Main_Figure_5

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,RColorBrewer, res
```

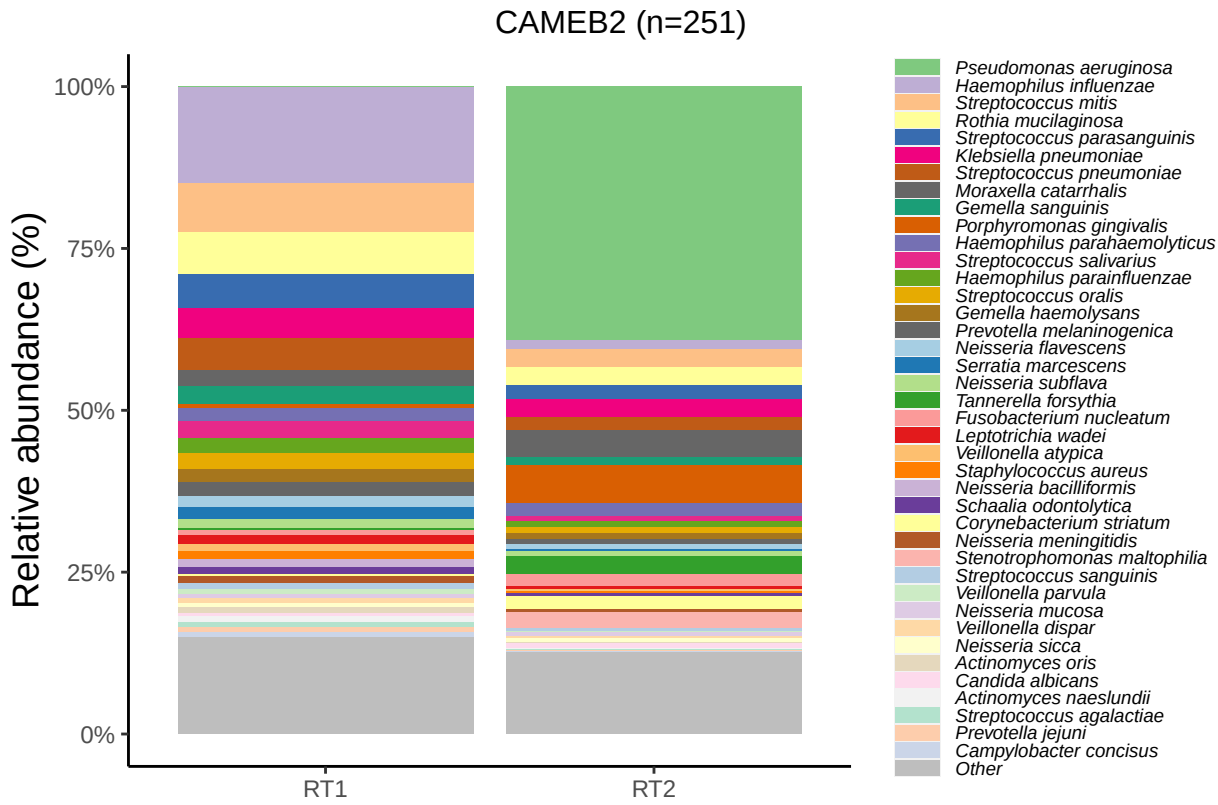
#Load and wrangle taxonomic data

```
##Plot the Taxonomy data####
#wrangle Metagenomic Taxonomy data
#load relative abundance data on top50 taxa
MetaG<-read.csv("../Data/R_input_files/CAMEB2_bacteria_top50.csv") %>%
  as_tibble() #%>%
#filter(ExacerbatorState != "NA") %>%
#filter(Matching == "Matched" )

#Load Master datafile####
Master <-read.csv("../Data/R_input_files/Clinical_AMR_Microbiome_R2.csv") %>%
  as_tibble()
MetaG$SC_AMR_alt<-Master[which(Master$Matching != "NA"),]$SC_AMR_alt
MetaG$Reads<-Master[which(Master$Matching != "NA"),]$ReadsNonHuman
MetaG$Aetiology_short<-Master[which(Master$Matching != "NA"),]$Aetiology_short
MetaG$SampleID <- factor(MetaG$SampleID, levels = MetaG$SampleID[order(MetaG$Reads)])
MetaG$Country<-Master[which(Master$Matching != "NA"),]$Country
MetaG$Country <- factor(MetaG$Country, levels=c("SG", "KL", "DD", "MI"))
MetaG$Aetiology_short <- factor(MetaG$Aetiology_short, levels =c("idiopathic", "postInfect", "postTB",

MetaG<-melt(MetaG, id.vars = c("SampleID", "Country",  "Continent", "Matching","ExacerbatorState",  "FE

MetaG$variable <- factor(gsub("\\.", " ", MetaG$variable), levels= gsub("\\.", " ", levels(MetaG$variabl

#cleaner but messes with colours.
#MetaG_long <- MetaG %>%
#  pivot_longer(cols = Pseudomonas.aeruginosa:Other,
#               names_to = "variable",
#               values_to = "value")

n <- 41
qual_col_pals = brewer.pal.info[brewer.pal.info$category == 'qual',]
col_vector = unlist(mapply(brewer.pal, qual_col_pals$maxcolors, rownames(qual_col_pals)))
```

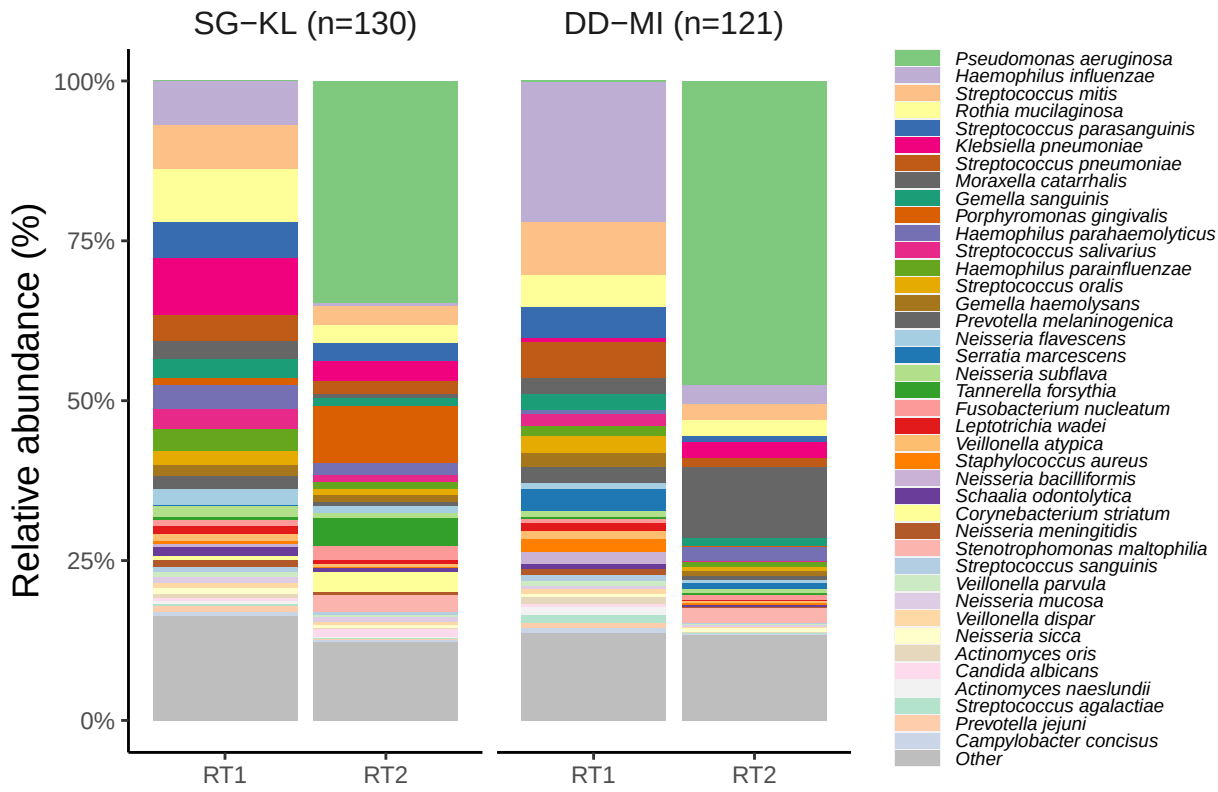
```r
col_vector_spec<-replace(col_vector, 41, "grey")

#Figure 5A
Taxa_Geo_Fig5a<-ggplot(data=MetaG[which(MetaG$SC_AMR_alt != "0"),],aes(x=as.factor(SC_AMR_alt), y=value
  scale_fill_manual(values = col_vector_spec) +
  geom_bar(aes(), stat="identity", position = "fill" )+
  scale_y_continuous(labels = scales::percent)+
  scale_x_discrete(labels = c('RT1','RT2'))+
  ggtitle("CAMEB2 (n=251)") +
  theme(
    legend.position="right",
    #axis.text=element_blank(),
    #axis.title=element_blank(),
    axis.title=element_text(size=14),
    #axis.text.x = element_blank(),
    #axis.text.x = element_text(angle = 90),
    panel.background = element_rect(fill = NA),
    axis.line = element_line(size = 0.5, colour = "black"),
    legend.title = element_blank(),
    legend.text = element_text(face = "italic", size = 7),
    legend.key.height = unit(1, "mm"))+
  xlab("")+
  ylab("Relative abundance (%)")+
  #facet_wrap(~MetaG[which(MetaG$SC_AMR_alt != "0"),]$SC_AMR_alt, scales="free_x",nrow = 1)+
  theme(plot.title = element_text(hjust = 0.5, size = 12),
        plot.title.position = "plot",
        #plot.title.margin = margin(b = 10),
        strip.background = element_rect(
          color="white", fill="white", size=1, linetype="solid"),
        strip.text.x = element_text(size = 12)
  )+
  guides(fill=guide_legend(ncol=1), size = .1)

variable_names_a <- list(
  "Asia" = "SG-KL (n=130)",
  "Europe" = "DD-MI (n=121)"
)

variable_labeller <- function(variable,value){
  return(variable_names_a[value])
}

#Figure 5B
Taxa_Geo_Fig5b<-ggplot(data=MetaG[which(MetaG$SC_AMR_alt != "0"),],aes(x=as.factor(SC_AMR_alt), y=value
  scale_fill_manual(values = col_vector_spec) +
  geom_bar(aes(), stat="identity", position = "fill" )+
  scale_y_continuous(labels = scales::percent)+
  scale_x_discrete(labels = c('RT1','RT2'))+
  theme(legend.position="right",
        #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(size = 0.5, colour = "black"),
        legend.title = element_blank(),
        legend.text = element_text(face = "italic", size = 7),
        legend.key.height = unit(1, "mm"))+
  xlab("")+
  ylab("Relative abundance (%)")+
  facet_wrap(~MetaG[which(MetaG$SC_AMR_alt != "0"),]$Continent, scales="free_x",nrow = 1,labeller=varia
  theme(
    strip.background = element_rect(
      color="white", fill="white", size=1, linetype="solid"),
    strip.text.x = element_text(size = 12)
  )+
  guides(fill=guide_legend(ncol=1), size = .1)

variable_names_b<- list(
  "SG" = "SG (n=95)",
  "KL" = "KL (n=35)",
  "DD" = "DD (n=96)",
  "MI" = "MI (n=25)"
)

variable_labeller <- function(variable,value){
  return(variable_names_b[value])
}

#Figure 5C
Taxa_Geo_Fig5c<-ggplot(data=MetaG[which(MetaG$SC_AMR_alt != "0"),],aes(x=as.factor(SC_AMR_alt), y=value
  scale_fill_manual(values = col_vector_spec) +
  geom_bar(aes(), stat="identity", position = "fill" )+
  scale_y_continuous(labels = scales::percent)+
  scale_x_discrete(labels = c('RT1','RT2'))+
  theme(legend.position="right",
        #axis.text=element_blank(),
        #axis.title=element_blank(),
        axis.title=element_text(size=14),
        #axis.text.x = element_blank(),
        #axis.text.x = element_text(angle = 90),
        panel.background = element_rect(fill = NA),
        axis.line = element_line(size = 0.5, colour = "black"),
        legend.title = element_blank(),
        legend.text = element_text(face = "italic", size = 7),
        legend.key.height = unit(1, "mm"))+
  xlab("")+
  ylab("Relative abundance (%)")+
  facet_wrap(~MetaG[which(MetaG$SC_AMR_alt != "0"),]$Country, scales="free_x",nrow = 1,labeller=variable
  theme(
    strip.background = element_rect(
      color="white", fill="white", size=1, linetype="solid"),
    strip.text.x = element_text(size = 12)
  )+
  guides(fill=guide_legend(ncol=1), size = .1)
```

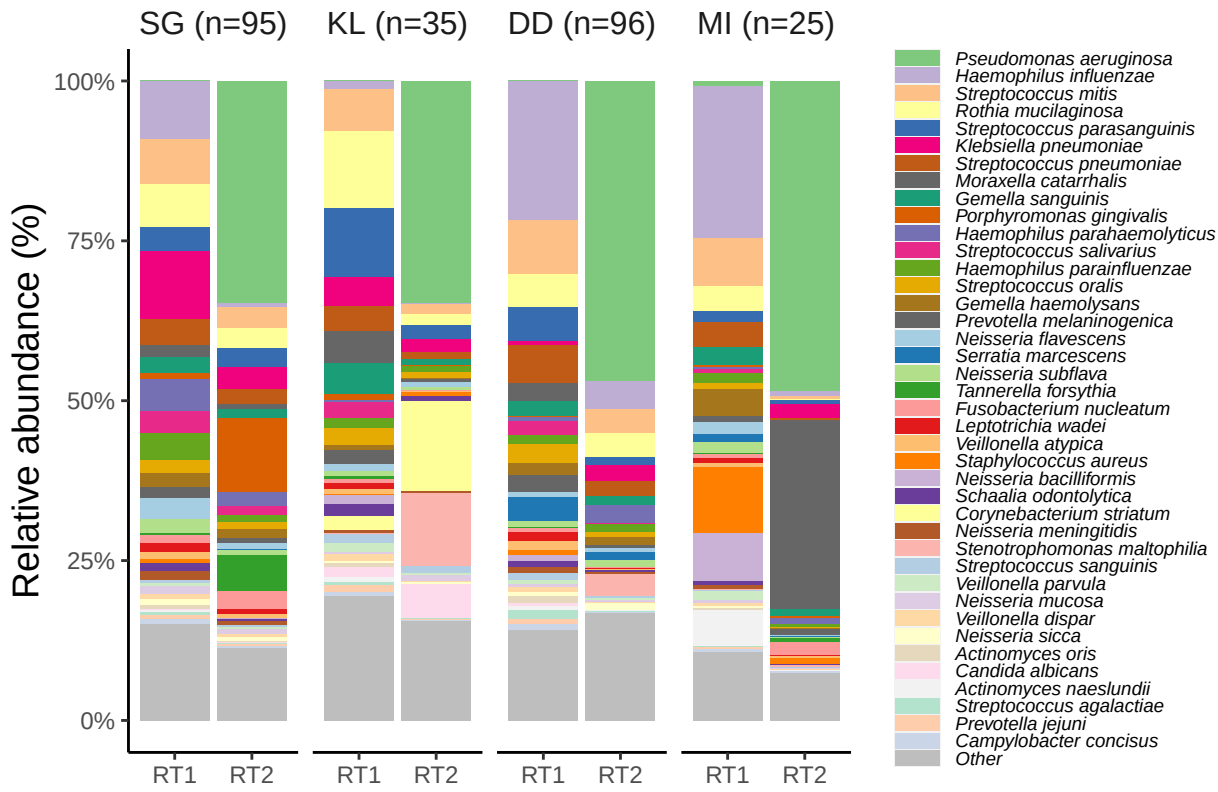
```
Taxa_Geo_Fig5a
```




```
Taxa_Geo_Fig5b
```




```
Taxa_Geo_Fig5c
```



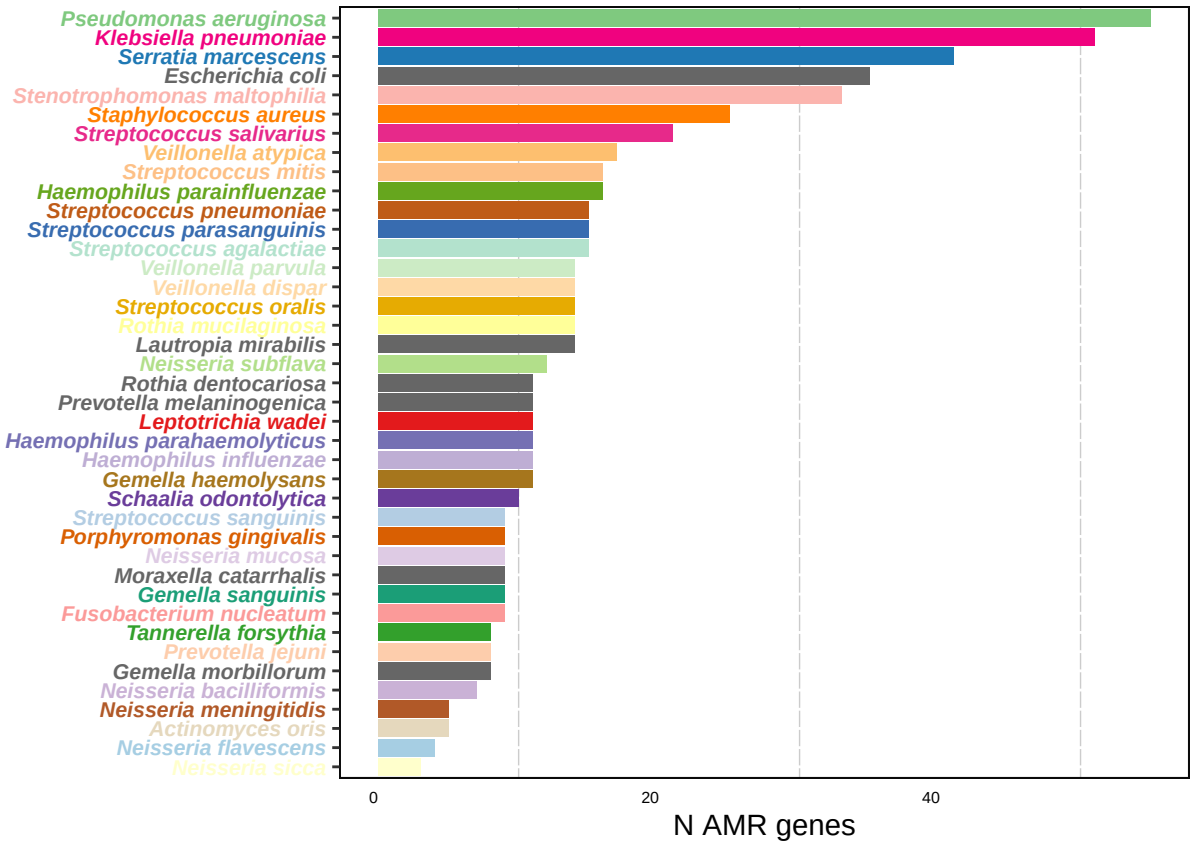

```
## BIPARTITE GRAPHS ####
### Species-AMR ####
#### Project name and working directory ####
library(phyloseq)
library(ggplot2)
library(tidyr)
library(RColorBrewer)
project_name <- "CAMEB2"
wkdir <- file.path("../Data/R_input_files/Ivan_AMR_analysis_2020", project_name)

indir <- file.path(wkdir, "analysis", "AMR-CONTIGS", "unweighted-bacteria-phages-plasmids")
outdir <- indir

# AMR-contigs
df <- read.csv(file.path(indir, "AMR-contigs-Species-abundance.csv"), row.names= 1, check.names= FALSE)
df <- df[df$Type=="Chromosome", ]
df <- df[df$Species!="Unknown sp.", ]

SpeciesAMR <- data.frame(Species= gsub(" - .*", "", unique(df$SpeciesAMR)), AMR= gsub(".* - ", "", uniq
AMRcountsPerSpecies <- data.frame(sort(table(SpeciesAMR$Species))); colnames(AMRcountsPerSpecies) <- c(
SpeciesCountsPerAMR <- data.frame(sort(table(SpeciesAMR$AMR))); colnames(SpeciesCountsPerAMR) <- c("AMR

kaiju<-readRDS("../Data/R_input_files/Ivan_AMR_analysis_2020/CAMEB2/analysis/KAIJU/ps_Species.RData")

ps.prop <- transform_sample_counts(kaiju, function(otu) {if (sum(otu)==0) otu else 100*otu/sum(otu)})
otutable<-as.data.frame(otu_table(ps.prop))
```

```r
z=colSums(otutable>0.1) #filter to reduced number of taxa
sel_col=row.names(as.data.frame(z[z>=(0.1*(nrow(otutable)))])) #In 5% patients prevalent
otutable<-otutable[sel_col]
remove(sel_col,z)
otutable["Sample_ID"]<-row.names(otutable)
otutable<-gather(otutable, Bacteriome, value, -Sample_ID)
taxa.means<-aggregate(value~Bacteriome, FUN=mean, data =otutable)
taxa.means<-taxa.means[
  with(taxa.means, order(-value)),
]
taxa.order<-as.vector(taxa.means$Bacteriome)
taxa.order.top50<-taxa.order[1:42]

#write.csv(otutable, file.path(wkdir, 'SP_otutable.csv'))
#write.csv(SpeciesAMR, file.path(outdir, "SpeciesAMR.csv"), row.names= FALSE)
#write.csv(AMRcountsPerSpecies, file.path(outdir, "SpeciesAMR-AMRcountsPerSpecies.csv"))
#write.csv(SpeciesCountsPerAMR, file.path(outdir, "SpeciesAMR-SpeciesCountsPerAMR.csv"))

# Barplot SpeciesAMR-AMRcountsPerSpecies
# ------------------------------------------------------------------------------------------

## Read files
colTaxa <- readRDS(file.path(wkdir, "analysis", "KAIJU", "colTaxa_Species.RData"))
names(colTaxa) <- gsub("_", " ", names(colTaxa))

##alt colours [original CAMEB2 colours]
qual_col_pals = brewer.pal.info[brewer.pal.info$category == 'qual',]
col_vector = unlist(mapply(brewer.pal, qual_col_pals$maxcolors, rownames(qual_col_pals)))
col_vector_spec<-replace(col_vector, 44, "grey")
cols_AMR<-col_vector[1:21]
names(cols_AMR) <- gsub("\\.", " ", colnames(Master[43:63]))

data <- read.csv(file.path(indir, "SpeciesAMR-AMRcountsPerSpecies.csv"))
data <- data[data$Species %in% as.list(taxa.order.top50), ]
data <- merge(data, taxa.means, by.x = "Species",
              by.y = "Bacteriome", all.x = TRUE, all.y = FALSE)
data$Species <- factor(data$Species, levels = data$Species[order(data$Count)])
colSpecies <- append((rep("gray40", nrow(data)-20)),rev(col_vector[1:20]))
names(colSpecies) <- as.vector(data$Species[order(data$value)])
colSpecies<-colSpecies[order(match(names(colSpecies),data$Species[order(data$Count)]))]

colSpecies2 <- col_vector_spec[1:40]
names(colSpecies2) <- levels(MetaG$variable)[1:40]
colSpecies[intersect(names(colSpecies), names(colSpecies2))] <- colSpecies2[intersect(names(colSpecies)
setdiff(names(colSpecies), names(colSpecies2))
```

```
## [1] "Gemella morbillorum" "Rothia dentocariosa" "Lautropia mirabilis"
## [4] "Escherichia coli"
```

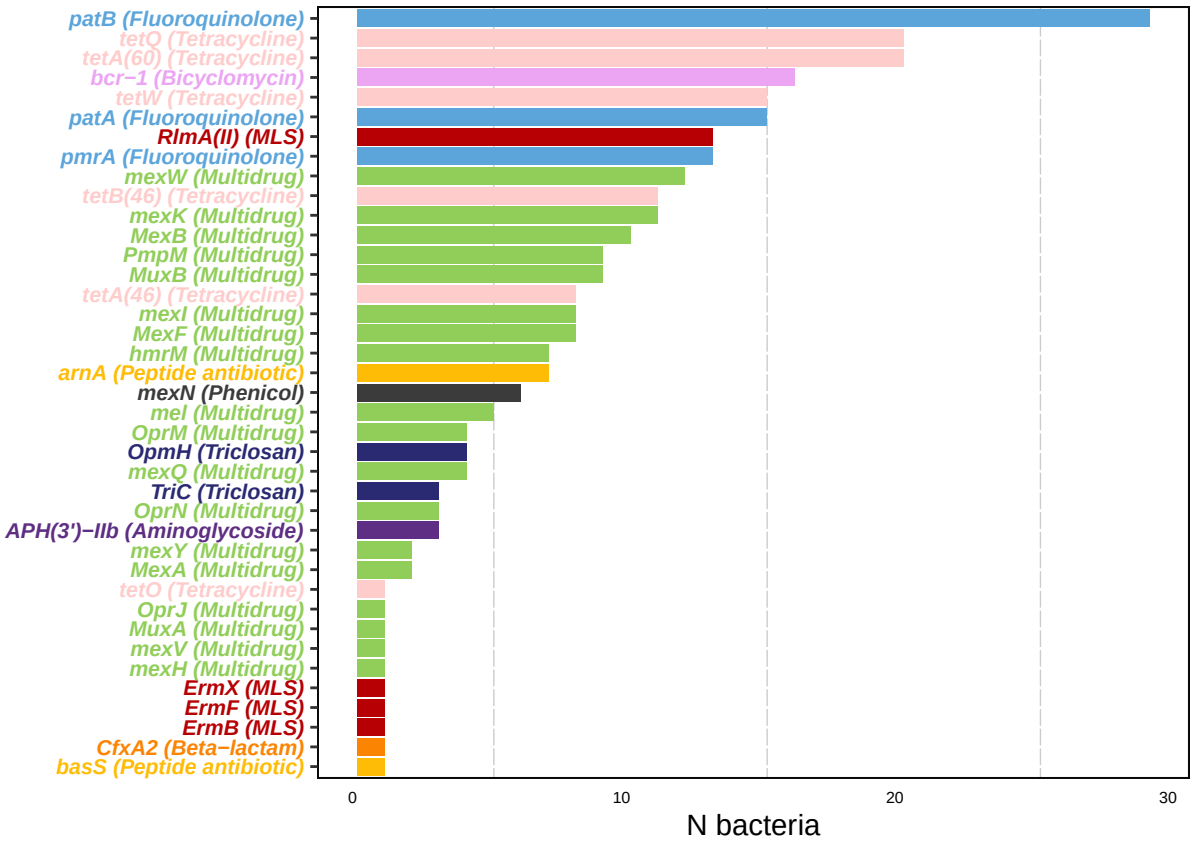
```r
#colSpecies["Esc... coli"] <- "..."
#colSpecies["Esc... coli"] <- "..."
#colSpecies["Esc... coli"] <- "..."

## Plotting
```

```
p <- ggplot(data, aes(x=Species, y=Count, fill=Species)) +
  geom_bar(stat = "identity") + #ggtitle("Number of AMR genes associated with bacterial species") +
  coord_flip() + xlab("") + ylab("N AMR genes") +
  theme(legend.position = "none",
        axis.ticks.x = element_blank(),
        axis.text.x = element_text(color="black", size= 6, angle= 0, hjust= 1),
        axis.text.y = element_text(face="bold.italic", size= 8, colour=colSpecies),
        panel.background = element_rect(fill = "white"),
        panel.border = element_rect(color="gray4", fill=NA),
        #panel.grid.major = element_line(color="gray80"),
        panel.grid.minor = element_line(color="gray80")
  ) + scale_x_discrete(position = "bottom") +
  scale_fill_manual(name= "Species", values=colSpecies)
p
```




```
#ggsave(file=file.path(outdir, "SpeciesAMR-AMRcountsPerSpecies.png"), width = 120, height = 300, units =

#--------------------

shortbread<-readRDS("../Data/R_input_files/Ivan_AMR_analysis_2020/CAMEB2/analysis/AMR/ps_Gene.RData")
ps.prop <- transform_sample_counts(shortbread, function(otu) {if (sum(otu)==0) otu else 100*otu/sum(otu
otutable<-as.data.frame(otu_table(ps.prop))
z=colSums(otutable>0.1) #filter to reduced number of taxa
sel_col=row.names(as.data.frame(z[z>=(0.1*(nrow(otutable)))])) #In 5% patients prevalent
otutable<-otutable[sel_col]
remove(sel_col,z)
```

```
otutable["Sample_ID"]<-row.names(otutable)
otutable<-gather(otutable, Resistome, value, -Sample_ID)
taxa.means<-aggregate(value~Resistome, FUN=mean, data =otutable)
taxa.means<-taxa.means[
  with(taxa.means, order(-value)),
]
ps <- readRDS(file.path(wkdir, "analysis", "AMR", "ps_Gene.RData"))
taxa.means$Resistome <- c(paste0(tax_table(ps)[setdiff(taxa.means$Resistome, "Others"),"Gene"], " (", ta
taxa.order<-as.vector(taxa.means$Resistome)
taxa.order.top50<-taxa.order[1:50] #40 genes in final graph

# Barplot SpeciesAMR-SpeciesCountsPerAMR
# ------------------------------------------------------------------------------------------

colTaxa <- readRDS(file.path(wkdir, "analysis", "AMR", "colTaxa_Gene.RData"))
#names(colTaxa) <- c(tax_table(ps)[setdiff(names(colTaxa), "Others"),"Gene"])
names(colTaxa) <- c(paste0(tax_table(ps)[setdiff(names(colTaxa), "Others"),"Gene"], " (", tax_table(ps)

data <- read.csv(file.path(indir, "SpeciesAMR-SpeciesCountsPerAMR.csv"))
#data$AMR <- factor(gsub(" \\(.*", "", as.vector(data$AMR)), levels= gsub(" \\(.*", "", as.vector(data$
data$AMR <- factor(as.vector(data$AMR), levels= as.vector(data$AMR))
data <- data[data$AMR %in% as.list(taxa.order.top50), ]
data <- merge(data, taxa.means, by.x = "AMR",
              by.y = "Resistome", all.x = TRUE, all.y = FALSE)
#data$AMR <- factor(data$AMR, levels = data$AMR[order(data$Count)])
colAMR <- rep("gray40", nrow(data))
names(colAMR) <- as.vector(data$AMR)
colAMR[names(colTaxa)] <- colTaxa
colAMR<-colAMR[order(match(names(colAMR),data$AMR[order(data$Count)]))]

colAMR2 <- c("#5CA5DB", "#ffcccc", "#B60004", "#91CE59", "#fc8403", "#EBA5F3","#FFBC06","#3B3B3B","#2A2
colAMR2 <- sapply(1:length(colAMR), function(x) if (!gsub("\\)", "", gsub(".*\\(", "", names(colAMR)[x]]
names(colAMR2) <- names(colAMR)
colAMR <- colAMR2

#data<-data[!grepl("Streptomyces rishiriensis parY", data$AMR),]

p2 <- ggplot(data, aes(x=AMR, y=Count, fill=AMR)) +
  geom_bar(stat = "identity") + #ggtitle("Number of bacterial species associated with AMR genes") +
  coord_flip() + xlab("") + ylab("N bacteria") +
  theme(legend.position = "none",
        axis.ticks.x = element_blank(),
        axis.text.x = element_text(color="black", size= 6, angle= 0, hjust= 1),
        axis.text.y = element_text(face="bold.italic", size= 8, colour=colAMR),
        panel.background = element_rect(fill = "white"),
        panel.border = element_rect(color="gray4", fill=NA),
        #panel.grid.major = element_line(color="gray80"),
        panel.grid.minor = element_line(color="gray80")
  ) +
  scale_fill_manual(name= "AMR", values=colAMR)

p2
```




```
Fig_5<-ggarrange(p, p2, nrow=1)
pdf(file = "../Data/R_output_files/Fig_5DE.pdf",   # The directory you want to save the file in
    width = 12, # The width of the plot in inches
    height = 8)
Fig_5
dev.off()
```

```
## pdf
##   2
```