E. coli abundance comparisons

Kelsey Jesser

7/11/2023

```r
#rm(list = ls())
```

#load libraries

```r
library(readr)
library(ggplot2)
library(tidyr)
library(dplyr)
library(reshape2)
library(gplots)
library(ggpubr)
library(ggthemr)
library(forcats)
library(ggpmisc)
library(cowplot)
library(ggtext)
```

#set theme

```r
ggthemr('fresh')
```

#import data

```r
qPCR<-read.csv("qpCR abundances v2.csv")
isolate<-read.csv("Isolate abundances v3.csv")
```

#a few calculations

```r
#percent E. coli by qPCR (uida/total bac copies)
qPCR$perc_E.coli<-qPCR$UIDA_copies_ng/qPCR$Total_Bacteria_copies_ng * 100

#convert isolate rel abundances to percents
isolate$MG.abundance.perc<-isolate$MG.abundance *100
```

#filter and melt data for symp/asymp abundance barplots

```r
#qPCR
qPCR_melt<-qPCR %>% subset(Include=='y' & pathotype=="pathpos") %>%
  melt(id.vars=c("Sample_ID", "cc", "symp.asymp", "pathotype", "UIDA_stdev"), variable.name=("measuremen
```

```r
qPCR_melt$value<-as.numeric(qPCR_melt$value)

#isolate
isolate_melt<-isolate %>% subset(Include=="yes") %>%
  melt(id.vars=c("E..coli.isolate.ID", "Sample_ID", "cc", "symp.asymp", "Pathotype.match", "Rotavirus.")
  subset(measurement=="MG.abundance.perc")
isolate_melt$value<-as.numeric(isolate_melt$value)
```

#qPCR plots

```r
#uida (total E. coli target)
#barplot
df_uida<-qPCR_melt %>%
  subset(measurement=="UIDA_copies_ng") %>%
  drop_na("value")

df_uida.mean = df_uida %>%
  group_by(symp.asymp) %>%
  mutate(mean = mean(value))

uida_bar<-qPCR_melt %>%
  subset(measurement=="UIDA_copies_ng") %>%
  drop_na("value")%>%
  arrange(value) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, value)) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Sample_ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"))+
    ggtitle("E. coli qPCR abundance")+
    ylab("E. coli abundance (uida copies/ng)")+
    xlab(NULL)+
    geom_errorbar(data=df_uida.mean, aes(Sample_ID, ymax = mean, ymin = mean), size=0.5, linetype = "lon
    geom_errorbar(aes(ymin=value-UIDA_stdev, ymax=value+UIDA_stdev, width=0.3), color="black")+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(a

#boxplot
uida_box<-qPCR_melt %>%
  subset(measurement=="UIDA_copies_ng") %>%
  drop_na("value")%>%
  ggplot(aes(fill=symp.asymp, y=value, x=symp.asymp))+
    geom_boxplot()+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"))+
    ylab("% abundance")+
    xlab(NULL)+
    #stat_compare_means(method="wilcox.test", label.x=0.53)+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(),  legend.position="none", 

#perc E.coli (uida/total back 16S target)
#barplot
df_perc_E.coli<-qPCR_melt %>%
  subset(measurement=="perc_E.coli") %>%
  drop_na("value")
```

```r
df_perc_E.coli.mean = df_perc_E.coli %>%
  group_by(symp.asymp) %>%
  mutate(mean = mean(value))

perc_E.coli_bar<-qPCR_melt %>%
  subset(measurement=="perc_E.coli") %>%
  drop_na("value")%>%
  arrange(value) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, value)) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Sample_ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    ggtitle(expression(paste(italic("E. coli")," qPCR abundance")))+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_perc_E.coli.mean, aes(Sample_ID, ymax = mean, ymin = mean), size=0.5, linetyp
   scale_y_continuous(breaks=seq(0,8,4))+
   theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(a

#boxplot
perc_E.coli_box<-qPCR_melt %>%
  subset(measurement=="perc_E.coli") %>%
  drop_na("value")%>%
  ggplot(aes(fill=symp.asymp, y=value, x=symp.asymp))+
    geom_boxplot()+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    xlab(NULL)+
    scale_y_continuous(breaks=seq(0,8,4))+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), legend.position="none", pa

uida_plot<- ggdraw() +
    draw_plot(uida_bar) +
    draw_plot(uida_box, x = 0.675, y = 0.55, width = 0.25, height = 0.325)

uida_plot
```
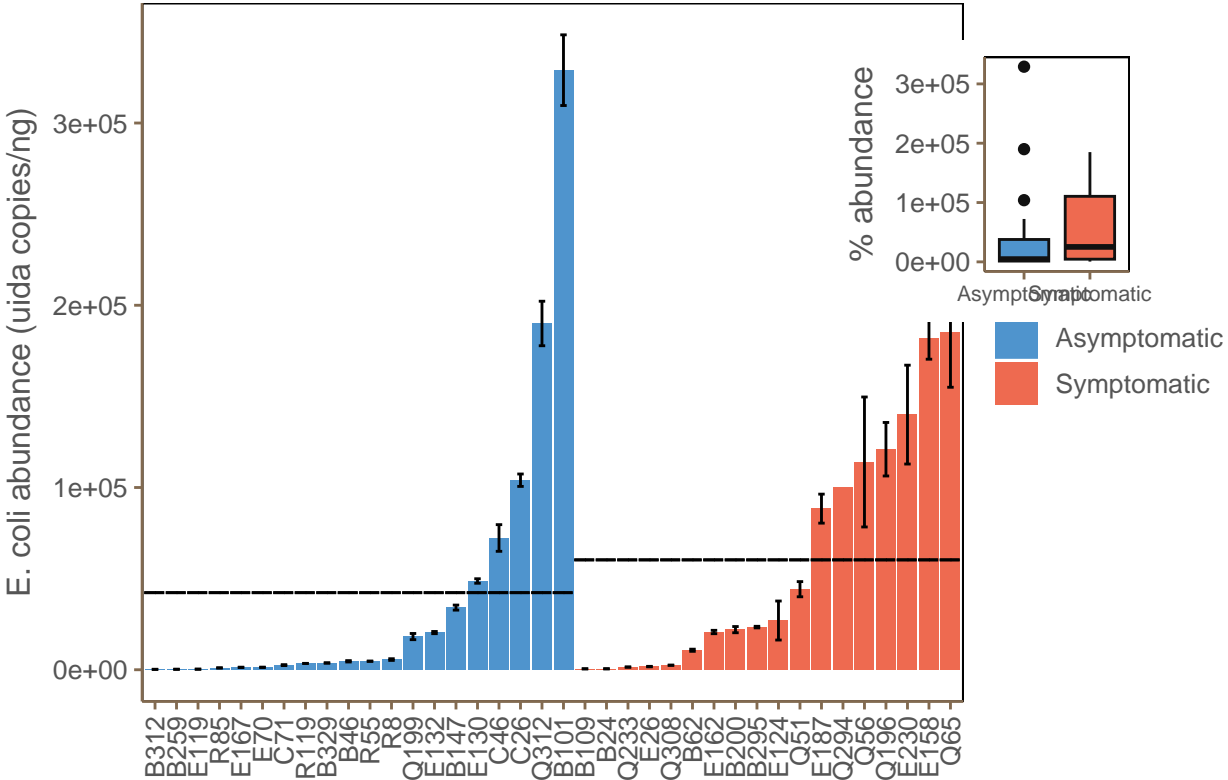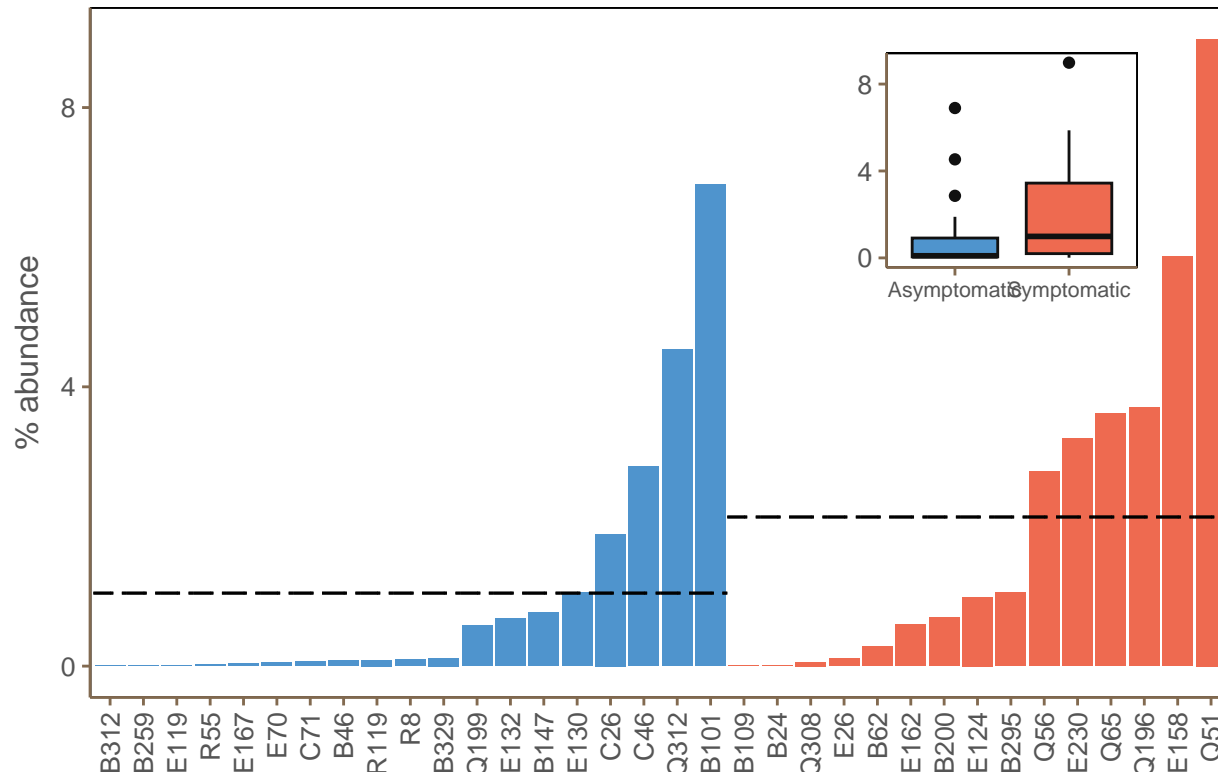
E. coli qPCR abundance


```
perc_E.coli_plot<- ggdraw() +
    draw_plot(perc_E.coli_bar) +
    draw_plot(perc_E.coli_box, x = 0.675, y = 0.55, width = 0.25, height = 0.325)

perc_E.coli_plot
```

E. coli qPCR abundance

#Isolate abundance plots

```
#isolate abundance
#barplot
df_isolate.mean = isolate_melt %>%
  group_by(symp.asymp) %>%
  mutate(mean = mean(value))

isolate_bar<-isolate_melt %>%
  #arrange(value) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, value)) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Sample_ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infection
    ggtitle("DEC isolate metagenome abundance")+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_isolate.mean, aes(Sample_ID, ymax = mean, ymin = mean), size=0.5, linetype =
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(
    theme(axis.line.x.bottom=element_line(color="black"))+
    theme(axis.line.y.left=element_line(color="black"))


#boxplot
isolate_box<-isolate_melt %>%
```

5

```
  ggplot(aes(fill=symp.asymp, y=value, x=symp.asymp))+
    geom_boxplot()+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    xlab(NULL)+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), legend.position="none", p
    theme(axis.line = element_line(color="black", size =.1), axis.ticks=element_line(color="black"))

isolate_plot<-ggdraw() +
    draw_plot(isolate_bar) +
    draw_plot(isolate_box, x = 0.675, y = 0.55, width = 0.25, height = 0.325)

isolate_plot
```

DEC isolate metagenome abundance
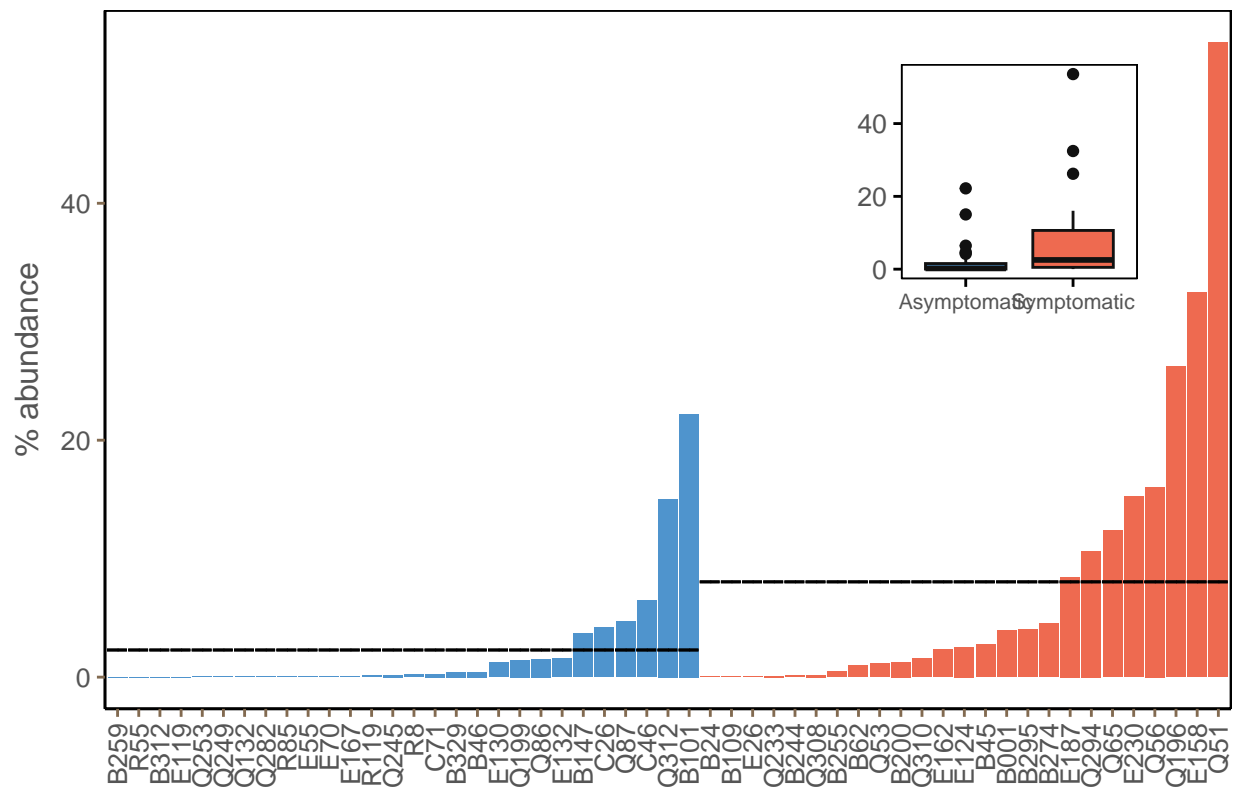
#qPCR stats

```
#Total E. coli (uida gene copies) symp.asymp comparison
qPCR_melt_symp_uida<-qPCR_melt %>%
  subset(symp.asymp=="Symptomatic" & measurement=="UIDA_copies_ng") %>%
  drop_na("value")
qPCR_melt_asymp_uida<-qPCR_melt %>%
  subset(symp.asymp=="Asymptomatic" & measurement=="UIDA_copies_ng") %>%
  drop_na("value")

wilcox.test(qPCR_melt_symp_uida$value, qPCR_melt_asymp_uida$value)
```

##

```
##  Wilcoxon rank sum exact test
##
## data:  qPCR_melt_symp_uida$value and qPCR_melt_asymp_uida$value
## W = 231, p-value = 0.1412
## alternative hypothesis: true location shift is not equal to 0
```

```r
#Percent E. coli symp.asymp comparison
qPCR_melt_symp_perc<-qPCR_melt %>%
  subset(symp.asymp=="Symptomatic" & measurement=="perc_E.coli") %>%
  drop_na("value")

#Percent E. coli symp.asymp comparison
qPCR_melt_asymp_perc<-qPCR_melt %>%
  subset(symp.asymp=="Asymptomatic" & measurement=="perc_E.coli") %>%
  drop_na("value")

wilcox.test(qPCR_melt_symp_perc$value, qPCR_melt_asymp_perc$value)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  qPCR_melt_symp_perc$value and qPCR_melt_asymp_perc$value
## W = 189, p-value = 0.1114
## alternative hypothesis: true location shift is not equal to 0
```

#isolate stats

```r
#isolate
isolate_melt_symp<-isolate_melt %>%
  subset(symp.asymp=="Symptomatic") %>%
  drop_na("value")
isolate_melt_asymp<-isolate_melt %>%
  subset(symp.asymp=="Asymptomatic") %>%
  drop_na("value")

wilcox.test(isolate_melt_symp$value, isolate_melt_asymp$value)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  isolate_melt_symp$value and isolate_melt_asymp$value
## W = 506, p-value = 0.004936
## alternative hypothesis: true location shift is not equal to 0
```

#qPCR and isolate metagenome abundance comparisons

```r
#join abundance data
isolate_qPCR<- left_join(isolate, qPCR, by=c("Sample_ID"))

#formula
formula <- y ~ x
```

```
#percent E. coli
isolate_qPCR_scat_perc<-ggplot(isolate_qPCR, aes(x=perc_E.coli, y=MG.abundance.perc))+
  geom_point(size=3)+
  xlab(expression(paste(italic("E. coli")," qPCR abundance (%)")))+
  ylab("DEC isolate metagenome abundance (%)")+
  geom_smooth(method=lm, linetype="dashed", color="black", se=FALSE)+
  stat_poly_eq(aes(label = paste(..rr.label..)), formula = formula, parse = TRUE, label.y.npc=.99, label
  stat_fit_glance(method = 'lm', method.args = list(formula = formula), geom = 'text', aes(label = past
  theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), legend.position="none", pan
isolate_qPCR_scat_perc
```
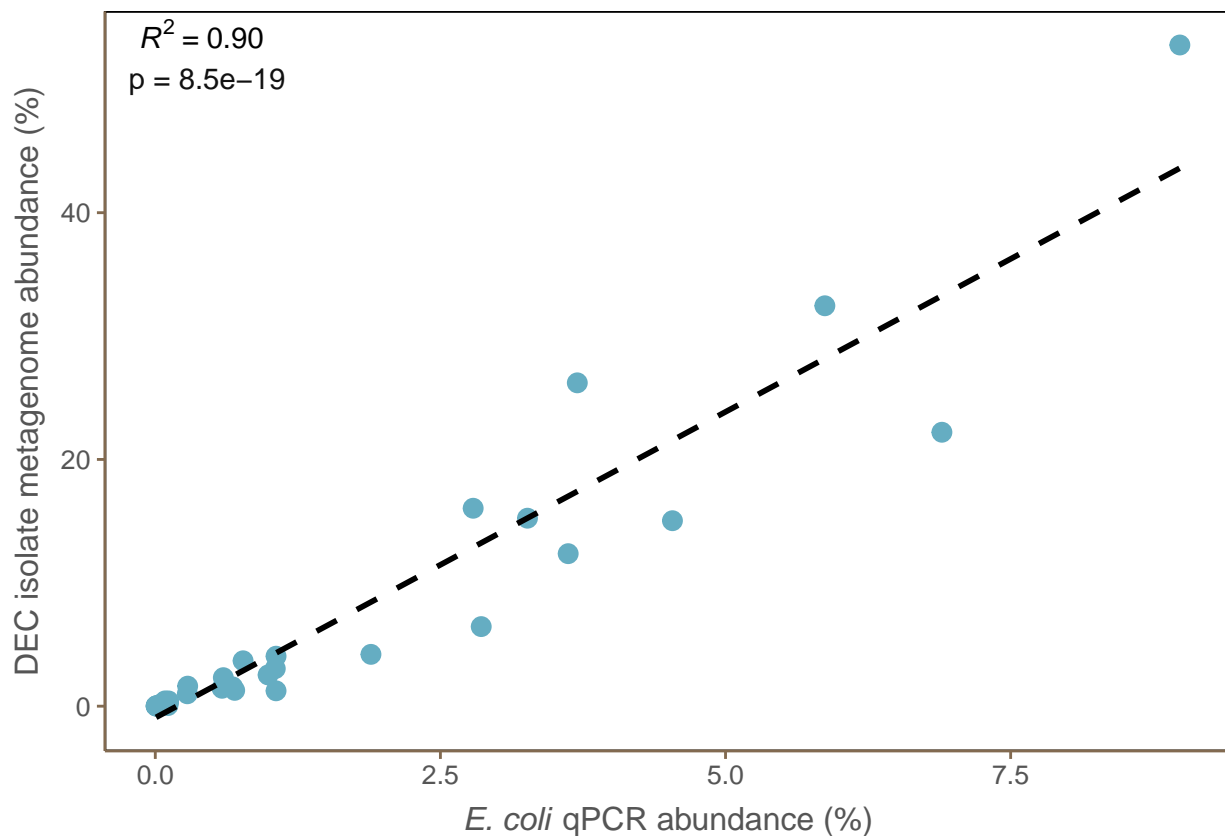

#extract legend

```
isolate_legend<-isolate_melt %>%
  #arrange(value) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, value)) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Sample_ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    #scale_fill_discrete(labels=c("Symptomatic DEC infections"="coral2", "Asymptomatic DEC infections"=
    ggtitle("DEC isolate metagenome abundance")+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_isolate.mean, aes(Sample_ID, ymax = mean, ymin = mean), size=0.5, linetype = 
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(
```

```
legend<-get_legend(isolate_legend)
l<-as_ggplot(legend)
l
```
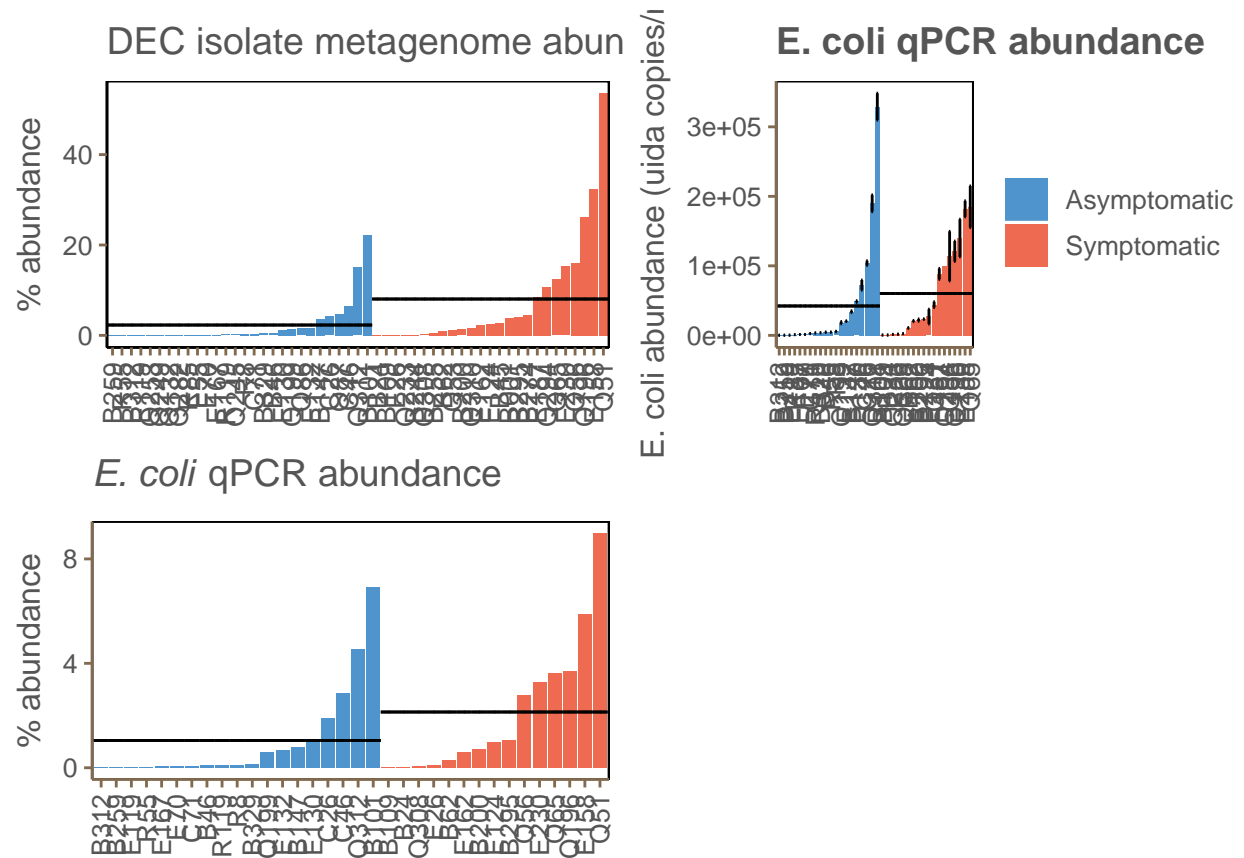
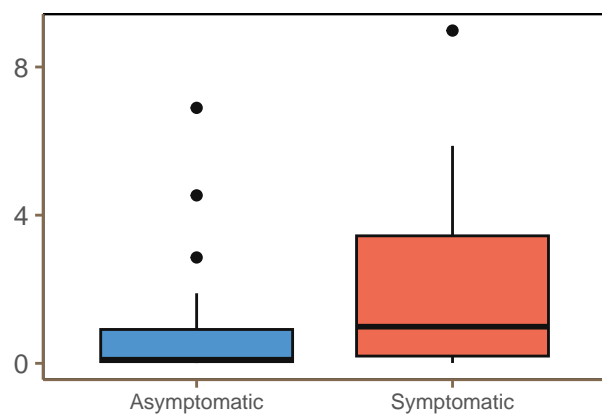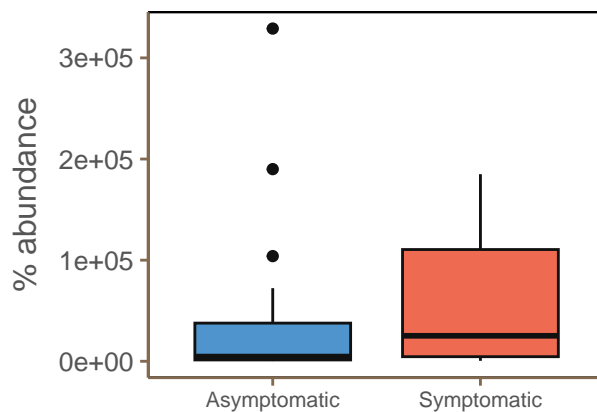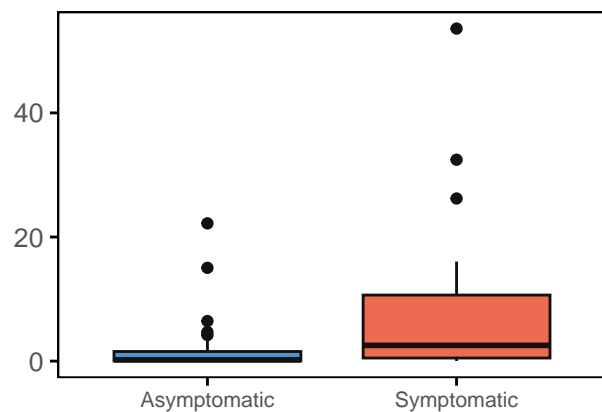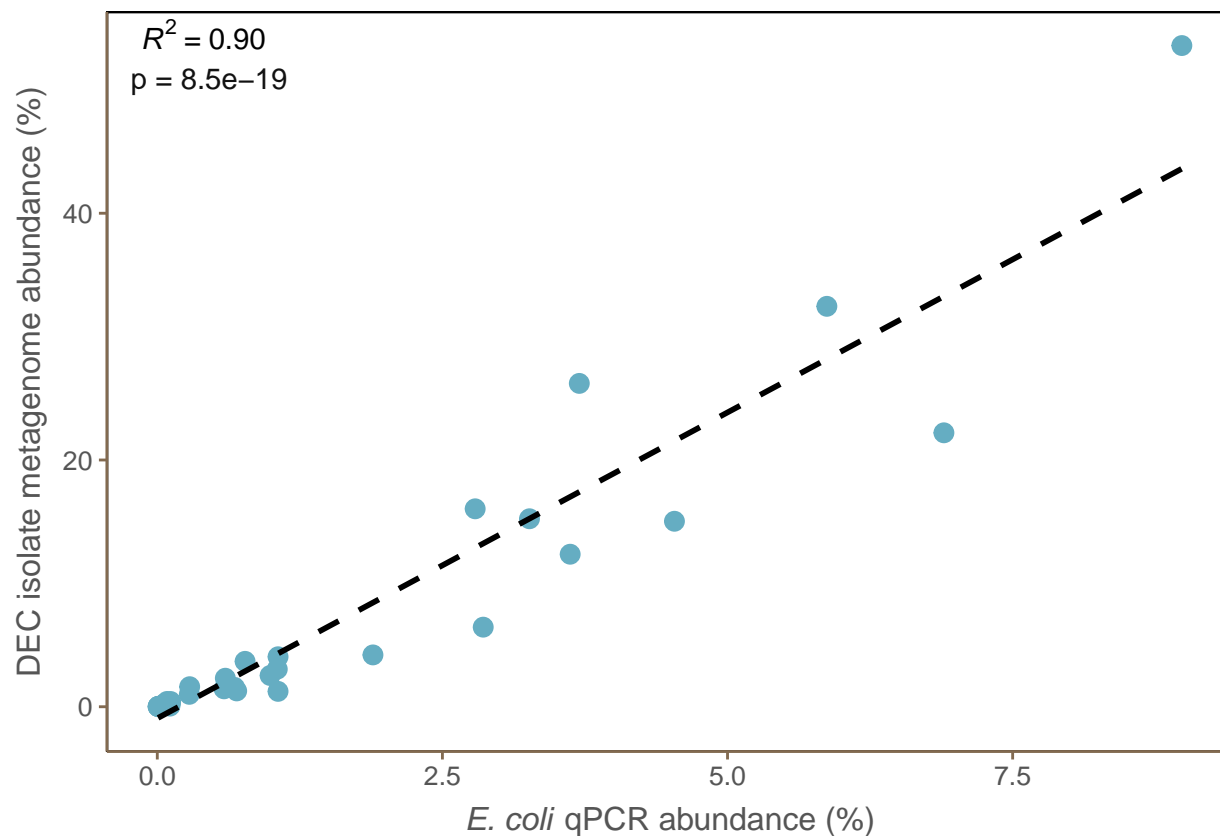■ Symptomatic DEC infections   ■ Asympotmatic DEC infections

#multiplots

```
#barplots
plot_grid(isolate_bar, uida_bar, perc_E.coli_bar, ncol=2)
```

```r
#boxplots
plot_grid(isolate_box, uida_box, perc_E.coli_box, ncol=2)
```

```
#comparisons
isolate_qPCR_scat_perc
```

$R^2 = 0.90$
p = 8.5e−19

DEC isolate metagenome abundance (%)

E. coli qPCR abundance (%)

```
#inset multiplot
p<-plot_grid(isolate_plot,
             perc_E.coli_plot,
             l,
             ncol=1, align="hv",
             labels=c("A", "B", ""),
             label_size=20,
             rel_heights = c(1, 1, 0.25),
             rel_widths=c(1, .9, 1))
q<-plot_grid(p)+
    draw_text("Wilcoxon, p=0.0049", x = 0.1725, y = .925, size = 10)+
    draw_text("Wilcoxon, p=0.11", x=0.175, y=0.475, size = 10)
q
```

A DEC isolate metagenome abundance

Wilcoxon, p=0.0049

B *E. coli* qPCR abundance

Wilcoxon, p=0.11

Symptomatic DEC infections   Asympotmatic DEC infections
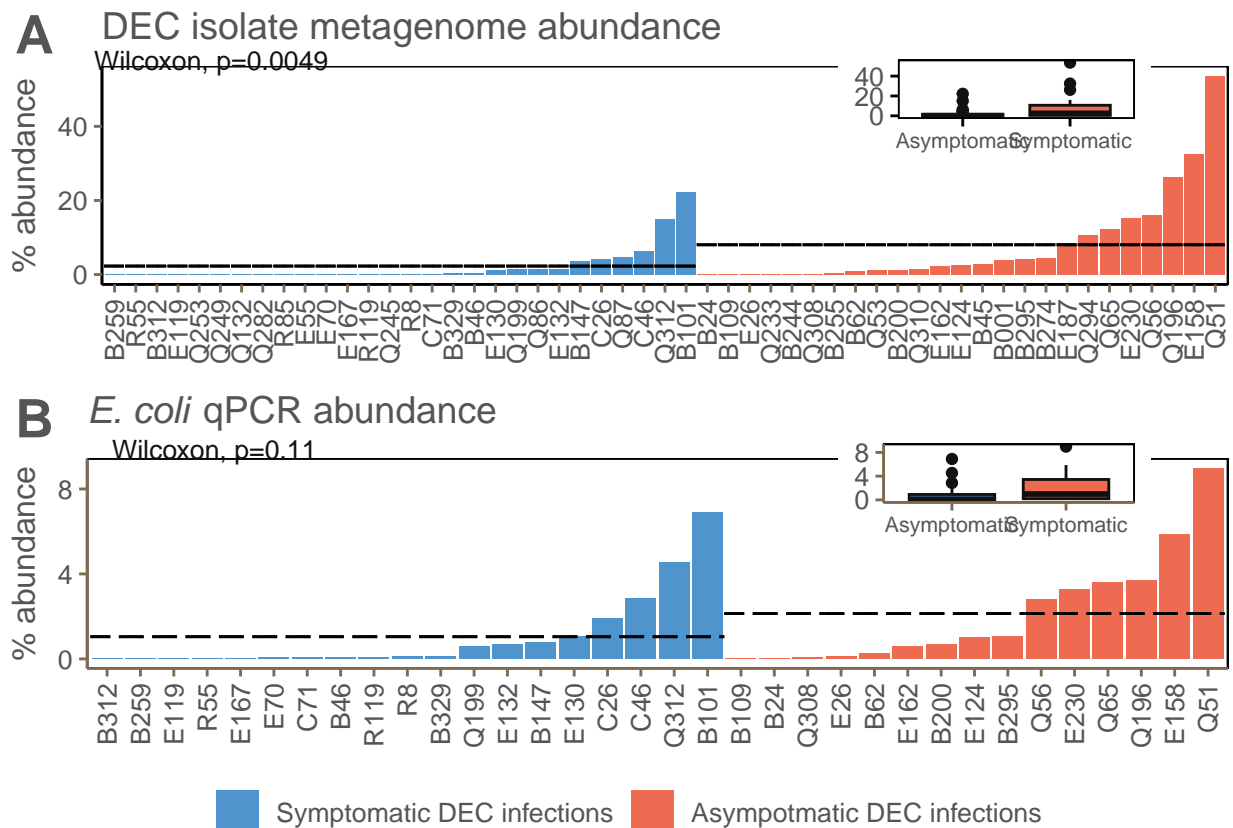
#nicely formatted isolate barplot

```
isolate_bar<-isolate_melt %>%
  #arrange(value) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, value)) %>%
  mutate(Sample_ID=fct_reorder(Sample_ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Sample_ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), labels=c("Asymptoma
    #ggtitle("DEC isolate metagenome abundance")+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_isolate.mean, aes(Sample_ID, ymax = mean, ymin = mean), size=0.5, linetype = 
    scale_y_continuous(expand=c(0,0), limits=c(0,60), breaks=seq(0, 60, by=25))+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(
    theme(axis.line = element_line(color="black", size =.1), axis.ticks=element_line(color="black"))

isolate_plot<-ggdraw() +
    draw_plot(isolate_bar) +
    draw_plot(isolate_box, x = 0.675, y = 0.625, width = 0.25, height = 0.325)

isolate_plot<-plot_grid(isolate_plot)+draw_text("Wilcoxon, p=0.0049", x = 0.17, y = .95, size = 10)
isolate_plot
```
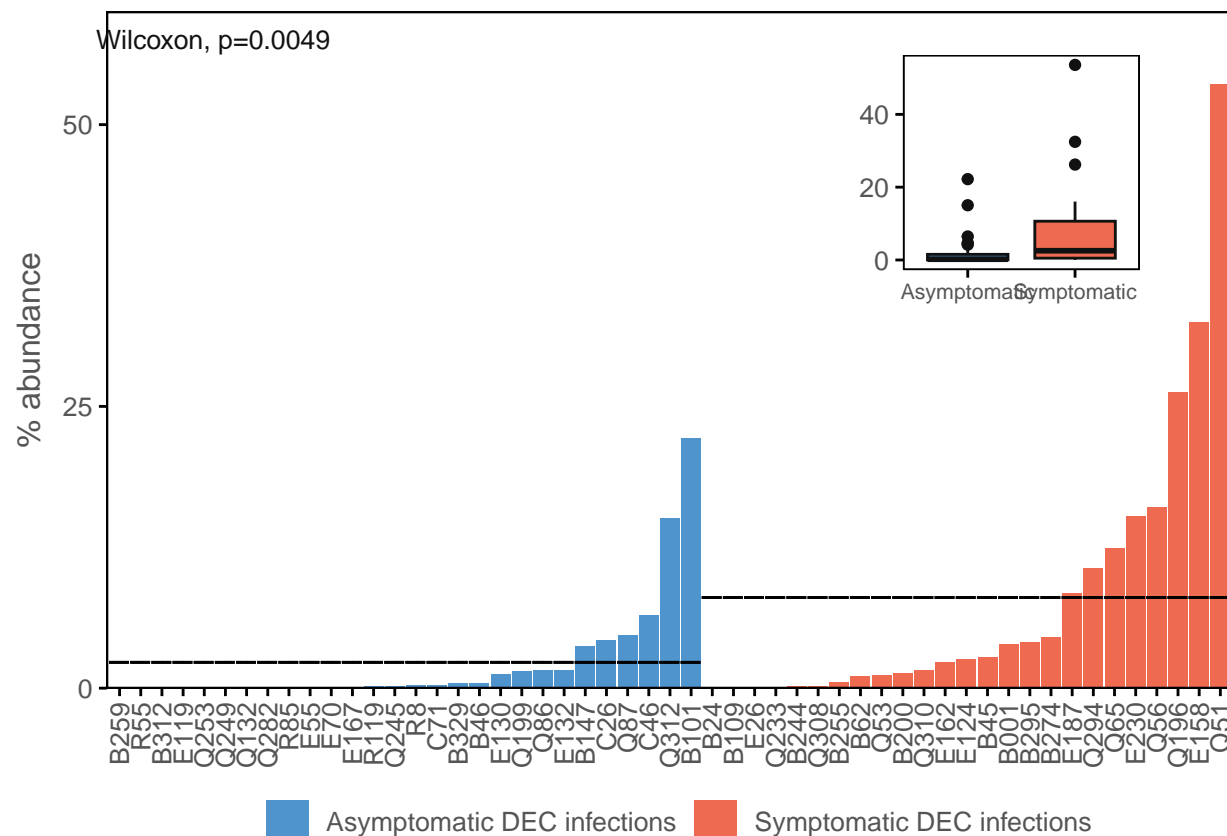
#Subset metagenome isolate abundances by age category

```r
ages<-read.csv("ages_metadata_final.csv")%>%
  select(c(Alternate.Sample.ID, Age.years, Age.months))

isolate_melt_age<-isolate_melt%>%
  rename(Alternate.Sample.ID=Sample_ID)%>%
  left_join(ages, by="Alternate.Sample.ID")

isolate_melt_adult<-isolate_melt_age%>%
  mutate(Age.years=as.numeric(Age.years))%>%
  subset(Age.years>=15) #only one sample, no additional analyses

isolate_melt_child<-isolate_melt_age%>%
  mutate(Age.years=as.numeric(Age.years))%>%
  subset(Age.years<=15)%>%
  subset(Age.years>=5)

isolate_melt_toddler<-isolate_melt_age%>%
  mutate(Age.years=as.numeric(Age.years))%>%
  subset(Age.years<=4)%>%
  subset(Age.years>=2)

isolate_melt_infant<-isolate_melt_age%>%
  mutate(Age.years=as.numeric(Age.years))%>%
  subset(Age.years<=1)%>%
```

```
    subset(Age.years>=0)
```

#Compare isolate abundances in metagenomes for child samples

```
#isolate abundance
#barplot
df_isolate.mean = isolate_melt_child %>%
  group_by(symp.asymp) %>%
  mutate(mean = mean(value))

isolate_bar_child<-isolate_melt_child %>%
  #arrange(value) %>%
  mutate(Alternate.Sample.ID=fct_reorder(Alternate.Sample.ID, value)) %>%
  mutate(Alternate.Sample.ID=fct_reorder(Alternate.Sample.ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Alternate.Sample.ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    ggtitle("DEC isolate metagenome abundance in children (ages 5-15y)")+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_isolate.mean, aes(Alternate.Sample.ID, ymax = mean, ymin = mean), size=0.5, l
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(
    theme(axis.line.x.bottom=element_line(color="black"))+
    theme(axis.line.y.left=element_line(color="black"))+
  ylim(0, 2.5)

#boxplot
isolate_box_child<-isolate_melt_child %>%
  ggplot(aes(fill=symp.asymp, y=value, x=symp.asymp))+
    geom_boxplot()+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infection
    xlab(NULL)+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), legend.position="none", p
    theme(axis.line = element_line(color="black", size =.1), axis.ticks=element_line(color="black"))

isolate_plot_child<-ggdraw() +
    draw_plot(isolate_bar_child) +
    draw_plot(isolate_box_child, x = 0.6, y = 0.55, width = 0.25, height = 0.3)

#stats
isolate_melt_symp_child<-isolate_melt_child %>%
  subset(symp.asymp=="Symptomatic") %>%
  drop_na("value")
isolate_melt_asymp_child<-isolate_melt_child %>%
  subset(symp.asymp=="Asymptomatic") %>%
  drop_na("value")

wilcox.test(isolate_melt_symp_child$value, isolate_melt_asymp_child$value)


##
##  Wilcoxon rank sum exact test
##
## data:  isolate_melt_symp_child$value and isolate_melt_asymp_child$value
```

```
## W = 12, p-value = 0.25
## alternative hypothesis: true location shift is not equal to 0
```

```
isolate_plot_child<- isolate_plot_child+draw_text("Wilcoxon, p=0.25", x = 0.17, y = .875, size = 10)
```

#Compare isolate abundances in metagenomes for toddler samples

```r
#isolate abundance
#barplot
df_isolate.mean = isolate_melt_toddler %>%
  group_by(symp.asymp) %>%
  mutate(mean = mean(value))

isolate_bar_toddler<-isolate_melt_toddler %>%
  #arrange(value) %>%
  mutate(Alternate.Sample.ID=fct_reorder(Alternate.Sample.ID, value)) %>%
  mutate(Alternate.Sample.ID=fct_reorder(Alternate.Sample.ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Alternate.Sample.ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    ggtitle("DEC isolate metagenome abundance in toddlers (ages 2-5y)")+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_isolate.mean, aes(Alternate.Sample.ID, ymax = mean, ymin = mean), size=0.5, l
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(a
    theme(axis.line.x.bottom=element_line(color="black"))+
    theme(axis.line.y.left=element_line(color="black"))

#boxplot
isolate_box_toddler<-isolate_melt_toddler %>%
  ggplot(aes(fill=symp.asymp, y=value, x=symp.asymp))+
    geom_boxplot()+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    xlab(NULL)+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), legend.position="none", pa
    theme(axis.line = element_line(color="black", size =.1), axis.ticks=element_line(color="black"))

isolate_plot_toddler<-ggdraw() +
    draw_plot(isolate_bar_toddler) +
    draw_plot(isolate_box_toddler, x = 0.65, y = 0.6, width = 0.25, height = 0.3)

#stats
isolate_melt_symp_toddler<-isolate_melt_toddler %>%
  subset(symp.asymp=="Symptomatic") %>%
  drop_na("value")
isolate_melt_asymp_toddler<-isolate_melt_toddler %>%
  subset(symp.asymp=="Asymptomatic") %>%
  drop_na("value")

wilcox.test(isolate_melt_symp_toddler$value, isolate_melt_asymp_toddler$value)
```

```
##
##  Wilcoxon rank sum exact test
```

```
##
## data:  isolate_melt_symp_toddler$value and isolate_melt_asymp_toddler$value
## W = 61, p-value = 0.07701
## alternative hypothesis: true location shift is not equal to 0
```

```
isolate_plot_toddler<- isolate_plot_toddler+draw_text("Wilcoxon, p=0.077", x = 0.18, y = .875, size = 1(
```

#Compare isolate abundances in metagenomes for infant samples

```
#isolate abundance
#barplot
df_isolate.mean = isolate_melt_infant %>%
  group_by(symp.asymp) %>%
  mutate(mean = mean(value))

isolate_bar_infant<-isolate_melt_infant %>%
  #arrange(value) %>%
  mutate(Alternate.Sample.ID=fct_reorder(Alternate.Sample.ID, value)) %>%
  mutate(Alternate.Sample.ID=fct_reorder(Alternate.Sample.ID, symp.asymp)) %>%
  ggplot(aes(fill=symp.asymp, y=value, x=Alternate.Sample.ID))+
    geom_bar(stat="identity")+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infectio
    ggtitle("DEC isolate metagenome abundance in infants (ages 0-2 y)")+
    ylab("% abundance")+
    xlab(NULL)+
    geom_errorbar(data=df_isolate.mean, aes(Alternate.Sample.ID, ymax = mean, ymin = mean), size=0.5, l
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), axis.text.x=element_text(a
    theme(axis.line.x.bottom=element_line(color="black"))+
    theme(axis.line.y.left=element_line(color="black"))

#boxplot
isolate_box_infant<-isolate_melt_infant %>%
  ggplot(aes(fill=symp.asymp, y=value, x=symp.asymp))+
    geom_boxplot()+
    scale_fill_manual(values=c("Symptomatic"="coral2", "Asymptomatic"="steelblue3"), name="DEC infection
    xlab(NULL)+
    theme(panel.grid.major=element_blank(), panel.grid.minor=element_blank(), legend.position="none", pa
    theme(axis.line = element_line(color="black", size =.1), axis.ticks=element_line(color="black"))

isolate_plot_infant<-ggdraw() +
    draw_plot(isolate_bar_infant) +
    draw_plot(isolate_box_infant, x = 0.65, y = 0.575, width = 0.25, height = 0.3)

#stats
isolate_melt_symp_infant<-isolate_melt_infant %>%
  subset(symp.asymp=="Symptomatic") %>%
  drop_na("value")
isolate_melt_asymp_infant<-isolate_melt_infant %>%
  subset(symp.asymp=="Asymptomatic") %>%
  drop_na("value")

wilcox.test(isolate_melt_symp_infant$value, isolate_melt_asymp_infant$value)
```

```
##
```

```
##  Wilcoxon rank sum exact test
##
## data:  isolate_melt_symp_infant$value and isolate_melt_asymp_infant$value
## W = 121, p-value = 0.1546
## alternative hypothesis: true location shift is not equal to 0
```

```
isolate_plot_infant<-isolate_plot_infant+draw_text("Wilcoxon, p=0.15", x = 0.17, y = .875, size = 10)
```

#multiplot by age

```
r<-plot_grid(isolate_plot_infant,
             isolate_plot_toddler,
             isolate_plot_child,
             l,
             ncol=1, align="hv",
             labels=c("A", "B", "C", ""),
             label_size=20,
             rel_heights = c(1, 1, 1, 0.25),
             rel_widths=c(1, 1, 1, 1))
r
```