R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

##Data import

```
df = read.csv("F:/ZLF/CAP/data/relative_data/dfall-220617.csv",row.names=1)
metadata = read.csv('F:/ZLF/CAP/data/relative_data/metadata-all-220617.csv',row.names = 1)
```

##library

```
#library(tidyverse)
library(vegan)

## 载入需要的程辑包：permute

## 载入需要的程辑包：lattice

## This is vegan 2.5-7

library(reshape2)
library(dplyr)

##
## 载入程辑包：'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(stringr)
library(phyloseq)
library(RColorBrewer)
library(tidyr)

##
## 载入程辑包：'tidyr'

## The following object is masked from 'package:reshape2':
##
##     smiths

library(ggsignif)
library(tinytex)

library(survminer)
```

```
## Warning: 程辑包'survminer'是用 R 版本 4.1.3 来建造的

## 载入需要的程辑包：ggpubr

library(survival)

## Warning: 程辑包'survival'是用 R 版本 4.1.3 来建造的

##
## 载入程辑包：'survival'

## The following object is masked from 'package:survminer':
##
##     myeloma
```

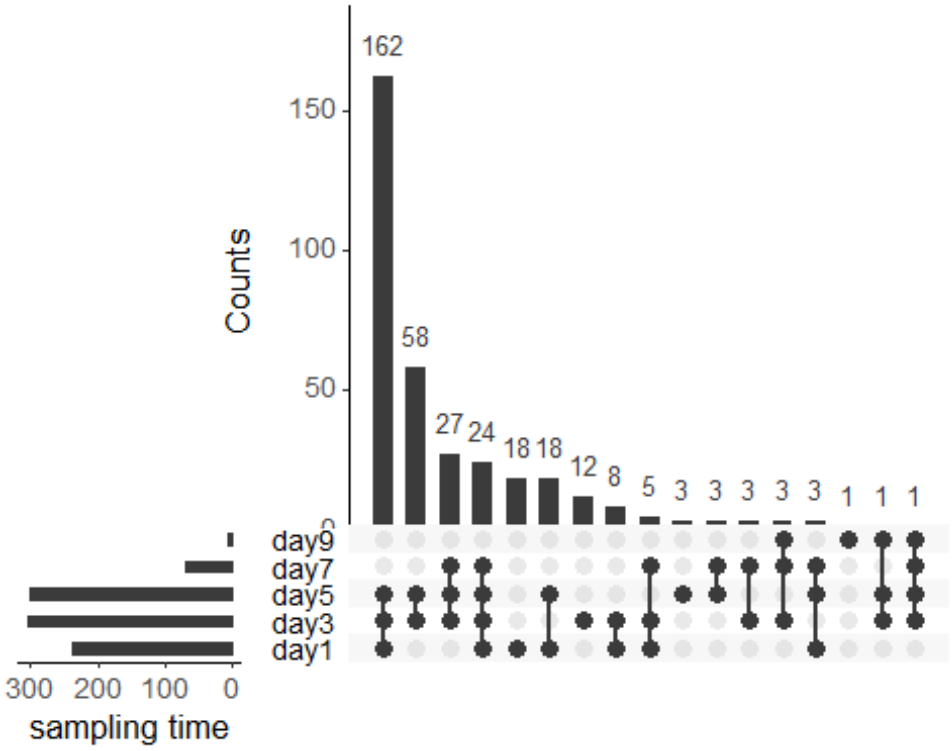
##FUNCTION REVERSIER

```
severityOfCluster <- function(metadata,sev){
  #print severe or non severe to calculate severe or nonsevere proportion of each cluster
  cluster_severity <- data.frame(cluster = 1:10,mild = rep(NA,10),severe = rep(NA,10))
  for (i in cluster_severity$cluster) {
    cluster_severity[i,'severe'] <- filter(metadata,pam_10_cluster == i) %>% filter(sever
e_case == 'yes') %>% nrow()
    cluster_severity[i,'mild'] <- filter(metadata,pam_10_cluster == i) %>% filter(severe_
case == 'no') %>% nrow()
  }
  cluster_severity$severe_proportion <- cluster_severity$severe/((cluster_severity$mild)+
(cluster_severity$severe))
  cluster_severity$nonsevere_proportion <- cluster_severity$mild/((cluster_severity$mild)
+(cluster_severity$severe))
  cluster_severerate <- select(cluster_severity,cluster,paste0(sev,'_proportion'))
  if(sev == 'severe')
  {cluster_severerate[11,] <- c('11',0)}else if(sev == 'nonsevere')
  {cluster_severerate[11,] <- c('11',1)}
  #cluster_severerate$cluster <- factor(cluster_severerate$cluster,levels = c(1:11))
  #sub.meta$pam_10_cluster<-factor(sub.meta$pam_10_cluster,levels = c(1:11))
  colnames(cluster_severerate)[1] <- 'pam_10_cluster'
  return(cluster_severerate)
}

disTance = function(sub.meta,sub.df,dis_method,bind_info1,bind_info2){
  OTU = phyloseq::otu_table(sub.df, taxa_are_rows = F)
  physeq = phyloseq(OTU)

  jsd=phyloseq::distance(physeq, method = dis_method)
  jsd1 = as.matrix(jsd) %>% as.matrix() %>% melt() %>% filter(as.character(Var1) != as.ch
aracter(Var2))

  metavar1 <- select(sub.meta,raw_id,bind_info1)
  metavar2 <- select(sub.meta,raw_id,bind_info2)
  colnames(metavar1)[1]<-c('Var1')
  colnames(metavar2)[1]<-c('Var2')

  jsd2 <- left_join(jsd1,metavar1,by = "Var1")
  jsd2 <- left_join(jsd2,metavar2,by = "Var2")
```

```
  return(jsd2)
}
```

##Figure 1 ###Figure 1A

```
library(UpSetR)

##
## 载入程辑包：'UpSetR'

## The following object is masked from 'package:lattice':
##
##     histogram

for (i in c(1,3,5,7,9)) {
  assign(paste0('day',i),filter(metadata,d == i) %>% rownames() %>% str_split('d') %>% as.
data.frame() %>% t() %>% as.data.frame() %>% select(V1))
}

f1a_plot = list(    day1 = day1$V1,
                    day3 = day3$V1,
                    day5 = day5$V1,
                    day7 = day7$V1,
                    day9 = day9$V1
)


f1a <- upset(fromList(f1a_plot),
             nsets = 8,
             order.by = "freq",
             sets = c('day1','day3','day5','day7','day9'),
             keep.order = TRUE,
             point.size = 3,
             line.size = 1,
             mainbar.y.label = "Counts",
             sets.x.label = "sampling time",
             mb.ratio = c(0.7,0.3),
             text.scale = c(1.5, 1.5, 1.5, 1.5, 1.5, 1.5),
)

f1a
```



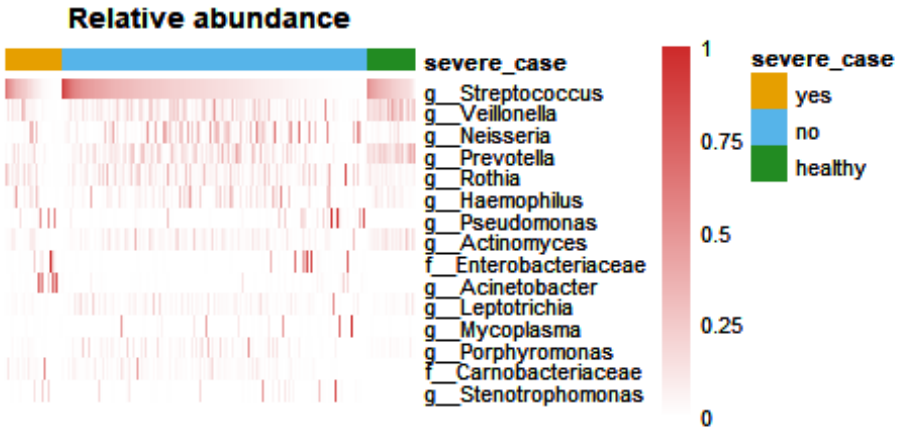
###Figure 1B

```
# Figure 1B ------------------------------------------------------------
library(pheatmap)
metadata_cap <- filter(metadata,subject != 'healthy' & subject != 'nc' & subject != 'NCPC
R')
df_cap <- df[,rownames(metadata_cap)]

#top15 microbes of CAP
sub.df <- df_cap
sub.meta <- metadata_cap
sub.df$sum<-rowSums(sub.df)
sub.df<-sub.df[order(sub.df$sum,decreasing = TRUE),]
captop15 <- rownames(sub.df)[1:15]

##relative abundance of all samples except NC
sub.meta <- metadata %>% filter(subject != 'nc' & subject != 'NCPCR')
sub.df <- df[,rownames(sub.meta)]
sub.df$sum<-rowSums(sub.df)
sub.df<-sub.df[order(sub.df$sum,decreasing = TRUE),]
sub.df<-select(sub.df,-sum)
sub.df<-sub.df[captop15,]
sub.df['others',]<-1-colSums(sub.df)

##ordered by Streptococcus
sub.df<-sub.df[,order(sub.df['g__Streptococcus',],decreasing = TRUE)]

## Warning in xtfrm.data.frame(x): cannot xtfrm data frames

annotation = metadata[colnames(sub.df),]%>%select(severe_case) %>% arrange(severe_case)

for (ci in c('yes','no','healthy')) {
```

```
  mid <- filter(sub.meta,severe_case == ci)
  assign(paste0(ci),select(sub.df,rownames(mid)))
  assign(paste0(ci),get(ci)[,order(get(ci)['g__Streptococcus',],decreasing = TRUE)])
}
```

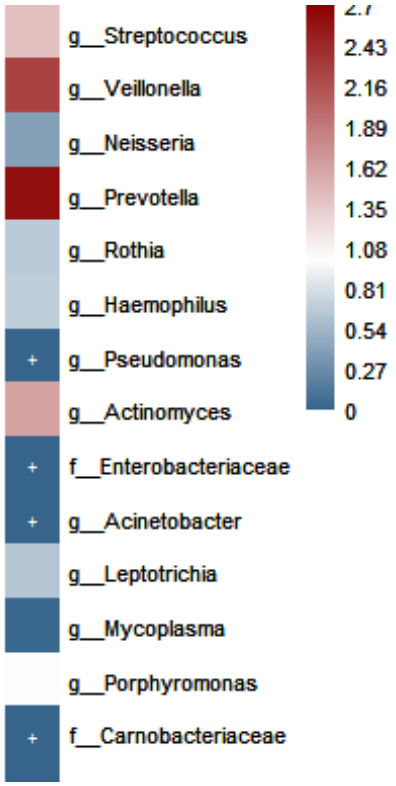
```
## Warning in xtfrm.data.frame(x): cannot xtfrm data frames

## Warning in xtfrm.data.frame(x): cannot xtfrm data frames

## Warning in xtfrm.data.frame(x): cannot xtfrm data frames
```

```
sub.df1<-cbind(yes,no,healthy)
sub.df1<-sub.df1[rownames(sub.df1)[1:nrow(sub.df1)-1],]

f1b_plot <- sub.df1


sub.meta$severe_case <- factor(sub.meta$severe_case,levels = c('yes','no','healthy'))
ann_colors = list(severe_case = c(yes = "#E69F00", no = "#56B4E9",healthy = 'forestgreen
'))

bk = c(seq(0,1,by=0.001))

f1b<-pheatmap(f1b_plot,
              color = colorRampPalette(c( "white", "firebrick3"))(length(colnames(sub.df
1))),
              cellwidth = 0.15, cellheight = 8,
              fontsize=8,
              show_colnames=FALSE,
              legend = TRUE,
              legend_breaks=seq(0,1,0.25),
              breaks = bk,
              cluster_row = FALSE,
              cluster_cols = FALSE,
              main = 'Relative abundance',
              border = F,
              annotation_col = annotation,
              annotation_colors = ann_colors
)
f1b
```




```
dff1a<-data.frame(median = rep(NA,15))
rownames(dff1a)<-rownames(sub.df1)

for (i in 1:nrow(sub.df1)) {
  dff1a$median[i] <- median(sub.df1[i,] %>% as.numeric())
}

for (i in 1:nrow(sub.df1)) {
  dfcap <- select(sub.df1,-starts_with('SRR'))
  dfhea <- select(sub.df1,starts_with('SRR'))
  dff1a$capmedian[i] <- median(dfcap[i,] %>% as.numeric())
  dff1a$heamedian[i] <- median(dfhea[i,] %>% as.numeric())
  dff1a$capmean[i] <- mean(dfcap[i,] %>% as.numeric())
  dff1a$heamean[i] <- mean(dfhea[i,] %>% as.numeric())
}

dff1a$foldchange <- dff1a$heamean/dff1a$capmean
dff1a$foldchangech <- dff1a$capmean/dff1a$heamean
f1b_plot_2 <- dff1a


bk = c(seq(0,2.7,by=0.0027))
f1b_2<-pheatmap(select(f1b_plot_2 ,foldchange),
                color = c(colorRampPalette(colors = c("steelblue4","white"))(floor(length
(bk)*(1/2.7))),colorRampPalette(colors = c( "white","darkred"))(ceiling(length(bk)*(1.7/2.
7)))),
                cellwidth = 20, cellheight = 20,
                fontsize=8,
                show_colnames=FALSE,
                legend = TRUE,
                legend_breaks=seq(0,2.7,0.27),
                breaks=bk,
```

```
                cluster_row = FALSE,
                cluster_cols = FALSE,
                main = 'Relative abundance',
                border = F,
                display_numbers = matrix(ifelse(select(dff1a,foldchange) == 0, "+", ""),
nrow(select(dff1a,foldchange))),
                number_color = "white",
                annotation_colors = ann_colors

)

f1b_2
```



###Figure 1D

```
sub.meta <- metadata %>% filter(subject != 'nc' & subject != 'NCPCR')
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()

f1d_plot <-disTance(sub.meta,sub.df,'jsd','severe_case','pam_10_cluster') %>%
  filter(severe_case == 'healthy')

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(bind_info1)` instead of `bind_info1` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(bind_info2)` instead of `bind_info2` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

colnames(f1d_plot)[4] <- 'severity'
colnames(f1d_plot)[3]<-'distance_to_healthy'
```

```
f1d_plot$pam_10_cluster <- f1d_plot$pam_10_cluster %>% as.character()%>% factor(levels =
as.character(1:11))
cluster_severerate = severityOfCluster(metadata,sev = 'severe')

f1d_plot <- left_join(f1d_plot,cluster_severerate,by = "pam_10_cluster")
f1d_plot$severe_proportion <- as.numeric(f1d_plot$severe_proportion)

f1d<-
  ggplot(f1d_plot,aes(x=pam_10_cluster,y=distance_to_healthy))+
  geom_violin(aes(alpha = severe_proportion),fill = 'forestgreen',width = 1.5)+
  geom_boxplot(aes(alpha = severe_proportion),fill = 'forestgreen',width = 0.1,
               position = position_identity(),outlier.size = 0.1)+
  scale_alpha_continuous(limits = c(0,0.8))+
  geom_signif(comparisons = list(c("1", "11"),c('2','11'),c('3','11'),c("4", "11"),c('5',
'11'),c('6','11'),c("7", "11"),c('8','11'),c('9','11'),c('10','11')),map_signif_level=T,t
est = "wilcox.test")+
  labs(title="distance to healthy",size=11) +
  theme_bw()+
  theme(plot.margin=unit(c(2,0.5,2,0.5),'cm'))
f1d

## Warning: position_dodge requires non-overlapping x intervals
```

###Figure 1E

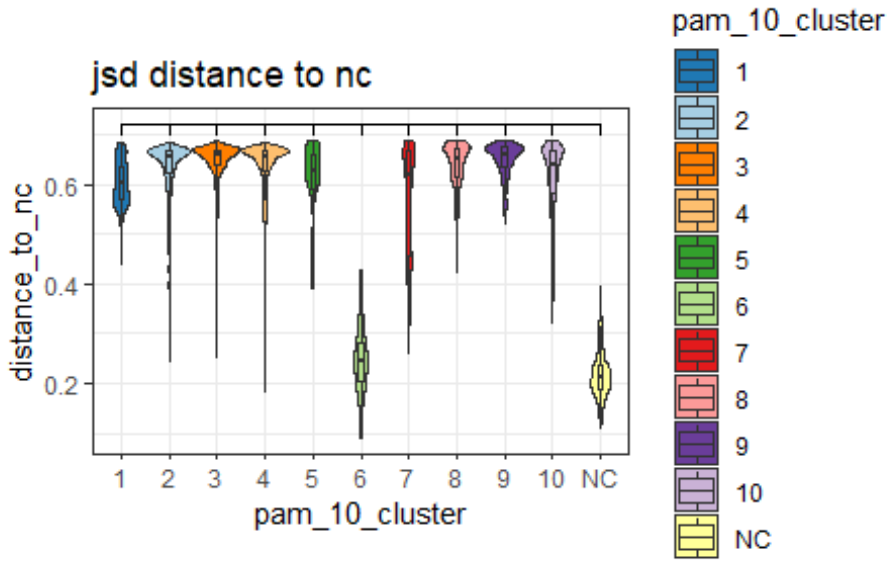
```
sub.meta <- metadata %>% filter(subject != 'nc' & subject != 'NCPCR') %>% select(raw_id,p
am_10_cluster,Shannon)

cluster_severerate = severityOfCluster(metadata,sev = 'severe')
sub.meta$pam_10_cluster <- sub.meta$pam_10_cluster %>% as.character()
f1e_plot <- left_join(sub.meta,cluster_severerate)

## Joining, by = "pam_10_cluster"
```

```
f1e_plot$severe_proportion = as.numeric(f1e_plot$severe_proportion)
f1e_plot$pam_10_cluster <- factor(f1e_plot$pam_10_cluster,levels = 1:11 %>% as.character
())

f1e<-
  ggplot(f1e_plot,aes(x=pam_10_cluster,y=Shannon,alpha = severe_proportion))+
  geom_violin(aes(alpha = severe_proportion),fill = 'forestgreen',width = 1)+
  geom_boxplot(aes(alpha = severe_proportion),fill = 'forestgreen',width = 0.1,position =
 position_identity(),outlier.size = 0.1)+
  scale_alpha_continuous(limits = c(0,0.8))+
  geom_signif(comparisons = list(c("1", "11"),c('2','11'),c('3','11'),c("4", "11"),c('5',
'11'),c('6','11'),c("7", "11"),c('8','11'),c('9','11'),c('10','11')),map_signif_level=T,t
est = "wilcox.test")+
  labs(title="shannnon of each cluster",size=11) +
  theme_bw()+
  theme(plot.margin=unit(c(2,0.5,2,0.5),'cm'))
f1e
```

###Figure 1G

```
sub.meta <- metadata %>% filter(subject != 'healthy')
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()

dis_to_nc <- disTance(sub.meta,sub.df,'jsd','subject',c('subject','pam_10_cluster')) %>%
  filter(subject.x == 'nc'|subject.x == 'NCPCR')

dis_to_nc[dis_to_nc[,'subject.y'] %in% c('nc','NCPCR'),'pam_10_cluster'] = 'NC'

colnames(dis_to_nc)[3]<-'distance_to_nc'
#colnames(dis_to_nc)[5]<-'cluster'

dis_to_nc$pam_10_cluster<-factor(dis_to_nc$pam_10_cluster,levels = c(1,2,3,4,5,6,7,8,9,10,
```

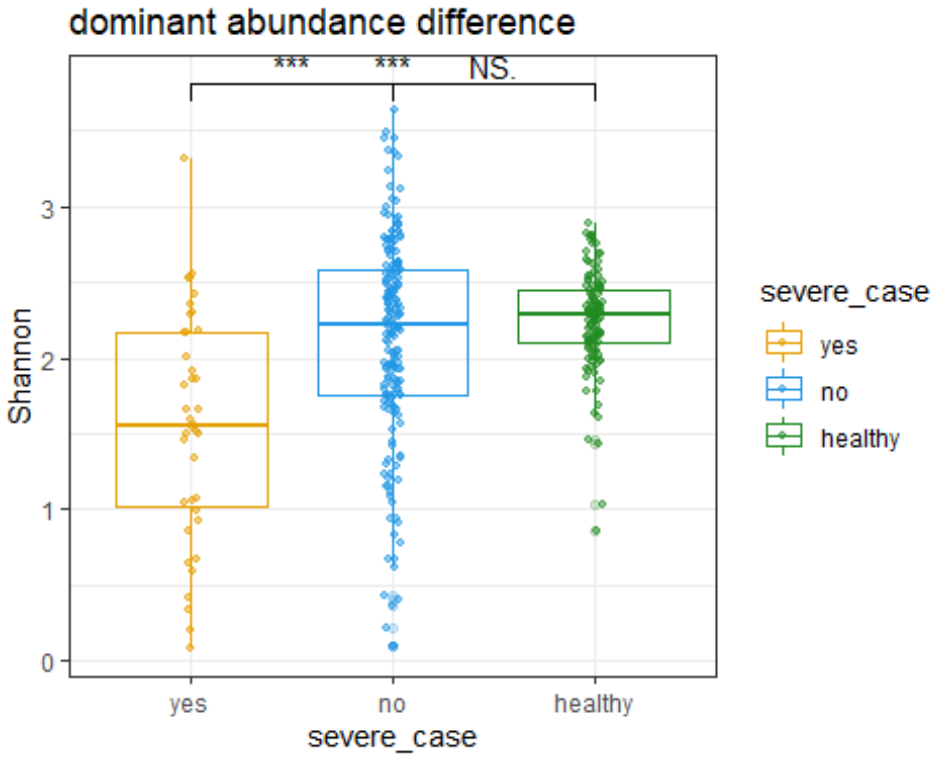
```
'NC') %>% as.character())

f1f_plot <- dis_to_nc[!is.na(dis_to_nc[,'pam_10_cluster']),]

#pal <- c(rev(brewer.pal(11,'Paired'),rev(brewer.pal(7,8,'Paired'))))
pal <- c("#1F78B4","#A6CEE3","#FF7F00","#FDBF6F","#33A02C","#B2DF8A","#E31A1C","#FB9A99",
"#6A3D9A","#CAB2D6","#FFFF99")

pal <- c("#1F78B4","#A6CEE3","#FF7F00","#FDBF6F","#33A02C","#B2DF8A","#E31A1C","#FB9A99",
"#6A3D9A","#CAB2D6","#FFFF99")

#dis_to_nc[is.na(dis_to_nc[,'cluster']),'Var2'] %>% unique()
f1f<-
  ggplot(f1f_plot,aes(pam_10_cluster,y=distance_to_nc))+
  geom_violin(aes(fill = pam_10_cluster),width = 1,)+
  geom_boxplot(aes(fill = pam_10_cluster),width = 0.1,position = position_identity(),outl
ier.shape = NA)+
  scale_color_manual(values = pal)+
  scale_fill_manual(values = pal)+
  geom_signif(comparisons = list(c("1", "NC"),c('2','NC'),c('3','NC'),c("4", "NC"),c('5',
'NC'),c('6','NC'),c("7", "NC"),c('8','NC'),c('9','NC'),c('10','NC')),map_signif_level=T,t
est = "wilcox.test")+
  labs(title="jsd distance to nc",size=11) +
  theme_bw()+
  theme(plot.margin=unit(c(2,0.5,2,0.5),'cm'))
f1f
```



##Figure 2 ###Figure 2A

```
sub.meta <- filter(metadata,d == 1 | d == 'healthy') %>% filter(severe_case == 'yes'|seve
re_case == 'no'|severe_case == 'healthy')
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()
```

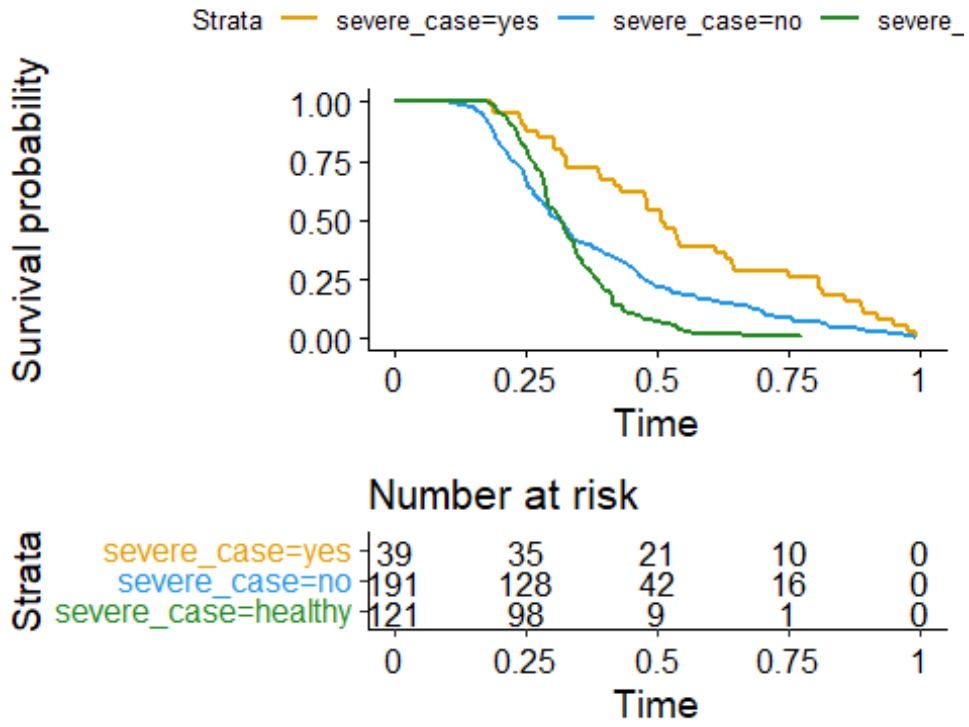
```
f2a_plot <- sub.meta
f2a_plot$severe_case<-factor(f2a_plot$severe_case,levels = c('yes','no','healthy'))

f2a<-
  ggplot(f2a_plot,aes(x=severe_case,y=Shannon))+
  geom_boxplot(aes(color = severe_case),alpha =0.2,weight = 3)+
  geom_point(aes(fill = severe_case,color = severe_case),position = position_jitterdodge
(0.2),alpha = 0.5,shape = 20,size = 1.6)+
  scale_color_manual(values = c('#E69F00','124','forestgreen'))+
  scale_fill_manual(values = c('#E69F00','124','forestgreen'))+
  geom_signif(comparisons = list(c("yes", "no"), c("no", "healthy"),c("yes", "healthy")),
map_signif_level=T,test = "wilcox.test")+
  labs(title="dominant abundance difference",size=15) +
  theme_bw()
f2a
```



###Figure 2B

```
sub.meta <- filter(metadata,d == 1 | d == 'healthy') %>% filter(severe_case == 'yes'|seve
re_case == 'no'|severe_case == 'healthy')
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()

mid<-apply(sub.df,1,max) %>% as.data.frame()
mid$raw_id<-rownames(mid)
colnames(mid) <- c('abundance','raw_id')
mid1<-apply(sub.df,1,function(t) colnames(sub.df)[which.max(t)]) %>% as.data.frame()
mid1$raw_id<-rownames(mid1)
colnames(mid1) <- c('taxonomy','raw_id')

dom_microbe<-full_join(mid,mid1)
```

```
## Joining, by = "raw_id"

rownames(dom_microbe)<-dom_microbe$raw_id

metainfo<-select(sub.meta,raw_id,severe_case,d)
metainfo$raw_id<-rownames(metainfo)

dom_microbe<-left_join(dom_microbe,metainfo)

## Joining, by = "raw_id"

rownames(dom_microbe)<-dom_microbe$raw_id

dom_microbe$severe_case<-factor(dom_microbe$severe_case,levels = c('yes','no','healthy'))

#write.csv(dom_microbe,'F:/ZLF/CAP/paper_structure/figure2/dominant_microbe0621.csv')

f2b_plot <- dom_microbe
f2b_plot$status = 1

attach(f2b_plot)
#Surv(abundance,status)
fit <- survfit(Surv(abundance,status) ~ severe_case, data = f2b_plot)
ggsurvplot(fit,data = f2b_plot,
           risk.table = TRUE,
           palette = c('#E69F00','124','forestgreen'),
           tables.height = 0.36,
           ggtheme = theme_survminer())
```




```
##Log rank test
sn <- filter(f2b_plot,severe_case == 'yes'|severe_case == 'no')
sh <- filter(f2b_plot,severe_case == 'yes'|severe_case == 'healthy')
```

```r
for (i in c('sn','sh')) {
  attach(get(i))
  b <- coxph(Surv(abundance,status) ~ severe_case,data = get(i))
  print(i)
  print(summary(b))
}
```

```
## The following objects are masked from f2b_plot:
##
##     abundance, d, raw_id, severe_case, status, taxonomy
```

```
## [1] "sn"
## Call:
## coxph(formula = Surv(abundance, status) ~ severe_case, data = get(i))
##
##   n= 230, number of events= 230
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## severe_caseno       0.5921    1.8078   0.1777 3.332 0.000861 ***
## severe_casehealthy     NA        NA   0.0000    NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## severe_caseno           1.808     0.5531     1.276     2.561
## severe_casehealthy         NA         NA        NA        NA
##
## Concordance= 0.559  (se = 0.014 )
## Likelihood ratio test= 12.51  on 1 df,   p=4e-04
## Wald test            = 11.1  on 1 df,   p=9e-04
## Score (logrank) test = 11.41  on 1 df,   p=7e-04
```

```
## The following objects are masked from get(i) (pos = 3):
##
##     abundance, d, raw_id, severe_case, status, taxonomy
```

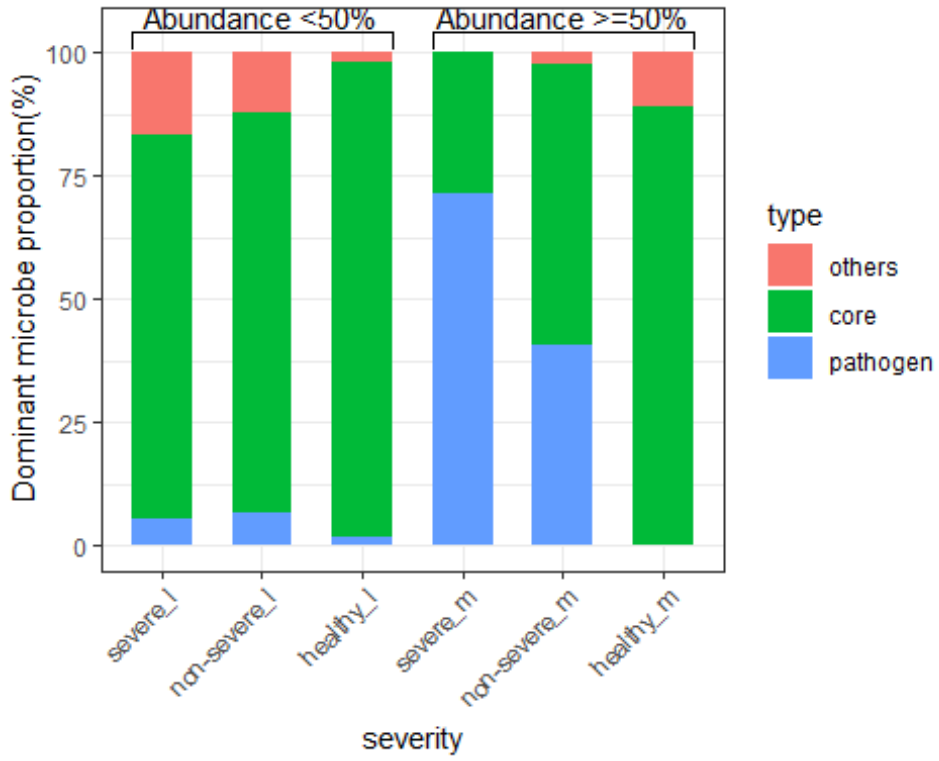
```
## The following objects are masked from f2b_plot:
##
##     abundance, d, raw_id, severe_case, status, taxonomy
```

```
## [1] "sh"
## Call:
## coxph(formula = Surv(abundance, status) ~ severe_case, data = get(i))
##
##   n= 160, number of events= 160
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## severe_caseno           NA        NA   0.0000    NA       NA
## severe_casehealthy 1.3390    3.8150   0.2241 5.976 2.29e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## severe_caseno              NA         NA        NA        NA
## severe_casehealthy      3.815     0.2621     2.459     5.919
```

```
##
## Concordance= 0.6  (se = 0.021 )
## Likelihood ratio test= 43.41  on 1 df,   p=4e-11
## Wald test            = 35.71  on 1 df,   p=2e-09
## Score (logrank) test = 39.11  on 1 df,   p=4e-10
```

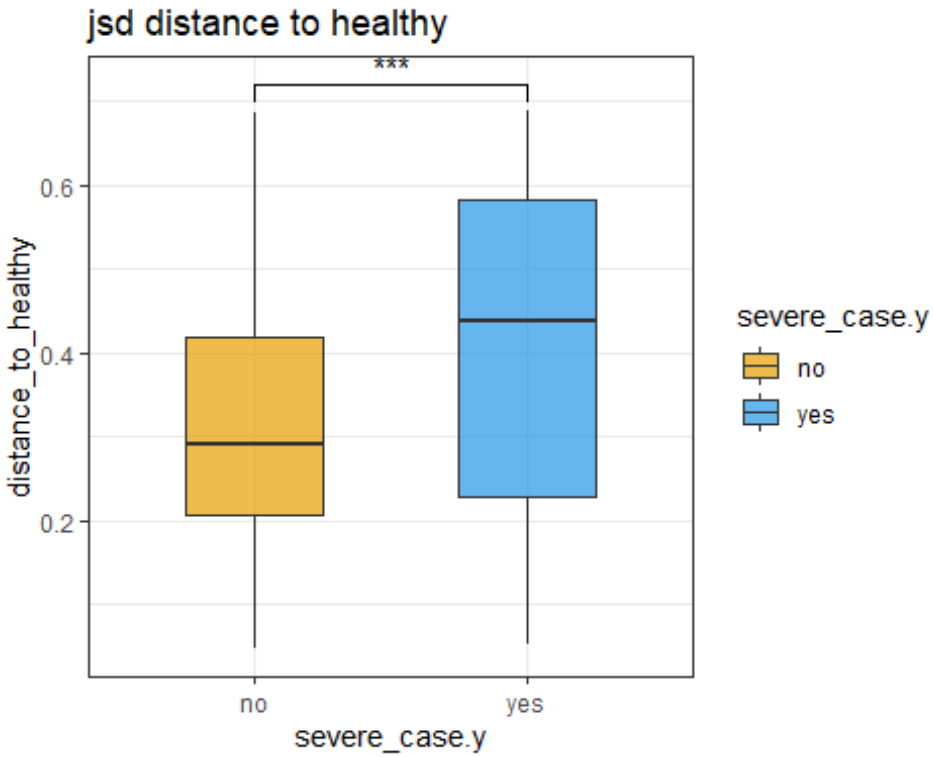
###Figure 2C

```
#f2c_plot <- Dominant microbe type proportion (0.5)

#write.csv(plot_5,'F:/ZLF/CAP/paper_structure/figure2/f1c_0621.csv')
f2c_plot <- read.csv('F:/ZLF/CAP/paper_structure/figure2/f1c_0727.csv',row.names = 1)
f2c_plot$group <- factor(f2c_plot$group,levels = c("severe_l","non-severe_l","healthy_l",
"severe_m","non-severe_m","healthy_m" ))
f2c_plot$type <- factor(f2c_plot$type,levels = rev(c('pathogen','core','others')))
col <- brewer.pal(3,'Set1')



f2c <-
  ggplot(f2c_plot,mapping=aes(x=group,y=proportion*100,fill=type))+
  geom_col(position = "stack",width = 0.6)+
  labs(x = 'severity', y = 'Dominant microbe proportion(%)') +
  #geom_signif(annotations = c('Abundance <50%','Abundance >=50%'), y_position = c(rep(10
4,2)),xmin = c(0.7,3.7),xmax = c(2.3,5.3))+
  geom_signif(annotations = c('Abundance <50%','Abundance >=50%'), y_position = c(rep(104,
2)),xmin = c(0.7,3.7),xmax = c(3.3,6.3))+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
  xlab(NULL)+
  scale_fill_manual(values =  rev(col))

## NULL

f2c
```



###Figure 2D

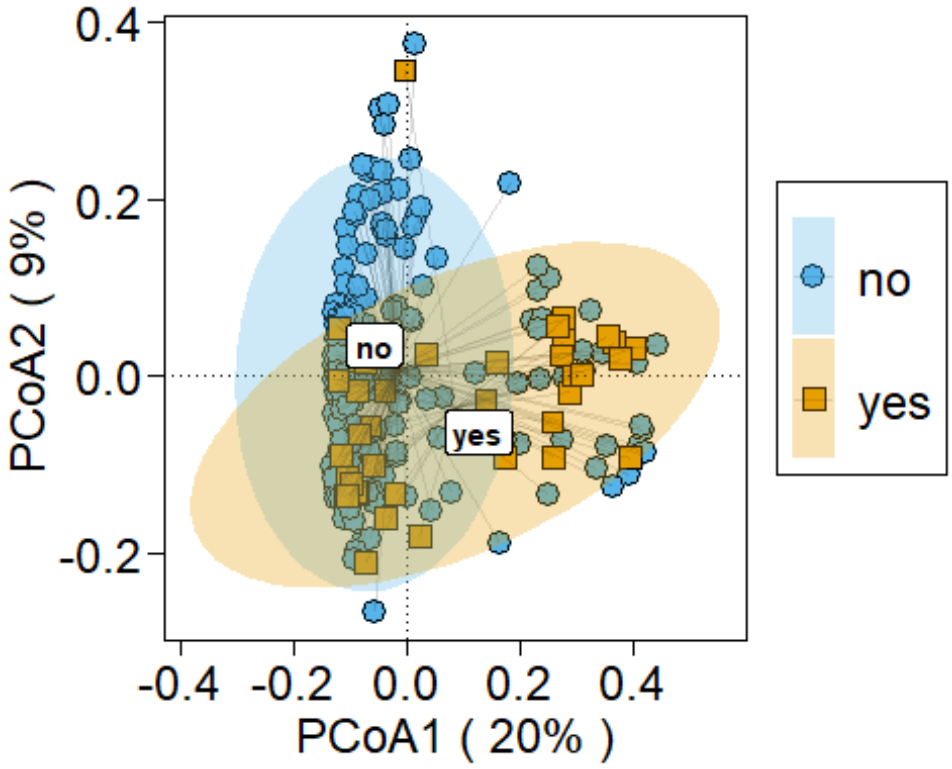
```
sub.meta <- metadata %>% filter(subject != 'nc' & subject != 'NCPCR') %>% filter(d == 1 |
 d == 'healthy')
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()

plot_f2d <- disTance(sub.meta,sub.df,'jsd','severe_case','severe_case') %>% filter(severe
_case.x == 'healthy') %>% filter(severe_case.y == 'yes' | severe_case.y == 'no')


colnames(plot_f2d)[3]<-'distance_to_healthy'


f2d <-
  ggplot(plot_f2d,aes(x=severe_case.y,y=distance_to_healthy))+
  #geom_violin(aes(color = cluster,fill = cluster),alpha=0.7)+
  geom_boxplot(aes(fill = severe_case.y),width = 0.5,position = position_identity(),alpha
=0.7)+
  scale_color_manual(values = c('#E69F00','124','forestgreen'))+
  scale_fill_manual(values = c('#E69F00','124','forestgreen'))+
  geom_signif(comparisons = list(c("yes", "no")),map_signif_level=T,test = "wilcox.test")
+
  labs(title="jsd distance to healthy",size=11) +
  #  xlab(paste("PC1 ( ",pc1,"%"," )",sep="")) +
  #  ylab(paste("PC2 ( ",pc2,"%"," )",sep=""))+
  theme_bw()
#legend.position = 'NONE')

#xlab(NULL)+ylab(NULL)

f2d
```



###Figure 2E

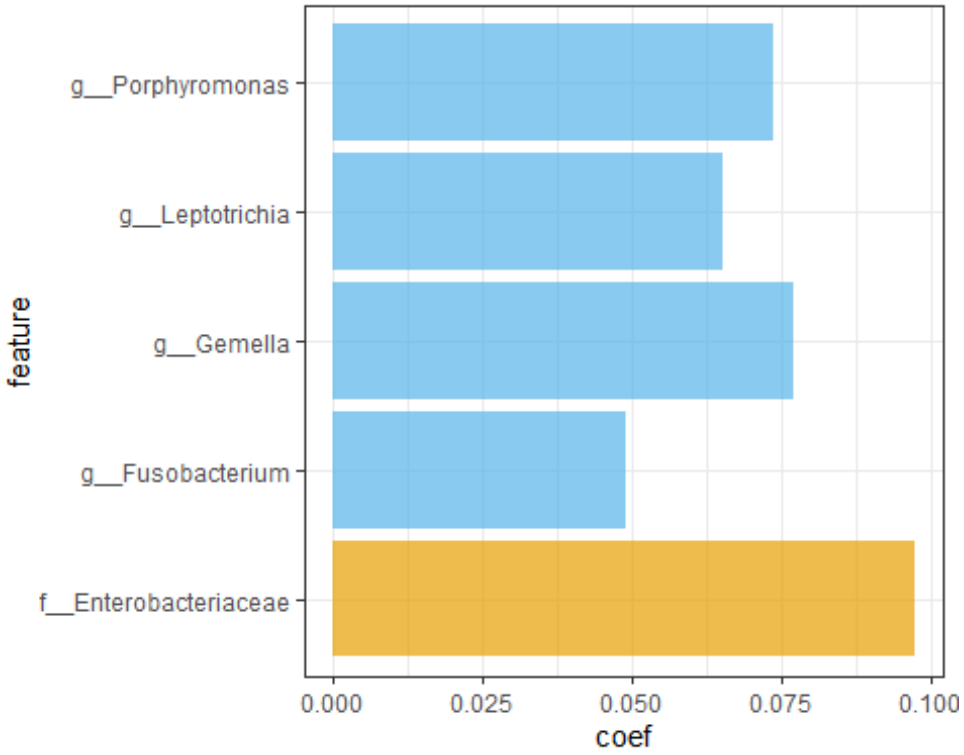
```
sub.meta <- metadata %>% filter(severe_case == 'yes' | severe_case == 'no') %>% filter(d
== 1)
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()

val = 'severe_case'

OTU = phyloseq::otu_table(sub.df, taxa_are_rows = F)
physeq = phyloseq(OTU)
jsd=phyloseq::distance(physeq, method = "jsd")
jsd.cord = cmdscale(jsd,k=2,eig = T)

#制作绘图文件
PC1 = jsd.cord$points[,1]
PC2 = jsd.cord$points[,2]
plotdata <- data.frame(rownames(jsd.cord$points),PC1,PC2,sub.meta[,val])
colnames(plotdata) <-c("sample","PC1","PC2","group")


#用于填充样本点的颜色
cbbPalette <- c( "#56B4E9","#E69F00", "#009E73", "#F0E442","red","grey")
#样本点的边框颜色
Palette <- c("#000000", "#000000","#000000","#000000","#000000","#000000")
#用于绘制横纵坐标 label 的文本，以显示解释比例
eigen.vals.jsd = jsd.cord$eig
last_one = sum(eigen.vals.jsd>0)
pc1 <-floor(eigen.vals.jsd[1]*100/sum(eigen.vals.jsd[1:last_one]))
pc2 <-floor(eigen.vals.jsd[2]*100/sum(eigen.vals.jsd[1:last_one]))

pich=rep(c(21:24),3)
```

```
f2e_plot <- merge(plotdata,aggregate(cbind(mean.x=PC1,mean.y=PC2)~group,plotdata,mean),by
="group")

f2e = ggplot(f2e_plot, aes(PC1, PC2,)) +
  geom_point(aes(colour=group,shape=group,fill=group),size=4)+
  geom_segment(aes(x=mean.x,y=mean.y,xend=PC1, yend=PC2,color =  group),alpha=0.15)+
  scale_shape_manual(values=pich)+
  scale_colour_manual(values=Palette)+
  scale_fill_manual(values=cbbPalette)+
  xlab(paste("PCoA1 ( ",pc1,"%"," )",sep="")) +
  ylab(paste("PCoA2 ( ",pc2,"%"," )",sep=""))+
  theme(text=element_text(size=15))+
  geom_vline(aes(xintercept = 0),linetype="dotted")+
  geom_hline(aes(yintercept = 0),linetype="dotted")+
  theme(panel.background = element_rect(fill='white', colour='black'),
        panel.grid=element_blank(),
        axis.title = element_text(color='black',size=10),
        axis.ticks.length = unit(0.4,"lines"), axis.ticks = element_line(color='black'),
        axis.line = element_line(colour = "black"),
        axis.title.x=element_text(colour='black', size=18),
        axis.title.y=element_text(colour='black', size=18),
        axis.text=element_text(colour='black',size=18),
        legend.title=element_blank(),
        legend.text=element_text(size=18),
        legend.key=element_blank(),
        legend.background = element_rect(colour = "black"),
        legend.key.height=unit(1.6,"cm"))+
  theme(plot.title = element_text(size=34,colour = "black",hjust = 0.5,face = "bold")) +
  stat_ellipse(aes(fill = group),geom = "polygon",level = 0.95,alpha = 0.3)+
  ggrepel::geom_label_repel(data=unique(select(f2e_plot ,mean.x,mean.y,group)),
                            aes(mean.x,mean.y,color=group),
                            #label=c('quit','non-smoke','somke'),
                            label=c(unique(f2e_plot $group)),
                            #fontface="bold",show.legend = F,box.padding = 0,size=1.5)
                            fontface="bold",show.legend = F,box.padding = 0,size=4)
f2e
```



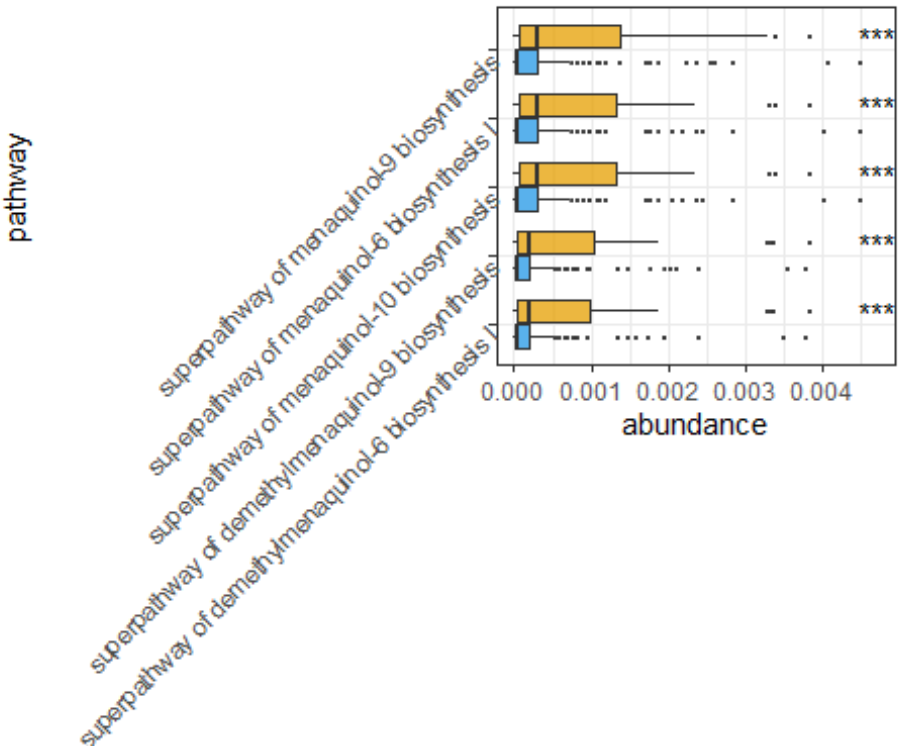
###Figure 2G

```
#f2g_plot <- Severity related microbes
f2g_plot <- read.csv('F:/ZLF/CAP/paper_structure/figure2/f2g_0621.csv',row.names = 1)
#write.csv(f2g_plot,'F:/ZLF/CAP/paper_structure/figure2/f2g_0621.csv')

f2g<-ggplot(f2g_plot,aes(x=coef,y=feature))+
  geom_bar(stat = "identity",fill = ifelse(f2g_plot$severity == 'yes',"#E69F00","#56B4E9
"),alpha = 0.7)+
  theme_bw()
f2g
```



###Figure 2H

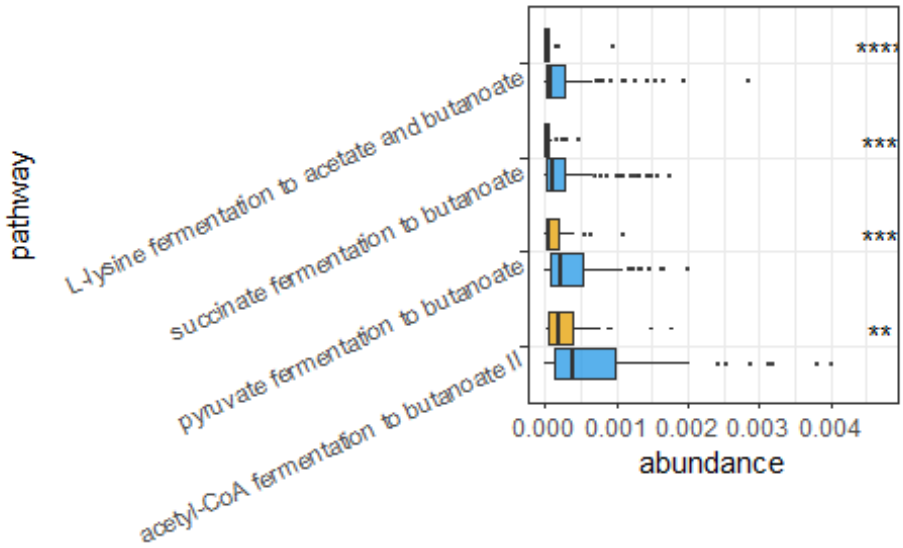
```
pathway_abundance <- read.csv("F:/ZLF/CAP/paper_structure/figure6/pathway/pathway_rel.csv
",row.names = 1)

sub.meta <- filter(metadata,d == 1) %>% filter(severe_case == 'yes' | severe_case == 'no')
sub.df <- pathway_abundance[,rownames(sub.meta)] %>% t() %>% as.data.frame()
sub.df$raw_id <- rownames(sub.df)

f2h_plot <- left_join(select(sub.meta,raw_id, severe_case),sub.df) %>%
  gather(key="pathway",value="abundance",c('superpathway of menaquinol-9 biosynthesis','s
uperpathway of menaquinol-10 biosynthesis','superpathway of menaquinol-6 biosynthesis I',
                                           'superpathway of demethylmenaquinol-6 biosynth
esis I','superpathway of demethylmenaquinol-9 biosynthesis')) %>%
  dplyr::select(raw_id,pathway,abundance,everything())

## Joining, by = "raw_id"

f2h<-ggplot(data = f2h_plot,aes(x = pathway, y = abundance))+
  geom_boxplot(aes(fill = severe_case),outlier.shape=7,outlier.size = 0.0001,alpha = 0.75)
+
  scale_fill_manual(values = c('124','#E69F00'))+
  stat_compare_means(aes(group = severe_case), method = "wilcox.test",label = "p.signif",
label.y = 0.0047)+
  coord_flip()+
  theme_bw()+
  theme(axis.text.y = element_text(angle = 45, vjust = 1))+
  theme(plot.margin=unit(c(0.5,0.5,4,0.5),'cm'))+
  theme(legend.position = 'none')
  #ggtitle('Menaquinol/Demethylmenaquinol Biosynthesis')
f2h
```



###Figure 2I

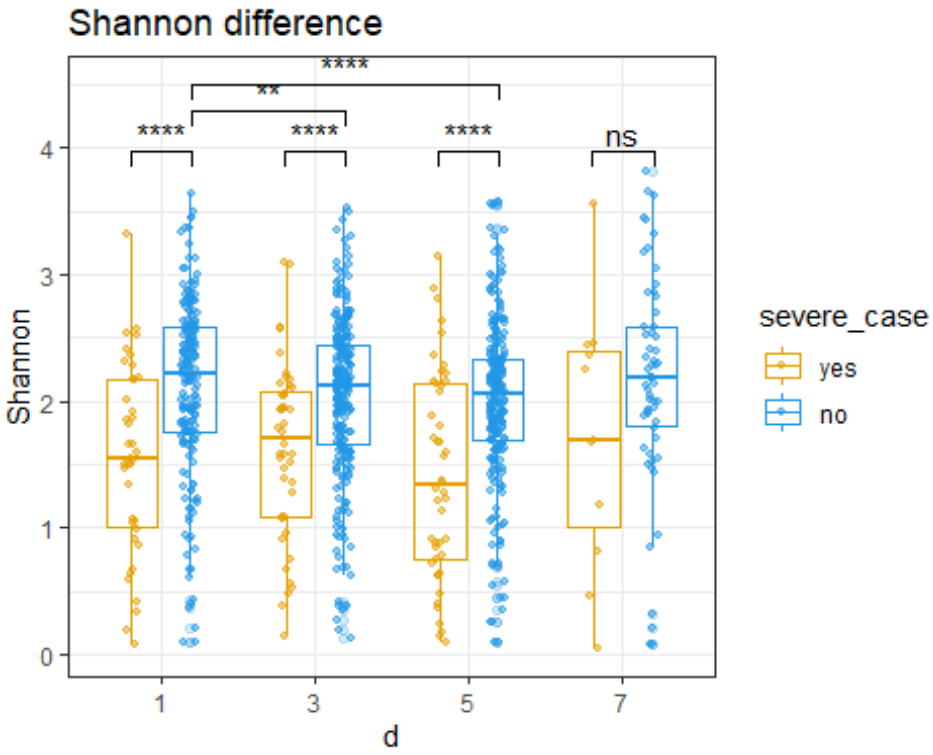
```
pathway_abundance <- read.csv("F:/ZLF/CAP/paper_structure/figure6/pathway/pathway_rel.csv
",row.names = 1)

sub.meta <- filter(metadata,d == 1) %>% filter(severe_case == 'yes' | severe_case == 'no')
sub.df <- pathway_abundance[,rownames(sub.meta)] %>% t() %>% as.data.frame()
sub.df$raw_id <- rownames(sub.df)

f2i_plot <- left_join(select(sub.meta,raw_id, severe_case),sub.df) %>%
  gather(key="pathway",value="abundance",c('acetyl-CoA fermentation to butanoate II','pyr
uvate fermentation to butanoate','succinate fermentation to butanoate','L-lysine fermenta
tion to acetate and butanoate')) %>%
  dplyr::select(raw_id,pathway,abundance,everything())

## Joining, by = "raw_id"

f2i_plot$pathway <- factor(f2i_plot$pathway,levels = c('acetyl-CoA fermentation to butano
ate II','pyruvate fermentation to butanoate','succinate fermentation to butanoate','L-lys
ine fermentation to acetate and butanoate'))


f2i<-ggplot(data = f2i_plot,aes(x = pathway, y = abundance))+
  geom_boxplot(aes(fill = severe_case),outlier.shape=7,outlier.size = 0.0001,alpha = 0.75)
+
  scale_fill_manual(values = c('124','#E69F00'))+
  stat_compare_means(aes(group = severe_case), method = "wilcox.test",label = "p.signif",
label.y = 0.0047)+
  coord_flip()+
  theme_bw()+
  theme(axis.text.y = element_text(angle = 25, vjust = 1))+
  theme(plot.margin=unit(c(2,0.5,2,0.5),'cm'))+
```

```
  theme(legend.position = 'none')
  #ggtitle('Fermentation to butanoate')
f2i
```



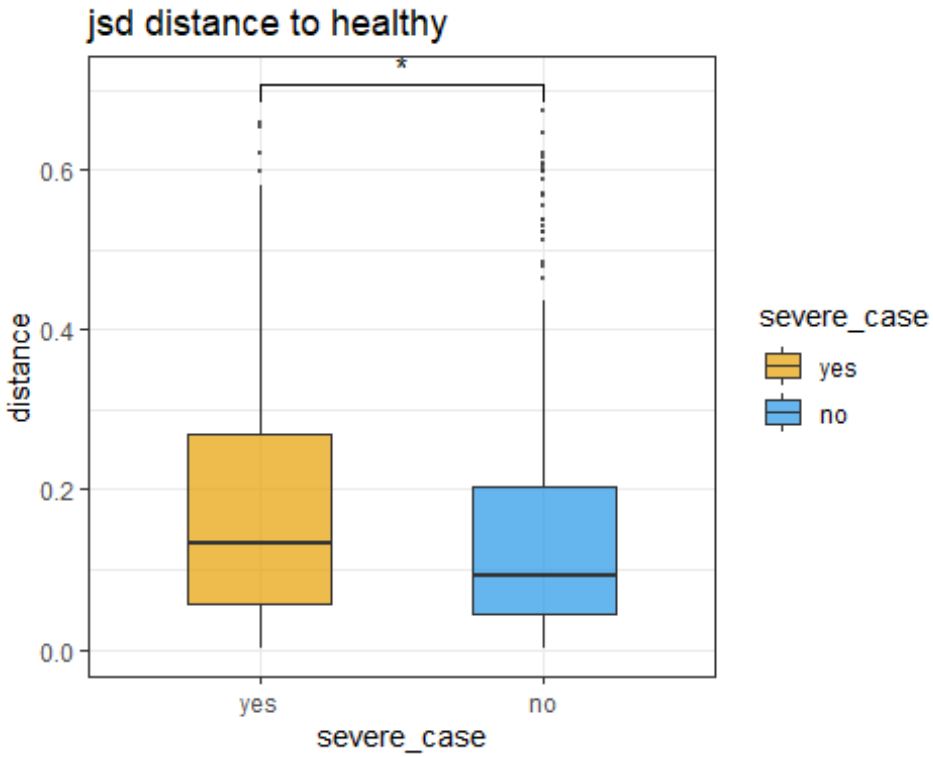
##Figure 3 ###Figure 3A

```
sub.meta <- filter(metadata, severe_case == 'yes' | severe_case == 'no') %>% filter(d != 9)

f3a_plot <- sub.meta
f3a_plot$severe_case <- factor(f3a_plot$severe_case,levels = c('no','yes'))
f3a_plot$severe_case <- factor(f3a_plot$severe_case,levels = c('yes','no'))

f3a<-
  ggplot(f3a_plot,aes(x=d,y=Shannon))+
  geom_boxplot(aes(color = severe_case),alpha =0.2,weight = 3)+
  geom_point(aes(fill = severe_case,color = severe_case),position = position_jitterdodge(0.2),alpha = 0.5,shape = 20,size = 1.6)+
  scale_color_manual(values = c('#E69F00','124'))+  scale_fill_manual(values = c('#E69F00','124'))+
  geom_signif(annotations = c('****','****','****','ns'), y_position = c(rep(3.975,4)),
              xmin = c(0.8,1.8,2.8,3.8),xmax = c(1.2,2.2,3.2,4.2))+
  geom_signif(annotations = c('**','****'), y_position = c(4.3,4.5),xmin = c(1.2,1.2),xmax = c(2.2,3.2))+
  labs(title="Shannon difference",size=15) +
  theme_bw()


f3a
```



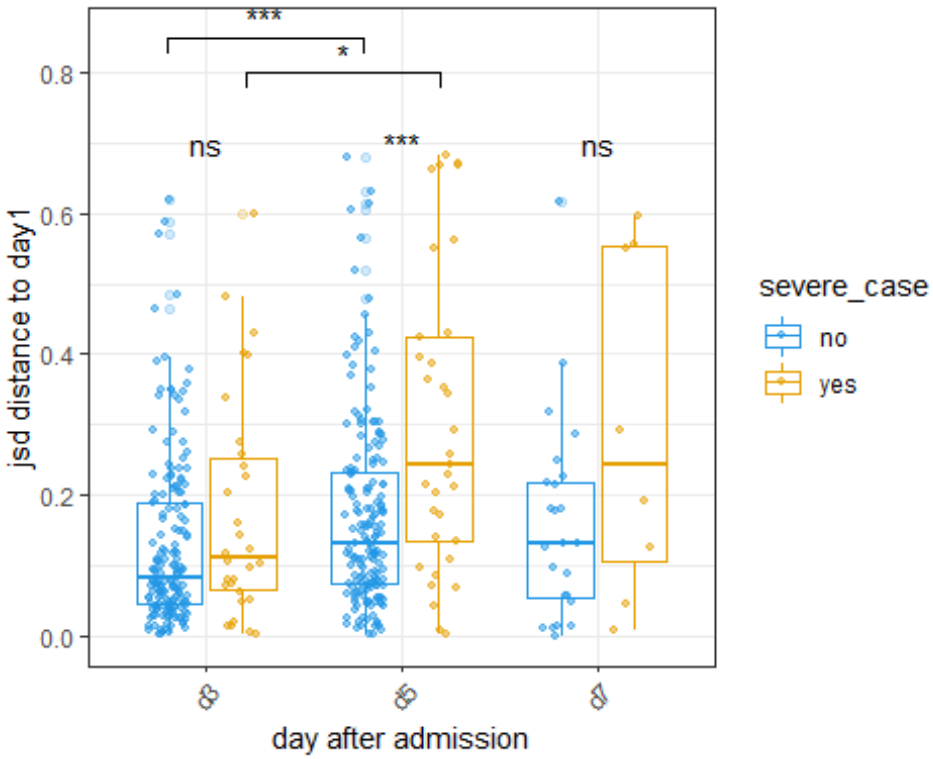
###Figure 3B

```
f3b_plot <- read.csv('F:/ZLF/CAP/paper_structure/figure3/jsd_220618.csv',row.names = 1)

f3b_plot$severe_case <- factor(f3b_plot$severe_case,levels = c('yes','no'))

f3b <-
  ggplot(f3b_plot,aes(x = severe_case,y = distance))+
  geom_boxplot(aes(fill = severe_case),width = 0.5,position = position_identity(),alpha=0.
7,outlier.size = 0.001)+
  scale_color_manual(values = c('#E69F00','124','forestgreen'))+
  scale_fill_manual(values = c('#E69F00','124','forestgreen'))+
  geom_signif(comparisons = list(c("yes", "no")),map_signif_level=T,test = "wilcox.test")
+
  labs(title="jsd distance to healthy",size=11) +
  theme_bw()

f3b
```



###Figure 3C

```
f3c_plot<-read.csv('F:/ZLF/CAP/paper_structure/figure3/f3c0621.csv',row.names = 1)

#f3c_plot<-JSD distance to day1

colnames(f3c_plot)[3:4] <- c('day','distance')

f3c<-
  ggplot(f3c_plot,aes(x=day,y=distance))+
  geom_boxplot(aes(color = severe_case),alpha =0.2,weight = 3)+
  geom_point(aes(fill = severe_case,color = severe_case),position = position_jitterdodge
(0.2),alpha = 0.5,shape = 20,size = 1.6)+

  scale_color_manual(values = c('124','#E69F00'))+
  scale_fill_manual(values = c('124','#E69F00'))+
  stat_compare_means(aes(group = severe_case), method = "wilcox.test",label = "p.signif")
+
  #geom_signif(annotations = c('ns','**','ns'), y_position = c(rep(0.75,3)),xmin = c(0.8,
1.8,2.8),xmax = c(1.2,2.2,3.2))+
  geom_signif(annotations = c('*','***'), y_position = c(0.8,0.85),
              xmin = c(1.2,0.8),xmax = c(2.2,1.8))+
  ylab("jsd distance to day1")+
  xlab("day after admission")+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
f3c
```

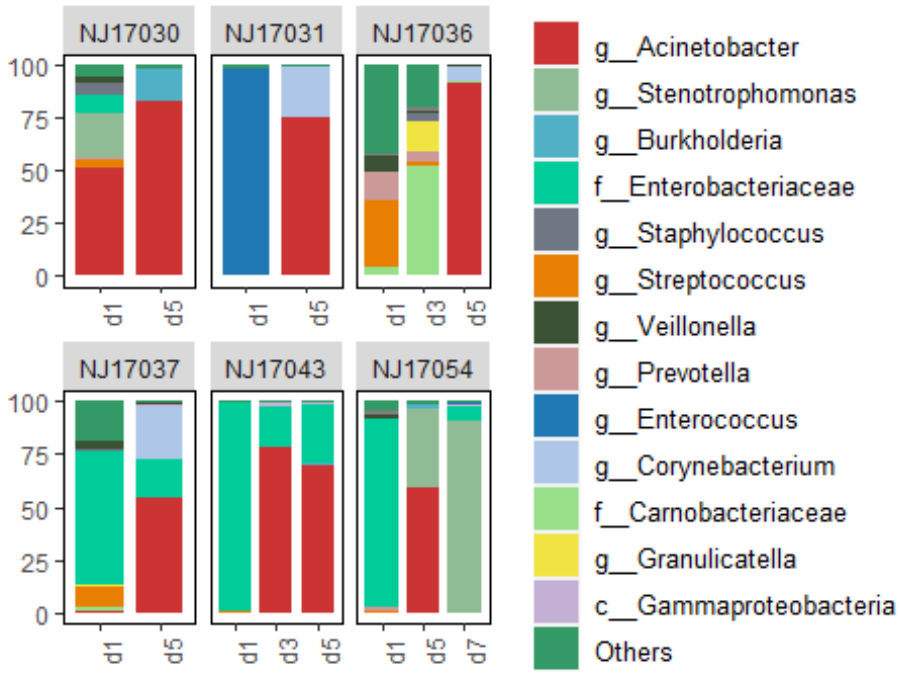


###Figure 3E

```
subj<-c("NJ17036", "NJ17030", "NJ17037", "NJ17043", "NJ17031", "NJ17054")
sub.meta <- filter(metadata,subject %in% subj)
sub.df <- df[,rownames(sub.meta)]


genus<-c()
for (i in subj) {
  taxonomy_table = sub.df
  meta<-filter(sub.meta,respiratory.support.invasive == 'yes')%>%select(raw_id,subject,re
spiratory.support.invasive)
  assign(paste0('meta',i),filter(meta,subject == i))
  rn<-sort(get(paste0('meta',i))$raw_id)
  taxonomy_table<-select(taxonomy_table,one_of(rn)) %>% decostand('total',2)
  taxonomy_table$sum<-rowSums(taxonomy_table)
  taxonomy_table<-taxonomy_table[order(taxonomy_table$sum,decreasing=TRUE),] %>% select(-
sum)
  taxonomy_table<-taxonomy_table[1:5,]
  taxonomy_table<-t(taxonomy_table) %>% as.data.frame()
  taxonomy_table<-arrange(taxonomy_table,desc(taxonomy_table[,1])) %>% t() %>% as.data.fr
ame() %>% select(rn)
  taxonomy_table$taxonomy<-rownames(taxonomy_table)
  assign(paste0('genus',i),taxonomy_table$taxonomy)
  genus<-c(genus,get(paste0('genus',i)))
}

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(rn)` instead of `rn` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```r
genus<-unique(genus)
genus<-genus[-11]



suball<-sort(sub.meta$raw_id)
taxonomy_table<-select(sub.df,one_of(suball)) %>% decostand('total',2)

taxonomy_table$sum<-rowSums(taxonomy_table)
taxonomy_table<-taxonomy_table[order(taxonomy_table$sum,decreasing=TRUE),]
taxonomy_table<-select(taxonomy_table,-sum)

metaall<-sub.meta
suball<-sort(sub.meta$raw_id)

taxonomy_table<-select(sub.df,one_of(suball)) %>% decostand('total',2)
taxonomy_table$sum<-rowSums(taxonomy_table)
taxonomy_table<-taxonomy_table[order(taxonomy_table$sum,decreasing=TRUE),] %>% select(-su
m)
taxonomy_table<-taxonomy_table[genus,]
taxonomy_table['Others', ] <- 1 - colSums(taxonomy_table)
taxonomy_table<-t(taxonomy_table) %>% as.data.frame()
taxonomy_table<-arrange(taxonomy_table,desc(taxonomy_table[,1])) %>% t() %>% as.data.fram
e() %>% select(suball)

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(suball)` instead of `suball` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

taxonomy_table$taxonomy<-rownames(taxonomy_table)
taxonomy_table$taxonomy<- factor(taxonomy_table$taxonomy,levels = rev(taxonomy_table$taxo
nomy))

taxonomy_table1<-melt(taxonomy_table,id.vars = "taxonomy",variable.name = "variable",valu
e.name = "value")


subject<-str_split(taxonomy_table1$variable,'d')%>%as.data.frame()%>%t()%>%as.data.frame()

subject<-c(subject$V1)
taxonomy_table1$subject<-subject

taxonomy_table1$variable<-str_sub(taxonomy_table1$variable,8,9)



f3e<-ggplot(taxonomy_table1,mapping=aes(x=variable,y=value*100,fill=taxonomy))+
  geom_col(position = "stack",width = 0.8)+
  #  labs(x = '', y = 'Relative Abundance(%)',title = "genus abundance") +
  labs(x = '', y = '',title = "") +
  #facet_grid(~subject,scales="free",space= "free" )+
  facet_wrap(~subject,scales="free_x",)+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  theme(legend.text = element_text(size = 10))+
  theme(panel.grid = element_blank(), panel.background = element_rect(color = 'black', fi
```

```
ll = 'transparent')) +
  theme(legend.title = element_blank())+
  scale_fill_manual(values =  c("g__Acinetobacter" = '#CC3333', "g__Stenotrophomonas" = "
#8FBC94", "g__Burkholderia" = "#4FB0C6", "f__Enterobacteriaceae"="#00CC99", "g__Staphyloc
occus"="#6E7783",
                                "g__Streptococcus" = "#e97f02", "g__Veillonella"="#3a5134
", "g__Prevotella"="#CC9999",
                                "g__Enterococcus"='#1f77b4', "g__Corynebacterium" = '#aec
7e8',
                                "f__Carnobacteriaceae" = '#98df8a', "g__Granulicatella" =
 "#F0E442",
                                "c__Gammaproteobacteria" = '#c5b0d5',"Others"="#339966"))

#scale_fill_manual(values =  rev(c('#ffbb78',"#339966",'#CC3333',"#8FBC94","#4FB0C6","#00
CC99","#6E7783","#e97f02","#3a5134","#99CCCC","#CC9999",'#1f77b4','#aec7e8','#ff7f0e','#2
ca02c','#98df8a','#d62728','#ff9896','#9467bd','#c5b0d5','#8c564b',"#999999","#0099CC")))
#scale_fill_manual(values =  rev(color))

f3e
```



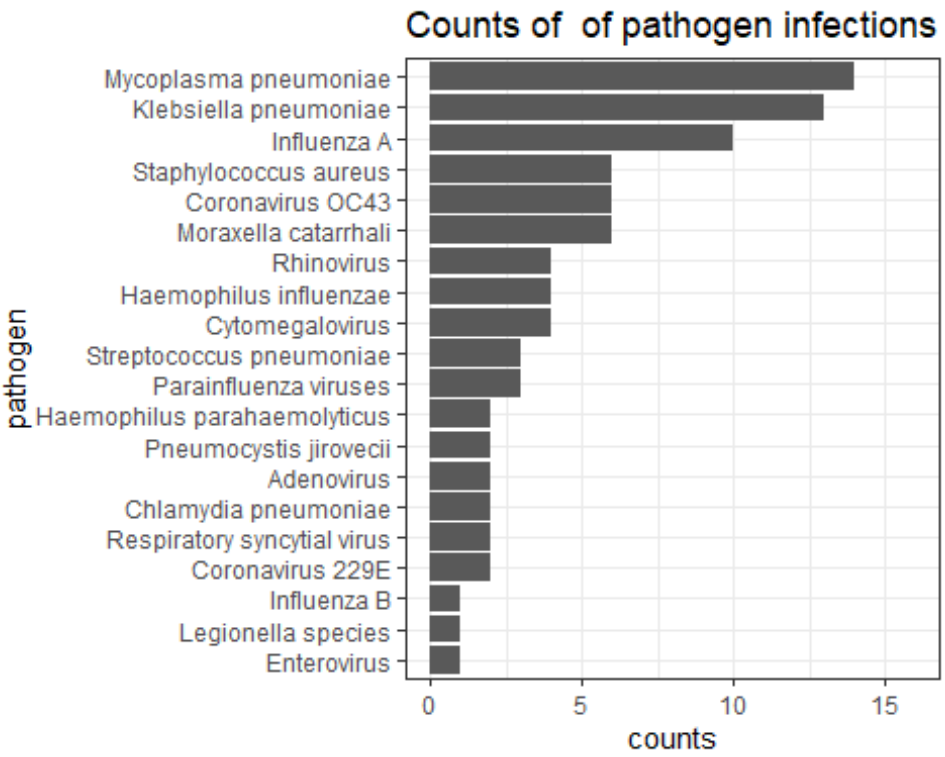
##Figure 4 ###pathogen data import

```
pathogens <- c('Mycoplasma.pneumoniae','Klebsiella.pneumoniae','Influenza.A')
sub.meta <- filter(metadata, dfirst == 'dfirst') %>% filter(type %in% pathogens)
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()


demy_df<-sub.df[rownames(filter(sub.meta,type == 'Mycoplasma.pneumoniae')),]
dekle_df<-sub.df[rownames(filter(sub.meta,type == 'Klebsiella.pneumoniae')),]
demy_df[,'g__Mycoplasma'] = 0
dekle_df[,'f__Enterobacteriaceae'] = 0
other_df <- sub.df[rownames(filter(sub.meta,type != 'Mycoplasma.pneumoniae' & type != 'Kl
```

```
ebsiella.pneumoniae')),]

df_deself <- rbind(demy_df,dekle_df) %>% rbind(other_df) %>% decostand('total',MARGIN = 1)


metahea <- filter(metadata,severe_case == 'healthy')
dfhea <- df[,rownames(metahea)] %>% t() %>% as.data.frame()
df_deself_hea <- rbind(dfhea,df_deself)
meta_deself_hea <- metadata[rownames(df_deself_hea),] %>% select(raw_id,city,severe_case,
Shannon,Mycoplasma.pneumoniae,Klebsiella.pneumoniae,Influenza.A)
meta_deself_hea <- left_join(meta_deself_hea,select(sub.meta,raw_id,type))

## Joining, by = "raw_id"

meta_deself_hea[is.na(meta_deself_hea[,'type']),'type'] <- 'healthy'
rownames(meta_deself_hea) <- meta_deself_hea$raw_id


# Bacteria Virus Mix
sub.meta <- filter(metadata, dfirst == 'dfirst')
sub.meta <- sub.meta[sub.meta[,'pathogen_type'] != 'na',]
sub.df <- df[,rownames(sub.meta)] %>% t() %>% as.data.frame()


metahea <- filter(metadata,severe_case == 'healthy')
dfhea <- df[,rownames(metahea)] %>% t() %>% as.data.frame()
df_pathogen_hea <- rbind(dfhea,sub.df)
meta_pathogen_hea <- metadata[rownames(df_pathogen_hea),] %>% select(raw_id,city,severe_c
ase,Shannon)
meta_pathogen_hea <- left_join(meta_pathogen_hea,select(sub.meta,raw_id,pathogen_type))

## Joining, by = "raw_id"

meta_pathogen_hea[is.na(meta_pathogen_hea[,'pathogen_type']),'pathogen_type'] <- 'healthy
'
rownames(meta_pathogen_hea) <- meta_pathogen_hea$raw_id
```

###Figure 4A

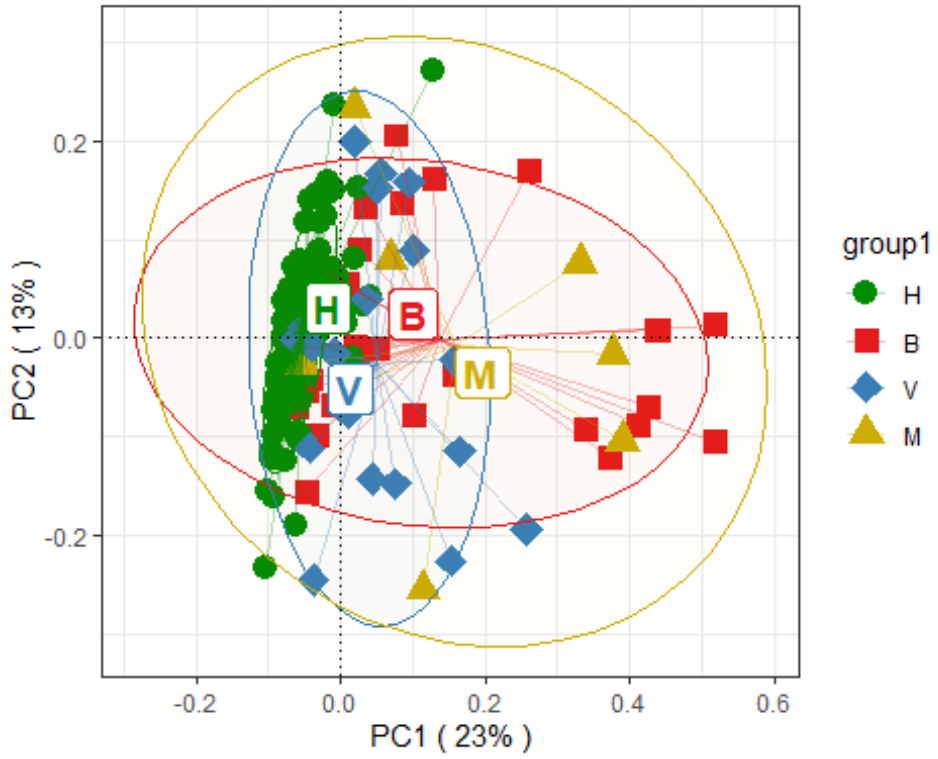
```
f4a_plot <- read.csv("F:/ZLF/CAP/downstream/relative_analysis/FTD_counts0621.csv")

f4a_plot <- f4a_plot[order(f4a_plot[,'counts']),]

f4a_plot$pathogen<-factor(f4a_plot$pathogen,levels = f4a_plot$pathogen)

f4a<-ggplot(data=f4a_plot,aes(x=counts,y=pathogen))+
  geom_bar(stat = "identity")+
  ggtitle("Counts of  of pathogen infections")+
  xlim(0,16)+
  theme_bw()
f4a
```



###Figure 4B

```
d1.meta = meta_pathogen_hea
d1.df = df_pathogen_hea
val = 'pathogen_type'
sub.meta = d1.meta[!is.na(d1.meta[,val]),]
sub.df = d1.df[rownames(sub.meta),]
OTU = phyloseq::otu_table(sub.df, taxa_are_rows = F)
physeq = phyloseq(OTU)
jsd=phyloseq::distance(physeq, method = "jsd")

jsd.cord = cmdscale(jsd,k=2,eig = T)


PC1 = jsd.cord$points[,1]
PC2 = jsd.cord$points[,2]
plotdata <- data.frame(rownames(jsd.cord$points),PC1,PC2,sub.meta[,val])
colnames(plotdata) <-c("sample","PC1","PC2","group")

plotdata1<-merge(plotdata,aggregate(cbind(mean.x=PC1,mean.y=PC2)~group,plotdata,mean),by=
"group")

cbbPalette <- c("green4","#E41A1C","#377EB8","gold3")
cbbPalette1 <- c("green4","#E41A1C","gold3","#377EB8")

Palette <- c("#000000", "#000000","#000000","#000000","#000000","#000000")

eigen.vals.jsd = jsd.cord$eig
last_one = sum(eigen.vals.jsd>0)
pc1 <-floor(eigen.vals.jsd[1]*100/sum(eigen.vals.jsd[1:last_one]))
pc2 <-floor(eigen.vals.jsd[2]*100/sum(eigen.vals.jsd[1:last_one]))
```

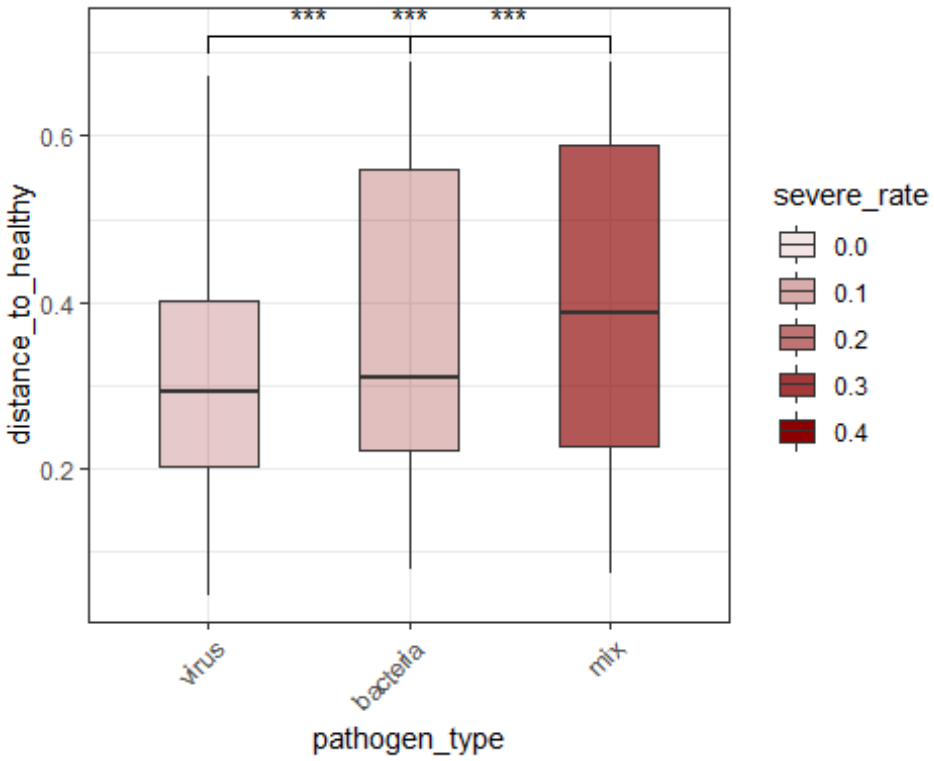
```
pich=rep(c(21:24),3)

core <- unique(select(plotdata1,mean.x,mean.y,group))
core$group1 <- c('B','H','M','V')

plotdata1$group1 <- NA
for (i in 1:nrow(plotdata1)) {
  plotdata1[i,'group1'] = str_sub(plotdata1[i,'group'],1,1) %>% toupper()
}
plotdata1$group1<-factor(plotdata1$group1,levels = c('H','B','V','M'))

f4b_plot <- plotdata1

p = ggplot(f4b_plot, aes(PC1, PC2)) +
  geom_point(aes(color=group1,shape=group1,fill=group1),size=4)+
  geom_segment(aes(x=mean.x,y=mean.y,xend=PC1, yend=PC2,color = group1),alpha=0.3)+
  stat_ellipse(aes(fill = group1,color = group1),show.legend = F, geom="polygon",alpha=0.
02,level = 0.90,type = "t",position = "identity")+
  scale_shape_manual(values=pich)+
  scale_colour_manual(values=Palette)+
  scale_fill_manual(values=cbbPalette)+
  xlab(paste("PC1 ( ",pc1,"%"," )",sep="")) +
  ylab(paste("PC2 ( ",pc2,"%"," )",sep=""))+
  theme(text=element_text(size=15))+
  geom_vline(aes(xintercept = 0),linetype="dotted")+
  geom_hline(aes(yintercept = 0),linetype="dotted")+
  scale_color_manual(values =cbbPalette) +
  theme_bw()+
  ggrepel::geom_label_repel(data=core,
                            aes(mean.x,mean.y,color=group1),
                            label=c(unique(plotdata1$group1)),
                            fontface="bold",show.legend = F,box.padding = 0,size=5
  )

## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.

p
```



###Figure 4C

```
val = 'pathogen_type'

sub.meta = meta_pathogen_hea[!is.na(meta_pathogen_hea[,val]),]
sub.df = df_pathogen_hea
#sub.df = df1[rownames(sub.meta),]
#View(sub.df)
OTU = phyloseq::otu_table(sub.df, taxa_are_rows = F)
physeq = phyloseq(OTU)
#jsd=phyloseq::distance(physeq, method = "jsd")
jsd=phyloseq::distance(physeq, method = "jsd")
jsd1 = as.matrix(jsd)
#jsd1[row(jsd1)>=col(jsd1)] <- NA
jsd1 = melt(as.matrix(jsd1))%>% filter(as.character(Var1) != as.character(Var2))

metavar1 <- select(sub.meta,raw_id,val)

## Note: Using an external vector in selections is ambiguous.
## i Use `all_of(val)` instead of `val` to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.

metavar2 <- select(sub.meta,raw_id,val)
colnames(metavar1)<-c('Var1','group1')
colnames(metavar2)<-c('Var2','group2')

jsd2 <- left_join(jsd1,metavar1)

## Joining, by = "Var1"

jsd2 <- left_join(jsd2,metavar2)
```

```
## Joining, by = "Var2"

hb <- filter(jsd2,group1 == 'healthy'&group2 == 'bacteria')
hv <- filter(jsd2,group1 == 'healthy'&group2 == 'virus')
hm <- filter(jsd2,group1 == 'healthy'&group2 == 'mix')
hbvm<-rbind(hb,hv) %>% rbind(hm)

colnames(hbvm)[3]<-'distance_to_healthy'
colnames(hbvm)[5]<-'pathogen_type'
hbvm$pathogen_type<-factor(hbvm$pathogen_type,levels = c('virus','bacteria','mix'))
severe_rate <- data.frame(pathogen_type = c('bacteria','virus','mix'),severe_rate = c(0.0
67,0.05,0.25))
hbvm1<-left_join(hbvm,severe_rate)

## Joining, by = "pathogen_type"

hbvm1$pathogen_type<-factor(hbvm1$pathogen_type,levels = c('virus','bacteria','mix'))

f4c_plot <- hbvm1

compaired = combn(c('virus','bacteria','mix'),2) %>% t()
comp = c()
for (i in c(1:3)) {
  mid = list(compaired[i,])
  comp = c(comp,mid)
}

f4c<-
  ggplot(f4c_plot,aes(x=pathogen_type,y=distance_to_healthy))+
  geom_boxplot(aes(alpha = severe_rate),fill = 'darkred',width = 0.5,position = position_
identity(),outlier.size = 0.5)+
  scale_alpha_continuous(limits = c(0,0.4))+
  geom_signif(comparisons = comp,map_signif_level=T,test = "wilcox.test")+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

f4c
```



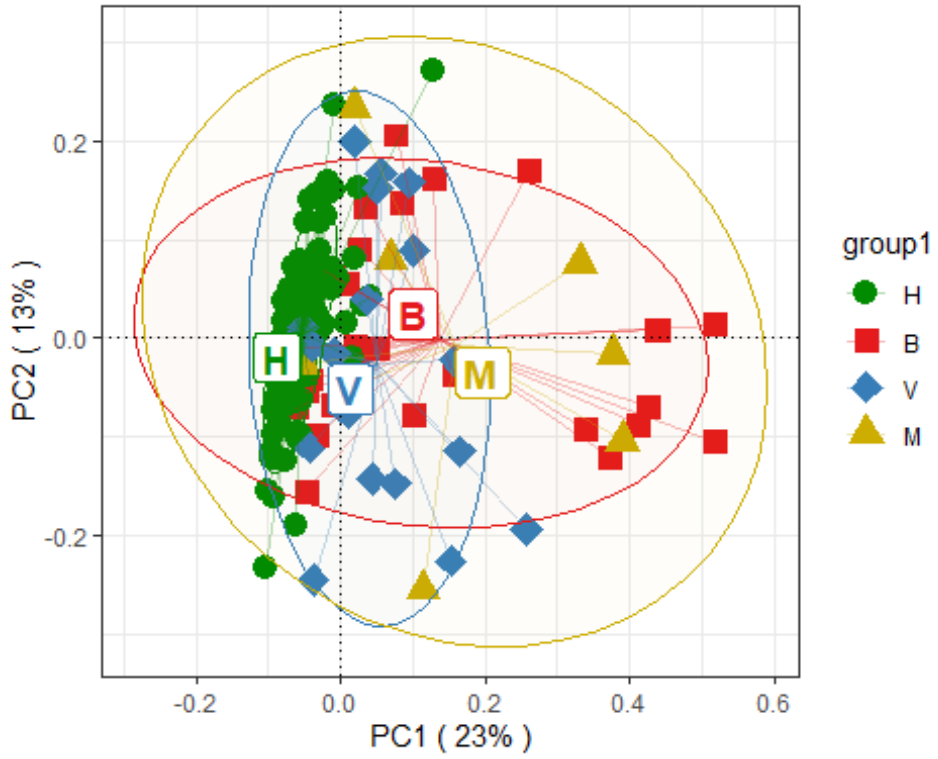
###Figure 4D

```
d1.meta = meta_pathogen_hea
d1.df = df_pathogen_hea
val = 'pathogen_type'
sub.meta = d1.meta[!is.na(d1.meta[,val]),]
sub.df = d1.df[rownames(sub.meta),]
OTU = phyloseq::otu_table(sub.df, taxa_are_rows = F)
physeq = phyloseq(OTU)
jsd=phyloseq::distance(physeq, method = "jsd")
#jsd=phyloseq::distance(physeq, method = "bray")
jsd.cord = cmdscale(jsd,k=2,eig = T)

PC1 = jsd.cord$points[,1]
PC2 = jsd.cord$points[,2]
plotdata <- data.frame(rownames(jsd.cord$points),PC1,PC2,sub.meta[,val])
colnames(plotdata) <-c("sample","PC1","PC2","group")

plotdata1<-merge(plotdata,aggregate(cbind(mean.x=PC1,mean.y=PC2)~group,plotdata,mean),by=
"group")

cbbPalette <- c( "#56B4E9","#E69F00", "#009E73", "purple1","red","sienna")
cbbPalette <- c("#E41A1C","#377EB8","#4DAF4A","#984EA3","#FF7F00","#A65628")
cbbPalette <- c("green4","#E41A1C","#377EB8","gold3")
cbbPalette1 <- c("green4","#E41A1C","gold3","#377EB8")

Palette <- c("#000000", "#000000","#000000","#000000","#000000","#000000")

eigen.vals.jsd = jsd.cord$eig
last_one = sum(eigen.vals.jsd>0)
pc1 <-floor(eigen.vals.jsd[1]*100/sum(eigen.vals.jsd[1:last_one]))
pc2 <-floor(eigen.vals.jsd[2]*100/sum(eigen.vals.jsd[1:last_one]))
```

```r
pich=rep(c(21:24),3)

core <- unique(select(plotdata1,mean.x,mean.y,group))
core$group1 <- c('B','H','M','V')
#core$group1<-factor(core,levels = c('I','M','K'))
plotdata1$group1 <- NA
for (i in 1:nrow(plotdata1)) {
  plotdata1[i,'group1'] = str_sub(plotdata1[i,'group'],1,1) %>% toupper()
}
plotdata1$group1<-factor(plotdata1$group1,levels = c('H','B','V','M'))

f4d_plot <- plotdata1

f4d = ggplot(f4d_plot, aes(PC1, PC2)) +
  geom_point(aes(color=group1,shape=group1,fill=group1),size=4)+
  geom_segment(aes(x=mean.x,y=mean.y,xend=PC1, yend=PC2,color = group1),alpha=0.3)+
  stat_ellipse(aes(fill = group1,color = group1),show.legend = F, geom="polygon",alpha=0.
02,level = 0.90,type = "t",position = "identity")+
  scale_shape_manual(values=pich)+
  scale_colour_manual(values=Palette)+
  scale_fill_manual(values=cbbPalette)+

  xlab(paste("PC1 ( ",pc1,"%"," )",sep="")) +
  ylab(paste("PC2 ( ",pc2,"%"," )",sep=""))+
  theme(text=element_text(size=15))+
  geom_vline(aes(xintercept = 0),linetype="dotted")+
  geom_hline(aes(yintercept = 0),linetype="dotted")+
  #geom_text_repel(data = species_lefse, aes(label = name,color = group),size = 5.5)+
  scale_color_manual(values =cbbPalette) +
  theme_bw()+
  ggrepel::geom_label_repel(data=core,
                            aes(mean.x,mean.y,color=group1),
                            label=c(unique(plotdata1$group1)),
                            fontface="bold",show.legend = F,box.padding = 0,size=5
  )
```

```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```

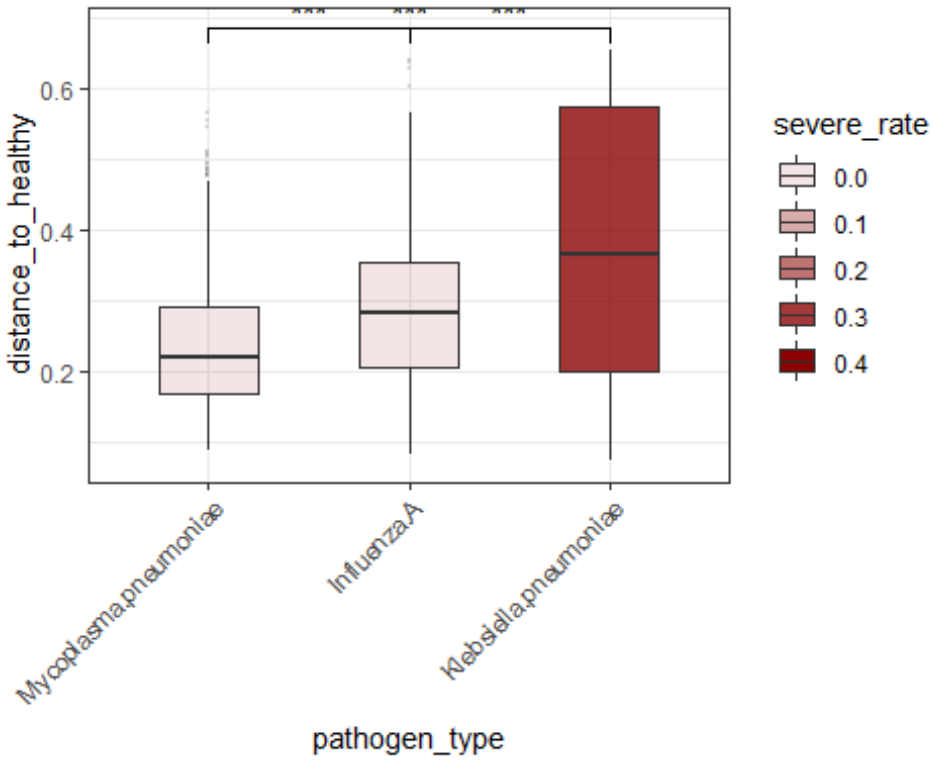
```r
f4d
```



###Figure 4E

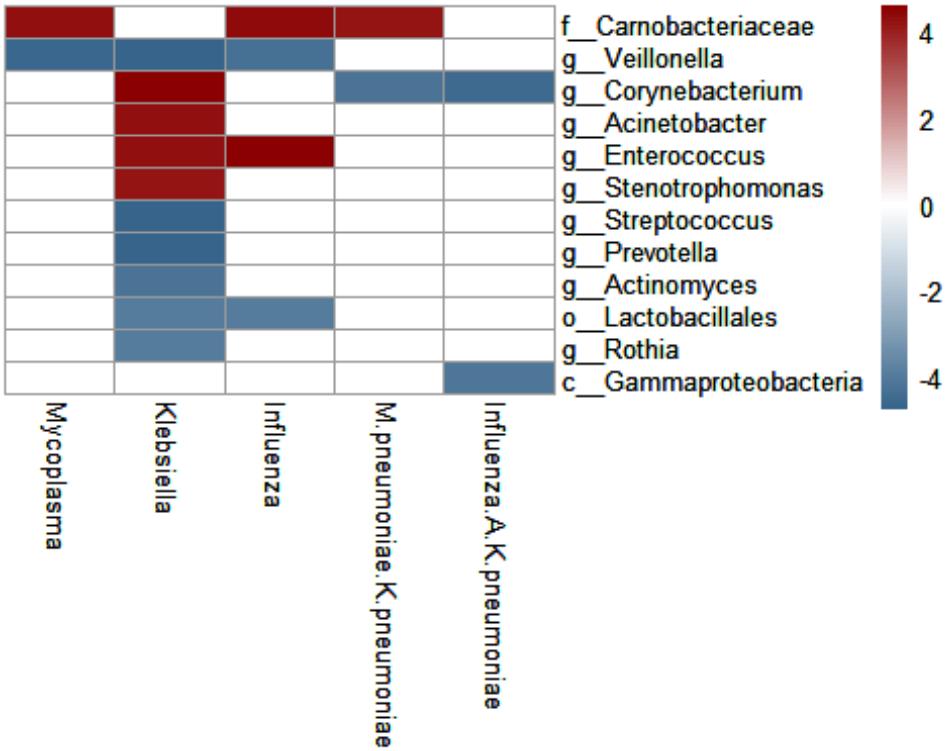
```
val = 'type'
sub.meta = meta_deself_hea[!is.na(meta_deself_hea[,val]),]
sub.df = df_deself_hea
#sub.df = df1[rownames(sub.meta),]
#View(sub.df)
OTU = phyloseq::otu_table(sub.df, taxa_are_rows = F)
physeq = phyloseq(OTU)
#jsd=phyloseq::distance(physeq, method = "jsd")
jsd=phyloseq::distance(physeq, method = "jsd")
jsd1 = as.matrix(jsd)
#jsd1[row(jsd1)>=col(jsd1)] <- NA
jsd1 = melt(as.matrix(jsd1))%>% filter(as.character(Var1) != as.character(Var2))

metavar1 <- select(sub.meta,raw_id,val)
metavar2 <- select(sub.meta,raw_id,val)
colnames(metavar1)<-c('Var1','group1')
colnames(metavar2)<-c('Var2','group2')

jsd2 <- left_join(jsd1,metavar1)

## Joining, by = "Var1"

jsd2 <- left_join(jsd2,metavar2)

## Joining, by = "Var2"

hm <- filter(jsd2,group1 == 'healthy'&group2 == 'Mycoplasma.pneumoniae')
hk <- filter(jsd2,group1 == 'healthy'&group2 == 'Klebsiella.pneumoniae')
hi <- filter(jsd2,group1 == 'healthy'&group2 == 'Influenza.A')
hmki<-rbind(hm,hk) %>% rbind(hi)
```

```r
colnames(hmki)[3]<-'distance_to_healthy'
colnames(hmki)[5]<-'pathogen_type'
hmki$pathogen_type<-factor(hmki$pathogen_type,levels = c("Mycoplasma.pneumoniae", "Klebsiella.pneumoniae", "Influenza.A"))
#hmki$pathogen_type<-factor(hmki$pathogen_type,levels = c("Mycoplasma.pneumoniae","Influenza.A", "Klebsiella.pneumoniae" ))

severe_rate1 <- data.frame(pathogen_type = pathogens,severe_rate = c(0,0.3077,0))

hmki1 <- left_join(hmki,severe_rate1)

## Joining, by = "pathogen_type"

hmki1$pathogen_type<-factor(hmki1$pathogen_type,levels = c("Mycoplasma.pneumoniae","Influenza.A", "Klebsiella.pneumoniae"))

f4e_plot <- hmki1

compaired = combn(c("Mycoplasma.pneumoniae", "Klebsiella.pneumoniae", "Influenza.A"),2)
 %>% t()
comp = c()
for (i in c(1:3)) {
  mid = list(compaired[i,])
  comp = c(comp,mid)
}



f4e<-
  ggplot(f4e_plot,aes(x=pathogen_type,y=distance_to_healthy))+
  #geom_violin(aes(color = cluster,fill = cluster),alpha=0.7)+
  #geom_boxplot(aes(fill = pathogen_type),width = 0.5,position = position_identity(),alpha=0.7)+
  geom_boxplot(aes(alpha = severe_rate),fill = 'darkred',width = 0.5,position = position_identity(),outlier.size = 0.5)+
  scale_alpha_continuous(limits = c(0,0.4))+
  geom_signif(comparisons = comp,map_signif_level=T,test = "wilcox.test")+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

f4e
```



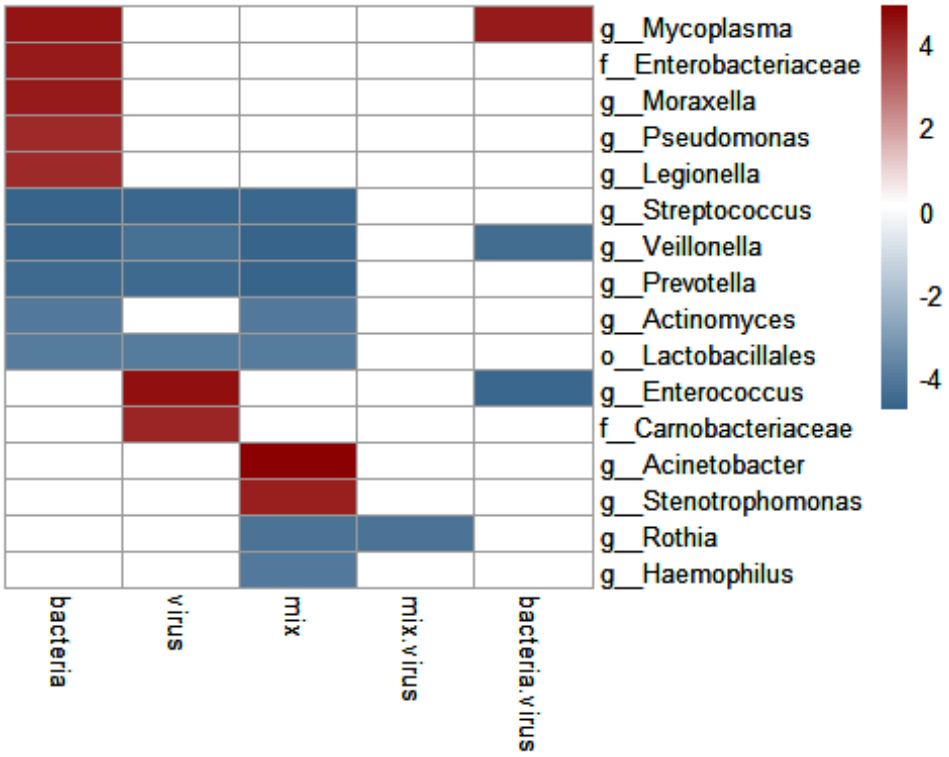
###Figure 4F

```
library(pheatmap)
kmi_lefse <- read.csv('F:/ZLF/CAP/paper_structure/ftd/kmi_lefse_0618.csv',row.names = 1)

f4f_plot <- kmi_lefse

bk = c(seq(-4.8,4.6,by=0.094))
f4f<-pheatmap(f4f_plot ,
              cluster_row = FALSE,
              cluster_cols = FALSE,
              #display_numbers = TRUE,
              na_col = "white",
              #color = colorRampPalette(c("steelblue4",'white',"darkred"))(100)
              color = c(colorRampPalette(colors = c( "steelblue4","white"))(floor(length(bk)
*(4.8/9.4))),colorRampPalette(colors = c( "white","darkred"))(ceiling(length(bk)*(4.6/9.
4)))),

              #color = colorRampPalette(c("steelblue4",'white',"darkred"))(100)
              #file="F:/ZLF/CAP/paper_structure/figure1/csjsddistance4.pdf",
)

f4f
```



###Figure 4G

```
bmv_lefse <- read.csv('F:/ZLF/CAP/paper_structure/ftd/bmv_lefse_0618.csv',row.names = 1)

f4g_plot <- bmv_lefse

bk = c(seq(-4.8,4.6,by=0.094))
f4g<-pheatmap(f4g_plot,
              cluster_row = FALSE,
              cluster_cols = FALSE,
              #display_numbers = TRUE,
              na_col = "white",
              #color = colorRampPalette(c("steelblue4",'white',"darkred"))(100)
              color = c(colorRampPalette(colors = c( "steelblue4","white"))(floor(length(bk)
*(4.8/9.4))),colorRampPalette(colors = c( "white","darkred"))(ceiling(length(bk)*(4.6/9.
4)))),

              #color = colorRampPalette(c("steelblue4",'white',"darkred"))(100)
              #file="F:/ZLF/CAP/paper_structure/figure1/csjsddistance4.pdf",
)

f4g
```



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.