae", "Clostridiaceae",
"Ruminococcaceae", "Oscillospiraceae", "Erysipelotrichaceae", "Lactobacillaceae", "Leuconostocaceae",
"Enterococcaceae", "Aerococcaceae", "Butyricicoccaceae","Pseudomonadaceae", "Enterobacteriaceae",
"Burkholderiaceae", "Xanthomonadaceae", "Caulobacteraceae", "Sphingomonadaceae", "Moraxellaceae",
"Akkermansiaceae", "Taxa < 2% abund."))

# Set color palette to accommodate the number of families
ColorTest = c("gray31", "yellow",

"springgreen4","olivedrab1","olivedrab4","yellowgreen","darkolivegreen","chartreuse","forestgreen",
              "dodgerblue","lightblue3","navy","cornflowerblue",
              "lightblue2","cyan","mediumblue","steelblue3","blue",      #these are same colors taken
from ceca
              "dodgerblue3","skyblue2","royalblue1",
              "deepskyblue1",                #these are same colors taken from ceca
              "purple1",
              "darkorchid4", "orchid1",      #these are same colors taken from ceca
              "darkorange3","sandybrown")

# Plot using ggplot
ggplot(data=Family.lung.barplot, aes(x=MouseID, y=Abundance, fill=fct_rev(Family))) +
  geom_bar(aes(), colour = "black", stat="identity", position="stack", width = 0.8) +
  scale_fill_manual(values=ColorTest)+
  facet_grid(~dpi, scales = "free",space = "free")+
  theme(legend.position="bottom") +
  guides(fill=guide_legend(nrow=5))+
  theme(legend.position="right") + guides(fill=guide_legend(ncol =1))+
  theme_bw()+
  theme(panel.grid = element_blank())+
  theme(strip.text = element_text(size=11, face="bold"))
```

The figure is then combined with activity score and viral titers in Adobe Illustrator