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 and 6A-D

Load packages

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

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

```
# Load the file RarifiedASVs.rds
Phylo.samples.final <- readRDS(file = "/Users/Phylo.samples.final.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")
```

Calculate the abundances, plot the boxplots and perform statistical analysis

```
# 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" | 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, 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")

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 ration 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.NoRare.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/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","blue", "darkorange2")
```

```
# 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")

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

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 for the most abundant families
Family.cecum.box <- Family.cecum %>%
  filter(Family == "Akkermansiaceae" | Family == "Lachnospiraceae" | Family == "Muribaculaceae")

# Relevel the families to be alphabetical
Family.cecum.box$Family <- factor(Family.cecum.box$Family, levels = c("Muribaculaceae",
"Lachnospiraceae", "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 == "Akkermansiaceae")
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 == "Muribaculaceae")
compare_means(Abundance ~ Group, data = Family.cecum.box.Special, method = "wilcox.test")
```

Now, I will graph the less abundant families. I will only show families.

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

# Relevel the families so that they are grouped by phylum
Family.cecum.less.box$Family <- factor(Family.cecum.less.box$Family, levels = c("Bacteroidaceae",
"Acholeplasmataceae", "Lactobacillaceae","Clostridiaceae","Erysipelotrichaceae",
"Monoglobaceae","Oscillospiraceae", "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, 15, by=5), 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 less abundant families
Family.cecum.less.box.Special <- Family.cecum.less.box %>%
  filter(Family == "Bacteroidaceae")
compare_means(Abundance ~ Group, data = Family.cecum.less.box.Special, method = "wilcox.test")

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 == "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 == "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 == "Erysipelotrichaceae")
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 == "Oscillospiraceae")
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: Taxonomic barplot of the cecum at the Family level

Now lets produce a graph for the taxonomic barplot

```
# Set color palette to accommodate the number of class
ColorTest = c("antiquewhite4","slategray2","navy", "bisque2", "red3","plum1","yellow","forestgreen",
"cyan","deeppink", "darkorchid4", "sandybrown")

# Set mouseID as a factor and order based on dpc
Phylum.NoRare.cecum$MouseID <- factor(Phylum.NoRare.cecum$MouseID, levels = c("5", "6", "23", "24",
"25", "34", "35", "36", "3", "4", "20", "21", "22", "31", "32", "33", "1", "2", "15", "17", "19"))

# Plot the graph using ggplot
ggplot(data=Phylum.NoRare.cecum, aes(x=MouseID, y=Abundance, fill=fct_reorder(Phylum, Abundance))) +
  geom_bar(aes(), colour = "black", stat="identity", position="stack", width = 0.9) +
  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"))
```

Figure 6A: 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")

# 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 == "Verrucomicrobiota" | Phylum == "Firmicutes" | Phylum == "Proteobacteria" | Phylum
== "Bacteroidota"| Phylum == "Actinobacteriota")

# 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 == "Verrucomicrobiota")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")

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 == "Proteobacteria")
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 == "Actinobacteriota")
compare_means(Abundance ~ Group, data = Phylum.lung.GC376.box.Special, method = "wilcox.test")
```

Figure 6B: 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.noRare.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.noRare.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/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("black","blue", "darkorange2")

# 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")
```

Fig 6C: 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 those sequences that have extremely small abundances could be sequencing artifacts.

```
# 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")

# Group taxa that have overall abundance of less than 1%
Family.lung.GC376$Family[Family.lung.GC376$Abundance < 0.02] <- "Taxa < 2% abund."
```

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" |  Family == "Staphylococcaceae")

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

# 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 == "Staphylococcaceae")
compare_means(Abundance ~ Group, data = Family.lung.GC376.box.Special, method = "wilcox.test")

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

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

Now, I will graph the less abundant families. I will only show families.

```
# Filter the families that have lower abundance
Family.lung.GC376.less.box <- Family.lung.GC376 %>%
  filter(Family == "Bacteroidaceae" | Family == "Peptostreptococcaceae" | Family == "Clostridiaceae" |
Family == "Ruminococcaceae" | Family == "Oscillospiraceae" | Family == "Erysipelotrichaceae" | Family
== "Lactobacillaceae" | Family == "Enterococcaceae" | Family == "Aerococcaceae" | Family ==
"Pseudomonadaceae" | Family == "Enterobacteriaceae" | Family == "Moraxellaceae" | Family ==
"Akkermansiaceae")

# Order the families
Family.lung.GC376.less.box$Family <- factor(Family.lung.GC376.less.box$Family, levels =
c("Bacteroidaceae", "Peptostreptococcaceae", "Clostridiaceae", "Ruminococcaceae", "Oscillospiraceae",
"Erysipelotrichaceae", "Lactobacillaceae", "Enterococcaceae", "Aerococcaceae", "Pseudomonadaceae",
"Enterobacteriaceae", "Moraxellaceae", "Akkermansiaceae"))

# 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, 15, by=5), 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
Family.lung.GC376.less.box.Special <- Family.lung.GC376.less.box %>%
  filter(Family == "Bacteroidaceae")
compare_means(Abundance ~ Group, data = Family.lung.GC376.less.box.Special, method = "wilcox.test")

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

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

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

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

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

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

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

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

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

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

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

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

Figure 6D: Taxonomic barplot of the lungs at the Family level

Now lets produce a graph for the taxonomic barplot

```
# Set the MouseID to a factor and assign the order
Family.lung.GC376$MouseID <- factor(Family.lung.GC376$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.GC376$Family <- factor(Family.lung.GC376$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.GC376, aes(x=MouseID, y=Abundance, fill=fct_rev(Family))) +
  geom_bar(aes(), colour = "black", stat="identity", position="stack", width = 0.85) +
  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"))
```