Main_Figure_4

Micheál Mac Aogáin

2023-06-13

#Knitr setup

```
knitr::opts_chunk$set(echo = TRUE)
```

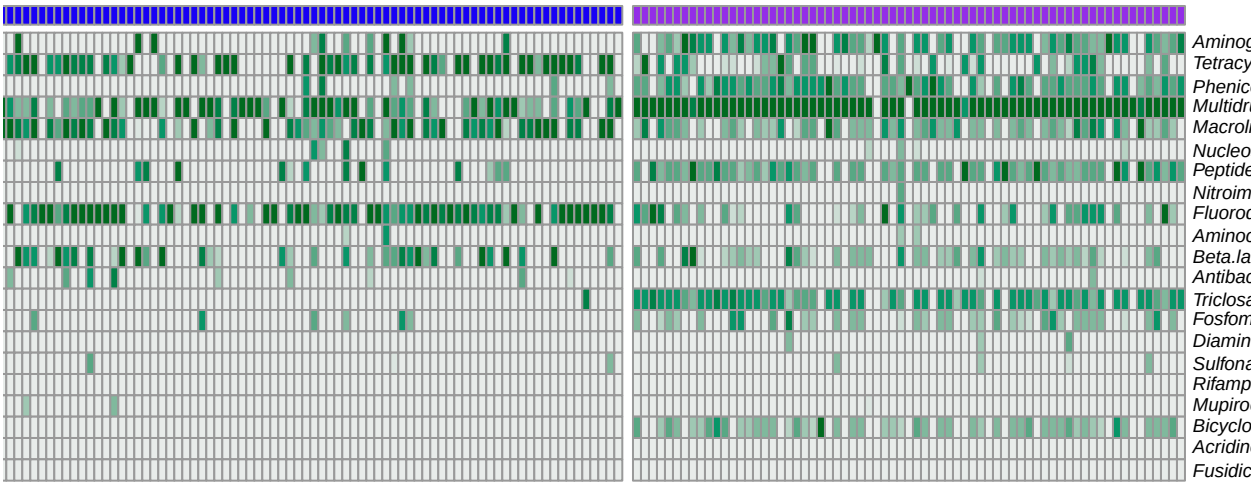
#Load required R packages

```
# Load required R packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(ggplot2, tidyverse, vegan, vcd, knitr,funrar,pheatmap,colorspace,ggpubr, dplyr)
```

#Figure 4A Heatmap by Ivan

```
#Data formatting for Heatmap figures
project_name <- "CAMEB2"
wkdir <- file.path("/Data/R_input_files/Ivan_AMR_analysis_2020", project_name)
#setwd(wkdir)

Omic <- "Microbiome-AMR"
groupvar <- "SC_AMR_alt"
#groupvar <- "Continent"
#my.labels <- c("H", "B")
my.labels <- c("1", "2")

#taxalevel
taxLevel <- "Drug"
##Read ps object####
ps <- readRDS(file.path("..",wkdir, "analysis", "AMR", paste0("ps_", taxLevel, ".RData")))

#.AMR data####
ab_data=make_relative(as.matrix(read.csv("../Data/R_input_files/AMR_CAMEB2_Drug.csv", row.names = 1)))*
ab_data[is.nan(ab_data)] <- 0
ab_data<-as.data.frame(ab_data)

#Master data####
Master <-read.csv("../Data/R_input_files//Clinical_AMR_Microbiome_R2.csv") %>%
  as_tibble() %>%
  dplyr::filter(SC_AMR_alt != 0)
ab_data <- subset(ab_data, rownames(ab_data) %in% Master$SampleID)
#RT1 and RT2

##Sample ordering####
ab_data$SC_AMR_alt<-ifelse(is.na(Master$SC_AMR_alt) == TRUE,  "H", Master$SC_AMR_alt)
#ab_data$SC_AMR_alt<-ifelse(is.na(Master$SC_AMR_alt) == TRUE,  "H", "B")
#ab_data$SC_AMR_alt <- factor(ab_data$SC_AMR_alt, levels = c("H","B")) #set ranking for order of bars
```

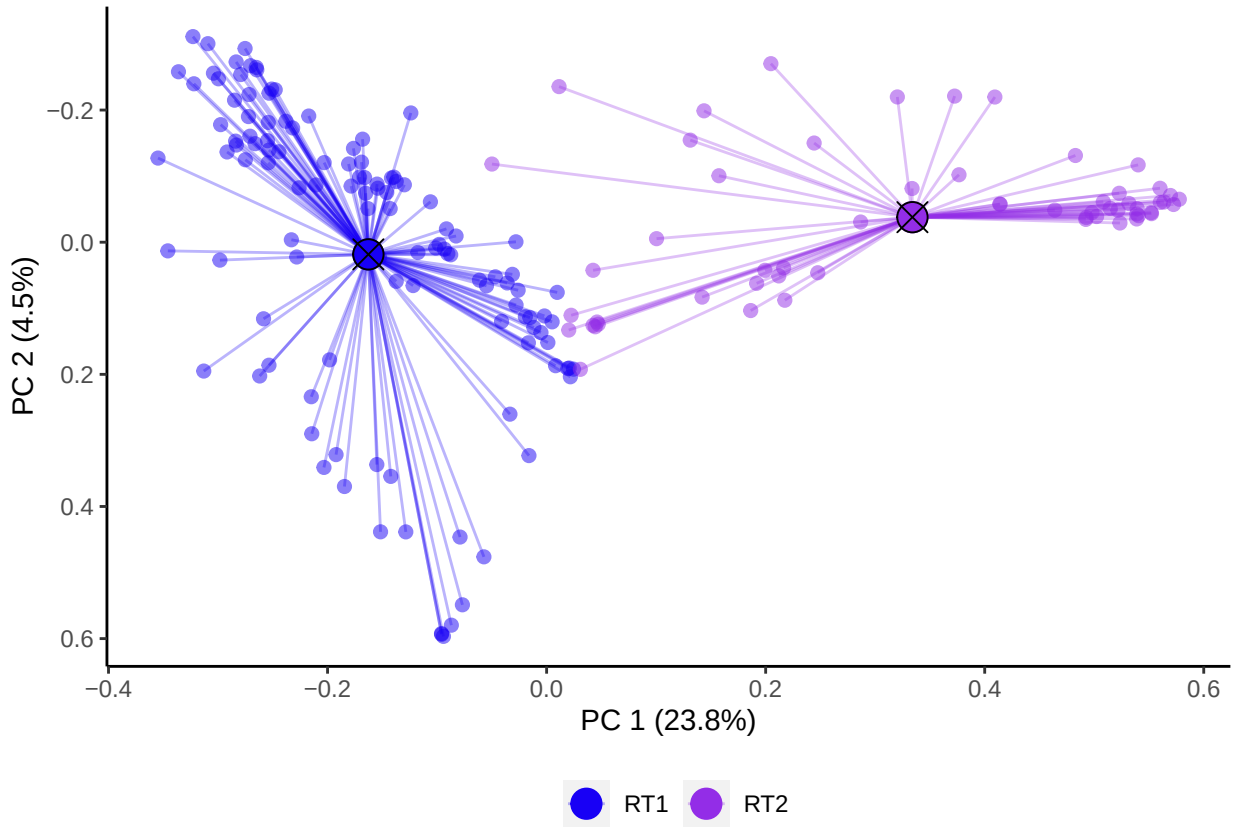
```
groupvar = "SC_AMR_alt"

sorted_group <- sort(as.vector(data.frame(ab_data)[ ,groupvar]), index.return= TRUE)

starts <- which(!duplicated(sorted_group$x))
sample_order <- c()
for (i in 1:length(starts)) {
  if (i!=length(starts)) {
    sample_order <- c(sample_order, sorted_group$ix[starts[i]:(starts[i+1]-1)])
  } else {
    sample_order <- c(sample_order, sorted_group$ix[starts[i]:length(sorted_group$ix)])
  }
}

#.AMR data reload####
ab_data_otu=make_relative(as.matrix(read.csv("../Data/R_input_files/AMR_CAMEB2_Drug.csv", row.names = 1
ab_data_otu[is.nan(ab_data_otu)] <- 0
ab_data_otu<-as.data.frame(ab_data_otu)
#ab_data_otu<-ab_data_otu[row.names(ab_data_otu) != "TBS672", , drop = FALSE]

ab_data_otu <- subset(ab_data_otu, rownames(ab_data_otu) %in% Master$SampleID)


##Prepare heatmap components####
lefsetaxa<-(ab_data_otu)
df <- t(ab_data_otu[sample_order, colnames(lefsetaxa)])
#rownames(df) <- c(tax_table(ps.prop)[rownames(df), taxLevel])
df <- log(df+1,2)

mymat <- as.matrix(data.frame(df))
mydf <- data.frame(row.names= as.vector(data.frame(Master)[sample_order, "SampleID"]), category= as.vec

##Color and text setting####
#lightenParams <- seq(0, 0.9999, 1/55)
#col1 <- lighten("#152CAD", lightenParams)[length(lightenParams):1]
col1 <- sequential_hcl(12, h = 150, c = 80, l = c(35, 95), rev = TRUE, power = 2)[c(3, rep(4, 2), rep(5
newnames <- lapply(rownames(mymat), function(x) bquote(italic(.(x))))
my.labelsCol <- c("#1800F5", "#DF9A8F", "#1B9E77", "#932DE7"); names(my.labelsCol) <- my.labels
ann_colors = list(`category` = my.labelsCol)

##Plotting heatmap####
pheatmap(mymat, color= col1, cluster_cols = FALSE, cluster_rows = FALSE, annotation_col = mydf, annotati
         show_colnames = FALSE, labels_row = as.expression(newnames), cellheight = 8, cellwidth = 3,
         gaps_col = starts[2:length(starts)]-1, breaks= seq(0, 4.5, 0.1),
         annotation_names_row= FALSE, annotation_names_col= FALSE, annotation_legend= FALSE, fontsize= 
```



#Figure 4DEF

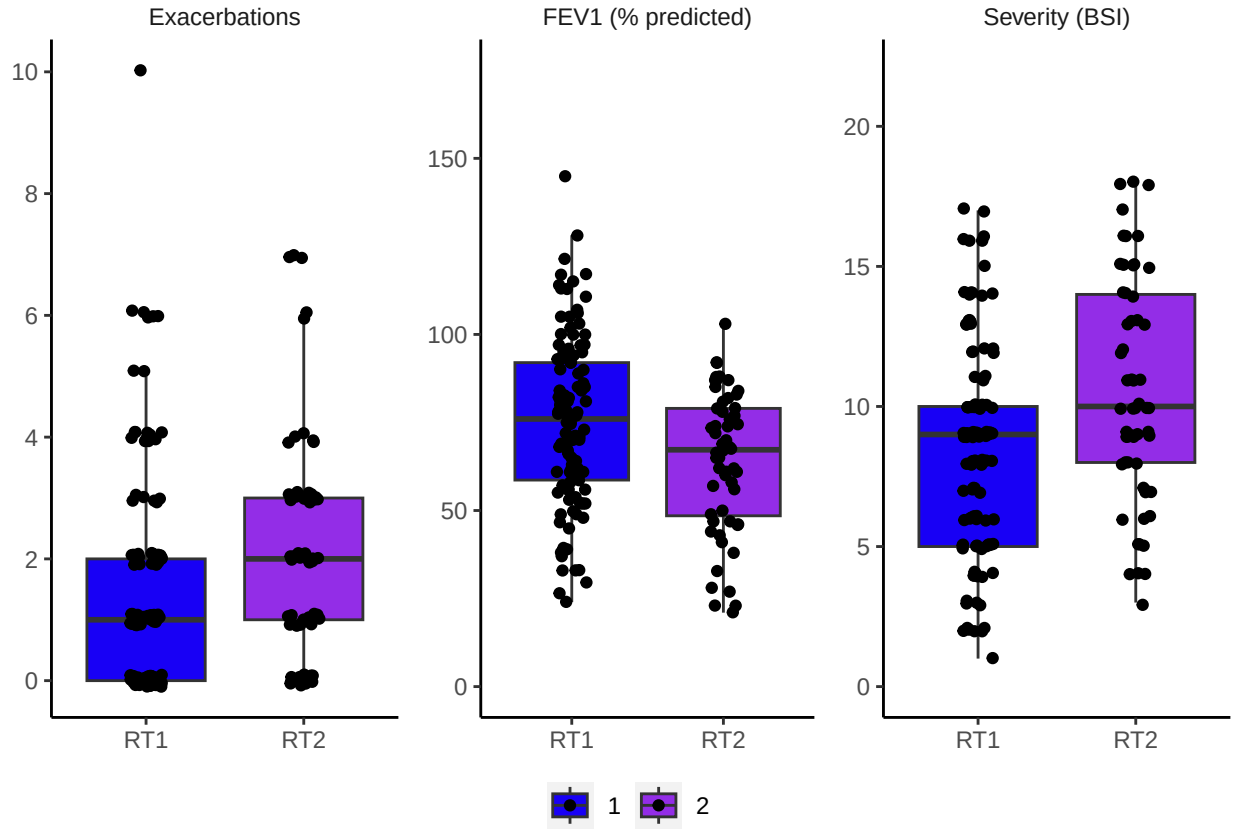
```
## RE-wranagle data for PCoA visualization####
AMR_diversity <- Master %>%
  as_tibble() %>%
  select(1:1,64:314) #for genes
  #select(1:1,43:63) #for amr drug class
NAMES_list <- AMR_diversity$SampleID
main_data <- AMR_diversity[AMR_diversity$SampleID %in% NAMES_list, ]
AMR_diversity<-as.matrix(AMR_diversity)
rownames(AMR_diversity) <- AMR_diversity[,1]
AMR_diversity = as.data.frame(subset(AMR_diversity, select = -c(SampleID) ))
AMR_diversity[] <- lapply(AMR_diversity, as.numeric)
#AMR_diversity<-AMR_diversity[row.names(AMR_diversity) != "TBS672", , drop = FALSE]

isZero <- base::rowSums(AMR_diversity) == 0
AMR_diversity<-AMR_diversity[!isZero,]

MasterVIZ = Master
MasterVIZ$select <- ifelse(MasterVIZ$SC_AMR_alt==0, "null", "Bronchiectasis")
MasterVIZ$select <- ifelse(is.na(MasterVIZ$select), "Non-diseased", MasterVIZ$select)
MasterVIZ$SC_AMR_alt <- ifelse(is.na(MasterVIZ$SC_AMR_alt), "Non-diseased", MasterVIZ$SC_AMR_alt)
AMRDiversityViz<-subset(MasterVIZ, select != "null")
#AMRDiversityViz<-AMRDiversityViz[AMRDiversityViz$SampleID != "TBS153", , drop = FALSE] #remove for gen

AMRDiversityViz_Geo<-subset(AMRDiversityViz, is.na(AMRDiversityViz$SC_AMR_alt) == FALSE & AMRDiversityVi
```

```r
AMR_diversity <- AMR_diversity[ row.names(AMR_diversity) %in% AMRDiversityViz_Geo$SampleID, ]

vegdist(AMR_diversity, "bray")-> Mbiome_PCoA
as.matrix(Mbiome_PCoA)->Mbiome_PCoA
BrayCurtMbiome=cmdscale(Mbiome_PCoA)
#ordiplot (BrayCurtMbiome, display = 'species', type = 'text')
BCords<-scores(BrayCurtMbiome)
BCords<-(as.data.frame(t(BCords)))
BCords<-as.data.frame(t(BCords))

AMRDiversityViz_Geo$Dim1<-BCords$Dim1
AMRDiversityViz_Geo$Dim2<-BCords$Dim2

AMRDiversityViz_Geo$Country <- factor(AMRDiversityViz_Geo$Country, levels = c("SG", "KL", "DD", "MI"))
AMRDiversityViz_Geo$Aetiology_short<- factor(AMRDiversityViz_Geo$Aetiology_short, levels=c("idiopathic"

#Figure 4#####
#AMR PCOA of Resistotypes BY SC_RESISTOTYPE
##Panel A####
##Panel B####
#AMRDiversityViz_Geo<-as.data.frame(merge(AMRDiversityViz_Geo, lab, by.x = 1, by.y = 0, all.x = TRUE)) 
gg <- data.frame(cluster=factor(AMRDiversityViz_Geo$SC_AMR_alt), x=AMRDiversityViz_Geo$Dim1, y=AMRDiver
# calculate group centroid locations
centroids <- aggregate(cbind(x,y)~cluster,data=gg,mean)
# merge centroid locations into ggplot dataframe
gg <- merge(gg,centroids,by="cluster",suffixes=c("",".centroid"))
# generate star plot...
PCA_RT<-ggplot(gg) +
  #scale_col_manual(values=c(16, 16, 16,16))+
  scale_linetype_identity() +
  geom_segment(aes(x=x.centroid, y=y.centroid, xend=x, yend=y, colour = cluster),alpha = 0.3)+
  geom_point(aes(x=x,y=y, colour = cluster), size = 2, alpha = 0.5) + #can add ",shape = shape" in aes
  #geom_point(aes(x=x,y=y, colour = cluster, shape = shape), size = 2) +
  geom_point(data=centroids, aes(x=x, y=y, color=cluster), size=5) +
  geom_point(data=centroids, aes(x=x, y=y, color=cluster), size=5, shape = 13, colour = "black") +
  scale_colour_manual(values = c("#1800F5","#932DE7"), labels = c("RT1", "RT2"))+
  labs(colour="",
       x = "PC 1 (23.8%)", y = "PC 2 (4.5%)")+
  theme(legend.position="bottom",
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
  )+
  scale_y_reverse()+
  #scale_x_reverse()+
  guides(colour = guide_legend(reverse = FALSE))
PCA_RT
```




```
##Panel DEF####
RT1_RT2_plot<-AMRDiversityViz %>%
  subset(select == "Bronchiectasis") %>%
  subset(Matching == "Matched")

Ex<-ggplot(data=RT1_RT2_plot, aes(as.factor(SC_AMR_alt), Exacerbations, group = as.factor(SC_AMR_alt), f
  scale_fill_manual(values=c("#1800F5","#932DE7"))+
  scale_y_continuous(breaks=seq(0,12,2))+
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(width = 0.1, height = 0.1)+
  facet_grid(. ~ "Exacerbations")+
  theme(legend.position="none",
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
        strip.background = element_rect( fill="white",size = 1))+
  labs(fill='Resistotype')+
  scale_x_discrete(labels = c('RT1', 'RT2'))

FEV1<-ggplot(data=RT1_RT2_plot, aes(as.factor(SC_AMR_alt), FEV1, group = as.factor(SC_AMR_alt), fill = a
  scale_fill_manual(values=c("#1800F5","#932DE7"))+
  expand_limits(y=c(0,175))+
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(width = 0.1, height = 0.1)+
```

```
  facet_wrap(~ "FEV1 (% predicted)")+
  theme(legend.position="none",
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
        strip.background = element_rect( fill="white",size = 1))+
  labs(fill='Resistotype')+
  scale_x_discrete(labels = c('RT1', 'RT2'))

BSI<-ggplot(data=RT1_RT2_plot, aes(as.factor(SC_AMR_alt), BSI, group = as.factor(SC_AMR_alt), fill = as
  scale_fill_manual(values=c("#1800F5","#932DE7"))+
  expand_limits(y=c(0,22))+
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(width = 0.1, height = 0.1)+
  facet_grid(. ~ "Severity (BSI)")+
  theme(legend.position="none",
    axis.title.x=element_blank(),
    axis.title.y=element_blank(),
    legend.title = element_blank(),
    axis.line = element_line(size = 0.5, colour = "black"),
    panel.background = element_rect(fill = NA),
    strip.background = element_rect(fill="white",size = 1))+
  labs(fill='Resistotype')+
  scale_x_discrete(labels = c('RT1', 'RT2'))

Figure_4DEF<-ggarrange(Ex,FEV1, BSI, ncol=3, nrow=1, common.legend = TRUE, legend="bottom")

Figure_4DEF
```




```
wilcox.test(RT1_RT2_plot$Exacerbations~RT1_RT2_plot$SC_AMR_alt)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  RT1_RT2_plot$Exacerbations by RT1_RT2_plot$SC_AMR_alt
## W = 2508.5, p-value = 0.006512
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(RT1_RT2_plot$BSI~RT1_RT2_plot$SC_AMR_alt)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  RT1_RT2_plot$BSI by RT1_RT2_plot$SC_AMR_alt
## W = 2451, p-value = 0.004426
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(RT1_RT2_plot$FEV1~RT1_RT2_plot$SC_AMR_alt)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  RT1_RT2_plot$FEV1 by RT1_RT2_plot$SC_AMR_alt
## W = 4095.5, p-value = 0.007875
## alternative hypothesis: true location shift is not equal to 0
```

```r
#Geographic prevalence of RT1 vs RT2
prop.table(table(Master$Country, Master$SC_AMR_alt), margin = 1) * 100
```

```
##
##             1        2
##   DD 81.92771 18.07229
##   KL 56.66667 43.33333
##   MI 50.00000 50.00000
##   SG 62.33766 37.66234
```