Main_Figure_1

Micheál Mac Aogáin

2023-11-23

#Load required R packages

```
#Load packages
if(!require("pacman")) install.packages("pacman")
pacman::p_load(pacman, ggplot2, tidyverse, tidyr, vegan, RColorBrewer, ggpubr)
```

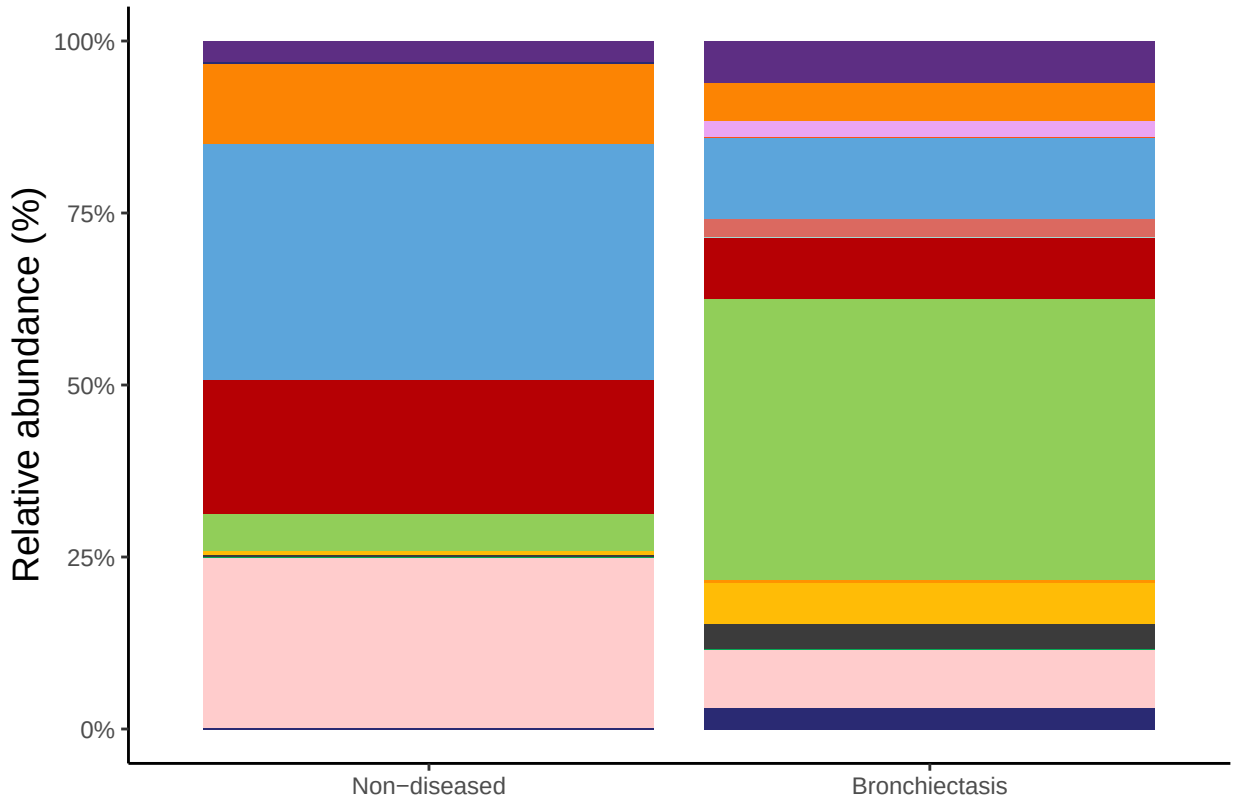
#Load data

```
##Master data cross-sectional####
Master <-read.csv("../Data/R_input_files//Clinical_AMR_Microbiome_R2.csv") %>%
  as_tibble()
Master$FEVfactor<-cut(Master$FEV1, breaks=c(0, 30, 50, 70, Inf))

##Longitudinal AMR data ####
MasterLT <-read.csv("../Data/R_input_files//LT_master_combined_8.0.csv")

###wrangle AMR data ####
AMRFam <- Master %>% #clinical variables + amr families
  as_tibble() %>%
  select(-29:-42,-64:-356)
AMRFam$FEVfactor<-cut(AMRFam$FEV1,  breaks=c(0, 30, 50, 70, Inf))
#set levels
AMRFam$ExacerbatorState <- factor(AMRFam$ExacerbatorState, levels=c("NonEx", "Exacerbator", "FreqEx"))
AMRFam$Country <- factor(AMRFam$Country, levels=c("SG", "KL", "DD", "MI"))
AMRFam$Aetiology_short <- factor(AMRFam$Aetiology_short, levels=c("idiopathic", "postInfect", "postTB",
AMRFam$SampleID <- factor(AMRFam$SampleID, levels = AMRFam$SampleID[order(AMRFam$SC_AMR_alt)])
AMRFam$FEVfactor<-fct_rev(AMRFam$FEVfactor)
AMRFam <- AMRFam %>%
  gather(Resistome, RPKM, starts_with(c("Acridine.dye", "Aminocoumarin.antibiotic", "Aminoglycoside", "A

AMRFam$CTRL<-ifelse(is.na(AMRFam$BSI), "CTRL", "PATIENT")

##Longitudinal data####
###Data wrangle####
AMRLT <- MasterLT %>%
  as_tibble() %>%
  select(-14:-230)
AMR_cols<-colnames(AMRLT[14:34])
AMRLT <- AMRLT %>%
  gather(AMR, RPKM, AMR_cols, -SampleSeqNo, -SputumSampleNo,  -TypeSamples, -TypeSamplesA,-TypeSamplesB
AMRLT$TmToNxtEx <- factor(AMRLT$TmToNxtEx , levels = c("MoreThan12w","LessThan12w"))
AMRLT$Exacerbations <- factor(AMRLT$Exacerbations , levels = c("NFE","FE"))
relapse.labs <- c(
  `LessThan12w` = "<12 w",
```

```
  `MoreThan12w` = ">12 w")
AMRLT$FEV170<-ifelse(AMRLT$FEV1 >70, ">70", "<70")
AMRLTctrols<-subset(AMRLT, is.na(TypeSamplesB))
```

#Figure 1 ##Figure 1A

```
##Stacked Barplot AMR rel abudance ####
HvsBE<-ggplot(data=AMRFam,aes(x=CTRL, y=RPKM, fill=Resistome))+
  geom_bar(aes(), stat="identity", position = "fill") +
  scale_fill_manual(values = c("#026EB8","#06A955","#5D2E83","#2A2A73","#fc8403","#EBA5F3","#fc5017","#
  scale_y_continuous(labels = scales::percent)+
  scale_x_discrete(labels = c('Non-diseased','Bronchiectasis'))+
  theme(legend.position="none",
        #axis.text=element_blank(),
        #axis.title=element_blank(),
        axis.title=element_text(size=14),
        #axis.text.x = element_blank(),
        #axis.text.x = element_text(angle = 90),
        panel.background = element_rect(fill = NA),
        axis.line = element_line(linewidth = 0.5, colour = "black"),
        legend.title = element_blank(),
        legend.text = element_text(face = "italic"))+
  guides(fill=guide_legend(ncol=1), size = .1)+
  xlab("")+
  ylab("Relative abundance (%)")+
  #facet_wrap(~AMRFam$CTRL, scales="free_x")+
  theme(
    strip.background = element_rect(
      color="white", fill="white", size=1, linetype="solid"),
    strip.text.x = element_text(size = 12)
  )
HvsBE
```



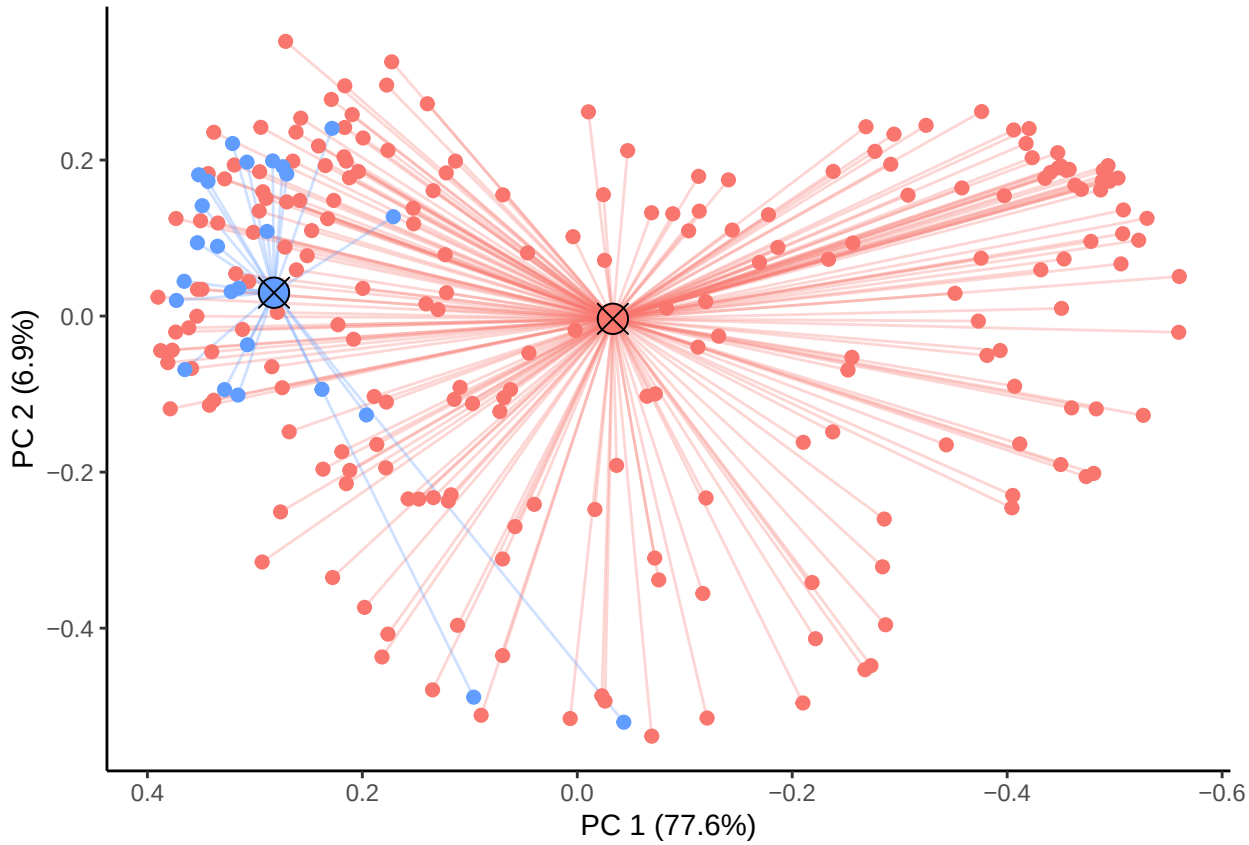
##Figure 1B

```
## PCA plot Non-diseased vs Bronchiectasis####
AMR_diversity <- Master %>%
  as_tibble() %>%
  #select(1:1,395:645) #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
#sum(isZero)#NO amr detected in 37 samples
AMR_diversity<-AMR_diversity[!isZero,]

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))

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$SampleID != "TBS153", , drop = FALSE] #remove for gene
AMRDiversityViz$Dim1<-BCords$Dim1
AMRDiversityViz$Dim2<-BCords$Dim2

#checking PC loadings
checkEig<-capscale(AMR_diversity ~1)
Eig <-eigenvals(checkEig)
print(Eig[1:2] / sum(Eig))
```

```
##       MDS1       MDS2
## 0.77623336 0.06917351
```

```
#AMR PCOA of Resistotypes BY SC_RESISTOTYPE
gg <- data.frame(cluster=factor(AMRDiversityViz$select), x=AMRDiversityViz$Dim1, y=AMRDiversityViz$Dim2
# 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...
BC<-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) + #can add ",shape = shape" in aes to introduce
  #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_shape_discrete(labels = c("Healthy", "Bronchiectasis"))+
  scale_colour_manual(values = c("#F8766D", "#619CFF"), labels = c("Bronchiectasis","Healthy"))+
  labs(colour="",
       x = "PC 1 (77.6%)", y = "PC 2 (6.9%)")+ #calculated loadings
  theme(legend.position="none",
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
  )+
  scale_x_reverse()+
  #scale_y_reverse()+ #add for gene level analysis
  guides(colour = guide_legend(reverse = T))
BC
```



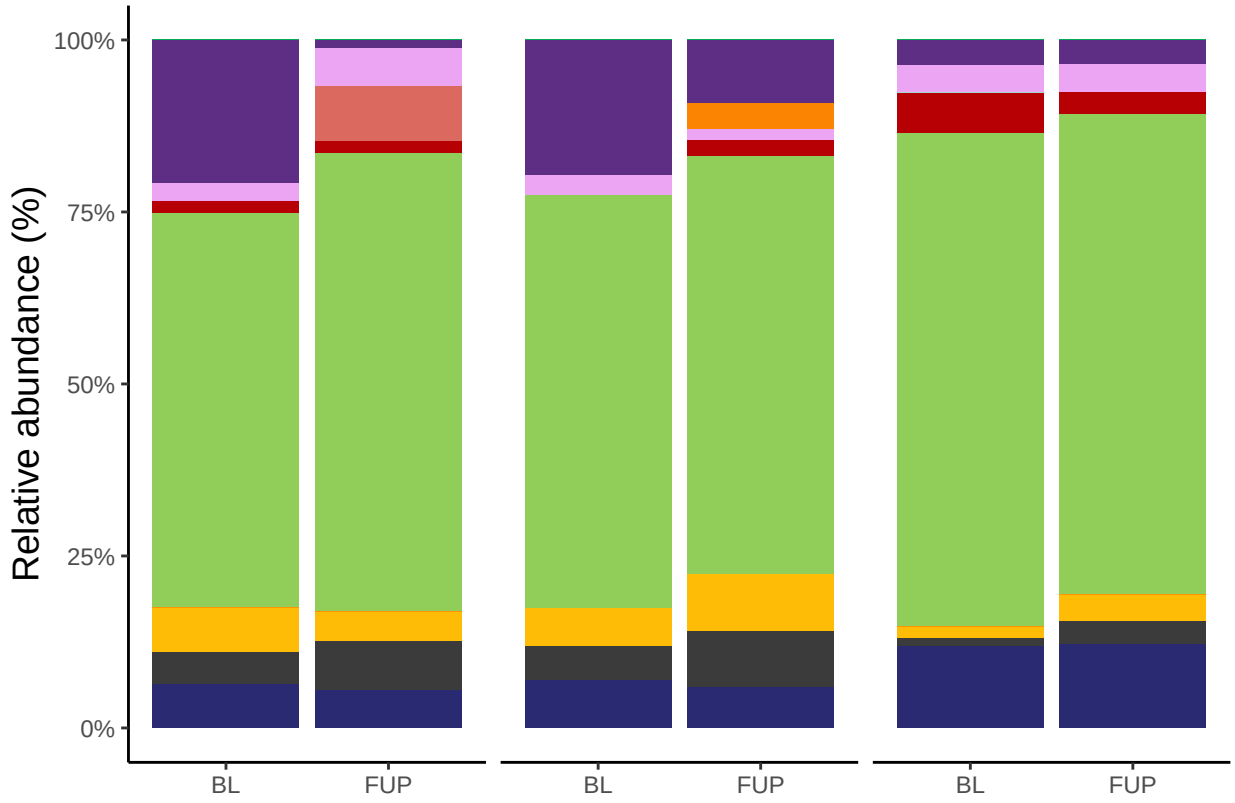

```
#Are patients distinct from healthy controls? Yes
adonis2(AMR_diversity ~ select, data=AMRDiversityViz, method="bray", permutations=999)
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = AMR_diversity ~ select, data = AMRDiversityViz, permutations = 999, method = "bray
##           Df SumOfSqs      R2      F Pr(>F)
## select     1    2.568 0.03602 8.8172  0.001 ***
## Residual 236   68.739 0.96398
## Total    237   71.307 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Figure 1C

```
STBL<-ggplot(data=AMRLTctrols,aes(x=SampleSeqNo, y=RPKM, fill=AMR))+
  geom_bar(aes(), stat="identity", position = 'fill') +
  scale_fill_manual(values = c("#026EB8","#06A955","#5D2E83","#2A2A73","#fc8403","#EBA5F3","#fc5017","#
  scale_x_discrete(labels = c('BL','FUP'))+
  scale_y_continuous(labels = scales::percent)+
  theme(#legend.position="none",
    #axis.text=element_blank(),
    axis.title=element_text(size=14),
    #axis.text.x = element_text(angle = 90),
```

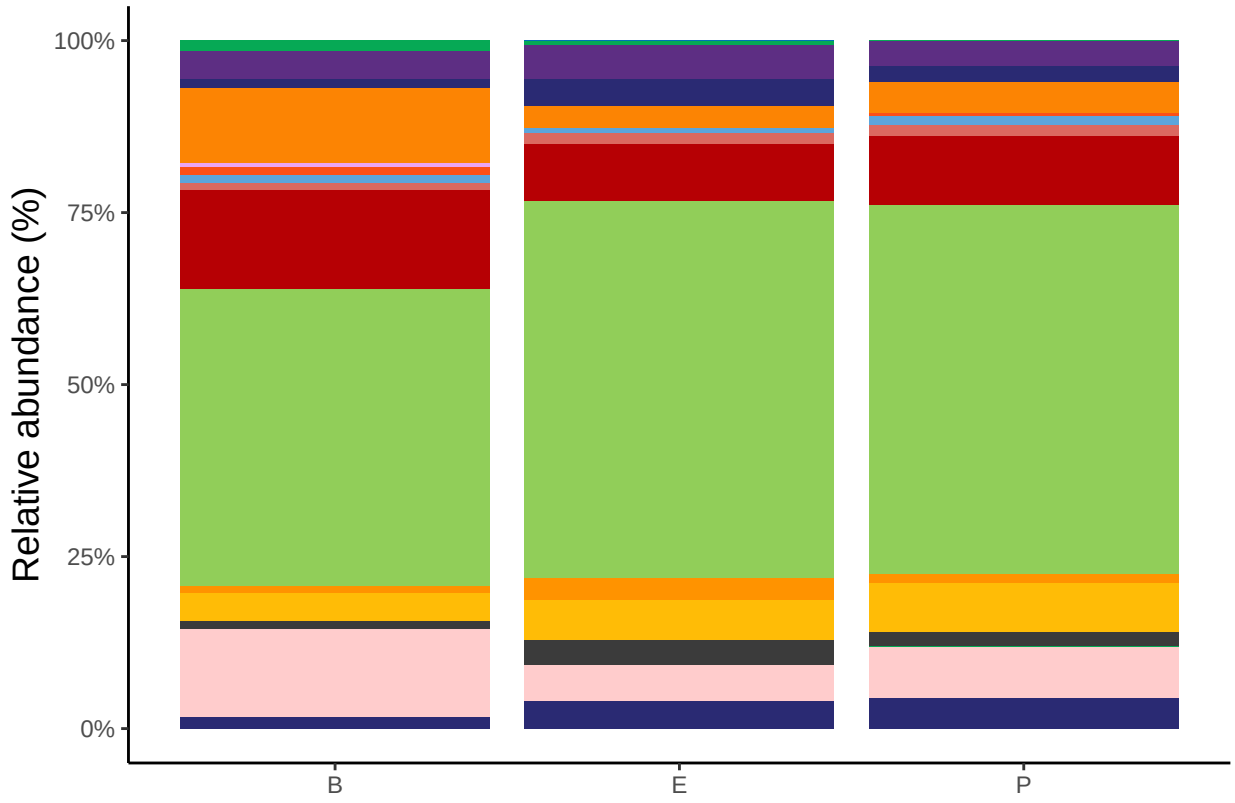
```
    panel.background = element_rect(fill = NA),
    axis.line = element_line(size = 0.5, colour = "black"))+
  #legend.title = element_blank(),
  #legend.text = element_text(face = "italic"))
  #guides(fill=guide_legend(ncol=1), size = 0.1)+
  xlab("")+
  ylab("Relative abundance (%)")+
  facet_grid(~AMRLTctrols$SputumSampleNo, scales="free_x")+
  theme(strip.background = element_rect(
    color="white", fill="white", size=1, linetype="solid"),
    strip.text.x = element_blank(),
    #legend.text=element_text(size=8)
    legend.position="none"
  )
STBL
```



##Figure 1D

```
#drop the controls stable
AMRLT<-subset(AMRLT, is.na(Severity) != TRUE)
D<-ggplot(data=AMRLT[which(AMRLT$TypeSamplesA !="NA"),],aes(x=TypeSamplesB, y=RPKM, fill=AMR))+
  geom_bar(aes(), stat="identity", position="fill") +
  scale_fill_manual(values = c("#026EB8","#06A955","#5D2E83","#2A2A73","#fc8403","#EBA5F3","#fc5017","#5
  scale_y_continuous(labels = scales::percent)+
  scale_x_discrete(labels = c('B','E', 'P'))+
  theme(#legend.position="none",
    #axis.text=element_blank(),
```

```
    axis.title=element_text(size=14),
    #axis.text.x = element_text(angle = 90),
    panel.background = element_rect(fill = NA),
    axis.line = element_line(size = 0.5, colour = "black"))+
  #legend.title = element_blank(),
  #legend.text = element_text(face = "italic"))
  #guides(fill=guide_legend(ncol=1), size = 0.1)+
  xlab("")+
  ylab("Relative abundance (%)")+
  #facet_grid(~AMRLT$TmToNxtEx, scales="free_x", labeller = as_labeller(relapse.labs))+
  theme(
    strip.background = element_rect(
      color="white", fill="white", size=1, linetype="solid"),
    strip.text.x = element_blank(),
    #legend.text=element_text(size=8)
    legend.position="none"
  )
D
```



##Figure 1E

```
### PCA ####
AMRLT_diversity <- MasterLT[which(MasterLT$TypeSamplesA !="NA"),] %>%
  as_tibble() %>%
  select(-14:-230)
#AMRLT_diversity <- select(MasterLT[which(MasterLT$TypeSamplesA !="NA"),], -2:-79, -231:-3621) #ugly 'w
AMRLT_diversity <- select(MasterLT[which(MasterLT$TypeSamplesA !="NA"),], -2:-230) #ugly 'which subsett
```

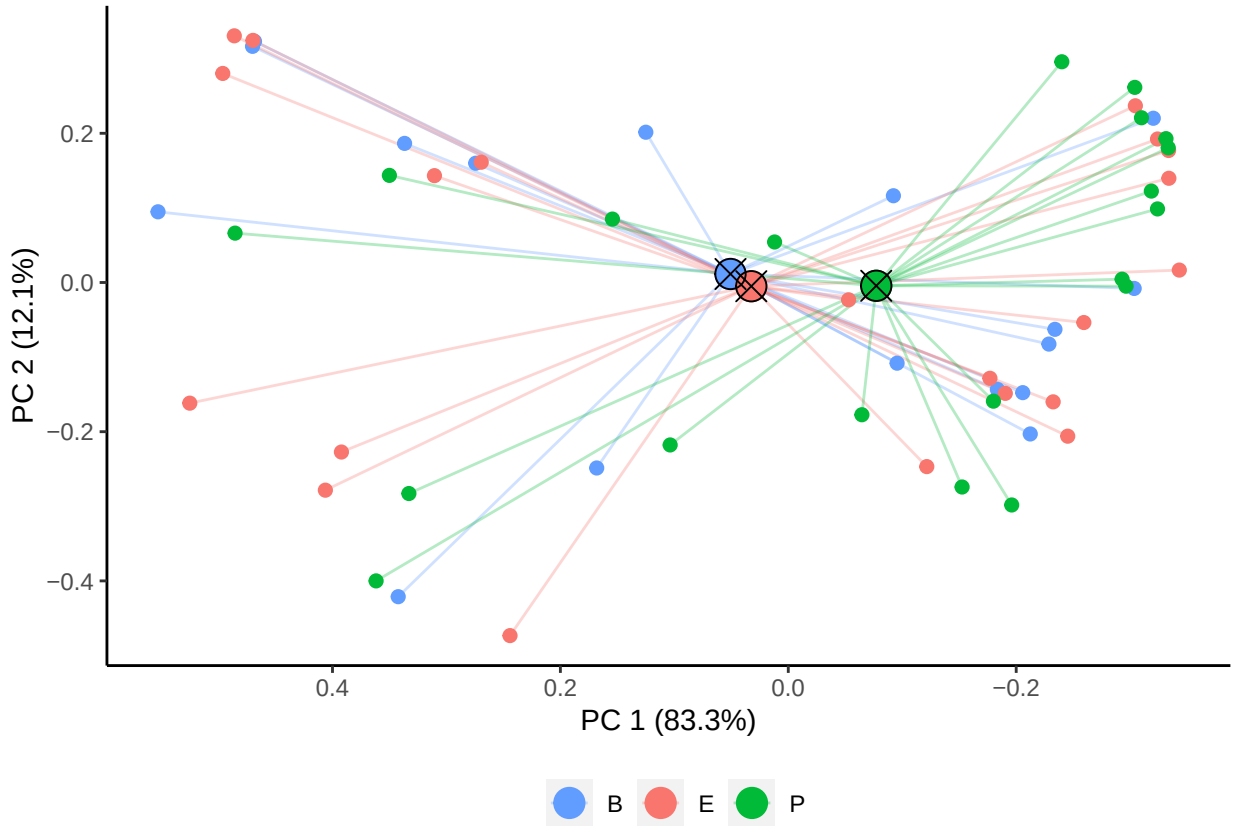
```r
NAMES_list <- head(MasterLT$SampleSeqNo, -6) #head is just to drop controls again n=6
main_dataLT <- AMRLT_diversity[AMRLT_diversity$SampleSeqNo %in% NAMES_list, ]
AMRLT_diversity<-as.matrix(AMRLT_diversity)
rownames(AMRLT_diversity) <- AMRLT_diversity[,1]
AMRLT_diversity = as.data.frame(subset(AMRLT_diversity, select = -c(SampleSeqNo) ))
AMRLT_diversity[] <- lapply(AMRLT_diversity, as.numeric)
isZero <- base::rowSums(AMRLT_diversity) == 0
AMRLT_diversity<-AMRLT_diversity[!isZero,]
vegdist(AMRLT_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))

LTDiversityViz<-MasterLT[which(MasterLT$TypeSamplesA !="NA"),] #drop controls -which subsetting
#LTDiversityViz$SampleSeqNo %in% row.names(BCords)
LTDiversityViz<-LTDiversityViz[ LTDiversityViz$SampleSeqNo %in% row.names(BCords) , ]

LTDiversityViz$Dim1<-BCords$Dim1
LTDiversityViz$Dim2<-BCords$Dim2

LTDiversityViz$FEV170<-ifelse(LTDiversityViz$FEV1 >70, ">70", "<70")
LTDiversityViz$FEV170<- factor(LTDiversityViz$FEV170 , levels = c(">70","<70"))

#AMR PCOA of Resistotypes BY sample type
gg <- data.frame(cluster=factor(LTDiversityViz$TypeSamplesB), x=LTDiversityViz$Dim1, y=LTDiversityViz$D
# 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...
E<-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) + #can add ",shape = shape" in aes to introduce
  #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_shape_discrete(labels = c("B", "E", "P"))+
  scale_colour_manual(values = c("#619CFF", "#F8766D", "#00BA38"), labels = c("B", "E", "P"))+
  labs(colour="",
       x = "PC 1 (83.3%)", y = "PC 2 (12.1%)")+
  theme(legend.position="bottom",
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
  )+  scale_x_reverse()
#+ggtitle("Timepoint")
#PERMANOVA - timepoint
adonis2(AMRLT_diversity~TypeSamplesB , data = LTDiversityViz, method = "bray",permutations=999, strata =
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Blocks:  strata
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = AMRLT_diversity ~ TypeSamplesB, data = LTDiversityViz, permutations = 999, method =
##              Df SumOfSqs      R2      F Pr(>F)
## TypeSamplesB  2   0.5016 0.02912 0.8249  0.502
## Residual     55  16.7226 0.97088
## Total        57  17.2242 1.00000
```

```
#checking PC %s
#rda(X = AMRLT_diversity, scale = TRUE)
checkEig<-capscale(AMRLT_diversity ~1)
Eig <-eigenvals(checkEig)
print(Eig[1:2] / sum(Eig))
```

```
##      MDS1      MDS2
## 0.8327988 0.1211846
```

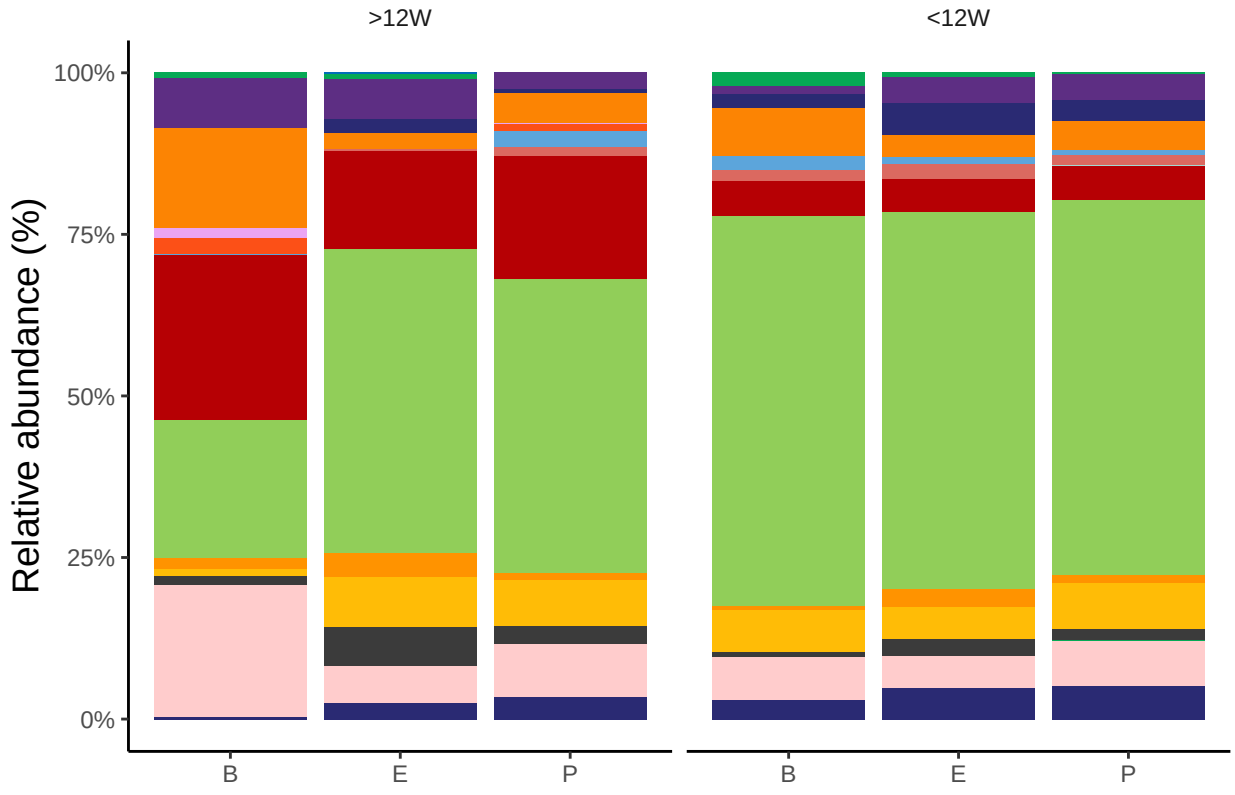
```
E
```



##Figure 1F

```
relapse.labs <- as_labeller(c(`LessThan12w` = "<12W", `MoreThan12w` = ">12W"))
F<-ggplot(data=AMRLT[which(AMRLT$TypeSamplesA !="NA"),],aes(x=TypeSamplesB, y=RPKM, fill=AMR))+
  geom_bar(aes(), stat="identity", position="fill") +
```

```
  scale_fill_manual(values = c("#026EB8","#06A955","#5D2E83","#2A2A73","#fc8403","#EBA5F3","#fc5017","#5
  scale_y_continuous(labels = scales::percent)+
  scale_x_discrete(labels = c('B','E', 'P'))+
  theme(legend.position="none",
        #axis.text=element_blank(),
        axis.title=element_text(size=14),
        #axis.text.x = element_text(angle = 90),
        panel.background = element_rect(fill = NA),
        axis.line = element_line(size = 0.5, colour = "black"))+
  #legend.title = element_blank(),
  #legend.text = element_text(face = "italic"))
  #guides(fill=guide_legend(ncol=1), size = 0.1)+
  xlab("")+
  ylab("Relative abundance (%)")+
  facet_grid(~AMRLT[which(AMRLT$TypeSamplesA !="NA"),]$TmToNxtEx, scales="free_x", labeller = relapse.la
  theme(
    strip.background = element_rect(
      color="white", fill="white", size=1, linetype="solid"),
    #strip.text.x = element_blank(),
    #legend.text=element_text(size=8)
    legend.position="none"
  )
F
```



##Figure 1G

```
#AMR PCOA of Resistotypes BY Time To Next Exacerbation
gg <- data.frame(cluster=factor(LTDiversityViz$TmToNxtEx), x=LTDiversityViz$Dim1, y=LTDiversityViz$Dim2
# 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...
G<-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, 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_shape_discrete(labels = c("B", "E", "P"))+
  scale_colour_manual(values = c("#F8766D", "#619CFF"), labels = c("<12 w", ">12 w"))+
  labs(colour="",
       x = "PC 1 (83.3%)", y = "PC 2 (12.1%)")+
  theme(legend.position="bottom",
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
  )+
  scale_x_reverse()
#PERMANOVA - timeToexacerbation
adonis2(AMRLT_diversity~TmToNxtEx , data = LTDiversityViz, method = "bray",  permutations=9999, strata =
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Blocks:  strata
## Permutation: free
## Number of permutations: 9999
##
## adonis2(formula = AMRLT_diversity ~ TmToNxtEx, data = LTDiversityViz, permutations = 9999, method =
##           Df SumOfSqs      R2     F Pr(>F)
## TmToNxtEx  1   0.7194 0.04177 2.441 0.0239 *
## Residual  56  16.5047 0.95823
## Total     57  17.2242 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Combine and print panels for Figure 1####

```
Figure_1top<-ggarrange(HvsBE,BC,STBL,
                  font.label = list(size = 5),
                  common.legend = FALSE, nrow = 1, ncol = 3) #this one

Figure_1bot<-ggarrange(D,NULL,F,E,NULL,G, font.label = list(size = 5),
                       common.legend = FALSE, widths = c(1, 0.1,1,1,0.1,1))

Figure_1<- ggarrange(Figure_1top, Figure_1bot, font.label = list(size = 5),
                           common.legend = FALSE,heights = c(0.35, 0.65), nrow =2)

# pdf(file = "../Data/R_output_files/Figure_1.pdf",   # The directory you want to save the file in
#     width = 10, # The width of the plot in inches
```

```
#   height = 12)
# Figure_1
# dev.off()
# Figure_1
```