AlphaDiversity_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 alpha diversity of the different groups within the ceca and the lungs. This will correspond to Figures 2A-D, 4A-C, 4F-H, 5A-D, and Supplemental Figures S1-4.

Load the needed packages

```
library(vegan)
library(phyloseq)
library(ggplot2)
library(ggpubr)
library(dplyr)
library(microbiome)
```

Load the R file from previous data processing

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

Before we perform alpha diveristy analysis we will need to rarify the data.

We will rarify the data using the vegan package

Before we rarify the data in vegan, I will convert the phyloseq object to a dataframe that is worke-able in vegan. We will do this by creating a function

```
# Create a function to convert the sample_data() within a phyloseq object to a vegan compatible data object
pssd2veg <- function(physeq) {
  sd <- sample_data(physeq)
  return(as(sd,"data.frame"))
}

# Use the function to create a dataframe of the sample data
sampledataveg<- pssd2veg(Phylo.samples.final)

# Create a function to convert the count table within a phyloseq object to a vegan compatible data object
psotu2veg <- function(physeq) {
  OTU <- otu_table(physeq)
  if (taxa_are_rows(OTU)) {
    OTU <- t(OTU)
  }
  return(as(OTU, "matrix"))
}

# Use the function to create a matrix of the ASV table. This will be a matrix that has ASV number as column headers and sample IDs as row names
ASVtableveg <- psotu2veg(Phylo.samples.final)
```

Rarify the SAMPLES in vegan

```
# Calculate total OTU counts for all samples (sum all columns)
total.counts <- apply(ASVtableveg, 1, sum)



# List all sample sums in descending order - useful for deciding what your sample cutoff should be
sort(total.counts, decreasing=TRUE)
```

```
# List the minimum coverage (sequences per sample)
min(total.counts)

# Change as appropriate for samples
sample.size <- 12000

# Subset sample.shared to remove samples with < sample.size sequences
sample.used <- ASVtableveg[total.counts>=sample.size,]

# Resample all samples to a total of sample.size sequences (I do not want to rarify again so i will
load sample.rare from rarified data previously analyzed)
sample.rare <- rrarefy(sample.used, sample=sample.size)

# Save sample.rare to a file so that i can reload the rarified file
saveRDS(object = sample.rare, file = "RarifiedASVs.rds")

# Load the file RarifiedASVs.rds
sample.rare <- readRDS(file = "/Users/RarifiedASVs.rds")
```

Figure 2A-C: Alpha and beta diversity of metrics of ceca samples

Alpha diversity

Estimating alpha diversity of microbial communities is problematic no matter what you do. The best is to subsample the libraries with replacement to estimate the species abundance of the real population while standardizing sampling effort.

Import the rarified data into phyloseq since that is how we are going to calculate the alpha diversity metrics

```
# Import rarified count into phyloseq
psrare <- otu_table(sample.rare, taxa_are_rows = FALSE)

# Import taxonomy file
taxa <- as.matrix(read.table("/Users/baseibert/asv_tax_no_contam_470.tsv", header=T, row.names=1,
check.names=F, sep="\t"))

# Import taxonomy table into phyloseq
tax = tax_table(taxa)

# Import metadata file
metadata <- read.csv("/Users/Sars_metadata2.csv", header=T, row.names=1, check.names=F)

# Import the sample metadata into phyloseq object
sample = sample_data(metadata)

# Merge the count table, taxonomy table and sample data
Phylo.rare  = merge_phyloseq(psrare, tax, sample)
Phylo.rare
```

Then, subset the phyloseq object into cecum samples that include only the 3 groups: PBS, Infected-3-Vehicle and Infected-5-Vehicle

```
# First filter only the cecum samples
Cecum.rare <- subset_samples(Phylo.rare, SampleType == "Cecum")
Cecum.rare

# Then filter based on groups
Cecum.group1 <- subset_samples(Cecum.rare, Group == "PBS" | Group == "Infected-3-Vehicle" | Group ==
"Infected-5-Vehicle")
Cecum.group1
```

Now, calculate the alpha diversity metrics and plot the graphs in ggplot

```
# Calculate diversity metrics in phyloseq
PhlyoRichness.cecum <- estimate_richness(Cecum.group1, measures = c("Observed", "Shannon",
"InvSimpson", "Fisher"))

# Make a column that contains the rownames so that we have a column of the SampleIDs
PhlyoRichness.cecum$SampleID <- rownames(PhlyoRichness.cecum)

# Join the diversity tables with metadata
PhlyoRichness.cecum.Meta = left_join(PhlyoRichness.cecum, metadata, by = "SampleID")

PhlyoRichness.cecum.Meta$Group<- factor(PhlyoRichness.cecum.Meta$Group, levels = c('PBS','Infected-3-
```

```
Vehicle', 'Infected-5-Vehicle'))

# Set the colors for all of the graphs
colorgroups = c("black","darkorange2", "blue")

# GRAPH OBSERVED ASVs

# Filter so that you only have observed values
PhlyoRichness.cecum.Meta.observed <- PhlyoRichness.cecum.Meta[,-c(2,3,4)]

# Make a boxplot of the observed values
ggplot(PhlyoRichness.cecum.Meta.observed, aes(x=Group, y=Observed, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(50, 300, by=50), limits=c(50, 300))+
  xlab("Group")+
  ggtitle("Observed ASVs")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Observed ~ Group,  data = PhlyoRichness.cecum.Meta.observed)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.cecum.Meta.shannon <- PhlyoRichness.cecum.Meta[,-c(1,3,4)]

# Make a boxplot of the shannon diversity values
ggplot(PhlyoRichness.cecum.Meta.shannon, aes(x=Group, y=Shannon, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(1, 7, by=1), limits=c(1, 7))+
  xlab("Group")+
  ggtitle("Shannon Diversity")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Shannon ~ Group,  data = PhlyoRichness.cecum.Meta.shannon)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.cecum.Meta.invsimp <- PhlyoRichness.cecum.Meta[,-c(1,2,4)]

# Make a boxplot of the inverse simpson diversity values
ggplot(PhlyoRichness.cecum.Meta.invsimp, aes(x=Group, y=InvSimpson, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(limits = c(0,40))+
  xlab("Group")+
  ggtitle("Inv Simpson")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(InvSimpson ~ Group,  data = PhlyoRichness.cecum.Meta.invsimp)
```

Figure 1S: Alpha diversity metrics of ceca across time

Lets look at the alpha diversity across groups and dpc. Since the mouse number is limited, we do not have enough statistical power to compare the groups across dpc. We will use this to look at trends.

```
# Set dpi as a factor with specific levels to determine the order
PhlyoRichness.cecum.Meta$dpi<- factor(PhlyoRichness.cecum.Meta$dpi, levels = c('2', '5', '14'))

# Set the colors for all of the graphs
colorgroups = c("black","darkorange2", "blue")

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.cecum.Meta.observed <- PhlyoRichness.cecum.Meta[,-c(2,3,4)]

# Plot individual groups across dpc but cannot do statistical tests since there are too few samples
(n=3)
ggplot(PhlyoRichness.cecum.Meta.observed, aes(x=dpi, y=Observed, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(50, 250, by=50), limits=c(50, 250))+
  ggtitle("Observed ASVs")+
  xlab("Days post challenge")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.cecum.Meta.shannon <- PhlyoRichness.cecum.Meta[,-c(1,3,4)]

# Plot individual groups across dpc but cannot do statistical tests since there are too few samples
(n=3)
ggplot(PhlyoRichness.cecum.Meta.shannon, aes(x=dpi, y=Shannon, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(1, 5, by=1), limits=c(1, 5))+
  ggtitle("Shannon Index")+
  xlab("Days post challenge")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# GRAPH Inv Simpson

# Filter so that you oly have shannon values
PhlyoRichness.cecum.Meta.invsimp <- PhlyoRichness.cecum.Meta[,-c(1,2,4)]

# Plot individual groups across dpc but cannot do statistical tests since there are too few samples
(n=3)
ggplot(PhlyoRichness.cecum.Meta.invsimp, aes(x=dpi, y=InvSimpson, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(limits = c(0,30))+
  ggtitle("Inv Simpson")+
  xlab("Days post challenge")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)
```

Figure 2S: Alpha diversity of cecum samples among time

Now lets look at the graphs across dpc of the infected groups. Therefore, I can perform statistical test with higher power.

Then, subset the phyloseq object into cecum samples that include only the infected groups: Infected-3-Vehicle and Infected-5-Vehicle

```
Cecum.group1.inf <- subset_samples(Cecum.group1, Group == "Infected-3-Vehicle" | Group == "Infected-5-Vehicle")
Cecum.group1.inf
```

Now, calculate the alpha diversity metrics and plot the graphs in ggplot

```
# Calculate diversity metrics in phyloseq for only infected groups
PhlyoRichness.cecum.inf <- estimate_richness(Cecum.group1.inf, measures = c("Observed", "Shannon", "InvSimpson", "Fisher"))

# Make a column that contains the rownames so that we have a column of the SampleIDs
PhlyoRichness.cecum.inf$SampleID <- rownames(PhlyoRichness.cecum.inf)

# Join the diversity tables with metadata
PhlyoRichness.cecum.inf.Meta = left_join(PhlyoRichness.cecum.inf, metadata, by = "SampleID")

PhlyoRichness.cecum.inf.Meta$dpi<- factor(PhlyoRichness.cecum.inf.Meta$dpi, levels = c('2', '5', '14'))

PhlyoRichness.cecum.inf.Meta$Group<- factor(PhlyoRichness.cecum.inf.Meta$Group, levels = c('PBS', 'Infected-3-Vehicle', 'Infected-5-Vehicle'))

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.cecum.inf.Meta.observed <- PhlyoRichness.cecum.inf.Meta[,-c(2,3,4)]

# Make a boxplot of the observed ASVS values
ggplot(PhlyoRichness.cecum.inf.Meta.observed, aes(x=dpi, y=Observed))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(50, 300, by=50), limits=c(50, 300))+
  ggtitle("Observed ASVs")+
  xlab("Days post challenge")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the Wilcoxon Sign test (non-parametric))
compare_means(Observed ~ dpi,  data = PhlyoRichness.cecum.inf.Meta.observed)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.cecum.Meta.shannon.inf <- PhlyoRichness.cecum.inf.Meta[,-c(1,3,4)]

# Make a boxplot of the shannon diversity values
ggplot(PhlyoRichness.cecum.Meta.shannon.inf, aes(x=dpi, y=Shannon))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(1, 7, by=1), limits=c(1, 7))+
  ggtitle("Shannon Diversity")+
  xlab("Days post challenge")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
```

```
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this
uses the Wilcoxon Sign test (non-parametric))
compare_means(Shannon ~ dpi,  data = PhlyoRichness.cecum.Meta.shannon.inf)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.cecum.Meta.invsimp.inf <- PhlyoRichness.cecum.inf.Meta[,-c(1,2,4)]

# Make a boxplot of the inverse simpson diversity values
ggplot(PhlyoRichness.cecum.Meta.invsimp.inf, aes(x=dpi, y=InvSimpson))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(limits = c(0,40))+
  ggtitle("Inv Simpson")+
  xlab("Days post challenge")+
   ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this
uses the Wilcoxon Sign test (non-parametric))
compare_means(InvSimpson ~ dpi,  data = PhlyoRichness.cecum.Meta.invsimp.inf)
```

Figure 4A-C and F-H: Alpha and Beta Diversity metrics of lung samples without/with antiviral GC-376

I would like to make alpha diversity graphs for Vehicle versus GC-376 for lungs to see if there is a difference among the two groups and if there is an antiviral effect

Figure 4A-C: Low dose Then, subset the phyloseq object into cecum samples that include only the 3 groups: PBS, Infected-3-Vehicle and Infected-5-Vehicle

```
# First subset the samples so it only includes lung samples
Lung.rare <- subset_samples(Phylo.rare, SampleType == "Lung")
Lung.rare

# Then subset the samples to include only the low dose vehicle and antiviral GC-376
Lung.group.3 <- subset_samples(Lung.rare, Group == "Infected-3-Vehicle" | Group == "Infected-3-GC376")
Lung.group.3
```

Now, calculate the alpha diversity metrics and plot the graphs in ggplot

```
# Calculate diversity metrics in phyloseq
PhlyoRichness.lungs.GC376 <- estimate_richness(Lung.group.3, measures = c("Observed", "Shannon",
"InvSimpson", "Fisher"))

# Make a column that contains the rownames so that we have a column of the SampleIDs
PhlyoRichness.lungs.GC376$SampleID <- rownames(PhlyoRichness.lungs.GC376)

# Join the diversity tables with metadata
PhlyoRichness.lungs.GC376.Meta = left_join(PhlyoRichness.lungs.GC376, metadata, by = "SampleID")

# Set Group to a factor with levels
PhlyoRichness.lungs.GC376.Meta$Group<- factor(PhlyoRichness.lungs.GC376.Meta$Group, levels =
c('Infected-3-Vehicle', 'Infected-3-GC376'))

# Set the colors for the graphs
colorgroups = c("darkorange2","forestgreen")

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.lungs.GC376.Meta.observed <- PhlyoRichness.lungs.GC376.Meta[,-c(2,3,4)]
```

```
# Plot the graph of observed ASVs
ggplot(PhlyoRichness.lungs.GC376.Meta.observed, aes(x=Group, y=Observed, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 300, by=50), limits=c(0, 300))+
  ggtitle("Observed ASVs")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Observed ~ Group,  data = PhlyoRichness.lungs.GC376.Meta.observed)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.shannon <- PhlyoRichness.lungs.GC376.Meta[,-c(1,3,4)]

# Plot the boxplot with shannon diversity
ggplot(PhlyoRichness.lungs.Meta.shannon, aes(x=Group, y=Shannon, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 6, by=1), limits=c(0, 6))+
  ggtitle("Shannon Diversity")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Shannon ~ Group,  data = PhlyoRichness.lungs.Meta.shannon)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.invsimp <- PhlyoRichness.lungs.GC376.Meta[,-c(1,2,4)]

# Plot the boxplot with inv simpson
ggplot(PhlyoRichness.lungs.Meta.invsimp, aes(x=Group, y=InvSimpson, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 50, by=10), limits=c(0, 50))+
  ggtitle("Inv Simpson")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(InvSimpson ~ Group,  data = PhlyoRichness.lungs.Meta.invsimp)
```

Figure 4F-H: High dose Then, subset the phyloseq object into cecum samples that include only the 3 groups: PBS, Infected-3-Vehicle and Infected-5-Vehicle

```
# Subset the samples to include only the low dose vehicle and antiviral GC-376
Lung.group.5 <- subset_samples(Lung.rare, Group == "Infected-5-Vehicle" | Group == "Infected-5-GC376")
Lung.group.5
```

Now, calculate the alpha diversity metrics and plot the graphs in ggplot

```
# Calculate diversity metrics in phyloseq
PhlyoRichness.lungs.GC376 <- estimate_richness(Lung.group.5, measures = c("Observed", "Shannon",
"InvSimpson", "Fisher"))
```

```r
# Make a column that contains the rownames so that we have a column of the SampleIDs
PhlyoRichness.lungs.GC376$SampleID <- rownames(PhlyoRichness.lungs.GC376)

# Join the diversity tables with metadata
PhlyoRichness.lungs.GC376.Meta = left_join(PhlyoRichness.lungs.GC376, metadata, by = "SampleID")

# Set Group to a factor with levels
PhlyoRichness.lungs.GC376.Meta$Group<- factor(PhlyoRichness.lungs.GC376.Meta$Group, levels = c('PBS',
'Mock-GC376', 'Infected-3-Vehicle', 'Infected-3-GC376', 'Infected-5-Vehicle', 'Infected-5-GC376'))

# Set the colors for the graphs
colorgroups = c("blue", "maroon1")

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.lungs.GC376.Meta.observed <- PhlyoRichness.lungs.GC376.Meta[,-c(2,3,4)]

# Plot the boxplot of the Observed ASVs
ggplot(PhlyoRichness.lungs.GC376.Meta.observed, aes(x=Group, y=Observed, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 300, by=50), limits=c(0, 300))+
  ggtitle("Observed ASVs")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Observed ~ Group,  data = PhlyoRichness.lungs.GC376.Meta.observed)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.shannon <- PhlyoRichness.lungs.GC376.Meta[,-c(1,3,4)]

# Plot the boxplot of Shannon diversity values
ggplot(PhlyoRichness.lungs.Meta.shannon, aes(x=Group, y=Shannon, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 6, by=1), limits=c(0, 6))+
  ggtitle("Shannon Diversity")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Shannon ~ Group,  data = PhlyoRichness.lungs.Meta.shannon)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.invsimp <- PhlyoRichness.lungs.GC376.Meta[,-c(1,2,4)]

# Plot the boxplot of the inv simpson values
ggplot(PhlyoRichness.lungs.Meta.invsimp, aes(x=Group, y=InvSimpson, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 50, by=10), limits=c(0, 50))+
  ggtitle("Inv Simpson")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
```

```
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(InvSimpson ~ Group,  data = PhlyoRichness.lungs.Meta.invsimp)
```

Figure 5A-C: Alpha and beta diversity metrics of lung samples

Then, subset the phyloseq object into cecum samples that include only the 3 groups: PBS, Infected-3-Vehicle and Infected-5-Vehicle

```
# Filter the lungs with only Mock, low and high infected antiviral groups
Lung.group.GC376 <- subset_samples(Lung.rare, Group == "Mock-GC376" | Group == "Infected-3-GC376" |
Group == "Infected-5-GC376")
Lung.group.GC376
```

Now, calculate the alpha diversity metrics and plot the graphs in ggplot

```
# Calculate diversity metrics in phyloseq
PhlyoRichness.lungs.GC376 <- estimate_richness(Lung.group.GC376, measures = c("Observed", "Shannon",
"InvSimpson", "Fisher"))

# Make a column that contains the rownames so that we have a column of the SampleIDs
PhlyoRichness.lungs.GC376$SampleID <- rownames(PhlyoRichness.lungs.GC376)

# Join the diversity tables with metadata
PhlyoRichness.lungs.GC376.Meta = left_join(PhlyoRichness.lungs.GC376, metadata, by = "SampleID")

# Convert Group into a factor and define the levels so that we can order it
PhlyoRichness.lungs.GC376.Meta$Group<- factor(PhlyoRichness.lungs.GC376.Meta$Group, levels = c('Mock-
GC376','Infected-3-GC376', 'Infected-5-GC376'))

# Define the colorgroups for all graphs
colorgroups = c("chocolate4","forestgreen","maroon1")

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.lungs.GC376.Meta.observed <- PhlyoRichness.lungs.GC376.Meta[,-c(2,3,4)]

# Graph the boxplot for the Observed ASVs
ggplot(PhlyoRichness.lungs.GC376.Meta.observed, aes(x=Group, y=Observed, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 300, by=50), limits=c(0, 300))+
  ggtitle("Observed ASVs")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Observed ~ Group,  data = PhlyoRichness.lungs.GC376.Meta.observed)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.shannon <- PhlyoRichness.lungs.GC376.Meta[,-c(1,3,4)]

# Plot the boxplot for Shannon diversity
ggplot(PhlyoRichness.lungs.Meta.shannon, aes(x=Group, y=Shannon, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 6, by=1), limits=c(0, 6))+
  ggtitle("Shannon Diversity")+
```

```
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Shannon ~ Group,  data = PhlyoRichness.lungs.Meta.shannon)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.invsimp <- PhlyoRichness.lungs.GC376.Meta[,-c(1,2,4)]

# Plot the boxplot for the inv simpson
ggplot(PhlyoRichness.lungs.Meta.invsimp, aes(x=Group, y=InvSimpson, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 50, by=10), limits=c(0, 50))+
  ggtitle("Inv Simpson")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(InvSimpson ~ Group,  data = PhlyoRichness.lungs.Meta.invsimp)
```

Figure S3: Alpha diversity metrics of lung samples across time

Lets look at the alpha diversity across groups and dpc. Since the mouse number is limited, we do not have enough statistical power to compare the groups across dpc. We will use this to look at trends.

```
# Set dpi as a factor with specific levels to determine the order
PhlyoRichness.lungs.GC376.Meta$dpi<- factor(PhlyoRichness.lungs.GC376.Meta$dpi, levels = c('2', '5',
'14'))

# Define the colorgroups for all graphs
colorgroups = c("chocolate4","forestgreen","maroon1")

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.lungs.GC376.Meta.observed <- PhlyoRichness.lungs.GC376.Meta[,-c(2,3,4)]

# Plot different groups across dpc but cannot do statistical tests since we only have 3 samples / group
ggplot(PhlyoRichness.lungs.GC376.Meta.observed, aes(x=dpi, y=Observed, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  ylab("Alpha Diversity Index")+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 250, by=50), limits=c(0, 250))+
  ggtitle("Observed ASVs")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.shannon <- PhlyoRichness.lungs.GC376.Meta[,-c(1,3,4)]

# Plot different groups across dpc but cannot do statistical tests since we only have 3 samples / group
ggplot(PhlyoRichness.lungs.Meta.shannon, aes(x=dpi, y=Shannon, color=Group))+
  geom_boxplot()+
  theme_bw()+
```

```
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 5, by=1), limits=c(0, 5))+
  ggtitle("Shannon Diversity")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.invsimp <- PhlyoRichness.lungs.GC376.Meta[,-c(1,2,4)]

# Plot different groups across dpc but cannot do statistical tests since we only have 3 samples / group
ggplot(PhlyoRichness.lungs.Meta.invsimp, aes(x=dpi, y=InvSimpson, color=Group))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  scale_color_manual(values = colorgroups)+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 40, by=10), limits=c(0, 40))+
  ggtitle("Inv Simpson")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  #stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)
```

Figure 4S: Alpha diversity metrics of lung samples across time

Now lets look at the graphs with only dpc of the infected groups. Therefore, I can perform statistical test with higher power.

Then, subset the phyloseq object into cecum samples that include only the infected groups: Infected-3-Vehicle and Infected-5-Vehicle

```
Lung.group.GC376.inf <- subset_samples(Lung.group.GC376, Group == "Infected-3-GC376" | Group == "Infected-5-GC376")
Lung.group.GC376.inf
```

Now, calculate the alpha diversity metrics and plot the graphs in ggplot

```
# Calculate diversity metrics in phyloseq for only infected groups
PhlyoRichness.lungs.GC376.inf <- estimate_richness(Lung.group.GC376.inf, measures = c("Observed", "Shannon", "InvSimpson", "Fisher"))

# Make a column that contains the rownames so that we have a column of the SampleIDs
PhlyoRichness.lungs.GC376.inf$SampleID <- rownames(PhlyoRichness.lungs.GC376.inf)

# Join the diversity tables with metadata
PhlyoRichness.lungs.GC376.Meta.inf = left_join(PhlyoRichness.lungs.GC376.inf, metadata, by = "SampleID")

# Set the dpi as a factor and set the levels to make sure dpc are in order
PhlyoRichness.lungs.GC376.Meta.inf$dpi<- factor(PhlyoRichness.lungs.GC376.Meta.inf$dpi, levels = c('2', '5', '14'))

# GRAPH OBSERVED ASVs

# Filter so that you oly have observed values
PhlyoRichness.lungs.GC376.Meta.inf.observed <- PhlyoRichness.lungs.GC376.Meta.inf[,-c(2,3,4)]

# Plot the boxplot of observed ASVs
ggplot(PhlyoRichness.lungs.GC376.Meta.inf.observed, aes(x=dpi, y=Observed))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 300, by=50), limits=c(0, 300))+
  ggtitle("Observed ASVs")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
```

```
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means
# Have to group the dpc together since there are too little samples to test individually per sample
compare_means(Observed ~ dpi,  data = PhlyoRichness.lungs.GC376.Meta.inf.observed)

# GRAPH SHANNON

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.shannon.inf <- PhlyoRichness.lungs.GC376.Meta.inf[,-c(1,3,4)]

# Plot the boxplot of the Shannon diversity values
ggplot(PhlyoRichness.lungs.Meta.shannon.inf, aes(x=dpi, y=Shannon))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 6, by=1), limits=c(0, 6))+
  ggtitle("Shannon Diversity")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(Shannon ~ dpi,  data = PhlyoRichness.lungs.Meta.shannon.inf)

# GRAPH InvSimpson

# Filter so that you oly have shannon values
PhlyoRichness.lungs.Meta.invsimp.inf <- PhlyoRichness.lungs.GC376.Meta.inf[,-c(1,2,4)]

# Plot the boxplot of the inverse simpson
ggplot(PhlyoRichness.lungs.Meta.invsimp.inf, aes(x=dpi, y=InvSimpson))+
  geom_boxplot()+
  theme_bw()+
  theme(panel.grid = element_blank())+
  theme(axis.title.y = element_text(vjust=2))+
  scale_y_continuous(breaks = seq(0, 50, by=10), limits=c(0, 50))+
  ggtitle("Inv Simpson")+
  xlab("Days post infection")+
  ylab("Alpha Diversity Index")+
  theme(panel.spacing = unit(2, "lines"))+
  stat_compare_means(method = "kruskal.test")+
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, alpha = 0.3)

# Statistics testing for comparison among means of the bray curtis for different groups (this uses the
Wilcoxon Sign test (non-parametric))
compare_means(InvSimpson ~ dpi,  data = PhlyoRichness.lungs.Meta.invsimp.inf)
```