# **R补充作图**

## ****导出特征表：****

导出FeatureTable[Frequency]对象为BIOM v2.1格式：

mkdir export-feature-table

qiime tools export \

> --input-path table-deblur.qza \

> --output-path ./export-feature-table/



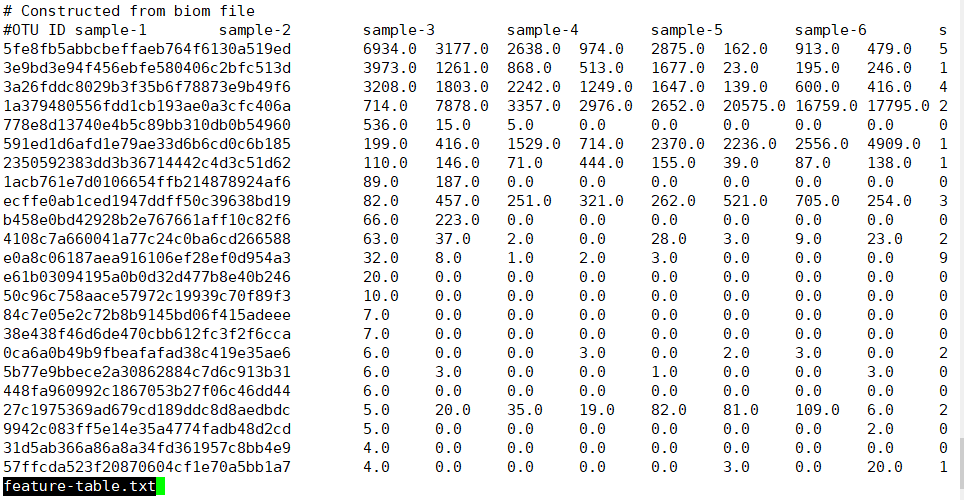
BIOM 2.1格式也是HDF5格式，为二进制，无法直接查看，必须转换为文本阅读。

biom转换biom为tsv格式：

biom convert -i feature-table.biom -o feature-table.txt --to-tsv

使用命令查看：

less -S feature-table.txt



## 将特征表导入Rstudio中进行抽平：(以下为R语言代码)

Vegan包

> library(vegan)

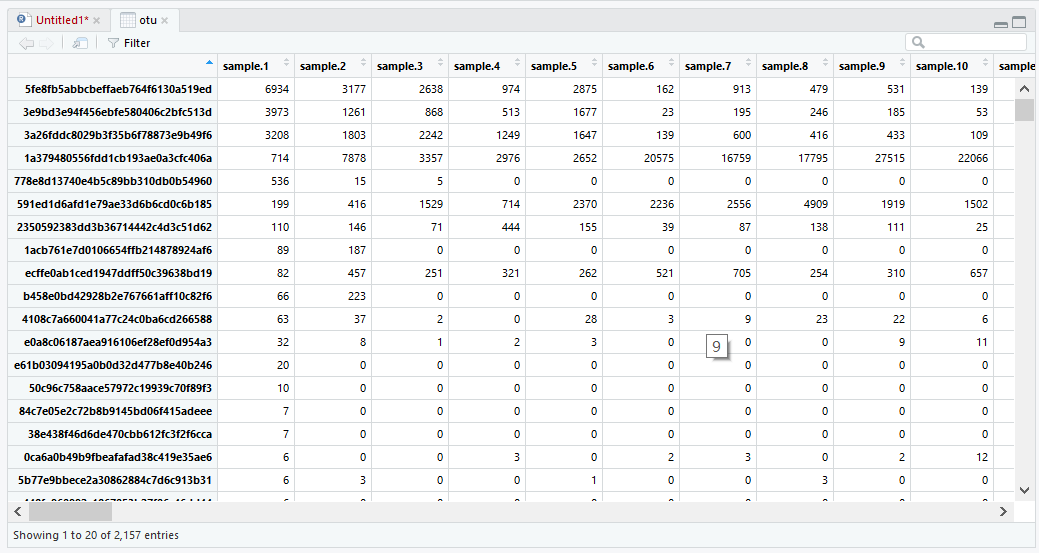
载入需要的程辑包：permute

载入需要的程辑包：lattice

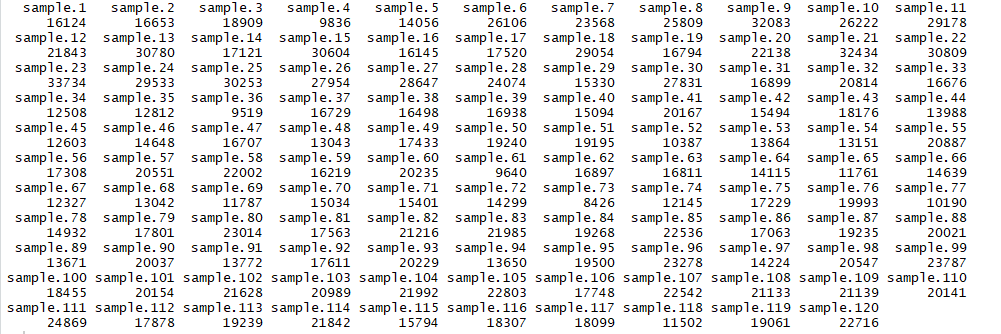
This is vegan 2.5-7

> otu = read.table('feature-table.txt', header=T, sep="\t", quote = "", row.names=1, comment.char="",stringsAsFactors = FALSE)

> View(otu)

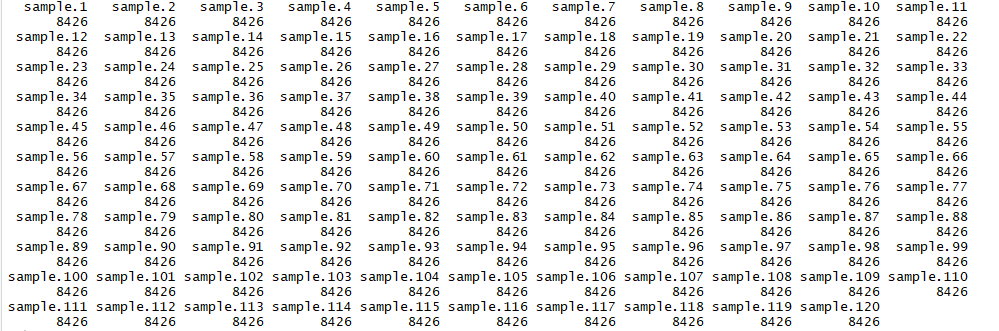


> colSums(otu)



> otu\_rare = as.data.frame(t(rrarefy(t(otu), min(colSums(otu)))))

> colSums(otu\_rare)



out表格转置以方便后续分析：

> View(otu\_rare)

Alpha多样性指数的计算

#计算香农指数

shannon=diversity(otu,"shannon")

#计算辛普森指数

simpson=diversity(otu,"simpson")

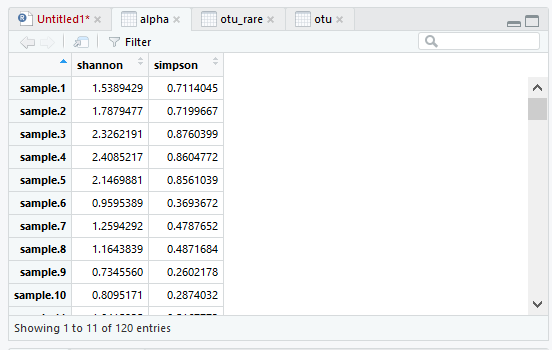
#合并数据

alpha=data.frame(shannon,simpson)

#储存结果

write.table(alpha,'alpha-summary.tsv',sep = '\t',quote=F)

结果如下：



可视化数据导入

#读取分组表格

map<-read.table('mapping\_file.txt',row.names=1,header=T,sep='\t',comment.char='',check.names=F)

#提取需要的分组，'Group1'为表中分组列名

group<-map['Group1']

#读取alpha多样性表

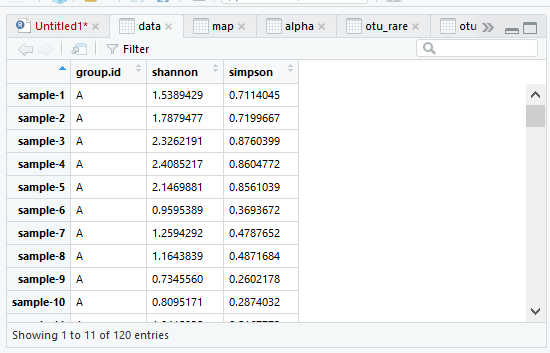
alpha<-read.table('alpha-summary.tsv',header = T,row.names = 1,sep = '\t')

#重排alpha的行的顺序，使其与group的样本id（行名）顺序一致

alpha<-alpha[match(rownames(group),rownames(alpha)),]

#合并两个表格

data<-data.frame(group,alpha)



### ggplot2进行可视化

以Shannon指数：

library(ggplot2)

# 加载R包ggplot2

alpha\_boxplot=ggplot(data, aes(x=group.id, y=shannon,fill=group.id))+

# 添加数据、xy值、 颜色参数给画图函数ggplot

geom\_boxplot()+

# 盒图

labs(title="Alpha diversity", x="Group", y="Shannon index")+

# 标题

theme(plot.title=element\_text(hjust=0.5), legend.title=element\_blank())

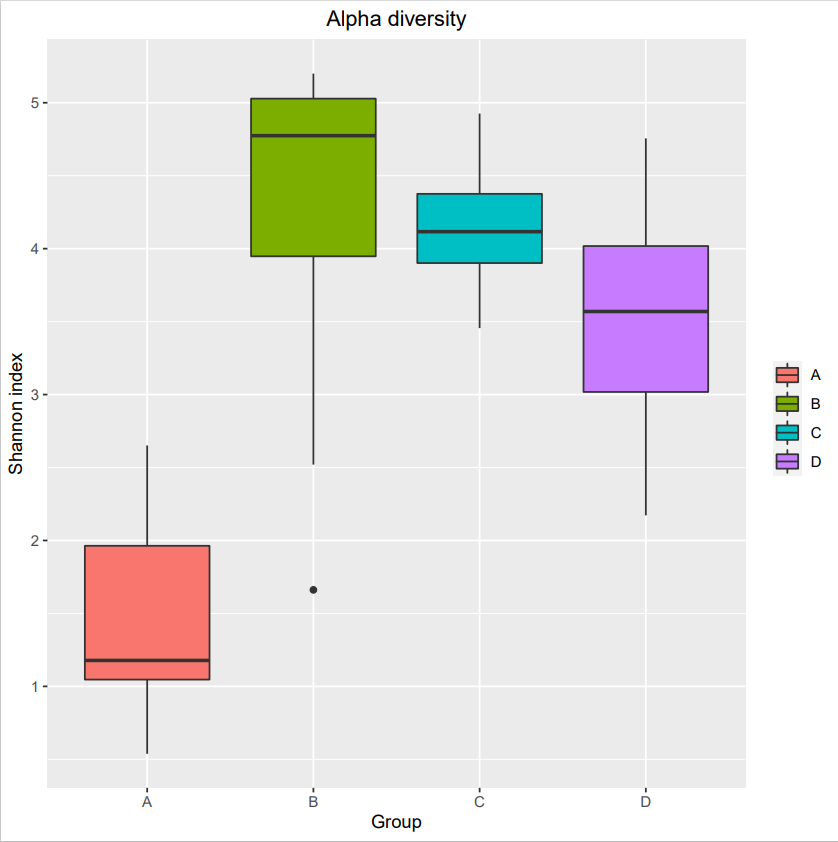
# 标题居中

pdf('result.pdf')

alpha\_boxplot

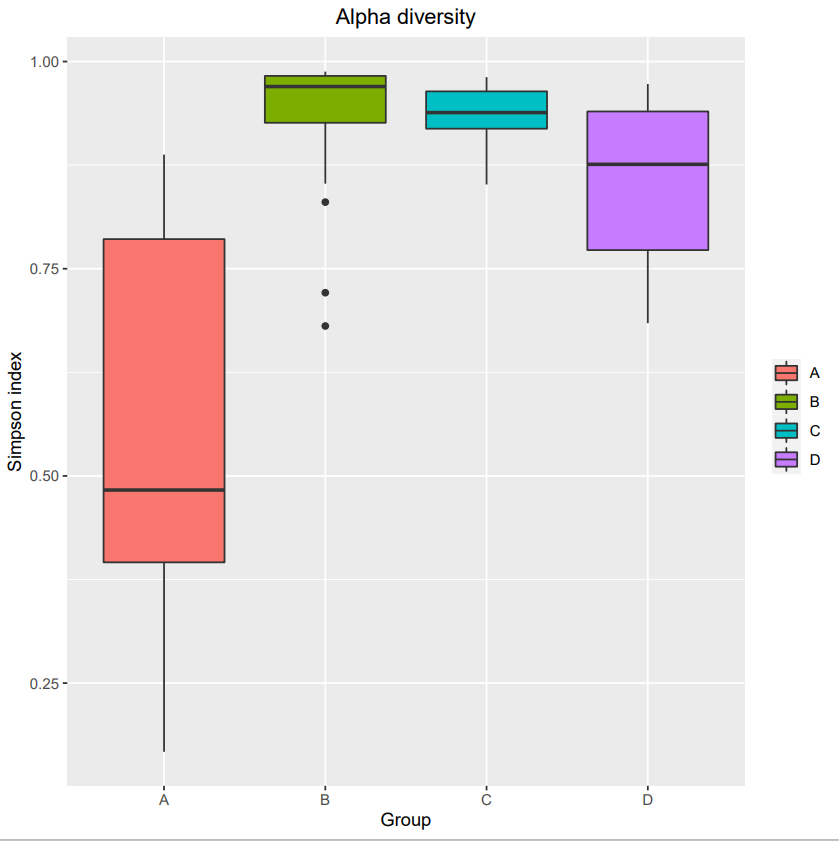
dev.off()

# 保存结果，打开result.pdf文件，结果如下：



以Simpson指数：

代码略



（A组回肠，B组结肠，C组盲肠，D组结肠）

### Beta多样性

library(vegan)

NMDS分析

##第 1 种模式，输入距离矩阵排序

#读取 OTU 丰度表

otu <- read.delim("feature-table.txt", row.names = 1, sep = '\t', stringsAsFactors = FALSE, check.names = FALSE)

otu <- data.frame(t(otu))

详情 ?vegdist

bray\_dis <- vegdist(otu\_rare, method = 'bray') #结果以 dist 数据类型存储

#NMDS 排序，定义 2 个维度，详情 ?metaMDS

nmds\_dis <- metaMDS(bray\_dis, k = 2)

#应力函数值，一般不大于 0.2 为合理

nmds\_dis$stress

#样方得分

nmds\_dis\_site <- data.frame(nmds\_dis$points)

nmds\_dis\_species <- wascores(nmds\_dis$points, otu\_rare)

library(ggplot2)

#主要展示 top10 丰度物种

abundance <- apply(otu\_rare, 2, sum)

abundance\_top10 <- names(abundance[order(abundance, decreasing = TRUE)][1:10])

species\_top10 <- data.frame(nmds\_dis\_species[abundance\_top10,1:2])

species\_top10$name <- rownames(species\_top10)

#添加分组信息 上传分组文件

nmds\_dis\_site$name <- rownames(nmds\_dis\_site)

map<-read.table("sample-metadata.txt",header=T,sep="\t",row.names=1)

#nmds\_dis\_site$group <- map$group

merged=merge(nmds\_dis\_site,map,by="row.names",all.x=TRUE)

color=c( "#3C5488B2","#00A087B2",

"#F39B7FB2","#91D1C2B2",

"#8491B4B2", "#DC0000B2",

"#7E6148B2","yellow",

"darkolivegreen1", "lightskyblue",

"darkgreen", "deeppink", "khaki2",

"firebrick", "brown1", "darkorange1",

"cyan1", "royalblue4", "darksalmon",

"darkgoldenrod1", "darkseagreen", "darkorchid")

#grid.col[row.names(data)] = color[1:dim(data)[1]]

library(ggplot2)

p <- ggplot(data = merged, aes(MDS1, MDS2)) +

geom\_point(size=2,aes(color = group.id,shape = group.id)) +

stat\_ellipse(aes(fill = group.id), geom = 'polygon', level = 0.95, alpha = 0.1, show.legend = FALSE) + #添加置信椭圆

scale\_color\_manual(values =color[1:length(unique(map$group))]) +

scale\_fill\_manual(values = color[1:length(unique(map$group))]) +

theme(panel.grid.major = element\_line(color = 'gray', size = 0.2), panel.background = element\_rect(color = 'black', fill = 'transparent'),

plot.title = element\_text(hjust = 0.5),legend.title = element\_blank()) +

#, legend.position = 'none'

geom\_vline(xintercept = 0, color = 'gray', size = 0.5) +

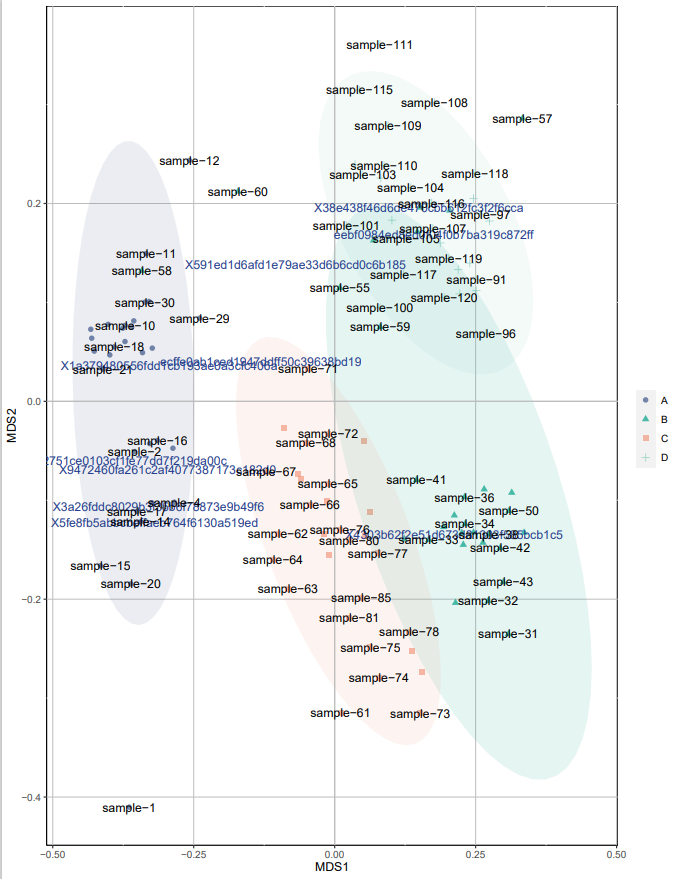
geom\_hline(yintercept = 0, color = 'gray', size = 0.5)+

geom\_text(data = species\_top10, aes(label = name), color ="royalblue4", size = 4)+

geom\_text(data =merged, aes(label = Row.names,x =MDS1, y = MDS2), size=4, check\_overlap = TRUE)

#geom\_text(data = species\_top10, aes(label = name), color = 'blue', size = textsize)

p



PCoA分析作图：

#导入作图包

library(vegan)

library(ggplot2)

#颜色选择

color=c( "#3C5488B2","#00A087B2",

"#F39B7FB2","#91D1C2B2",

"#8491B4B2", "#DC0000B2",

"#7E6148B2","yellow",

"darkolivegreen1", "lightskyblue",

"darkgreen", "deeppink", "khaki2",

"firebrick", "brown1", "darkorange1",

"cyan1", "royalblue4", "darksalmon",

"darkgoldenrod1", "darkseagreen", "darkorchid")

#读取otu数据文件

otu <- read.delim('feature-table.txt', row.names = 1, sep = '\t', stringsAsFactors = FALSE, check.names = FALSE)

otu <- data.frame(t(otu))

#根据物种组成计算样方距离，结果以 dist 数据类型存储

bray\_dis <- vegdist(otu, method = 'bray')

#样方排序坐标

pcoa <- cmdscale(bray\_dis, k = (nrow(otu) - 1), eig = TRUE)

site <- data.frame(pcoa$point)[1:2]

site$name <- rownames(site)

#读取分组文件

map<-read.table('sample-metadata.tsv',header=T,sep="\t",row.names=1)

site$group <- c(rep('A', 30), rep('B', 30), rep('C', 30),rep('D',30))

#将分组文件和数据文件以行名合并

merged=merge(site,map,by="row.names",all.x=TRUE)

species <- wascores(pcoa$points[,1:4], otu)

#主要展示 top10 丰度物种

#计算 top10 丰度物种

abundance <- apply(otu, 2, sum)

abundance\_top10 <- names(abundance[order(abundance, decreasing = TRUE)][1:10])

species\_top10 <- data.frame(species[abundance\_top10,1:2])

species\_top10$name <- rownames(species\_top10)

pcoa\_exp <- pcoa$eig/sum(pcoa$eig)

pcoa1 <- paste('PCoA axis1 :', round(100\*pcoa\_exp[1], 2), '%')

pcoa2 <- paste('PCoA axis2 :', round(100\*pcoa\_exp[2], 2), '%')

#ggplot2 作图

library(ggplot2)

p <- ggplot(data = merged, aes(X1, X2)) +

geom\_point(aes(color = group)) +

stat\_ellipse(aes(fill = group), geom = 'polygon', level = 0.95, alpha = 0.1, show.legend = FALSE) + #添加置信椭圆，注意不是聚???

scale\_color\_manual(values =color[1:length(unique(map$group))]) +

scale\_fill\_manual(values =color[1:length(unique(map$group))]) +

theme(panel.grid.major = element\_line(color = 'gray', size = 0.2), panel.background = element\_rect(color = 'black', fill = 'transparent'),

plot.title = element\_text(hjust = 0.5)) +

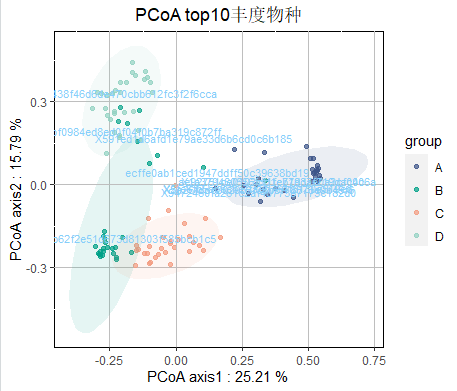
geom\_vline(xintercept = 0, color = 'gray', size = 0.5) +

geom\_hline(yintercept = 0, color = 'gray', size = 0.5) +

geom\_text(data = species\_top10, aes(label = name), color = "lightskyblue", size = 3) + #??? top10 丰度物种标签

labs(x = pcoa1, y = pcoa2, title = 'PCoA top10丰度物种')

p



### 物种组成分析

library("tidyr")

library("ggplot2")

library("ggsci")

library("reshape2")

setwd("C:/Users/lina1/Desktop/16S分析/齐鸣/不同肠段样品的16S分析/version2\_HMJZ/taxonomy/taxonomy\_level")

#####科水平的堆积图#######

table <- read.delim("level-5.txt")

rownames(table) <- table$INDEX

table<-table[,-1]

#统计在每个样本中top15的科，求并集，然后统计这些科出现在样本中的频次,求前12个科

genus\_top<- NULL

genus\_names\_table <-NULL

for (i in 1:ncol(table)){

i\_table <- table[,i]

names(i\_table)<-rownames(table)

tmp\_names <- names(sort(i\_table,decreasing = TRUE)[1:15]) #每个样本中前15

genus\_names\_table<-c(genus\_names\_table,tmp\_names)

}

genus\_top<-names(sort(table(genus\_names\_table),decreasing = TRUE)[1:11])

rm(list = c("i","genus\_names\_table","tmp\_names","i\_table"))

#将在genus\_top中的保留下来，其余的标记为others\_table，并求和记为others,然后合并sub\_table和others

sub\_table=table[genus\_top,]

others\_table=table[!rownames(table) %in% genus\_top,]

others=t(as.data.frame(colSums(others\_table)))

rownames(others)<-"others"

table=rbind(sub\_table,others)

rm(list = c("genus\_top","others","others\_table","sub\_table"))

table<-cbind(rownames(table),table)

names(table)[1]<-"names"

#调整分类信息

#按照分类信息进行分割,将未识别的列名记为unclassified,这一步出现报错正常

table=separate(data=table,col=names,into =c("tmp","F"),sep = "f\_\_")

table[is.na(table$F),"F"]<-"others"

rownames(table)<-table$F

table=table[,c(-1,-2)]

#求同组的均值

group\_table <- NULL

group\_list=c("HA","JA","MA","ZA")

for (i in group\_list){

group<-subset(table,select=which(grepl(i,colnames(table))))

group\_table<-cbind(group\_table,assign(paste0(i,"\_mean"),rowSums(group)/ncol(group)))

rm(list=paste0(i,"\_mean"))

}

colnames(group\_table)<-group\_list

#转换成丰度表

group\_table=apply(group\_table,2,function(x)x/sum(x))

#竖着排列

group\_table=melt(group\_table,id=rownames(group\_table))

names(group\_table)[1]='taxonomy'

names(group\_table)[2]='sample'

group\_table[grep("H",group\_table$sample),4]<-"ileum"

group\_table[grep("J",group\_table$sample),4]<-"colon"

group\_table[grep("M",group\_table$sample),4]<-"cecum"

group\_table[grep("Z",group\_table$sample),4]<-"rectum"

group\_table[grep("A",group\_table$sample),5]<-"A"

group\_table[grep("J",group\_table$sample),5]<-"B"

group\_table[grep("M",group\_table$sample),5]<-"C"

group\_table[grep("Z",group\_table$sample),5]<-"D"

names(group\_table)[4]='Gut\_region'

names(group\_table)[5]='Group'

group\_table$Gut\_region<-factor(group\_table$Gut\_region,levels = c("ileum","cecum","colon","rectum"))

#12个颜色配色

mypal<-c( "#E64B35FF","#0072B5FF","#4DBBD5FF","#20854EFF","#00A087FF","#3C5488FF","#F39B7FFF","#8491B4FF",

"#91D1C2FF","#DC0000FF","#7E6148FF","#E18727FF")

#可视化

p1<-ggplot(group\_table, aes(x=Group, fill=taxonomy, y=value\*100))+

geom\_col(position='stack',alpha=0.9)+

scale\_fill\_manual(values = mypal)+

labs(x='Samples', y='Relative Abundance (%)')+

scale\_y\_continuous(expand=c(0, 0))+

theme\_classic()+

theme(axis.text.x=element\_text(size=12,angle=45, hjust=1))+

theme(axis.text.y=element\_text(size=12))+

theme(legend.text= element\_text(size=10))+

theme(title = element\_text(size = 15))

p2<-p1+facet\_wrap(~Gut\_region,nrow = 1)

p2

ggsave(p2,filename = "silva\_Famliy\_stackPlot.pdf",width = 10.3,height = 8)

