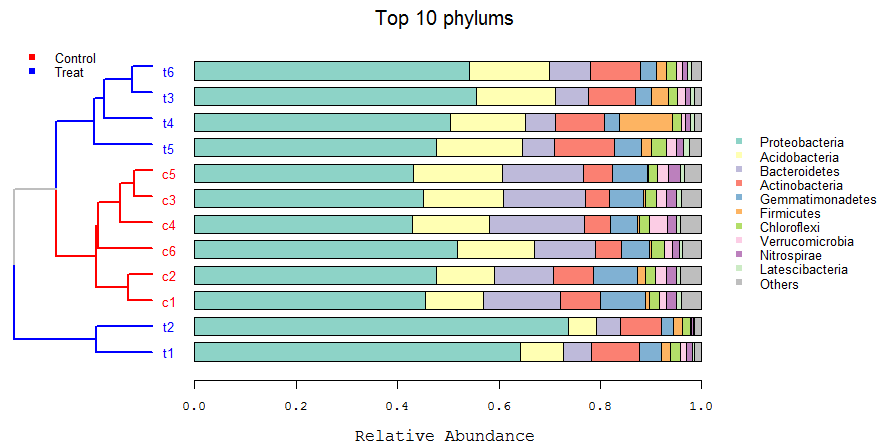
R语言绘制带聚类树的堆叠柱形图

聚类树与柱形图结合，即可反映样本或分组间的相似性，又能展示样本内的元素组成信息。

例如下图是一个在扩增子测序微生物群落分析中常见的统计图类型，在测序公司给的报告中通常都有这么一张图。聚类树表示了各样本或分组之间，物种丰度组成的相似性或相异程度；柱形图就是常见的物种堆叠柱形图，常用于表示代表性的门纲目科属等丰度，如下图中即展示了各样本中丰度排名前10的细菌门的丰度信息。



很容易看出，该图主体由两部分组成，[聚类树](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484616&idx=1&sn=b7ed9e20bac5d8a8b3b72ded3d28227f&chksm=97f5b4d0a0823dc6a365108ec7c5daf76a9402d28f1188b963fe307401eec3a52e8118e2570e&token=497100570&lang=zh_CN#rd)+[堆叠柱形图](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247483757&idx=1&sn=e3f6e387d6afef876dbd2af8b34e69dd&chksm=97f5b175a0823863f9bf0ccc3272d1360a0b12869511dc82a0704a9aeeb5f4fb9fdaff8ed89a&token=497100570&lang=zh_CN#rd)。

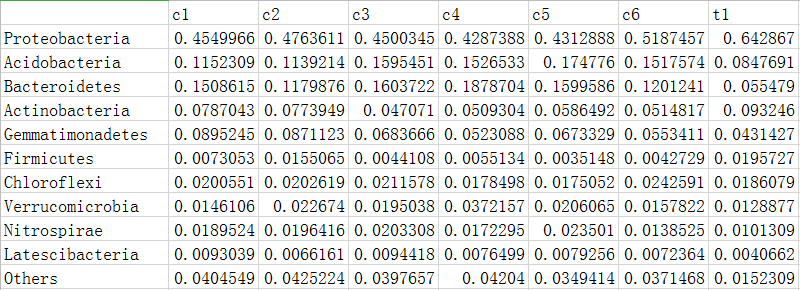
本篇分享怎样在R语言中绘制。

示例数据，R代码等的百度盘链接：

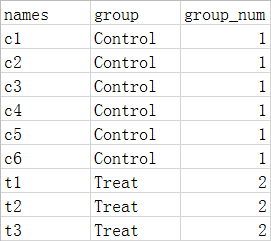
<https://pan.baidu.com/s/1mcn3XCma9_0C2LpDVytYsg>

## 作图示例文件

文件“phylum\_top10.txt”由16S高通量测序所得的物种丰度表计算得到，记录了丰度排名top10的主要细菌类群（在门水平统计，行）在各样本（12个样本，列）中的丰度信息。Top10丰度外的类群合并为Others。



“group.txt”为12个样本的分组信息。



通过这两个文件绘制带聚类树的堆叠柱形图。

## R语言作图

首先进行[层次聚类](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484593&idx=1&sn=d8160e2918b0aa4dd90cb10c2f82d266&chksm=97f5b4a9a0823dbf52ebacc1c1390c2047dfafcf217ae9afe34583dfe4a999621d6d274c4af2&token=497100570&lang=zh_CN#rd)，获得代表样本间物种组成相似性的聚类树，这一步很简单。

#层次聚类

#读取 OTU 丰度表

dat <- read.delim('phylum\_top10.txt', row.names = 1, sep = '\t', head = TRUE, check.names = FALSE)

#计算样本间距离，以群落分析中常用的 Bray-curtis 距离为例

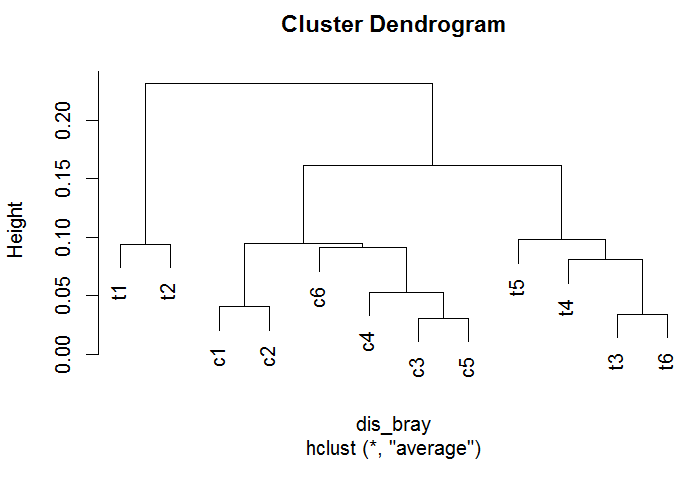
dis\_bray <- vegan::vegdist(t(dat), method = 'bray')

#层次聚类，以 UPGMA 为例

tree <- hclust(dis\_bray, method = 'average')

tree

plot(tree)



接下来就是对聚类树进行调整。

具体做法，首先定义一个画板，将聚类树放在画板的左侧，并按样本的已知分组信息给分支上色。聚类树的一些常见的可视化调整方法，可[参考前文](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247484616&idx=1&sn=b7ed9e20bac5d8a8b3b72ded3d28227f&chksm=97f5b4d0a0823dc6a365108ec7c5daf76a9402d28f1188b963fe307401eec3a52e8118e2570e&token=497100570&lang=zh_CN#rd)。

##聚类树绘制

#样本分组颜色、名称等

group <- read.delim('group.txt', row.names = 1, sep = '\t', head = TRUE, check.names = FALSE, stringsAsFactors = FALSE)

grp <- group[2]

group\_col <- c('red', 'blue')

names(group\_col) <- c('1', '2')

group\_name <- c('Control', 'Treat')

#样本分组标签

layout(t(c(1, 2, 2, 2, 3)))

par(mar = c(5, 2, 5, 0))

plot(0, type = 'n', xaxt = 'n', yaxt = 'n', frame.plot = FALSE, xlab = '', ylab = '',

xlim = c(-max(tree$height), 0), ylim = c(0, length(tree$order)))

legend('topleft', legend = group\_name, pch = 15, col = group\_col, bty = 'n', cex = 1)

#聚类树绘制，按分组给分支上色

treeline <- function(pos1, pos2, height, col1, col2) {

meanpos = (pos1[1] + pos2[1]) / 2

segments(y0 = pos1[1] - 0.4, x0 = -pos1[2], y1 = pos1[1] - 0.4, x1 = -height, col = col1,lwd = 2)

segments(y0 = pos1[1] - 0.4, x0 = -height, y1 = meanpos - 0.4, x1 = -height, col = col1,lwd = 2)

segments(y0 = meanpos - 0.4, x0 = -height, y1 = pos2[1] - 0.4, x1 = -height, col = col2,lwd = 2)

segments(y0 = pos2[1] - 0.4, x0 = -height, y1 = pos2[1] - 0.4, x1 = -pos2[2], col = col2,lwd = 2)

}

meanpos = matrix(rep(0, 2 \* length(tree$order)), ncol = 2)

meancol = rep(0, length(tree$order))

for (step in 1:nrow(tree$merge)) {

if(tree$merge[step, 1] < 0){

pos1 <- c(which(tree$order == -tree$merge[step, 1]), 0)

col1 <- group\_col[as.character(grp[tree$labels[-tree$merge[step, 1]],1])]

} else {

pos1 <- meanpos[tree$merge[step, 1], ]

col1 <- meancol[tree$merge[step, 1]]

}

if (tree$merge[step, 2] < 0) {

pos2 <- c(which(tree$order == -tree$merge[step, 2]), 0)

col2 <- group\_col[as.character(grp[tree$labels[-tree$merge[step, 2]],1])]

} else {

pos2 <- meanpos[tree$merge[step, 2], ]

col2 <- meancol[tree$merge[step, 2]]

}

height <- tree$height[step]

treeline(pos1, pos2, height, col1, col2)

meanpos[step, ] <- c((pos1[1] + pos2[1]) / 2, height)

if (col1 == col2) meancol[step] <- col1 else meancol[step] <- 'grey'

}



按分组给聚类树的分支或簇标记了颜色，放置在画板左侧的一小部分区域。

对于右侧区域，放置堆叠柱形图，用以展示top10丰度的细菌门在各样本中的丰度信息。堆叠柱形图单独的画法，可[参考该文](https://mp.weixin.qq.com/s?__biz=MzIxNzc1Mzk3NQ==&mid=2247483757&idx=1&sn=e3f6e387d6afef876dbd2af8b34e69dd&chksm=97f5b175a0823863f9bf0ccc3272d1360a0b12869511dc82a0704a9aeeb5f4fb9fdaff8ed89a&token=497100570&lang=zh_CN#rd)。

##堆叠柱形图

#样本顺序调整为和聚类树中的顺序一致

dat <- dat[ ,tree$order]

#物种颜色设置

phylum\_color <- c('#8DD3C7', '#FFFFB3', '#BEBADA', '#FB8072', '#80B1D3', '#FDB462', '#B3DE69', '#FCCDE5', '#BC80BD', '#CCEBC5', 'gray')

names(phylum\_color) <- rownames(dat)

#堆叠柱形图

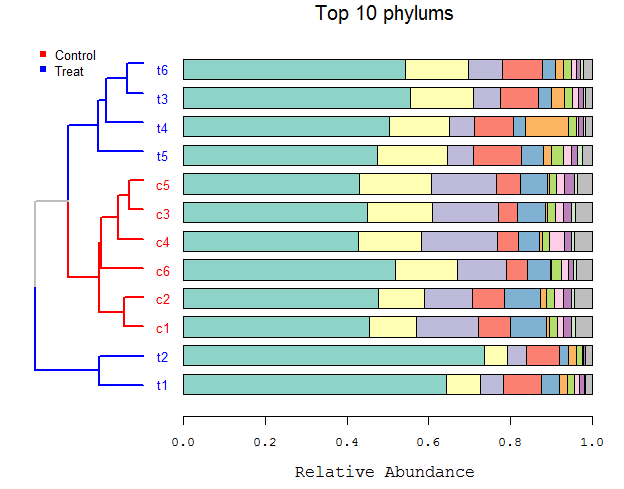
par(mar = c(5, 2, 5, 0))

tt <- barplot(as.matrix(dat), col = phylum\_color, space = 0.4, width = 0.7, cex.axis = 1, horiz = TRUE, cex.lab = 1.2,

xlab = 'Relative Abundance', yaxt = 'n', las = 1, ylim = c(0, ncol(dat)), family = 'mono')

mtext('Top 10 phylums', side = 3, line = 1, cex = 1)

text(x = -0.05, y = tt, labels = colnames(dat), col = group\_col[group[tree\_order, 2]], xpd = TRUE)



首先按聚类树中样本的顺序重新调整丰度表中样本的顺序，以保持二者能够对应，并定义颜色属性。之后绘制堆叠柱形图，放置在画板右侧区域。

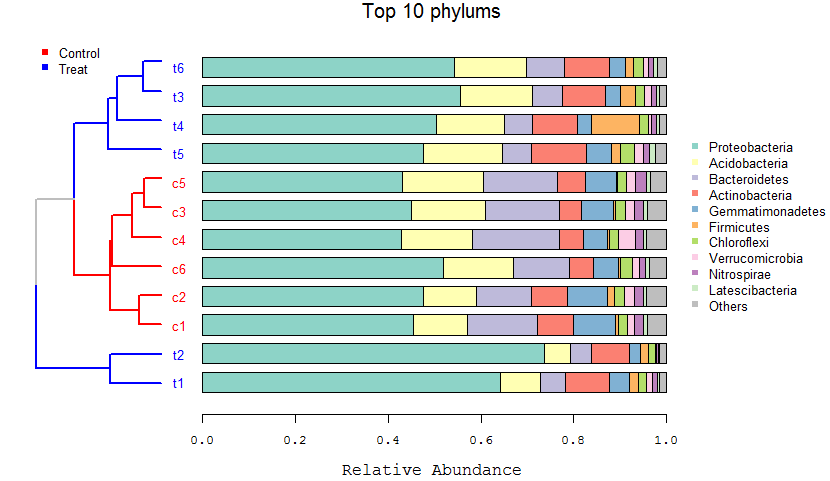
最后在最右侧添加柱形图图例，哪些颜色代表了哪种细菌类群。

#柱形图图例

par(mar = c(5, 1, 5, 0))

plot(0, type = 'n', xaxt = 'n', yaxt = 'n', bty = 'n', xlab = '', ylab = '')

legend('left', pch = 15, col = phylum\_color, legend = names(phylum\_color), bty = 'n', cex = 1)



这样这种带聚类的堆叠柱形图就搞定了。

## 其它说明

如果觉得使用R组合两张图比较麻烦，特别是细节部分不容易调整，不妨分别绘制聚类树和堆叠柱形图，然后使用AI、PS等工具将二者拼合在一起，可能相对方便许多。再略加修整，也会更好看。