Rmarkdown: Dietary carotenoid supplementation has long-term and community-wide effects on the amphibian skin microbiome

Alice Risely

29/3/23

Contents

Load packages

```
library(phyloseq)
library(ggplot2)
library(vegan)
library(dplyr)
library(scales)
library(grid)
library(reshape2)
library(ape)
library(gridExtra)
library(ade4)
library(plyr)
library(tidyr)
library(data.table)
library(stringr)
library(ggrepel)
library(r2glmm)
library(ggvenn)
library(viridis)
library(ggord)


library(GGally)
library(sjPlot)
library(performance)

library(ggpubr)
library(microbiome)
library(RColorBrewer)
library(gllvm)
library(EnvStats)
library(ggcorrplot)
```

Prevalence function

Load data

```
setwd("C:/Users/risel/Dropbox/Academic projects/Frog microbiome UOW/Frogs_UOW/Diet treatment project/Ana

frog_filtered <- readRDS("frog_final.RDS")
```

Create transformed datasets

```
frog_rare<-rarefy_even_depth(frog_filtered, sample.size = 6700)
```

Bar plots

```
frog_phylum<- tax_glom(frog_filtered, taxrank = "Phylum")
frog_phylum<- aggregate_top_taxa(frog_filtered, top = 10, level = "Phylum")
frog_phylum<- microbiome::transform(frog_phylum, "compositional")


### order

frog_order<- tax_glom(frog_filtered, taxrank = "Order")

frog_order<- microbiome::transform(frog_order, "compositional")
frog_order<-microbiome::core(frog_order, detection = 0, prevalence = 0.9)


##############################


ps1.com.fam <- microbiome::aggregate_top_taxa(frog_filtered, "Order", top = 22)

ps1.com.fam<- microbiome::transform(ps1.com.fam, "compositional")

plot.composition.relAbun <- microbiome::plot_composition(ps1.com.fam,
                                            sample.sort = "Treatment",
                                            x.label = "frog_id",
                                            group_by = "Treatment")


data.com <- plot.composition.relAbun$data
colnames(data.com)
```

```
## [1] "Tax"       "Sample"    "Abundance" "Group"     "xlabel"
```

```
data.com$Tax<-ifelse(data.com$Tax == "Fimbriimonadales" | data.com$Tax == "Propionibacteriales" |data.c


unique(data.com$Tax)
```

```
##  [1] "Aeromonadales"         "Azospirillales"        "Bacillales"
##  [4] "Bacteroidales"         "Betaproteobacteriales" "Chitinophagales"
##  [7] "Clostridiales"         "Corynebacteriales"     "Cytophagales"
## [10] "Enterobacteriales"     "Other"                 "Flavobacteriales"
## [13] "Fusobacteriales"       "Lactobacillales"       "Micrococcales"
## [16] "Pseudomonadales"       "Rhizobiales"           "Sphingobacteriales"
## [19] "Sphingomonadales"      "Xanthomonadales"
```

```
data.com$Tax<-factor(data.com$Tax, level = c("Flavobacteriales",
                                             "Bacteroidales",
                                             "Sphingobacteriales",
                                             "Chitinophagales",
                                             "Cytophagales",
```

```
                                                "Enterobacteriales",
                                                "Pseudomonadales",
                                                "Aeromonadales",
                                                "Xanthomonadales",
                                                "Betaproteobacteriales",
                                                "Rhizobiales",
                                                "Azospirillales",
                                                "Sphingomonadales",

                                                "Corynebacteriales",
                                                "Micrococcales",

                                                "Clostridiales",
                                                "Bacillales",
                                                "Lactobacillales",

                                                "Fusobacteriales",
                                                "Other"))


colors <- c("lightskyblue","skyblue4", "royalblue","darkslategray4", "cyan",

            "#E7E1EF", "#D4B9DA", "#C994C7", "#DF65B0" ,"#E7298A" ,"#CE1256", "#980043", "#67001F",

            "forestgreen","olivedrab2",

            "yellow","chocolate","bisque",

            "navy",

            "white")

barplot_order<-ggplot(data.com, aes(x = Sample, y = Abundance, fill = Tax))+
  geom_bar(position = "stack", stat = "identity", width = 1, col = "black")+
  scale_x_discrete(labels = data.com$xlabel, breaks = data.com$Sample)+
  facet_grid(~Group, scales = "free") + theme_bw()+
  scale_fill_manual(values = colors)+
   theme_bw(base_size = 14)+
  theme(axis.text.x = element_blank())+
  labs(fill = "Bacterial order")+
  ylab("Relative abundance")+
    theme(legend.key.size = unit(0.4, 'cm'))+
  scale_y_continuous(expand = c(0, 0), limits = c(0, 1))

#########################
#########################
#########################
#########################
#########################

plot.composition.relAbun <- microbiome::plot_composition(ps1.com.fam,
                                                sample.sort = "Sex",
                                                x.label = "frog_id",
```

```
                                          group_by = "Sex")


data.com <- plot.composition.relAbun$data
colnames(data.com)
```

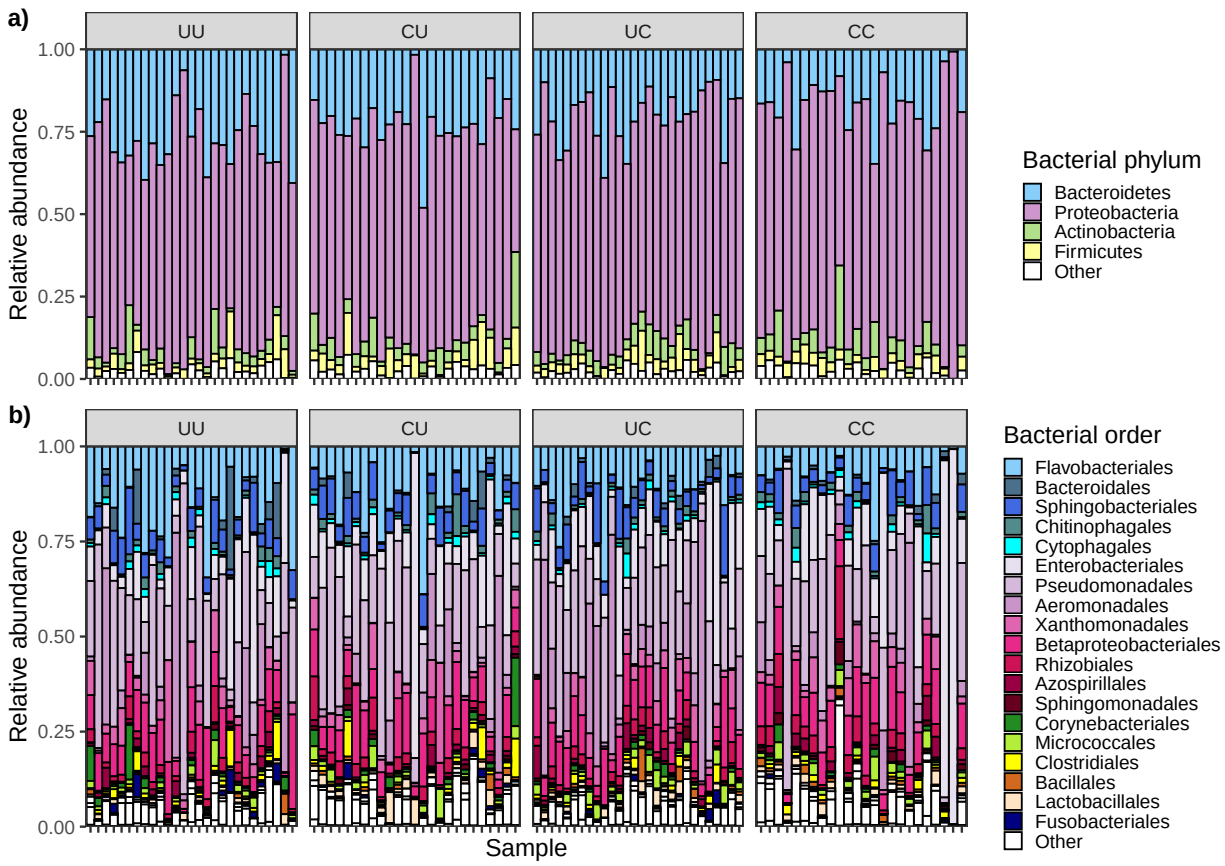
```
## [1] "Tax"       "Sample"    "Abundance" "Group"     "xlabel"
```

```
data.com$Tax<-ifelse(data.com$Tax == "Fimbriimonadales" | data.com$Tax == "Propionibacteriales" |data.co


unique(data.com$Tax)
```

```
##  [1] "Aeromonadales"         "Azospirillales"        "Bacillales"
##  [4] "Bacteroidales"         "Betaproteobacteriales" "Chitinophagales"
##  [7] "Clostridiales"         "Corynebacteriales"     "Cytophagales"
## [10] "Enterobacteriales"     "Other"                 "Flavobacteriales"
## [13] "Fusobacteriales"       "Lactobacillales"       "Micrococcales"
## [16] "Pseudomonadales"       "Rhizobiales"           "Sphingobacteriales"
## [19] "Sphingomonadales"      "Xanthomonadales"
```

```
data.com$Tax<-factor(data.com$Tax, level = c("Flavobacteriales",
                                             "Bacteroidales",
                                             "Sphingobacteriales",
                                             "Chitinophagales",
                                             "Cytophagales",

                                             "Enterobacteriales",
                                             "Pseudomonadales",
                                             "Aeromonadales",
                                             "Xanthomonadales",
                                             "Betaproteobacteriales",
                                             "Rhizobiales",
                                             "Azospirillales",
                                             "Sphingomonadales",

                                             "Corynebacteriales",
                                             "Micrococcales",

                                             "Clostridiales",
                                             "Bacillales",
                                             "Lactobacillales",

                                             "Fusobacteriales",
                                             "Other"))



# phylum level


ps1.com.phy <- microbiome::aggregate_top_taxa(frog_filtered, "Phylum", top = 4)
```

```r
ps1.com.phy<- microbiome::transform(ps1.com.phy, "compositional")

plot.composition.relAbun <- microbiome::plot_composition(ps1.com.phy,
                                                sample.sort = "Treatment",
                                                x.label = "frog_id",
                                                group_by = "Treatment")


data.com.phy <- plot.composition.relAbun$data


data.com.phy$Tax<-factor(data.com.phy$Tax, levels = c( "Bacteroidetes","Proteobacteria", "Actinobacteria


pal1<-c("lightskyblue", "plum3" , "#B2DF8A" , "#FFFF99", "white")


barplot_phylum<-ggplot(data.com.phy, aes(x = Sample, y = Abundance, fill = Tax))+
  geom_bar(position = "stack", stat = "identity", width = 1, col = "black")+
  scale_x_discrete(labels = data.com$xlabel, breaks = data.com$Sample)+
  facet_grid(~Group, scales = "free") + theme_bw()+
  scale_fill_manual(values = pal1)+
   theme_bw(base_size = 14)+
  theme(axis.text.x = element_blank())+
  theme(axis.title.x = element_blank())+
  labs(fill = "Bacterial phylum")+
  ylab("Relative abundance")+
    theme(legend.key.size = unit(0.4, 'cm'))+
  scale_y_continuous(expand = c(0, 0), limits = c(0, 1))


ggarrange(barplot_phylum, barplot_order, ncol = 1, align = "v", heights = c(1,1.2) , labels = c("a)", "
```




```
#ggsave("fig2.pdf")

#Saving 13.6 x 7.68 in image
```

Changes in relative abundance across treatment groups

```
frog_genus<- microbiome::aggregate_top_taxa(frog_rare, "Genus", top = 20)

taxtable<-data.frame(tax_table(frog_genus))
taxa_names(frog_genus)<-taxtable$Genus
taxa_names(frog_genus)
```

```
##  [1] "Acinetobacter"        "Aeromonas"            "Bacteroides"
##  [4] "Citrobacter"          "Comamonas"            "Delftia"
##  [7] "Enterobacter"         "Flavobacterium"       "Janthinobacterium"
## [10] "Methylobacterium"     "Myroides"             "Other"
## [13] "Pedobacter"           "Proteus"              "Providencia"
## [16] "Pseudomonas"          "Serratia"             "Sphingobacterium"
## [19] "Stenotrophomonas"     "uncultured bacterium" "Unknown"
```

```
#### table

UU_top<-subset_samples(frog_genus, Treatment == "UU") %>% transform("compositional")
UU_prev<-prevalence(UU_top)
UU_prev$Treatment<-"UU"


CU_top<-subset_samples(frog_genus, Treatment == "CU")%>% transform("compositional")
CU_prev<-prevalence(CU_top)
CU_prev$Treatment<-"CU"


UC_top<-subset_samples(frog_genus, Treatment == "UC") %>% transform("compositional")
UC_prev<-prevalence(UC_top)
UC_prev$Treatment<-"UC"


CC_top<-subset_samples(frog_genus, Treatment == "CC") %>% transform("compositional")
CC_prev<-prevalence(CC_top)
CC_prev$Treatment<-"CC"

prev_df<-rbind(UU_prev, UC_prev, CU_prev, CC_prev)
prev_df$Treatment <- factor(prev_df$Treatment, levels = c("UU", "CU", "UC", "CC"))

palx<-brewer.pal(12,"Paired")
paly<-brewer.pal(12,"Dark2")
pali<-brewer.pal(12,"Pastel1")

palz<-c(palx,paly, pali,palx,paly, pali)
scales::show_col(palz)
```

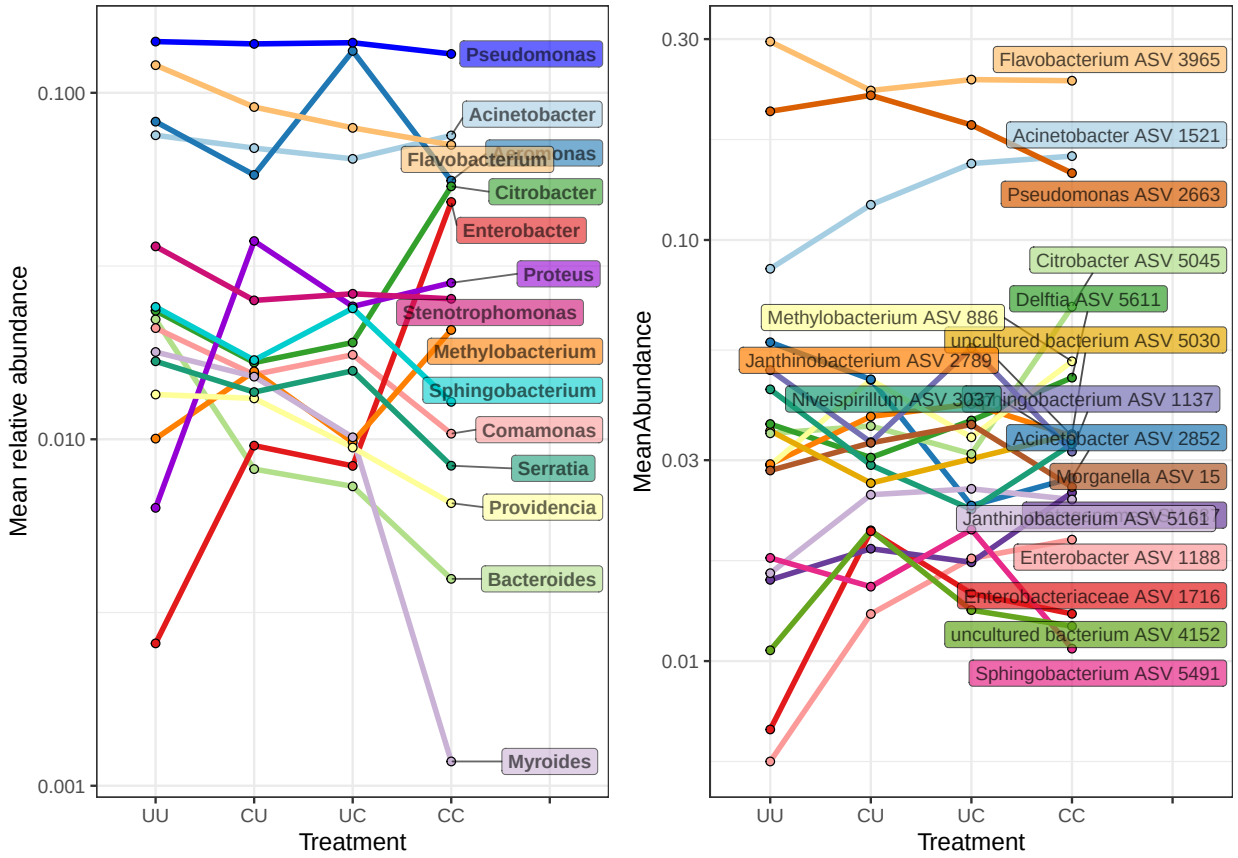

```
palz[12]<-"blue"
palz[14]<-"cyan3"
palz[15]<-"deeppink3"
palz[10]<-"darkviolet"

prev_df<-prev_df %>%  mutate(label = ifelse(Treatment == "CC", as.character(Genus), NA))
levels(prev_df$Treatment) <- c(levels(prev_df$Treatment),'') # add blank level

df<-subset(prev_df, Genus != "Other" & Genus !="Unknown" & Genus !="uncultured bacterium")
length(unique(df$Genus))
```

```
## [1] 18
```

```
p1<-ggplot(subset(prev_df, Genus != "Other" & Genus !="Unknown" & Genus !="uncultured bacterium" & Genus

  geom_line(aes(col = Genus), size = 1.5)+
  geom_point(aes(fill = Genus), size = 2, col = "black", pch = 21)+
  scale_color_manual(values = palz)+
  scale_fill_manual(values = palz)+
  scale_y_log10()+
  theme_bw(base_size = 14)+
  theme(legend.position = "none")+
  scale_x_discrete(drop=FALSE)+

  geom_label_repel(aes(label = label, fill = Genus),
```

```
                  alpha = 0.6,nudge_x = 1, fontface="bold",
                  na.rm = TRUE)+
  ylab("Mean relative abundance")




## ASV level ###
## ASV level ###
## ASV level ###
## ASV level ###


frog_asv<- core(frog_rare, detection = 0, prevalence = 0.94)


#### table

UU_top<-subset_samples(frog_asv, Treatment == "UU") %>% transform("compositional")
UU_prev<-prevalence(UU_top)
UU_prev$Treatment<-"UU"
UU_prev$ASV<-row.names(UU_prev)


CU_top<-subset_samples(frog_asv, Treatment == "CU")%>% transform("compositional")
CU_prev<-prevalence(CU_top)
CU_prev$Treatment<-"CU"
CU_prev$ASV<-row.names(CU_prev)



UC_top<-subset_samples(frog_asv, Treatment == "UC") %>% transform("compositional")
UC_prev<-prevalence(UC_top)
UC_prev$Treatment<-"UC"
UC_prev$ASV<-row.names(UC_prev)



CC_top<-subset_samples(frog_asv, Treatment == "CC") %>% transform("compositional")
CC_prev<-prevalence(CC_top)
CC_prev$Treatment<-"CC"
CC_prev$ASV<-row.names(CC_prev)



prev_df_asv<-rbind(UU_prev, UC_prev, CU_prev, CC_prev)
prev_df_asv$Treatment <- factor(prev_df_asv$Treatment, levels = c("UU", "CU", "UC", "CC"))

palx<-brewer.pal(12,"Paired")
paly<-brewer.pal(12,"Dark2")

palz<-c(palx,paly)
```

```
prev_df_asv<-prev_df_asv %>%  mutate(label = ifelse(Treatment == "CC", as.character(ASV), NA))
levels(prev_df_asv$Treatment) <- c(levels(prev_df_asv$Treatment),'') # add blank level

p2<-ggplot(prev_df_asv, aes(y = MeanAbundance, x = Treatment, group = ASV))+

  geom_line(aes(col = ASV), size = 1.5)+
  geom_point(aes(fill = ASV), size = 2, col = "black", pch = 21)+
  scale_color_manual(values = palz)+
  scale_fill_manual(values = palz)+
  scale_y_log10()+
  theme_bw(base_size = 14)+
  theme(legend.position = "none")+
  scale_x_discrete(drop=FALSE)+

  geom_label_repel(aes(label = label, fill = ASV),
                   alpha = 0.7,nudge_x = 1,
                   na.rm = TRUE)


ggarrange(p1, p2, ncol = 2)
```



Alpha diversity

```
alpha_df<-data.frame(estimate_richness(frog_rare))

alpha_df$Faiths_PD<-as.numeric(metagMisc::phyloseq_phylo_div(frog_rare, "PD")$PD)
row.names(alpha_df)<-sample_names(frog_filtered)

metadata<-data.frame(sample_data(frog_filtered))
metadata<-merge(metadata, alpha_df, by = 0)
summary(metadata$Observed)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    30.0   212.2   290.0   277.9   340.5   459.0
```

```
metadata$Treatment <- factor(metadata$Treatment, levels = c("UU", "CU", "UC", "CC"))

# scale variables

metadata$Mass_scaled<-as.numeric(scale(metadata$Mass))
metadata$Seq_depth_transformed<-as.numeric(scale(sqrt(metadata$Seq_depth)))
```

Model/distribution selection

```
## choose best distributions for 3 measures of alpha diversity

#EnvStats::distChoose(metadata$Observed)$decision
#EnvStats::distChoose(metadata$InvSimpson)$decision
#EnvStats::distChoose(metadata$Shannon)$decision

#### observed

observed_model<-glm(Observed~Treatment + Seq_depth_transformed+Sex+Mass_scaled, family= Gamma(link = "l

observed_model2<-glm(Observed~Treatment + Seq_depth_transformed+Sex+Mass_scaled, family= Gamma(link = "

observed_model3<-lm(Observed~Treatment + Seq_depth_transformed+Sex+Mass_scaled,  data = metadata) # bes


AIC(observed_model, observed_model2, observed_model3)
```

```
##                 df      AIC
## observed_model   8 1249.580
## observed_model2  8 1247.582
## observed_model3  8 1208.765
```

```
######## invnormal ###
######## invnormal ###
######## invnormal ###
```

```r
invsimp_model<-glm(InvSimpson~Treatment + Seq_depth_transformed+Sex, family= Gamma(link = "log"), data =

invsimp_model2<-glm(InvSimpson~Treatment + Seq_depth_transformed+Sex, family= Gamma(link = "identity"),

invsimp_model3<-lm(InvSimpson~Treatment + Seq_depth_transformed+Sex,  data = metadata)

AIC(invsimp_model, invsimp_model2, invsimp_model3)
```

```
##                df      AIC
## invsimp_model   7 827.3240
## invsimp_model2  7 827.1553
## invsimp_model3  7 847.4517
```

```r
##### shannon ########
##### shannon ########
##### shannon ########
##### shannon ########

shannon_model<-glm(Shannon~Treatment + Seq_depth_transformed+Sex+Mass_scaled, family= Gamma(link = "log

shannon_model2<-glm(Shannon~Treatment + Seq_depth_transformed+Sex+Mass_scaled, family= Gamma(link = "id

shannon_model3<-lm(Shannon~Treatment + Seq_depth_transformed+Sex+Mass_scaled,  data = metadata) # best

AIC(shannon_model, shannon_model2, shannon_model3)
```

```
##                df      AIC
## shannon_model   8 273.2405
## shannon_model2  8 273.0617
## shannon_model3  8 241.9187
```

Fitting final models

```r
# final model

observed_model<-lm(Observed~Treatment +Sex +  Seq_depth_transformed, data = metadata)

summary(observed_model)
```

```
##
## Call:
## lm(formula = Observed ~ Treatment + Sex + Seq_depth_transformed,
##     data = metadata)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -218.59  -50.19   14.72   54.80  199.56
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             258.40      16.47  15.690  < 2e-16 ***
## TreatmentCU              32.57      21.74   1.498   0.1374
## TreatmentUC              52.04      21.25   2.449   0.0161 *
## TreatmentCC              52.05      21.86   2.380   0.0192 *
## SexM                    -25.66      15.50  -1.656   0.1010
## Seq_depth_transformed    40.30       7.70   5.234 9.44e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 77.3 on 98 degrees of freedom
## Multiple R-squared:  0.2945, Adjusted R-squared:  0.2585
## F-statistic: 8.182 on 5 and 98 DF,  p-value: 1.733e-06
```

```
r2beta(observed_model, partial = TRUE, method = "sgv", data = metadata)
```

```
##                  Effect   Rsq upper.CL lower.CL
## 1                 Model 0.294    0.453    0.184
## 6 Seq_depth_transformed 0.218    0.363    0.096
## 3           TreatmentUC 0.058    0.172    0.003
## 4           TreatmentCC 0.055    0.168    0.002
## 5                  SexM 0.027    0.122    0.000
## 2           TreatmentCU 0.022    0.113    0.000
```

```
#### GET CONFIDENCE INTERVALS
summary<-data.frame(confint(observed_model))
summaryest<-summary(observed_model) # add estimate as column
summaryest<-data.frame(summaryest$coefficients)

summary$Estimate<-summaryest$Estimate
summary<-summary[c(1:4),] #keep just first 4 rows

summary[2:4,]<-summary[2:4,]+summary[1,3] # add intercept to estimates for forest and organic
row.names(summary)<-c("UU", "CU", "UC", "CC")
names(summary)<-c("CI_lower", "CI_upper", "Est")
summary$Treatment<-row.names(summary)
summary$Treatment<-factor(summary$Treatment, levels = c(c("UU", "CU", "UC", "CC")))


# standard errors

summary_coef1<-summary(observed_model)
summary_coef1$coefficients
```

```
##                       Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)           258.40078  16.469097 15.690039 1.753967e-28
## TreatmentCU            32.56699  21.741189  1.497939 1.373632e-01
## TreatmentUC            52.04228  21.247146  2.449378 1.608561e-02
## TreatmentCC            52.04955  21.865052  2.380491 1.922371e-02
```

```
## SexM                   -25.65911  15.497360 -1.655709 1.009797e-01
## Seq_depth_transformed  40.30494   7.700105  5.234337 9.439076e-07
```

```
summary_coef1<- data.frame(summary_coef1$coefficients)
names(summary_coef1)
```

```
## [1] "Estimate"   "Std..Error" "t.value"    "Pr...t.."
```

```
summary$SE_lower<-summary$Est-summary_coef1$Std..Error[1:4]
summary$SE_upper<-summary$Est+summary_coef1$Std..Error[1:4]

summary1<-summary

### ESIMATES AND CONFIDENCE INTERBALS ARE BACKTRANSFORMED IN PLOT USING EXP()

p1<-ggplot(summary1, aes( x =Treatment, y = Est))+
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper), width = 0, col = "grey", size = 1.5)+
  geom_errorbar(aes(ymin = SE_lower, ymax = SE_upper), width = 0,  size = 1.5)+
   geom_point( size = 4, pch = 21, fill = "black", col = "grey")+
  theme_light(base_size = 14)+
  ylab("Observed ASV richness")+
  theme(legend.position = "none")+
  xlab("")

############ inverse simpson  ##################
############ inverse simpson  ##################
############ inverse simpson  ##################
############ inverse simpson  ##################


# final model
invsimp_model<-glm(InvSimpson~Treatment +Sex+ Seq_depth_transformed, family= Gamma(link = "log"), data =


summary(invsimp_model)
```

```
##
## Call:
## glm(formula = InvSimpson ~ Treatment + Sex + Seq_depth_transformed,
##     family = Gamma(link = "log"), data = metadata)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9284  -0.5849  -0.1492   0.2561   1.6507
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.04680    0.14133  21.558   <2e-16 ***
## TreatmentCU             0.18525    0.18658   0.993   0.3232
## TreatmentUC             0.22551    0.18234   1.237   0.2191
## TreatmentCC             0.41501    0.18764   2.212   0.0293 *
## SexM                   -0.32308    0.13299  -2.429   0.0169 *
## Seq_depth_transformed   0.09862    0.06608   1.492   0.1388
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.4400446)
##
##     Null deviance: 54.452  on 103  degrees of freedom
## Residual deviance: 48.441  on  98  degrees of freedom
## AIC: 827.32
##
## Number of Fisher Scoring iterations: 6
```

```
r2beta(invsimp_model, partial = TRUE, method = "sgv", data = metadata)
```

```
##                   Effect   Rsq upper.CL lower.CL
## 1                  Model 0.119    0.284    0.053
## 5                   SexM 0.057    0.171    0.003
## 4           TreatmentCC 0.048    0.157    0.001
## 6 Seq_depth_transformed 0.022    0.112    0.000
## 3           TreatmentUC 0.015    0.098    0.000
## 2           TreatmentCU 0.010    0.084    0.000
```

```
#### GET CONFIDENCE INTERVALS FROM GAMMA

summary2<-data.frame(confint(invsimp_model))
summary2est<-summary(invsimp_model) # add estimate as column
summary2est<-data.frame(summary2est$coefficients)

summary2$Estimate<-summary2est$Estimate
summary2<-summary2[c(1:4),] #keep just first 4 rows

summary2[2:4,]<-summary2[2:4,]+summary2[1,3] # add intercept to estimates for forest and organic
row.names(summary2)<-c("UU", "CU", "UC", "CC")
names(summary2)<-c("CI_lower", "CI_upper", "Est")
summary2$Treatment<-row.names(summary2)
summary2$Treatment<-factor(summary2$Treatment, levels = c(c("UU", "CU", "UC", "CC")))


# standard errors

summary2_coef1<-summary(invsimp_model)
summary2_coef1$coefficients
```

```
##                           Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)            3.04680029 0.14133199 21.5577541 5.636062e-39
## TreatmentCU            0.18524507 0.18657522  0.9928707 3.232182e-01
## TreatmentUC            0.22550514 0.18233552  1.2367593 2.191321e-01
## TreatmentCC            0.41501154 0.18763817  2.2117651 2.930627e-02
## SexM                  -0.32308257 0.13299288 -2.4293223 1.694836e-02
## Seq_depth_transformed  0.09861764 0.06607959  1.4924070 1.388048e-01
```

```
summary2_coef1<- data.frame(summary2_coef1$coefficients)
names(summary2_coef1)
```

```
## [1] "Estimate"   "Std..Error" "t.value"    "Pr...t.."
```

```
summary2$SE_lower<-summary2$Est-summary2_coef1$Std..Error[1:4]
summary2$SE_upper<-summary2$Est+summary2_coef1$Std..Error[1:4]

### ESIMATES AND CONFIDENCE INTERBALS ARE BACKTRANSFORMED IN PLOT USING EXP()

p2<-ggplot(summary2, aes( x =Treatment, y = exp(Est)))+
  geom_errorbar(aes(ymin = exp(CI_lower), ymax = exp(CI_upper)), width = 0, col = "grey", size = 1.5)+
  geom_errorbar(aes(ymin = exp(SE_lower), ymax = exp(SE_upper)), width = 0,  size = 1.5)+
   geom_point( size = 4, pch = 21, fill = "black", col = "grey")+
  theme_light(base_size = 14)+
  ylab("Inverse Simpson Index")+
  theme(legend.position = "none")


################ shannon ##################
################ shannon ##################
################ shannon ##################
################ shannon ##################

# final model

shannon_model<-lm(Shannon~Treatment +Sex+ Seq_depth_transformed,  data = metadata)
summary(shannon_model)
```

```
##
## Call:
## lm(formula = Shannon ~ Treatment + Sex + Seq_depth_transformed,
##     data = metadata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6625 -0.3238  0.1315  0.4271  1.1983
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.76674    0.15735  23.938  < 2e-16 ***
## TreatmentCU            0.30217    0.20772   1.455  0.14896
## TreatmentUC            0.32622    0.20300   1.607  0.11127
## TreatmentCC            0.30646    0.20891   1.467  0.14559
## SexM                  -0.32411    0.14807  -2.189  0.03098 *
## Seq_depth_transformed  0.20587    0.07357   2.798  0.00619 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7386 on 98 degrees of freedom
## Multiple R-squared:  0.1467, Adjusted R-squared:  0.1032
## F-statistic:  3.37 on 5 and 98 DF,  p-value: 0.007501
```

```
r2beta(shannon_model, partial = TRUE, method = "sgv", data = metadata)
```

```
##                   Effect   Rsq upper.CL lower.CL
```

```
## 1                 Model 0.147    0.313    0.070
## 6 Seq_depth_transformed 0.074    0.196    0.007
## 5                   SexM 0.047    0.155    0.001
## 3           TreatmentUC 0.026    0.119    0.000
## 4           TreatmentCC 0.021    0.111    0.000
## 2           TreatmentCU 0.021    0.110    0.000
```

```
#### GET CONFIDENCE INTERVALS FROM GAMMA

summary3<-data.frame(confint(shannon_model))
summary3est<-summary(shannon_model) # add estimate as column
summary3est<-data.frame(summary3est$coefficients)

summary3$Estimate<-summary3est$Estimate
summary3<-summary3[c(1:4),] #keep just first 4 rows

summary3[2:4,]<-summary3[2:4,]+summary3[1,3] # add intercept to estimates for forest and organic
row.names(summary3)<-c("UU", "CU", "UC", "CC")
names(summary3)<-c("CI_lower", "CI_upper", "Est")
summary3$Treatment<-row.names(summary3)
summary3$Treatment<-factor(summary3$Treatment, levels = c(c("UU", "CU", "UC", "CC")))


# standard errors

summary3_coef1<-summary(shannon_model)
summary3_coef1$coefficients
```

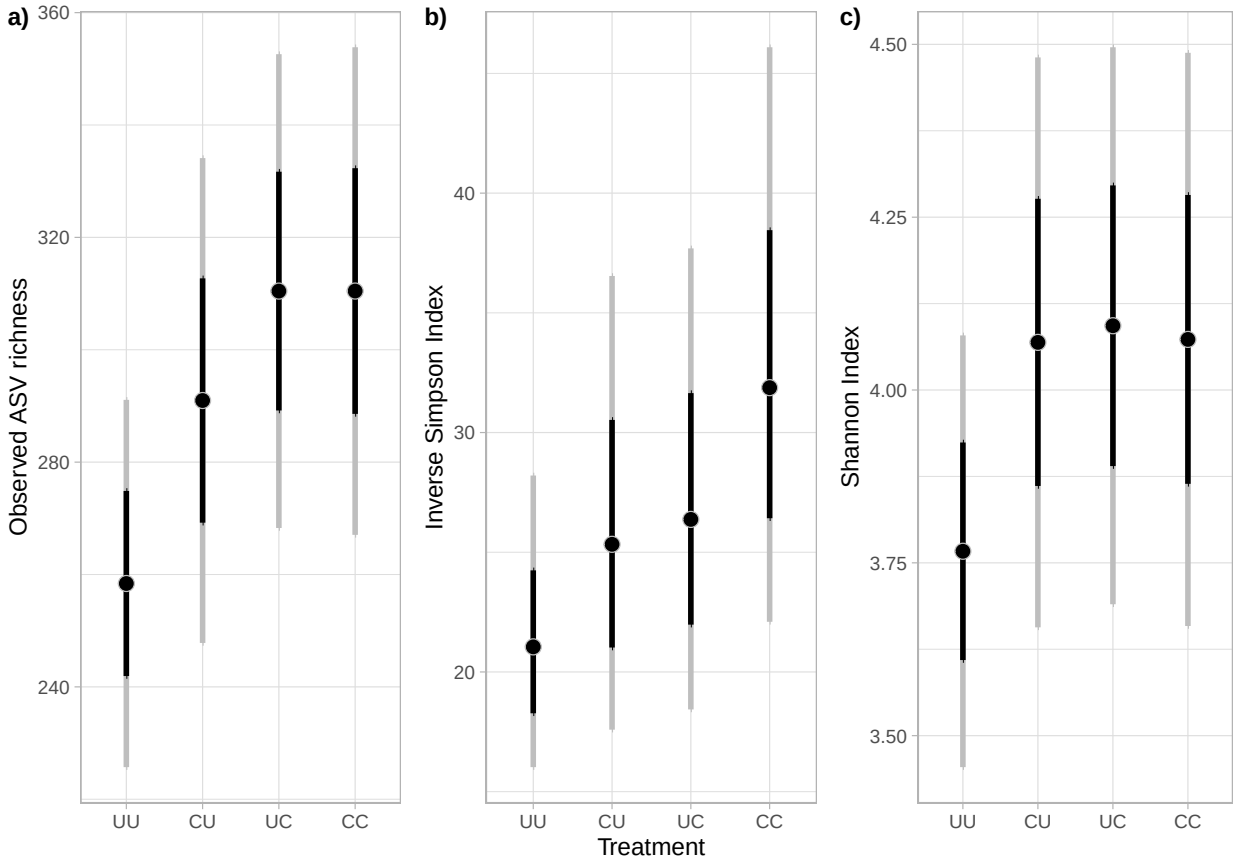
```
##                          Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)             3.7667387 0.15735222 23.938262 9.957611e-43
## TreatmentCU             0.3021697 0.20772386  1.454670 1.489562e-01
## TreatmentUC             0.3262249 0.20300358  1.606991 1.112737e-01
## TreatmentCC             0.3064599 0.20890729  1.466966 1.455871e-01
## SexM                   -0.3241057 0.14806786 -2.188900 3.097894e-02
## Seq_depth_transformed   0.2058650 0.07356983  2.798226 6.187398e-03
```

```
summary3_coef1<- data.frame(summary3_coef1$coefficients)
names(summary3_coef1)
```

```
## [1] "Estimate"   "Std..Error" "t.value"    "Pr...t.."
```

```
summary3$SE_lower<-summary3$Est-summary3_coef1$Std..Error[1:4]
summary3$SE_upper<-summary3$Est+summary3_coef1$Std..Error[1:4]

### ESIMATES AND CONFIDENCE INTERBALS ARE BACKTRANSFORMED IN PLOT USING EXP()

p3<-ggplot(summary3, aes( x =Treatment, y = Est))+
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper), width = 0, col = "grey", size = 1.5)+
  geom_errorbar(aes(ymin = SE_lower, ymax = SE_upper), width = 0,  size = 1.5)+
   geom_point( size = 4, pch = 21, fill = "black", col = "grey")+
  theme_light(base_size = 14)+
  ylab("Shannon Index")+
```

```
  theme(legend.position = "none")+
  xlab("")


ggpubr::ggarrange(p1, p2, p3, ncol = 3, labels = c("a)", "b)", "c)"))
```




```
tab_model(observed_model, invsimp_model, shannon_model,  pred.labels = c("Intercept [UU]", "Treatment [
```

a) Observed ASV richness

b) Inverse Simpson

c) Shannon

Predictors
Estimates
CI
p
Estimates
CI
p
Estimates
CI

p
Intercept [UU]
258.40
225.72 – 291.08
<0.001
21.05
16.02 – 28.20
<0.001
3.77
3.45 – 4.08
<0.001
Treatment [CU]
32.57
-10.58 – 75.71
0.137
1.20
0.84 – 1.74
0.321
0.30
-0.11 – 0.71
0.149
Treatment [UC]
52.04
9.88 – 94.21
0.016
1.25
0.88 – 1.79
0.216
0.33
-0.08 – 0.73
0.111
Treatment [CC]
52.05
8.66 – 95.44
0.019
1.51
1.05 – 2.19
0.027
0.31
-0.11 – 0.72
0.146
Sex [Male]
-25.66
-56.41 – 5.09
0.101
0.72
0.56 – 0.94
0.015
-0.32
-0.62 – -0.03
0.031
Sequencing depth
40.30
25.02 – 55.59

<0.001
1.10
0.97 – 1.26
0.136
0.21
0.06 – 0.35
0.006
Observations
104
104
104
R2 / R2 adjusted
0.294 / 0.259
0.138
0.147 / 0.103

```
#ggsave("Figures/Fig3.pdf")
```

Beta diversity - constrained ordination

Weighted Unifrac

```
#############


wunifrac_dist<-distance(frog_rare, method = "wunifrac")
otutable<-data.frame(t(frog_rare@otu_table@.Data))
metadata <- data.frame(sample_data(frog_rare))


Treatment <- metadata$Treatment
Seq_depth <- as.numeric(scale(metadata$Seq_depth))
Mass <-metadata$Mass
Sex<-as.factor(metadata$Sex)
Date<-as.factor(metadata$date_cat)

metadata<- metadata %>% mutate(MassCat = case_when((Mass <2.45 ~ "Light"),
                                          (Mass > 3 ~ "Heavy")))

metadata$MassCat <-ifelse(is.na(metadata$MassCat), "Average", metadata$MassCat)

MassCat <-as.factor(metadata$MassCat)


final_model<-capscale(wunifrac_dist ~
                    Treatment+
                    Seq_depth+
                      Sex,
                  env = metadata,
                  comm = otutable)
```

```r
# Note: including mass reduces effect of treatment - mechanism?

# weighted unifrac

anova_wunifrac<-anova.cca(final_model, by="terms")
anova_wunifrac
```

```
## Permutation test for capscale under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## Model: capscale(formula = wunifrac_dist ~ Treatment + Seq_depth + Sex, comm = otutable, env = metadat
##           Df SumOfSqs      F Pr(>F)
## Treatment  3  0.15413 1.7117  0.012 *
## Seq_depth  1  0.04426 1.4748  0.121
## Sex        1  0.06490 2.1624  0.016 *
## Residual  98  2.94136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
####### plot #####
####### plot #####
####### plot #####


## extract data from model

final_model_df<-scores(final_model)

# extract CAP scores

vectors_df<-data.frame(final_model_df$sites)
vectors_df$feature.id<-row.names(vectors_df)


# merge with info on dominant family

sample_metadata<-data.frame(sample_data(frog_rare))[,c("feature.id", "Treatment")]
vectors_df<-merge(vectors_df, sample_metadata, by = "feature.id")


#### add arrows #########
#### add arrows #########
#### add arrows #########
#### add arrows #########

centroids_df<-data.frame(final_model_df$centroids)
centroids_df<-centroids_df[1:6,]
centroids_df
```

```
##                    CAP1        CAP2
```

```
## TreatmentUU  0.59349632  0.02322374
## TreatmentCU  0.03285805 -0.10056055
## TreatmentUC -0.15652972  0.18997779
## TreatmentCC -0.51929248 -0.14301689
## SexF         0.03898003 -0.38958826
## SexM        -0.03091519  0.30898379
```

```
row.names(centroids_df)<-c("Treatment: UU", "Treatment: CU","Treatment: UC","Treatment: CC", "Sex:F", "S
```

```
### add taxa scores ########
### add taxa scores ########
### add taxa scores ########
### add taxa scores ########
```

```
species_scores<-data.frame(final_model_df$species)

summary(species_scores$CAP1)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.2273347 -0.0001671 -0.0000194  0.0000000  0.0000474  0.4229621
```

```
summary(species_scores$CAP2)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.0897663 -0.0001277 -0.0000143  0.0000000  0.0000390  0.5088769
```

```
species_scores<-subset(species_scores,  (CAP2 > 0.2 | CAP2 < -0.2) |  (CAP1 > 0.2 | CAP1 < -0.2))

species_scores$CAP1<-species_scores$CAP1 *4
species_scores$CAP2<-species_scores$CAP2 *4

############
############
############
############

vectors_wunifrac<-vectors_df
centroids_wunifrac<-centroids_df
species_wunifrac<-species_scores

# colour palette
pal<-pals::stepped3()[c(1,5,9,13)]
pal<-pals::tol()[c(1,3,4,12)]


plot_wunifrac<-ggplot(vectors_wunifrac, aes(x = CAP1, y = CAP2))+

   stat_ellipse(geom = "polygon", aes(fill = Treatment, col = Treatment), level = 0.9, alpha = 0.3, siz
```

```
  geom_point(aes(fill =Treatment), pch = 21, size = 3, alpha = 1, stroke = 1, col = "black")+

  theme_bw()+
 scale_fill_viridis(discrete = TRUE)+
 scale_color_viridis(discrete = TRUE)+

  # add arrows

  geom_segment(data=centroids_wunifrac[1:4,], aes(x = 0, y = 0, xend = CAP1*2, yend = CAP2*2),
    arrow = arrow(length = unit(0.5, "cm"), type = "closed"), lwd = 1,  col = "black")+
  ggrepel::geom_label_repel(data=centroids_wunifrac[1:4,],
    alpha = 0.9, col = "black", size = 4, fill = "yellow",
    aes(CAP1*2, CAP2*2, label = row.names(centroids_wunifrac[1:4,])))+

  theme_light(base_size = 14)+
  ggtitle("a) Treatment: WU")
```

Bray Curtis

```
#############

bray_dist<-distance(frog_rare, method = "bray")
otutable<-data.frame(t(frog_rare@otu_table@.Data))
metadata <- data.frame(sample_data(frog_rare))


Treatment <- metadata$Treatment
Seq_depth <- as.numeric(scale(metadata$Seq_depth))
Mass <-metadata$Mass
Sex<-as.factor(metadata$Sex)
Date<-as.factor(metadata$date_cat)

metadata<- metadata %>% mutate(MassCat = case_when((Mass <2.45 ~ "Light"),
                                          (Mass > 3 ~ "Heavy")))

metadata$MassCat <-ifelse(is.na(metadata$MassCat), "Average", metadata$MassCat)

MassCat <-as.factor(metadata$MassCat)



final_model<-capscale(bray_dist ~
                   Treatment+
                    Seq_depth+
                     Sex,

                env = metadata,
                comm = otutable)
```

```
# Note: including mass reduces effect of treatment - mechanism?

# weighted bray

anova_bray<-anova.cca(final_model, by="terms")
round(data.frame(anova_bray), 3)
```

```
##           Df SumOfSqs     F Pr..F.
## Treatment  3    1.052 1.393  0.028
## Seq_depth  1    0.363 1.441  0.067
## Sex        1    0.811 3.221  0.001
## Residual  98   24.679    NA     NA
```

```
round(data.frame(anova_wunifrac), 3)
```

```
##           Df SumOfSqs     F Pr..F.
## Treatment  3    0.154 1.712  0.012
## Seq_depth  1    0.044 1.475  0.121
## Sex        1    0.065 2.162  0.016
## Residual  98    2.941    NA     NA
```

```
####### plot #####
####### plot #####
####### plot #####


## extract data from model

final_model_df<-scores(final_model)


# extract CAP scores

vectors_df<-data.frame(final_model_df$sites)
vectors_df$feature.id<-row.names(vectors_df)


# merge with info on dominant family

sample_metadata<-data.frame(sample_data(frog_rare))[,c("feature.id", "Treatment")]
vectors_df<-merge(vectors_df, sample_metadata, by = "feature.id")



#### add arrows #########
#### add arrows #########
#### add arrows #########
#### add arrows #########

centroids_df<-data.frame(final_model_df$centroids)
centroids_df<-centroids_df[1:6,]
row.names(centroids_df)<-c("Treatment: UU", "Treatment: CU","Treatment: UC","Treatment: CC", "SexF", "Se
```

```
### add taxa scores ########
### add taxa scores ########


species_scores<-data.frame(final_model_df$species)

summary(species_scores$CAP1)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.9977338 -0.0000442  0.0000142  0.0000000  0.0002048  0.2277434
```

```
summary(species_scores$CAP2)
```

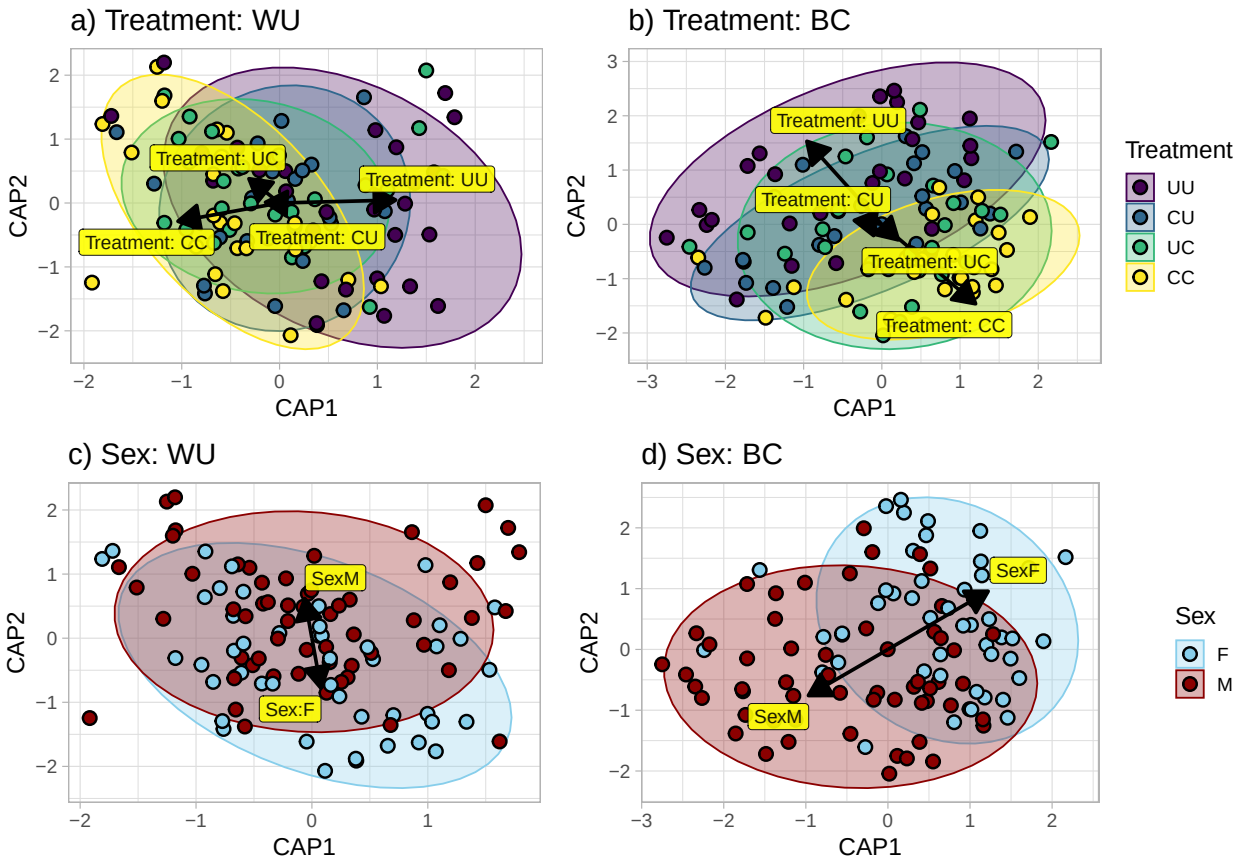
```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.5041614 -0.0001969 -0.0000298  0.0000000  0.0000506  0.6130853
```

```
species_scores<-subset(species_scores,  (CAP2 > 0.2 | CAP2 < -0.2) |  (CAP1 > 0.2 | CAP1 < -0.2))

species_scores$CAP1<-species_scores$CAP1 *3
species_scores$CAP2<-species_scores$CAP2 *3

############################
############################
############################
############################

plot_bray<-ggplot(vectors_df, aes(x = CAP1, y = CAP2))+

    stat_ellipse(geom = "polygon", aes(fill = Treatment, col = Treatment), level = 0.9, alpha = 0.3, si
  geom_point(aes(fill =Treatment), pch = 21, size = 3, alpha = 1, stroke = 1, col = "black")+


 scale_fill_viridis(discrete = TRUE)+
 scale_color_viridis(discrete = TRUE)+

  # add arrows

  geom_segment(data=centroids_df[1:4,], aes(x = 0, y = 0, xend = CAP1*2, yend = CAP2*2),
    arrow = arrow(length = unit(0.5, "cm"), type = "closed"), lwd = 1,  col = "black")+
  ggrepel::geom_label_repel(data=centroids_df[1:4,],
    alpha = 0.9, col = "black", size = 4, fill = "yellow",
    aes(CAP1*2, CAP2*2, label = row.names(centroids_df[1:4,])))+
  theme_light(base_size = 14)+
  ggtitle("b) Treatment: BC")


### plot together

treatment_plots<-ggarrange(plot_wunifrac, plot_bray,  common.legend = T, legend = "right")
```

Beta diversity by sex

```
sex_wunifrac<-ggplot(vectors_wunifrac, aes(x = CAP1, y = CAP2))+

   stat_ellipse(geom = "polygon", aes(fill = Sex, col = Sex), level = 0.9, alpha = 0.3, size = 0.5)+
  geom_point(aes(fill =Sex), pch = 21, size = 3, alpha = 1, stroke = 1, col = "black")+

  theme_bw()+
 scale_fill_manual(values = c("skyblue", "darkred"))+
 scale_color_manual(values = c("skyblue", "darkred"))+

  # add arrows

  geom_segment(data=centroids_wunifrac[5:6,], aes(x = 0, y = 0, xend = CAP1*2, yend = CAP2*2),
    arrow = arrow(length = unit(0.5, "cm"), type = "closed"), lwd = 1,  col = "black")+
  ggrepel::geom_label_repel(data=centroids_wunifrac[5:6,],
    alpha = 0.9, col = "black", size = 4, fill = "yellow",
    #hjust = c(0,1),
    #vjust = c(1,1),
    aes(CAP1*2, CAP2*2, label = row.names(centroids_wunifrac[5:6,])))+
  theme_light(base_size = 14)+
  ggtitle("c) Sex: WU")+
  theme(plot.margin = margin(0.2,0.8,0.2,0.2, "cm"))

##################
##################
##################
##################


sex_bray<- ggplot(vectors_df, aes(x = CAP1, y = CAP2))+

    stat_ellipse(geom = "polygon", aes(fill = Sex, col = Sex), level = 0.9, alpha = 0.3, size = 0.5)+
  geom_point(aes(fill =Sex), pch = 21, size = 3, alpha = 1, stroke = 1, col = "black")+


 scale_fill_manual(values = c("skyblue", "darkred"))+
 scale_color_manual(values = c("skyblue", "darkred"))+

  # add arrows

  geom_segment(data=centroids_df[5:6,], aes(x = 0, y = 0, xend = CAP1*2, yend = CAP2*2),
    arrow = arrow(length = unit(0.5, "cm"), type = "closed"), lwd = 1,  col = "black")+
  ggrepel::geom_label_repel(data=centroids_df[5:6,],
    alpha = 0.9, col = "black", size = 4, fill = "yellow",

    aes(CAP1*2, CAP2*2, label = row.names(centroids_df[5:6,])))+


  theme_light(base_size = 14)+
  ggtitle("d) Sex: BC")+
  theme(plot.margin = margin(0.2,1.3,0.2,0, "cm"))
```

```
sex_plots<-ggarrange(sex_wunifrac, sex_bray,  common.legend = T, legend = "right")


ggarrange(treatment_plots, sex_plots,  common.legend = T, legend = "right", nrow = 2, align = "v")
```




```
#ggsave("Figures/Fig4.pdf")
```

Joint-species distribution modelling

JSDM: Genus level

```
# https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13303

# final model

frog_genus<-tax_glom(frog_filtered, taxrank = "Genus")

frog_genus_core<-core(frog_genus, detection = 20, prevalence = 0.70)
#frog_genus_core<-microbiome::aggregate_top_taxa(frog_filtered, "Genus", top = 45)

frog_genus
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 661 taxa and 104 samples ]
## sample_data() Sample Data:       [ 104 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 661 taxa by 7 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 661 tips and 660 internal nodes ]
```

```
frog_genus_core
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 70 taxa and 104 samples ]
## sample_data() Sample Data:       [ 104 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 70 taxa by 7 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 70 tips and 69 internal nodes ]
```

```
sum(sample_sums(frog_genus_core))/sum(sample_sums(frog_genus))
```

```
## [1] 0.8882152
```

```
taxtable<-data.frame(tax_table(frog_genus_core))

frog_genus_core<-subset_taxa( frog_genus_core, Genus != "uncultured" )
frog_genus_core<-subset_taxa( frog_genus_core, Genus != "metagenome" )
frog_genus_core<-subset_taxa( frog_genus_core, Genus != "uncultured bacterium" )
frog_genus_core<-subset_taxa( frog_genus_core, Genus != "Other" )
frog_genus_core<-subset_taxa( frog_genus_core, Genus != "Unknown" )
frog_genus_core<-subset_taxa( frog_genus_core, Genus != "Allorhizobium-Neorhizobium-Pararhizobium-Rhizob
frog_genus_core<-subset_taxa( frog_genus_core, Genus != "CL500-29 marine group" )

taxa_keep<- taxa_names(frog_genus_core)

sum(sample_sums(frog_genus_core))/sum(sample_sums(frog_genus))
```

```
## [1] 0.8486052
```

```
taxtable<-data.frame(tax_table(frog_genus_core))

taxa_names(frog_genus_core)<- taxtable$Genus


## Extract relevant data for model

y <- data.frame(t(otu_table(frog_genus_core)))
X<-data.frame(sample_data(frog_genus_core))


X$Mass_scaled<-as.numeric(scale(X$Mass))
X$Seq_depth_scaled<-as.numeric(scale(sqrt(X$Seq_depth)))

X<-X[,c("Treatment", "Mass_scaled", "Seq_depth_scaled", "Sex")]
```

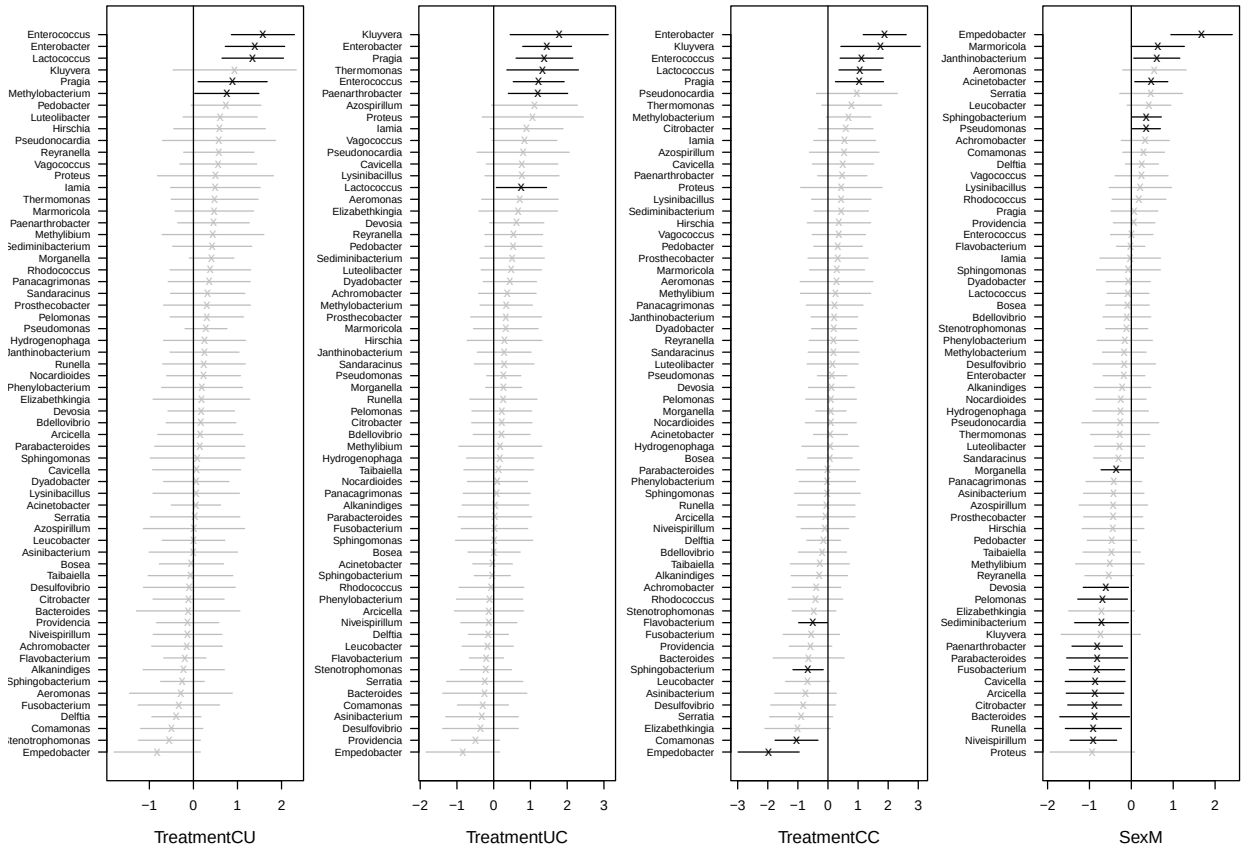
Model comparison

```
fit_1LV <- gllvm(y, X,
             num.lv = 1,
            formula = ~ Treatment+ Sex  + Seq_depth_scaled,
             family = "negative.binomial")

fit_2LV <- gllvm(y, X,
             num.lv = 2,
            formula = ~ Treatment+ Sex + Seq_depth_scaled,
             family = "negative.binomial")


fit_3LV <- gllvm(y, X,
             num.lv = 3,
            formula = ~ Treatment+ Sex  + Seq_depth_scaled,
             family = "negative.binomial")



# compare AIC

AIC( fit_1LV, fit_2LV, fit_3LV)
```

```
##          df      AIC
## fit_1LV 496 89140.63
## fit_2LV 557 88746.51
## fit_3LV 617 88433.39
```

Final model

```
fit <- gllvm(y, X,
             num.lv = 3,
            formula = ~ Treatment+ Sex + Seq_depth_scaled,
             family = "negative.binomial")

coefplot(fit, cex.ylab = 0.7, which.Xcoef = c(1:4))
```




```
dev.off()
```

```
## null device
##           1
```

```
cr1<-getResidualCor(fit)

cr2<-cor_pmat(cr1)

ggcorrplot(cr1, hc.order = TRUE,
   outline.col = "white",
   type = "full",
   ggtheme = ggplot2::theme_minimal,
   tl.cex = 7.5,
   p.mat = cr2,
   sig.level = 0.01,
  # show.diag = F,
   insig = "blank",
  # colors = c("#6D9EC1", "white", "#E46726"))
   colors = c("blue", "white", "red"))


corrplot::corrplot(cr1, type = "lower", order = "hclust", tl.cex = 0.7)
```

Extract estimates

```
df<-coef(fit)

est_df<-data.frame(df$Intercept)

est_df2<-data.frame(df$Xcoef) # choose columns of interest

est_df3<-merge(est_df, est_df2, by = 0)

head(est_df3)
```

```
##        Row.names df.Intercept  TreatmentCU TreatmentUC TreatmentCC       SexM
## 1  Achromobacter     4.359716 -0.145145212  0.37182683 -0.38397153  0.3340888
## 2  Acinetobacter     8.360056  0.056334951 -0.02965723  0.08274385  0.4764631
## 3      Aeromonas     8.216957 -0.283506259  0.71373349  0.28286986  0.5517377
## 4   Alkanindiges     6.158882 -0.221035705  0.04684757 -0.28515171 -0.2125896
## 5      Arcicella     5.523703  0.155417649 -0.13064746 -0.07288996 -0.8670685
## 6 Asinibacterium     4.632561 -0.001419577 -0.31915193 -0.74506265 -0.4201159
##   Seq_depth_scaled
## 1        0.9006572
## 2        0.7182654
## 3        0.3914608
## 4        0.7969600
## 5        0.7364781
## 6        0.9145002
```

```
# order genera

row.names(est_df3)<-est_df3$Row.names
est_df3<-est_df3[colnames(y),]

#put est_df3 into long format

names(est_df3)[1]<- "Genus"
names(est_df3)[2]<- "Intercept"

estimates_df <- gather(est_df3, Treatment, Estimate, names(est_df3)[2]:names(est_df3)[ncol(est_df3)], fa

############## extract confindence intervals  ####
############## extract confindence intervals  ####
############## extract confindence intervals  ####
############## extract confindence intervals  ####


confint_df<-data.frame(confint(fit))


confint_df<-rbind(confint_df[rownames(confint_df) %like% "Xcoef", ],
confint_df[rownames(confint_df) %like% "Intercept", ])

head(confint_df)
```

```
##                                        X2.5..   X97.5..
## Xcoef.TreatmentCU:Sphingobacterium -0.74542035 0.2449750
## Xcoef.TreatmentCU:Pedobacter       -0.05132282 1.5344731
## Xcoef.TreatmentCU:Bacteroides      -1.29548474 1.0552728
## Xcoef.TreatmentCU:Parabacteroides  -0.87823198 1.1707120
## Xcoef.TreatmentCU:Arcicella        -0.81123563 1.1220709
## Xcoef.TreatmentCU:Dyadobacter      -0.67712148 0.8061493
```

```
# add a column with correct variable level
variables<- colnames(est_df3)[3:ncol(est_df3)]
variables<-c(variables, "Intercept")
variables1<-rep(variables, nrow(est_df))
variables2<-variables1[order(match(variables1, variables))]

#confint_df$Treatment<-c(rep("UU", 40), rep("CU", 40), rep("UC", 40), rep("CC", 40))
confint_df$Treatment<-variables2


# column with taxa names. Should be automatically in the correct order but double check

confint_df$Genus<-rep(colnames(y), length(unique(confint_df$Treatment)))



# now have estimates, confidence intervals, aas seperate data frames, but they are in different formats

merged<-merge(estimates_df, confint_df, by = c("Treatment", "Genus"))


final_estimates_reduced<- merged
names(final_estimates_reduced)[4]<-"CI_lower"
names(final_estimates_reduced)[5]<-"CI_upper"

#final_estimates<- merged[,c(1,2,3,7,8)]
head(final_estimates_reduced)
```

```
##   Treatment          Genus Estimate CI_lower CI_upper
## 1 Intercept  Achromobacter 4.359716 3.762315 4.957116
## 2 Intercept  Acinetobacter 8.360056 7.926749 8.793363
## 3 Intercept      Aeromonas 8.216957 7.373062 9.060851
## 4 Intercept   Alkanindiges 6.158882 5.474104 6.843660
## 5 Intercept      Arcicella 5.523703 4.793851 6.253554
## 6 Intercept Asinibacterium 4.632561 3.869051 5.396071
```

```
unique(final_estimates_reduced$Treatment)
```

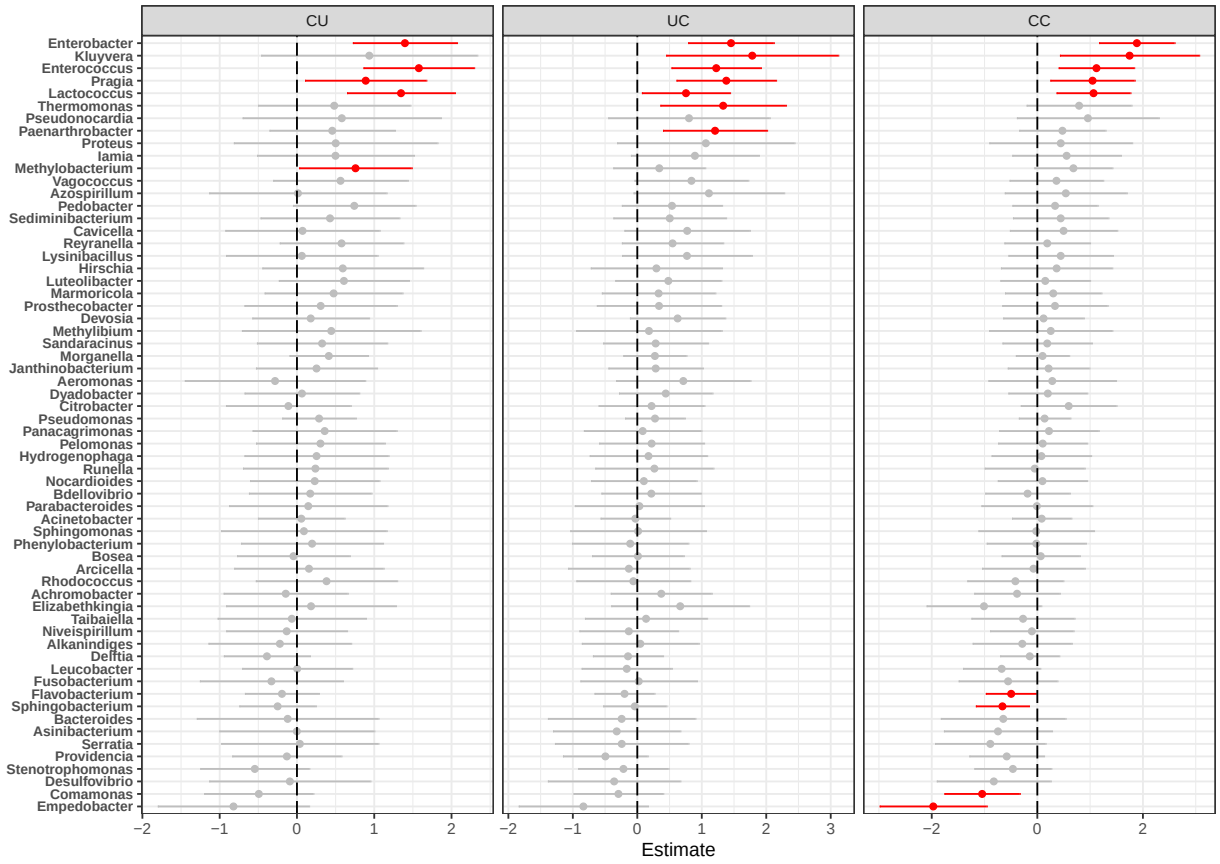
```
## [1] Intercept        Seq_depth_scaled SexM             TreatmentCC
## [5] TreatmentCU      TreatmentUC
## 6 Levels: Intercept TreatmentCU TreatmentUC TreatmentCC ... Seq_depth_scaled
```

```
final_estimates_reduced2<-subset(final_estimates_reduced, Treatment != "SexM" & Treatment != "Seq_depth
```

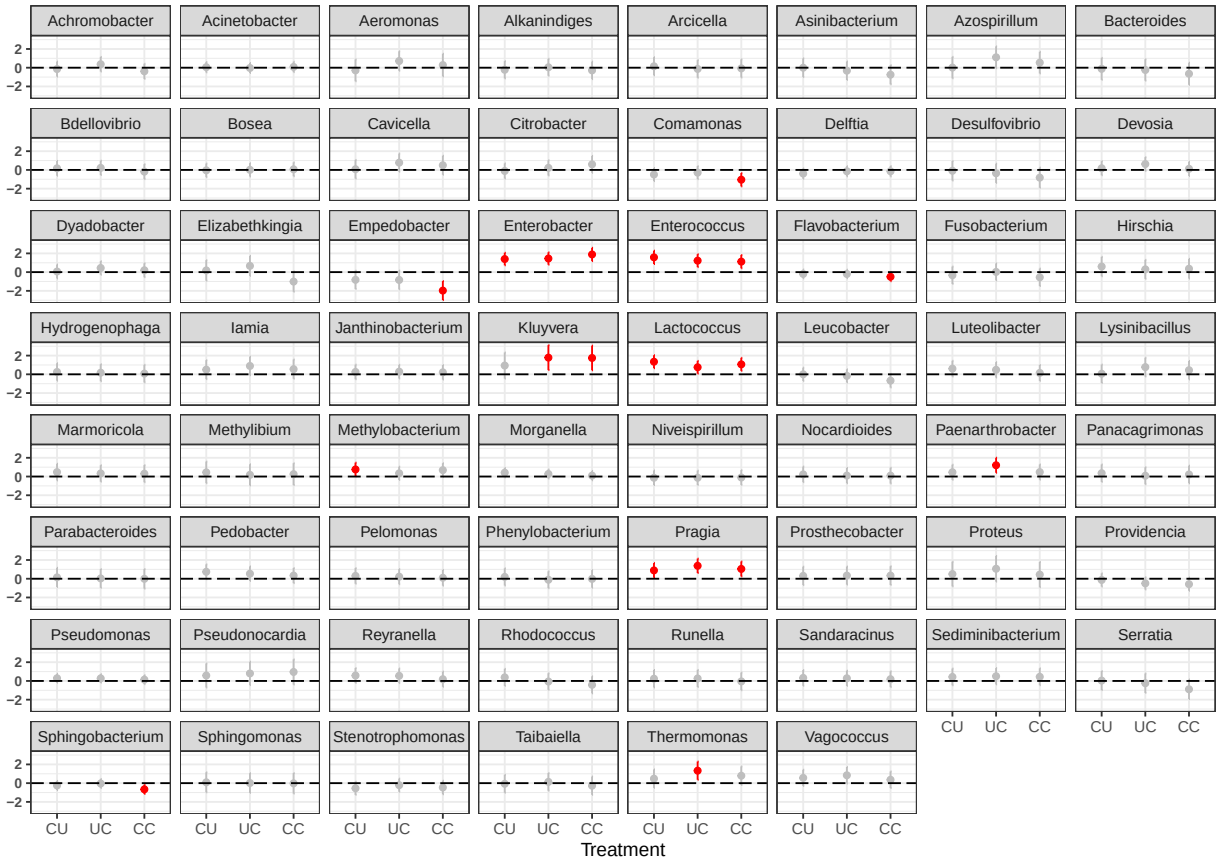
```
## add significance

final_estimates_reduced2$Sig<- !data.table::between(0, final_estimates_reduced2$CI_lower, final_estimat


final_estimates_reduced2$Genus<-factor(final_estimates_reduced2$Genus)

levels(final_estimates_reduced2$Treatment)
```

```
## [1] "Intercept"        "TreatmentCU"      "TreatmentUC"      "TreatmentCC"
## [5] "SexM"             "Seq_depth_scaled"
```

```
final_estimates_reduced2$Treatment<-factor(final_estimates_reduced2$Treatment)
levels(final_estimates_reduced2$Treatment)<-c("UU", "CU", "UC", "CC")



ggplot(subset(final_estimates_reduced2, Treatment != "UU"), aes(y = reorder(Genus, Estimate), x = Estima
  geom_point()+
  facet_wrap(~Treatment, nrow = 1, scales = "free_x")  +
  geom_errorbarh(aes(xmin = CI_lower, xmax = CI_upper, col = Sig), height = 0, size = 0.5)+
  geom_vline(xintercept = 0, linetype = "longdash")+
  theme_bw()+
  scale_color_manual(values = c("grey", "red"))+
  theme(axis.title.y = element_blank())+
  theme(axis.text.y = element_text(face="bold", size = 8))+
  theme(legend.position = "none")+
 theme(plot.margin=unit(c(0.2,0.2, 0.2, 0.6),"cm"))
```

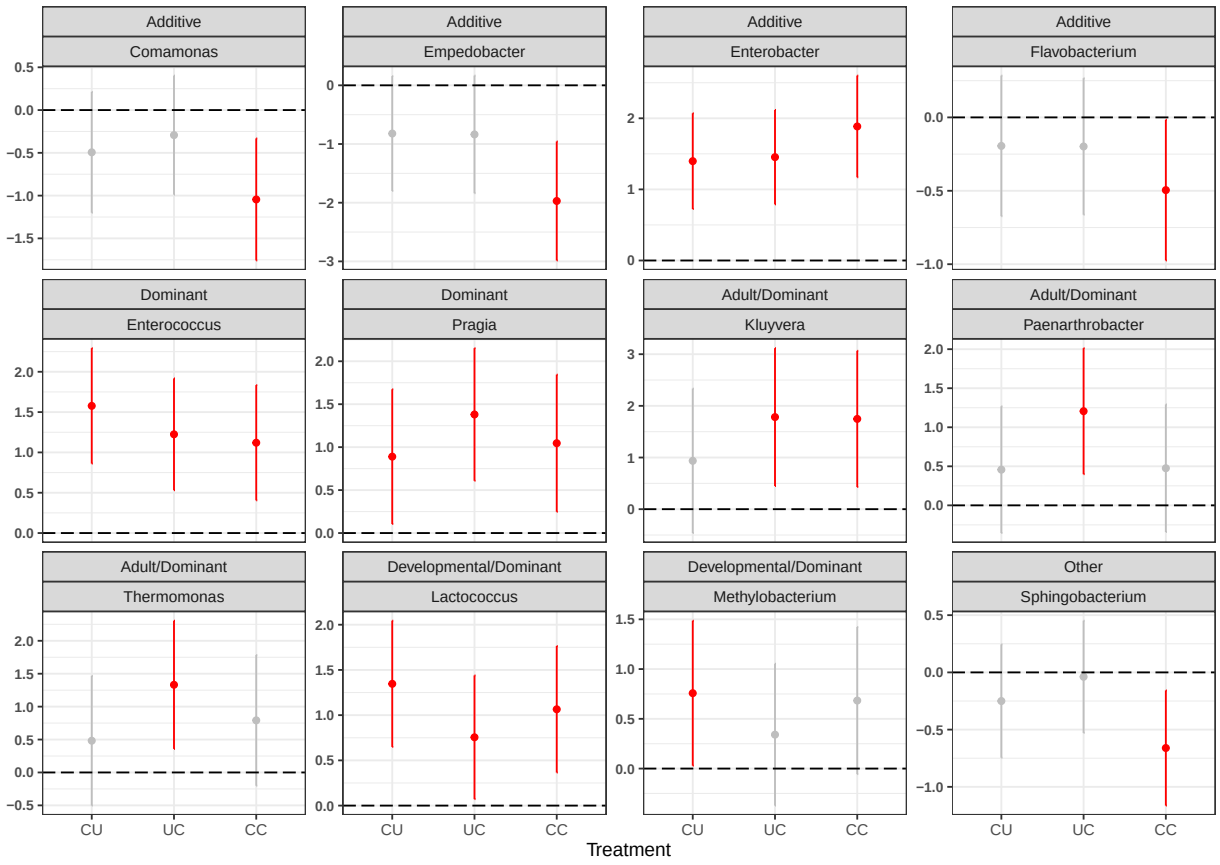



```
ggplot(subset(final_estimates_reduced2, Treatment != "UU"), aes(y = Estimate, x = Treatment, col = Sig))
  geom_point()+
  facet_wrap(~Genus)  +
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper, col = Sig), width = 0, size = 0.5)+
  geom_hline(yintercept = 0, linetype = "longdash")+
  theme_bw()+
  scale_color_manual(values = c("grey", "red"))+
  theme(axis.title.y = element_blank())+
  theme(axis.text.y = element_text(face="bold", size = 8))+
  theme(legend.position = "none")+
 theme(plot.margin=unit(c(0.2,0.2, 0.2, 0.6),"cm"))
```




```
## plot only significant

estimates_sig<- subset(final_estimates_reduced2,  Genus == "Comamonas" |  Genus == "Empedobacter" |  Ge

estimates_sig<-estimates_sig%>%arrange(Treatment, Genus)


estimates_sig$Genus<-factor(estimates_sig$Genus)


estimates_sig$Effect_type <- rep(c("Additive",
                                   "Additive",
                                   "Additive",
                                   "Dominant",
                                   "Additive",
                                   "Adult/Dominant",
                                   "Developmental/Dominant",
                                   "Developmental/Dominant",
                                   "Adult/Dominant",
                                   "Dominant",
                                   "Other",
                                   "Adult/Dominant"), 4)

estimates_sig$Effect_type <-factor(estimates_sig$Effect_type , levels = c("Additive", "Dominant", "Adul
```

```
ggplot(subset(estimates_sig, Treatment != "UU"), aes(y = Estimate, x = Treatment, col = Sig))+
  geom_point()+
  facet_wrap(~Effect_type+Genus, scales = "free_y")  +
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper, col = Sig), width = 0, size = 0.5)+
  geom_hline(yintercept = 0, linetype = "longdash")+
  theme_bw()+
  scale_color_manual(values = c("grey", "red"))+
  theme(axis.title.y = element_blank())+
  theme(axis.text.y = element_text(face="bold", size = 8))+
  theme(legend.position = "none")+
 theme(plot.margin=unit(c(0.2,0.2, 0.2, 0.6),"cm"))
```



JSDM: ASV level

```
##### ASV level ##########
##### ASV level ##########

frog_core<-core(frog_filtered, detection = 30, prevalence = 0.70)
taxanames<-taxa_names(frog_core)

sum(taxa_sums(frog_core))/sum(taxa_sums(frog_filtered))
```

```
## [1] 0.619076
```

```
y <- data.frame(t(otu_table(frog_core)))
X<-data.frame(sample_data(frog_core))


X$Mass_scaled<-as.numeric(scale(X$Mass))
X$Seq_depth_scaled<-as.numeric(scale(sqrt(X$Seq_depth)))

X<-X[,c("Treatment","Sex", "Mass_scaled", "Seq_depth_scaled")]

X$Treatment<-factor(X$Treatment, levels = c("UU", "CU", "UC", "CC"))
```

Model comparison

```
fit <- gllvm(y, X,
             num.lv = 2,
            formula = ~ Treatment+  Sex +Seq_depth_scaled,
             family = "negative.binomial")

#############

fit1 <- gllvm(y, X,
             num.lv = 2,
            formula = ~ Treatment+ Mass_scaled  +Seq_depth_scaled,
             family = "negative.binomial")

####################

fit2 <- gllvm(y, X,
             num.lv = 3,
            formula = ~ Treatment+  Sex +Seq_depth_scaled,
             family = "negative.binomial")

#############

fit3 <- gllvm(y, X,
             num.lv = 3,
            formula = ~ Treatment+ Mass_scaled  +Seq_depth_scaled,
             family = "negative.binomial")

####################


AIC(fit, fit1, fit2, fit3)
```

```
##       df      AIC
## fit  602 94089.03
## fit1 602 94105.22
## fit2 667 93885.90
## fit3 667 93912.83
```

Final model

```
fit <- gllvm(y, X,
             num.lv = 3,
            formula = ~ Treatment+ Sex  +Seq_depth_scaled,
             family = "negative.binomial")
dev.off()
```

```
## null device
##           1
```

```
coefplot(fit, cex.ylab = 0.7, which.Xcoef = c(1:4))

dev.off()
```

```
## null device
##           1
```

```
### extract estimates and CIs ###
### extract estimates and CIs ###
### extract estimates and CIs ###
### extract estimates and CIs ###


df<-coef(fit)

est_df<-data.frame(df$Intercept)

est_df2<-data.frame(df$Xcoef) # choose columns of interest

est_df3<-merge(est_df, est_df2, by = 0)

head(est_df3)
```

```
##                Row.names df.Intercept TreatmentCU   TreatmentUC TreatmentCC
## 1 Acinetobacter.ASV.1521     7.383608  0.25525755  0.4665879834   0.6692832
## 2 Acinetobacter.ASV.2852     6.671637  0.71515837  0.1376968970   0.2024135
## 3     Aeromonas.ASV.2598     6.427880  0.46211908  1.7048394235   1.4745674
## 4  Alkanindiges.ASV.1996     6.070336 -0.25918528  0.0975411237  -0.4852898
## 5     Arcicella.ASV.1236     5.639950  0.03884309 -0.1386843431  -0.1298194
## 6         Bosea.ASV.1971     4.821218 -0.09732116 -0.0004311626  -0.3085474
##         SexM Seq_depth_scaled
## 1  0.7866433        0.7406464
## 2 -0.6017362        0.5416664
## 3  0.4522108        0.6717120
## 4 -0.1195859        0.7937631
## 5 -0.9857491        0.7498585
## 6 -0.1229307        0.7854580
```

```
# order genera

row.names(est_df3)<-est_df3$Row.names
est_df3<-est_df3[colnames(y),]

#put est_df3 into long format

names(est_df3)[1]<- "Genus"
names(est_df3)[2]<- "Intercept"

estimates_df <- gather(est_df3, Treatment, Estimate, names(est_df3)[2]:names(est_df3)[ncol(est_df3)], fa

############## extract confindence intervals  ####
############## extract confindence intervals  ####
############## extract confindence intervals  ####
############## extract confindence intervals  ####


confint_df<-data.frame(confint(fit))


confint_df<-rbind(confint_df[rownames(confint_df) %like% "Xcoef", ],
confint_df[rownames(confint_df) %like% "Intercept", ])

head(confint_df)
```

```
##                                                X2.5..   X97.5..
## Xcoef.TreatmentCU:Sphingobacterium.ASV.5491 -0.5331054 0.5307501
## Xcoef.TreatmentCU:Sphingobacterium.ASV.1137 -0.7470763 0.2878809
## Xcoef.TreatmentCU:Pedobacter.ASV.5042        0.4143237 2.0760981
## Xcoef.TreatmentCU:Pedobacter.ASV.3494       -0.4207676 1.9527305
## Xcoef.TreatmentCU:Pedobacter.ASV.4760       -0.9885924 0.2212875
## Xcoef.TreatmentCU:Arcicella.ASV.1236        -0.6697596 0.7474458
```

```
# add a column with correct variable level
variables<- colnames(est_df3)[3:ncol(est_df3)]
variables<-c(variables, "Intercept")
variables1<-rep(variables, nrow(est_df))
variables2<-variables1[order(match(variables1, variables))]

#confint_df$Treatment<-c(rep("UU", 40), rep("CU", 40), rep("UC", 40), rep("CC", 40))
confint_df$Treatment<-variables2


# column with taxa names. Should be automatically in the correct order but double check

confint_df$Genus<-rep(colnames(y), length(unique(confint_df$Treatment)))



# now have estimates, confidence intervals, aas seperate data frames, but they are in different formats

merged<-merge(estimates_df, confint_df, by = c("Treatment", "Genus"))
```

```
final_estimates_reduced<- merged
names(final_estimates_reduced)[4]<-"CI_lower"
names(final_estimates_reduced)[5]<-"CI_upper"

#final_estimates<- merged[,c(1,2,3,7,8)]
head(final_estimates_reduced)
```

```
##   Treatment                  Genus Estimate CI_lower CI_upper
## 1 Intercept Acinetobacter.ASV.1521 7.383608 6.897726 7.869491
## 2 Intercept Acinetobacter.ASV.2852 6.671637 5.989827 7.353447
## 3 Intercept     Aeromonas.ASV.2598 6.427880 5.618830 7.236930
## 4 Intercept  Alkanindiges.ASV.1996 6.070336 5.461446 6.679226
## 5 Intercept     Arcicella.ASV.1236 5.639950 5.102228 6.177671
## 6 Intercept         Bosea.ASV.1971 4.821218 4.363990 5.278445
```

```
unique(final_estimates_reduced$Treatment)
```

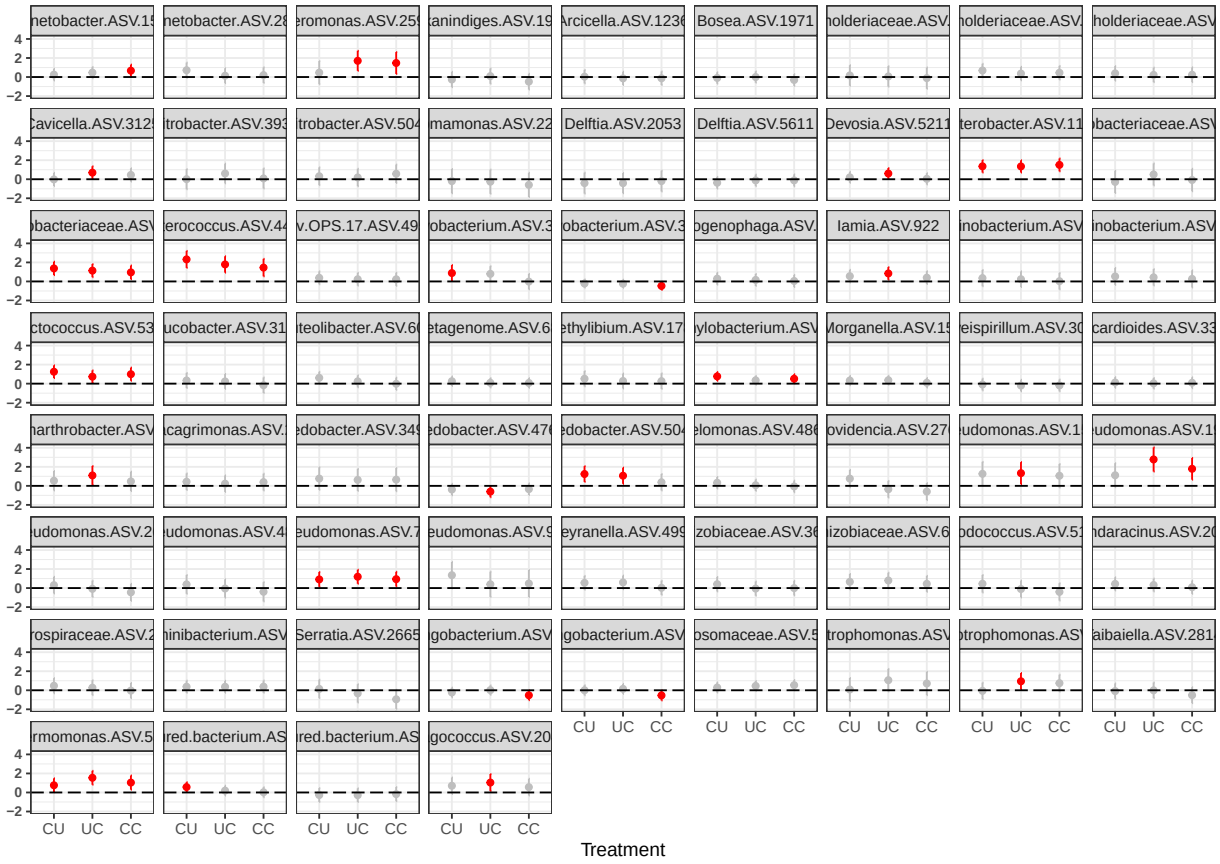
```
## [1] Intercept        Seq_depth_scaled SexM             TreatmentCC
## [5] TreatmentCU      TreatmentUC
## 6 Levels: Intercept TreatmentCU TreatmentUC TreatmentCC ... Seq_depth_scaled
```

```
final_estimates_reduced2<-subset(final_estimates_reduced, Treatment != "SexM" & Treatment != "Seq_depth

## add significance

final_estimates_reduced2$Sig<- !data.table::between(0, final_estimates_reduced2$CI_lower, final_estimat


final_estimates_reduced2$Genus<-factor(final_estimates_reduced2$Genus)

levels(final_estimates_reduced2$Treatment)
```

```
## [1] "Intercept"        "TreatmentCU"      "TreatmentUC"      "TreatmentCC"
## [5] "SexM"             "Seq_depth_scaled"
```

```
final_estimates_reduced2$Treatment<-factor(final_estimates_reduced2$Treatment)
levels(final_estimates_reduced2$Treatment)<-c("UU", "CU", "UC", "CC")



ggplot(subset(final_estimates_reduced2, Treatment != "UU"), aes(y = reorder(Genus, Estimate), x = Estima
  geom_point()+
  facet_wrap(~Treatment, nrow = 1, scales = "free_x")  +
  geom_errorbarh(aes(xmin = CI_lower, xmax = CI_upper, col = Sig), height = 0, size = 0.5)+
  geom_vline(xintercept = 0, linetype = "longdash")+
  theme_bw()+
  scale_color_manual(values = c("grey", "red"))+
  theme(axis.title.y = element_blank())+
  theme(axis.text.y = element_text(face="bold", size = 8))+
  theme(legend.position = "none")+
 theme(plot.margin=unit(c(0.2,0.2, 0.2, 0.6),"cm"))
```

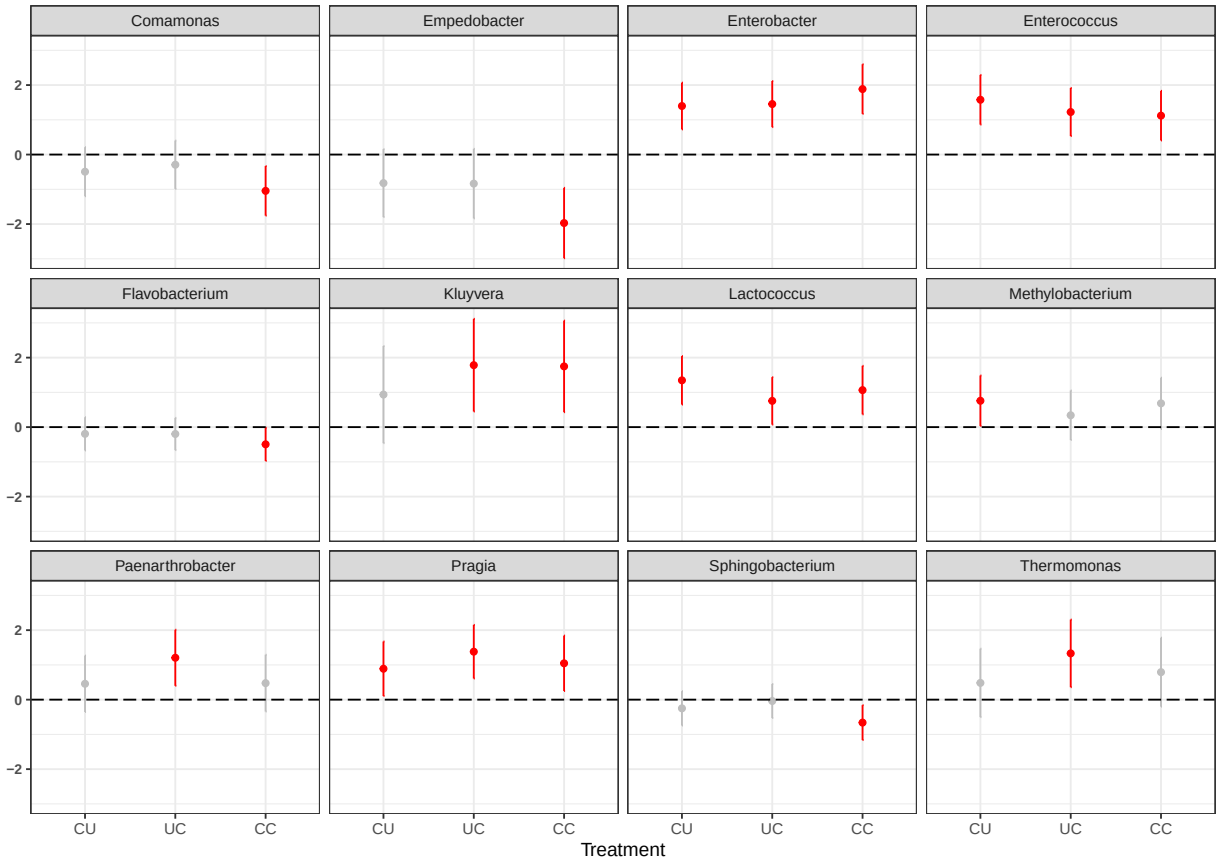

```
ggplot(subset(final_estimates_reduced2, Treatment != "UU"), aes(y = Estimate, x = Treatment, col = Sig))
  geom_point()+
  facet_wrap(~Genus)  +
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper, col = Sig), width = 0, size = 0.5)+
  geom_hline(yintercept = 0, linetype = "longdash")+
  theme_bw()+
  scale_color_manual(values = c("grey", "red"))+
  theme(axis.title.y = element_blank())+
  theme(axis.text.y = element_text(face="bold", size = 8))+
  theme(legend.position = "none")+
 theme(plot.margin=unit(c(0.2,0.2, 0.2, 0.6),"cm"))
```




```
# only significant ASVs

sig_asvs<-unique((subset(final_estimates_reduced2, Sig == "Significant"))$ASV)
final_estimates_reduced2$Keep<-final_estimates_reduced2$Genus %in% sig_asvs

estimates_sig<-subset(treatment_ci_df, Keep == TRUE)
```

```
## Error in subset(treatment_ci_df, Keep == TRUE): object 'treatment_ci_df' not found
```

```
ggplot(subset(estimates_sig, Treatment != "UU" ), aes(y = Estimate, x = Treatment, col = Sig))+
  geom_point()+
  facet_wrap(~Genus)  +
  geom_errorbar(aes(ymin = CI_lower, ymax = CI_upper, col = Sig), width = 0, size = 0.5)+
  geom_hline(yintercept = 0, linetype = "longdash")+
  theme_bw()+
  scale_color_manual(values = c("grey", "red"))+
  theme(axis.title.y = element_blank())+
  theme(axis.text.y = element_text(face="bold", size = 8))+
  theme(legend.position = "none")+
 theme(plot.margin=unit(c(0.2,0.2, 0.2, 0.6),"cm"))
```


Comamonas
Empedobacter
Enterobacter
Enterococcus
Flavobacterium
Kluyvera
Lactococcus
Methylobacterium
Paenarthrobacter
Pragia
Sphingobacterium
Thermomonas
2
0
−2
CU
UC
CC
Treatment