Procrustes analysis

library(readxl)

library(vegan)

library(ggplot2)

library(ade4)

microbiome <- read\_excel("Microbiome.xlsx", sheet = 1)

metabolites <- read\_excel("Metabolites.xlsx", sheet = 1)

microbiome <- as.data.frame(microbiome)

metabolites <- as.data.frame(metabolites)

rownames(microbiome) <- microbiome[,1]

microbiome <- microbiome[,-1]

rownames(metabolites) <- metabolites[,1]

metabolites <- metabolites[,-1]

common\_samples <- intersect(rownames(microbiome), rownames(metabolites))

microbiome <- microbiome[common\_samples, ]

metabolites <- metabolites[common\_samples, ]

micro\_hell <- decostand(microbiome, "hellinger")

meta\_z <- scale(metabolites)

micro\_dist <- vegdist(micro\_hell, method = "bray")

micro\_pcoa <- cmdscale(micro\_dist, k = nrow(micro\_hell)-1, eig = TRUE)

meta\_dist <- dist(meta\_z, method = "euclidean")

meta\_pcoa <- cmdscale(meta\_dist, k = nrow(meta\_z)-1, eig = TRUE)

keep\_axes <- function(pcoa, threshold = 0.8) {

eig <- pcoa$eig

cum\_prop <- cumsum(eig[eig>0]) / sum(eig[eig>0])

k <- sum(cum\_prop < threshold) + 1

return(pcoa$points[, 1:k])

}

micro\_coord <- keep\_axes(micro\_pcoa)

meta\_coord <- keep\_axes(meta\_pcoa)

proc\_result <- protest(

X = micro\_coord,

Y = meta\_coord,

permutations = 9999

)

proc\_fit <- proc\_result$Yrot

proc\_ref <- proc\_result$X

plot\_df <- data.frame(

Sample = rownames(proc\_ref),

Micro\_X = proc\_ref[,1],

Micro\_Y = proc\_ref[,2],

Meta\_X = proc\_fit[,1],

Meta\_Y = proc\_fit[,2]

)

ggplot(plot\_df) +

geom\_segment(aes(x = Micro\_X, y = Micro\_Y,

xend = Meta\_X, yend = Meta\_Y),

arrow = arrow(length = unit(0.2, "cm")),

color = "gray70") +

geom\_point(aes(Micro\_X, Micro\_Y), color = "#1B9E77", size = 3, alpha = 0.8) +

geom\_point(aes(Meta\_X, Meta\_Y), color = "#D95F02", size = 3, alpha = 0.8) +

geom\_text(aes(x = (Micro\_X+Meta\_X)/2, y = (Micro\_Y+Meta\_Y)/2,

label = Sample), size = 3, vjust = 1.5) +

labs(title = paste("Procrustes Analysis (M^2 =", round(proc\_result$ss, 4),

"R =", round(proc\_result$t0, 4)),

subtitle = paste("p-value =", proc\_result$signif),

x = "Dimension 1",

y = "Dimension 2") +

scale\_color\_manual(values = c("Microbiome" = "#1B9E77", "Metabolites" = "#D95F02")) +

theme\_minimal() +

theme(legend.position = "top")

capture.output(print(proc\_result), file = "Procrustes\_Results.txt")

ggsave("Procrustes\_Plot.png", width = 10, height = 8, dpi = 300)

MIMOSA2

if (!require("BiocManager", quietly = TRUE))

install.packages("BiocManager")

BiocManager::install("mimosa")

library(mimosa)

library(data.table)

species\_abundances <- fread("Species\_composition.csv")

metabolite\_data <- fread("Metabolite.csv")

config <- config\_mimosa2(

database = "EMBL",

method = "mimosa2",

kmin = 0.1

)

model <- build\_metabolic\_model(

config = config,

data = species\_abundances

)

results <- mimosa(

metabolomics = metabolite\_data,

metabolomic\_format = "table",

config = config,

model = model

)

plot\_mimosa2(results)

SEM analysis

library(nlme)

library(lme4)

library(piecewiseSEM)

library(QuantPsyc)

setwd('E:/360MoveData/Users/dell/Desktop/test\_R')

Rt <- read.csv("20250417a\_R.csv", header = TRUE)

result=aov(N\_foliar~group,data=Rt )

summary(result)

model1 <- lm(N\_foliar~ Shannon+ Sphe3, Rt)

coefs(model1, standardize ='scale')

sp\_Shannon <- summary(model1)$coefficients[2,1]

sp\_Sphe3 <- summary(model1)$coefficients[3,1]

sp <- sp\_Shannon \* Rt$Shannon+ sp\_Sphe3 \* Rt$Sphe3

Rt$sp <- sp

summary(lm(N\_foliar ~ sp,Rt))

coefs(lm(N\_foliar ~ sp,Rt))

vif(lm(N\_foliar ~ Shannon+ Sphe3,Rt))

model2 <- lm(N\_foliar ~ K10944+K26139+K03385+K02591, Rt)

coefs(model2, standardize ='scale')

sn\_K10944 <- summary(model2)$coefficients[2,1]

sn\_K26139 <- summary(model2)$coefficients[3,1]

sn\_K03385 <- summary(model2)$coefficients[4,1]

sn\_K02591 <- summary(model2)$coefficients[5,1]

sn <- sn\_K10944\* Rt$K10944+ sn\_K26139 \* Rt$K26139+ sn\_K03385 \* Rt$K03385+sn\_K02591 \* Rt$K02591

Rt$sn <- sn

summary(lm(N\_foliar ~ sn,Rt))

coefs(lm(N\_foliar ~ sn,Rt))

vif(lm(N\_foliar ~ K10944+K26139+K03385+K02591,Rt)) # 应 < 5

model3 <- lm(N\_foliar ~ Uridine+Ticrynafen+Dehydrocostus+Undecalactone, Rt)

coefs(model3, standardize ='scale')

Alpha\_Uridine <- summary(model3)$coefficients[2,1]

Alpha\_Ticrynafen <- summary(model3)$coefficients[3,1]

Alpha\_Dehydrocostus <- summary(model3)$coefficients[4,1]

Alpha\_Undecalactone <- summary(model3)$coefficients[5,1]

Alpha<- Alpha\_Uridine \* Rt$Uridine+ Alpha\_Ticrynafen \* Rt$Ticrynafen+Alpha\_Dehydrocostus \* Rt$Dehydrocostus+Alpha\_Undecalactone \* Rt$Undecalactone

Rt$Alpha <- Alpha

summary(lm(N\_foliar ~ Alpha,Rt))

coefs(lm(N\_foliar ~ Alpha,Rt))

vif(lm(N\_foliar ~ Uridine+Ticrynafen+Dehydrocostus+Undecalactone,Rt)) # 应 < 5

model4<- lm(N\_foliar ~ TN+NO3+NH4, Rt)

coefs(model4, standardize ='scale')

se\_TN <- summary(model4)$coefficients[2,1]

se\_NO3 <- summary(model4)$coefficients[3,1]

se\_NH4 <- summary(model4)$coefficients[4,1]

se <- se\_TN\* Rt$TN+se\_NO3\* Rt$NO3+se\_NH4\* Rt$NH4

Rt$se <- se

summary(lm(N\_foliar ~ se,Rt))

coefs(lm(N\_foliar ~ se,Rt))

vif(lm(N\_foliar ~ TN+NH4+NO3, Rt))

final\_model <- psem(

lm(N\_foliar ~se + Alpha+sn+group, data = Rt),

lm(Alpha ~ sp + sn, data = Rt),

lm(se ~ sp+group, data = Rt),

lm(sn ~group,data = Rt),

lm(sp ~ group, data = Rt)

)

summary(final\_model, .progressBar = FALSE)

plot(final\_model)

pca\_soilN <- prcomp(Rt[, c("TN", "NO3")], scale = TRUE)

Rt$SoilN\_PC1 <- pca\_soilN$x[, 1]

model\_optimized <- psem(

lm(N\_foliar ~ se + sn, data = Rt),

lm(Alpha ~ sp , data = Rt),

lm(se ~ sp + group, data = Rt),

lm(sn ~ group+Alpha, data = Rt),

lm(sp ~ group, data = Rt)

)

summary(model\_optimized, .progressBar = FALSE)

plot(model\_optimized)

Phylogenetic tree construction

setwd('E:/360MoveData/Users/dell/Desktop/test\_R')

sequences <- readDNAStringSet("sequence.fasta")

aligned\_sequences <- AlignSeqs(sequences)

phy\_data <- phyDat(as.matrix(aligned\_sequences), type = "DNA")

dist\_matrix <- dist.ml(phy\_data)

nj\_tree <- NJ(dist\_matrix)

write.tree(nj\_tree, file = "bacteria\_tree.nwk")

tree <- read.tree("bacteria\_tree.nwk")

library(dplyr)

Source <- readxl::read\_excel("20250507\_R.xlsx", sheet ="Source")

THXXSY <- readxl::read\_excel("20250507\_R.xlsx", sheet ="THXXSY")

Phylum <- readxl::read\_excel("20250507\_R.xlsx", sheet ="Phylum")

Genus <- readxl::read\_excel("20250507\_R.xlsx", sheet ="Genus")

Phylum <- Phylum %>% dplyr::rename(label = Strain)

Genus<- Genus %>% dplyr::rename(label = Strain)

phylum\_list <-split(Phylum$label, Phylum$phylum)

p <- ggtree(nj\_tree, layout ='fan', open.angle =30)

p1 <- groupOTU(p, phylum\_list,'phylum') + aes(color=Phylum) + guides(colour ="none")

phylum\_list <- split(Phylum$label, as.character(Phylum$phylum))

missing\_tips <- setdiff(unlist(phylum\_list), nj\_tree$tip.label)

if (length(missing\_tips) > 0) {

warning("以下 tip 在树中不存在: ", paste(missing\_tips, collapse = ", "))

}

library(ggtree)

library(dplyr)

Phylum <- readxl::read\_excel("20250507\_R.xlsx", sheet = "Phylum") %>%

dplyr::rename(label = Strain)

phylum\_list <- split(Phylum$label, Phylum$phylum)

grouped\_tree <- groupOTU(nj\_tree, phylum\_list, group\_name = "Phylum")

p1 <- ggtree(grouped\_tree, layout = 'fan', open.angle = 30) +

aes(color = Phylum) +

guides(color = "none")

print(p1)

p2 <- rotate\_tree(p1, angle =150) + geom\_aline(aes(color = Phylum), linetype ='longdash', linewidth =0.8, size =1, show.legend = TRUE) + ggplot2::scale\_color\_manual(values= c("Firmicutes"="#C6DCB9","Proteobacteria"="#B6E2DC"), na.value ='red') +guides(colour ="none") + new\_scale\_fill() + new\_scale\_color()

p2

phylum\_1 <- Phylum %>% dplyr::rename(phylum\_1 = phylum)

p3 <- p2 + geom\_fruit(data = phylum\_1, geom = geom\_bar,stat='identity',width =0.95,aes(y= label,x=10, fill= phylum\_1), pwidth =0.5 ) + geom\_tiplab(align = TRUE, offset = 0, linetype ="blank", size = 5) +scale\_fill\_manual(values= c("Firmicutes"="#C6DCB9","Proteobacteria"="#B6E2DC")) + guides(fill ="none") + new\_scale\_fill() + new\_scale\_color()

p3

p4 <- p3 + geom\_fruit( data = THXXSY,geom = geom\_bar,stat='identity', width =0.1,aes(y= Strain,x= THXXSY),fill ="#D2BCDE", pwidth =0.5, offset =0.08,axis.params =list(axis ="x",text.size =1.5, title ="THXXSY",line.size =0,line.color ="white", nbreak =3), grid.params = list(linetype =2,size =0.5)) +new\_scale\_fill() +new\_scale\_color()

p4

p5 <- p4 +geom\_fruit(data = Genus,geom = geom\_bar,stat='identity',offset = -0.2, width =1,aes(y= label,x=0.1, fill = genus),pwidth =0.3 ) +scale\_fill\_manual(values= c("Bacillus"="#8FDBF3","Lysinibacillus"="#98F4E0","Priestia"="#EEC2E5","Pseudomonas"="#FFCFD1" )) + guides(fill = guide\_legend(title ="Genus")) + new\_scale\_fill() + new\_scale\_color()

p5

dat = p5$data

dat <- dat[grep("^Strain", dat$label), ]

dat$Phylum <- as.character(dat$Phylum)

names(dat)

Phylum\_2 <- dat %>%

group\_by(Phylum) %>% arrange(y, .labelby\_group = TRUE) %>%

mutate(n = n(),

rank = row\_number(),

phylum = ifelse(n %%2==1& rank == ceiling(n/2) | (n %%2==0& rank == n/2),Phylum, "") ) %>% ungroup() %>%

select(-n, -rank)

Phylum\_2 <- na.omit(Phylum\_2)

Strain <- Phylum\_2[Phylum\_2$phylum !='', ]$label

Phylum\_3 <- Phylum\_2 %>%select(parent, label, phylum, angle, node)

colnames(Phylum\_3) <- c('parent','label','phylum\_1','angle\_degree','node')

unique(Phylum\_3$phylum\_1)

tmp <- Phylum\_3 %>% filter(phylum\_1 %in% c('Firmicutes','Proteobacteria'))

ALBERT <- tmp$label

Phylum\_3[!Phylum\_3$label %in% ALBERT, ]$phylum\_1 =''

Phylum\_3 <- Phylum\_3[Phylum\_3$label %in% ALBERT, ]

custom\_ranges <- list(180:270)

Phylum\_3 <- visbuilder::calculate\_rotate\_angle(df = Phylum\_3, angle\_col ="angle\_degree", range\_list = custom\_ranges)

Phylum\_3 <- Phylum\_3 %>% arrange(label)

P6 <- p5 + geom\_fruit( data = Phylum\_3, geom = geom\_text, mapping = aes(y= label,x=0, label = phylum\_1), size =4.5, angle = Phylum\_3$alb\_angle\_degree\_ert, color ="black", offset =0.22)

P6