# Main_Figure_3

Micheál Mac Aogáin

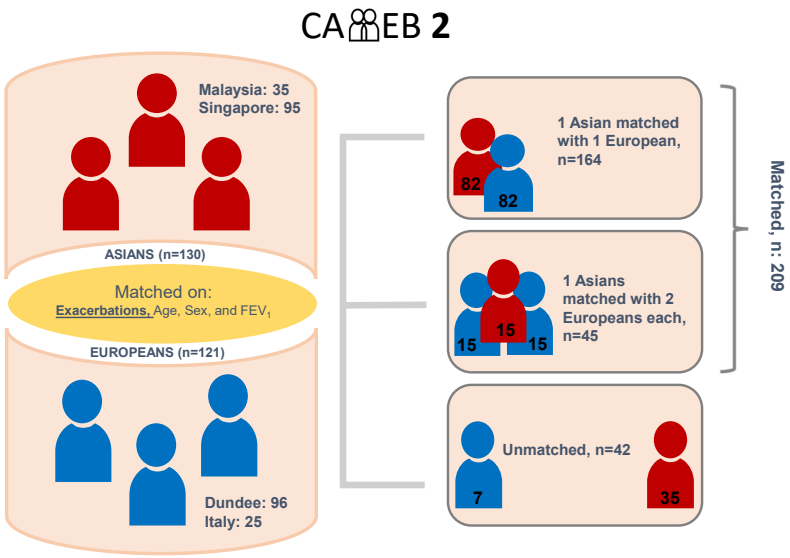
2023-11-23

#Load required R packages

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

#Figure 3A - Schematic illustration

```
knitr::include_graphics("../Data/R_input_files/Figure3A_TJ_schematic.pdf")
```




#Load data

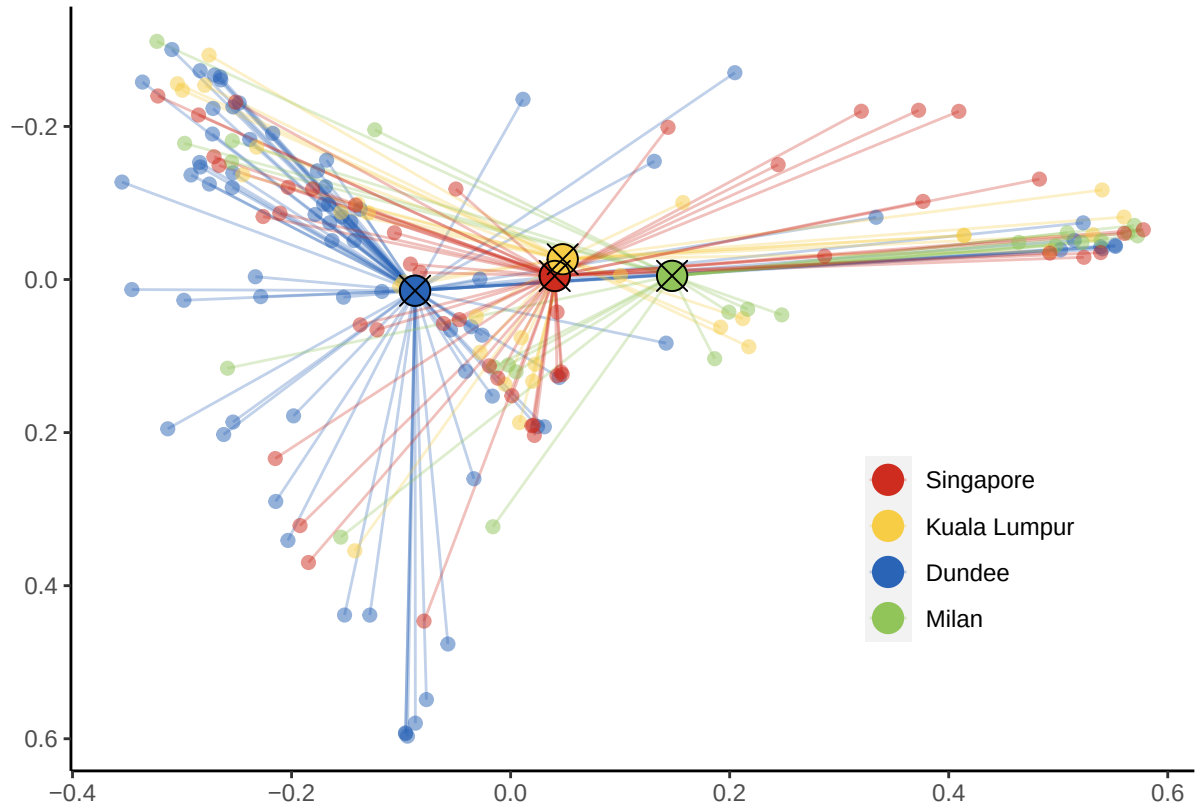
```
##Master data cross-sectional####
Master <-read.csv("../Data/R_input_files//Clinical_AMR_Microbiome_R2.csv") %>%
  as_tibble()
Master$FEVfactor<-cut(Master$FEV1, breaks=c(0, 30, 50, 70, Inf))
```

```
###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", "

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

#Figure 3B

```
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 gene

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

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
```

```r
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"

##AMR PCOA of Resistotypes BY Country####
gg <- data.frame(cluster=factor(AMRDiversityViz_Geo$Country), x=AMRDiversityViz_Geo$Dim1, y=AMRDiversity
# 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_Geo<-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(labels = c("Singapore", "Kuala Lumpur", "Dundee", "Milan"), values = c("#CD2C1E",
  labs(colour="",
       x = "PC 1 (23.8%)", y = "PC 2 (4.5%)")+
  theme(legend.position=c(0.8,0.3),
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
        panel.background = element_rect(fill = NA),
  )+
  scale_y_reverse()+
  xlab("")+
  ylab("")
PCA_Geo
```




```
#PERMANOVA ethnicity
adonis2(AMR_diversity ~ Ethnicity, data=AMRDiversityViz_Geo, 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 ~ Ethnicity, data = AMRDiversityViz_Geo, permutations = 999, method =
##            Df SumOfSqs      R2      F Pr(>F)
## Ethnicity   5    2.684 0.03694 1.2889  0.026 *
## Residual  168   69.977 0.96306
## Total     173   72.661 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#PERMANOVA geography + ethnicity
adonis2(AMR_diversity ~ Country+Ethnicity, data=AMRDiversityViz_Geo, 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 ~ Country + Ethnicity, data = AMRDiversityViz_Geo, permutations = 99
##            Df SumOfSqs      R2      F Pr(>F)
## Country     3    2.845 0.03915 2.3059  0.002 **
```

```
## Ethnicity   4    1.554 0.02139 0.9450  0.628
## Residual  166   68.262 0.93946
## Total     173   72.661 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

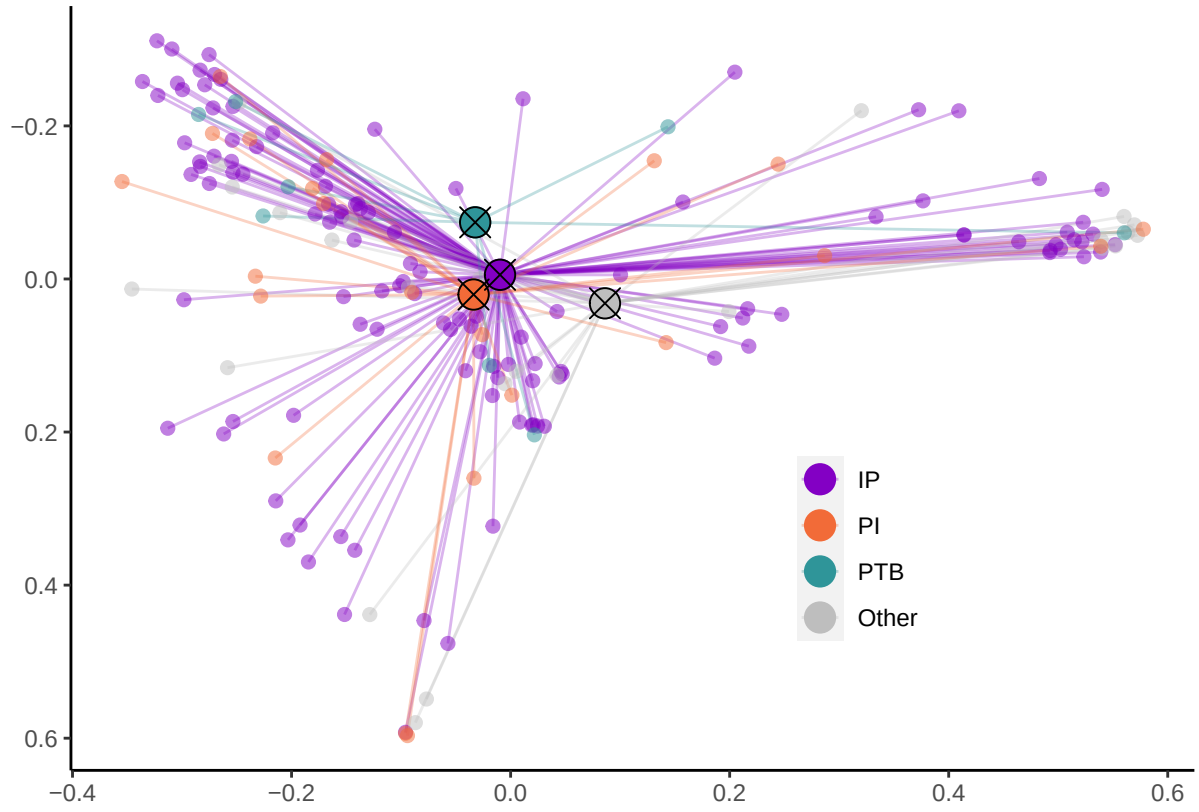
```
#PERMANOVA geography + ethnicity in Asian only
#Subset AMRDiversityViz_Geo based on 'Country'
AMRDiversityViz_Geo_subset <- subset(AMRDiversityViz_Geo, Country == "SG" | Country == "KL")

#Match SampleIDs between AMR_diversity and AMRDiversityViz_Geo_subset
matched_samples <- rownames(AMR_diversity) %in% AMRDiversityViz_Geo_subset$SampleID

#Use which() to get the indices of matched samples
matched_indices <- which(matched_samples)

#Subset AMR_diversity based on the matched_indices
AMR_diversity_subset <- AMR_diversity[matched_indices, ]

#run PERMANOVA geography + ethnicity
adonis2(AMR_diversity_subset ~ Ethnicity, data = AMRDiversityViz_Geo_subset, method = "bray", permutati
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = AMR_diversity_subset ~ Ethnicity, data = AMRDiversityViz_Geo_subset, permutations =
##           Df SumOfSqs      R2      F Pr(>F)
## Ethnicity  4   1.6015 0.05079 0.9364  0.639
## Residual  70  29.9287 0.94921
## Total     74  31.5302 1.00000
```

#Figure 3C

```
##AMR PCOA of Resistotypes BY Aetiology####
gg <- data.frame(cluster=factor(AMRDiversityViz_Geo$Aetiology_short), x=AMRDiversityViz_Geo$Dim1, y=AMRI
# 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_Aet<-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(labels = c("IP", "PI", "PTB", "Other"), values = c("#8300c4","#F26B38","#2F9599",
  labs(colour="",
       x = "PC 1 (23.8%)", y = "PC 2 (4.5%)")+
  theme(legend.position=c(0.7,0.3),
        legend.title = element_blank(),
        axis.line = element_line(size = 0.5, colour = "black"),
```

```
        panel.background = element_rect(fill = NA),
  )+
  scale_y_reverse()+
  xlab("")+
  ylab("")
PCA_Aet
```




```
#PERMANOVA Aetiology
adonis2(AMR_diversity ~ Aetiology_short, data=AMRDiversityViz_Geo, 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 ~ Aetiology_short, data = AMRDiversityViz_Geo, permutations = 999, me
##                   Df SumOfSqs      R2      F Pr(>F)
## Aetiology_short    3    1.338 0.01842 1.0632  0.326
## Residual         170   71.323 0.98158
## Total            173   72.661 1.00000
```

```
#PERMANOVA covariate analysis
AMRDiversityViz_Geo <- AMRDiversityViz_Geo %>% filter(!is.na(BMI))
AMRDiversityViz_Geo <- AMRDiversityViz_Geo %>% filter(!is.na(Number.of.lobes.affected))
AMRDiversityViz_Geo <- AMRDiversityViz_Geo %>% filter(!is.na(PriorHospitalisations))
AMRDiversityViz_Geo <- AMRDiversityViz_Geo %>% filter(!is.na(MMRC.score))
```

```r
filtered_samples <- AMRDiversityViz_Geo[, 1]  # Assuming first column is SampleID
sample_ids_filtered <- as.character(filtered_samples[[1]]) #Extract Sample IDs as a vector
common_samples <- intersect(sample_ids_filtered, rownames(AMR_diversity))#If row names of AMR_diversity

AMR_diversity_filtered <- AMR_diversity[rownames(AMR_diversity) %in% common_samples, ] #Filter AMR_dive

AMRDiversityViz_Geo$other.org<-ifelse(AMRDiversityViz_Geo$Other.organisms.isolated.in.sputum != "1" & Al

adonis2(AMR_diversity_filtered ~ Country + BMI + Aetiology_short+MMRC.score+Number.of.lobes.affected+Pri
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = AMR_diversity_filtered ~ Country + BMI + Aetiology_short + MMRC.score + Number.of.l
##                            Df SumOfSqs      R2      F Pr(>F)
## Country                     3    2.758 0.03966 2.2748  0.001 ***
## BMI                         1    0.436 0.00627 1.0791  0.309
## Aetiology_short             3    1.683 0.02420 1.3879  0.023 *
## MMRC.score                  1    0.814 0.01171 2.0146  0.015 *
## Number.of.lobes.affected    6    2.584 0.03716 1.0657  0.265
## PriorHospitalisations       1    0.373 0.00537 0.9241  0.528
## other.org                   1    0.456 0.00655 1.1278  0.284
## ICS.use                     1    0.313 0.00450 0.7750  0.760
## Long.term.antibiotics       1    0.313 0.00450 0.7742  0.766
## Residual                  148   59.809 0.86008
## Total                     166   69.539 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#INVESTIGATE COLINEARITY
#Create a contingency table for
mytable <- table(AMRDiversityViz_Geo$Aetiology_short, AMRDiversityViz_Geo$Country)
mytable
```

```
##
##              SG KL DD MI
##   idiopathic 25 27 48 16
##   postInfect  6  0 16  0
##   postTB      6  0  0  0
##   other       2  2 12  7
```

```r
#Compute  Cramer's V statistic
assocstats(mytable)
```

```
##                     X^2 df   P(> X^2)
## Likelihood Ratio 44.700  9 1.0483e-06
## Pearson          41.171  9 4.6564e-06
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.445
## Cramer's V        : 0.287
```