

# R 中的生物多样性分析

加拿大 Steve Kembel 原著

张金龙编译

## 目录

<b>1</b>	<b>背景</b>	<b>2</b>
<b>2</b>	<b>怎样使用课程材料</b>	<b>2</b>
<b>3</b>	<b>生物多样性数据导入 R</b>	<b>2</b>
3.1	群落数据 . . . . .	3
3.2	性状数据 . . . . .	6
3.3	进化树（或称为系统树） . . . . .	8
<b>4</b>	<b>生物多样性数据的可视化及汇总</b>	<b>11</b>
4.1	物种丰富度和多样性 . . . . .	11
4.2	等级聚类 . . . . .	13
4.3	排序 . . . . .	14
4.4	添加环境信息和性状数据到排序中 . . . . .	19
<b>5</b>	<b>性状与进化树</b>	<b>20</b>
5.1	系统发育信号 . . . . .	20
5.2	性状进化的可视化 . . . . .	22
<b>6</b>	<b>系统发育与性状多样性</b>	<b>24</b>
6.1	系统发育多样性 . . . . .	24
6.2	系统发育 beta 多样性 . . . . .	33
6.3	系统发育和功能性状排序 . . . . .	35

## 1 背景

本课程中，我们将学习如何分析阿尔伯塔草原的一组植物群落数据。数据中包括阿尔伯塔几个地点草原植物盖度、功能性状和系统发育关系等，

数据详细介绍参见以下论文: S.W. Kembel and J.F. Cahill, Jr. 2011. Independent evolution of leaf and root traits within and among temperate grassland plant communities. PLoS ONE 6(6): e19992. ([doi:10.1371/journal.pone.0019992](https://doi.org/10.1371/journal.pone.0019992)).

## 2 怎样使用课程材料

本课程覆盖了导入数据，在 R 中分析生物多样性的全部内容。如果要学习全部课程，可以从 R 入门部分开始。如果已经有 R 的基础，则可以直接学习本讲义。本课程需要用到 picante 程序包，以及数据映像（即.Rdata），后者包括练习用的全部数据，有了这些数据，运行后面的命令就可直接得到结果。

## 3 生物多样性数据导入 R

首先，查看一下我们需要的程序包是否能加载。需要用的程序包有 ape, picante, vegan. 由于 picante 是依赖另外两个程序包的，所以加载 picante 的时候，另外两个程序包也会自动加载。

设定工作路径，请注意 Windows 和 MacOS 以及 Linux 操作系统上，路径斜杠的方向

```
#setwd("/Users/jinlong/Dropbox/04\ to\ review/kembel/")  
setwd("C:\\Users\\jllzhang\\Dropbox\\04 to review\\kembel")  
library(picante)
```

```
## Loading required package: ape
```

```
## Loading required package: vegan
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-2

## Loading required package: nlme
```

### 3.1 群落数据

群落数据包括每个地点，或者每块样地物种的个体数，或者相对多度等。在本课程的数据中，多度数据是每个不同生境 20m\*20m 样方中植物的相对盖度。

群落数据的格式为 `data.frame`，行表示样地，列表示物种。练习数据已经包括 `comm` 数据了，所以直接用就行了。如果数据是保存在 `csv` 文件中，则需要用 `read.csv` 函数读取，格式如下：

```
# 每一行代表一块样地，因此行名就是样地编号，
# 第一行作为各列名称，所以这里设定 header = TRUE。

comm <- read.csv("grassland.community.csv", header = TRUE, row.names = 1)
```

用以上方法将数据读取到 R 中，注意行名和列名并不是数据本身，而是各行列的名称。行列有了名称，在后续就更容易操作和展示。群落数据的各行列有了名称，才方便和其他类型的数据关联。

```
# 先查看 comm 的数据类型
class(comm)

## [1] "data.frame"

# 查看该 data.frame 是由多少行列组成的。(rows x columns)
dim(comm)
```

```
## [1] 27 76
```

```
# 查看行名
rownames(comm)
```

```
## [1] "mix-0-1" "mix-0-2" "mix-0-3" "mix-0-4" "mix-0-5" "mix-0-6"
## [7] "mix-0-7" "fes-K-8" "fes-K-9" "fes-K-10" "fes-K-11" "fes-K-12"
## [13] "fes-K-13" "fes-K-14" "fes-K-15" "fes-K-16" "fes-K-17" "mix-H-18"
## [19] "mix-H-19" "mix-H-20" "mix-H-21" "mix-H-22" "mix-H-23" "mix-H-24"
## [25] "mix-H-25" "mix-H-26" "mix-H-27"
```

```
# 查看前六列的名
```

```
head(colnames(comm))
```

```
## [1] "Antennaria_parvifolia"
## [2] "Artemisia_cana"
## [3] "Artemisia_frigida"
## [4] "Symphyotrichum_ericoides_var._ericoides"
## [5] "Bouteloua_gracilis"
## [6] "Carex_filifolia"
```

```
# 提取数据的子集, 1 到 5 行, 1 到 5 列数据
```

```
comm[1:5, 1:5]
```

```
##           Antennaria_parvifolia Artemisia_cana Artemisia_frigida
## mix-0-1              10              10              50
## mix-0-2               0              10              50
## mix-0-3              20              20              30
## mix-0-4               0               0               0
## mix-0-5               0              10               0
##           Symphyotrichum_ericoides_var._ericoides Bouteloua_gracilis
## mix-0-1                      10                      70
## mix-0-2                      10                      90
## mix-0-3                      10                      60
## mix-0-4                       0                      90
## mix-0-5                       0                     100
```

本数据中, 多度其实是样地中某种植物的盖度。多元统计中的很多方法对于多度很敏感, 因此可以将绝对多度转换为相对多度。用 `vegan` 中的函数即可完成。

```
# 查看每个样方中盖度之和， 如果数据为个体数，
# 则表示每个样方中所有种的总个体数
apply(comm, 1, sum)
```

```
## mix-0-1 mix-0-2 mix-0-3 mix-0-4 mix-0-5 mix-0-6 mix-0-7 fes-K-8
##      640      630      710      350      400      650      560      960
## fes-K-9 fes-K-10 fes-K-11 fes-K-12 fes-K-13 fes-K-14 fes-K-15 fes-K-16
##      960      980      830      980      980      830      640     1080
## fes-K-17 mix-H-18 mix-H-19 mix-H-20 mix-H-21 mix-H-22 mix-H-23 mix-H-24
##      710      440      590      540      340      420      400      600
## mix-H-25 mix-H-26 mix-H-27
##      540      590      420
```

```
# 除以每个样方的盖度和， 将盖度转换为相对盖度
comm <- decostand(comm, method = "total")
```

```
# 每个样方的总盖度
apply(comm, 1, sum)
```

```
## mix-0-1 mix-0-2 mix-0-3 mix-0-4 mix-0-5 mix-0-6 mix-0-7 fes-K-8
##      1      1      1      1      1      1      1      1
## fes-K-9 fes-K-10 fes-K-11 fes-K-12 fes-K-13 fes-K-14 fes-K-15 fes-K-16
##      1      1      1      1      1      1      1      1
## fes-K-17 mix-H-18 mix-H-19 mix-H-20 mix-H-21 mix-H-22 mix-H-23 mix-H-24
##      1      1      1      1      1      1      1      1
## mix-H-25 mix-H-26 mix-H-27
##      1      1      1
```

```
# 查看转换后的数据
comm[1:5, 1:5]
```

```
##      Antennaria_parvifolia Artemisia_cana Artemisia_frigida
## mix-0-1      0.01562500      0.01562500      0.07812500
## mix-0-2      0.00000000      0.01587302      0.07936508
```

```
## mix-0-3          0.02816901    0.02816901    0.04225352
## mix-0-4          0.00000000    0.00000000    0.00000000
## mix-0-5          0.00000000    0.02500000    0.00000000
##      Symphyotrichum_ericoides_var._ericoides Bouteloua_gracilis
## mix-0-1                      0.01562500    0.10937500
## mix-0-2                      0.01587302    0.14285714
## mix-0-3                      0.01408451    0.08450704
## mix-0-4                      0.00000000    0.25714286
## mix-0-5                      0.00000000    0.25000000
```

### 3.2 性状数据

trait 包括每个种叶和根的数据。读取到 R 中的方法与群落数据一样，但是在性状数据中，每个种是一行，每个性状是一列。

```
traits <- read.csv("species.traits.csv", header = TRUE, row.names = 1)
```

```
# 查看性状数据的前 6 列
```

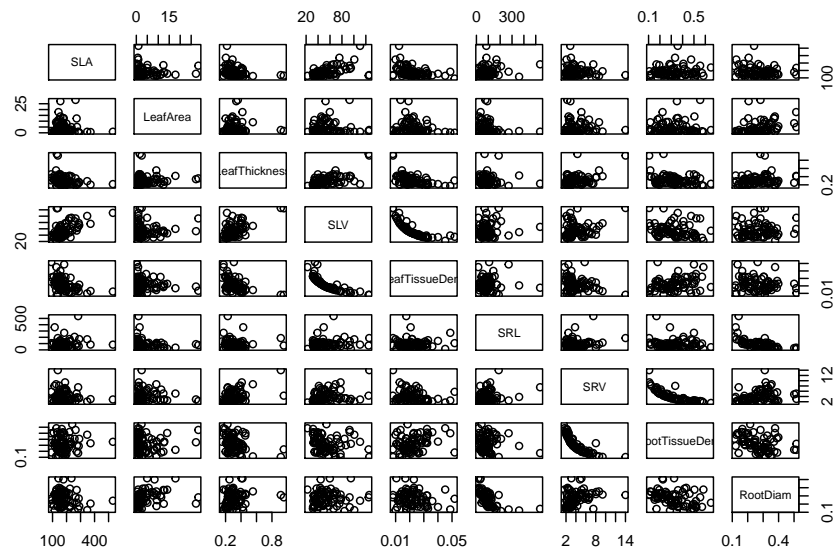
```
head(traits)
```

```
##              SLA  LeafArea LeafThickness    SLV
## Achillea_millefolium 140.2663  9.275390    0.4163333 59.56525
## Allium_textile      137.7006  2.445361    0.9147222 125.69496
## Amelanchier_alnifolia 156.1014 14.064856    0.2900000 45.45227
## Androsace_occidentalis 257.2050 0.274745    0.2535000 84.22189
## Antennaria_neglecta 171.0442  1.731990    0.2810000 48.14442
## Antennaria_parvifolia 193.8718 0.317200    0.2466667 47.64151
##              LeafTissueDens    SRL    SRV RootTissueDens
## Achillea_millefolium    0.018085428 74.14570 5.038776    0.2553510
## Allium_textile          0.008137136 187.85485 14.013757    0.1049986
## Amelanchier_alnifolia    0.022841232 20.87560 2.518939    0.5039683
## Androsace_occidentalis    0.017706402 207.45582 3.291592    0.4071078
## Antennaria_neglecta      0.020920233 124.73397 6.710526    0.1593750
## Antennaria_parvifolia    0.021048340 44.93859 4.003997    0.2504167
##              RootDiam
```

```
## Achillea_millefolium 0.3123600
## Allium_textile 0.3107833
## Amelanchier_alnifolia 0.3760667
## Androsace_occidentalis 0.1148714
## Antennaria_neglecta 0.2749500
## Antennaria_parvifolia 0.3495500
```

```
# 绘制两两相关图
```

```
pairs(traits)
```

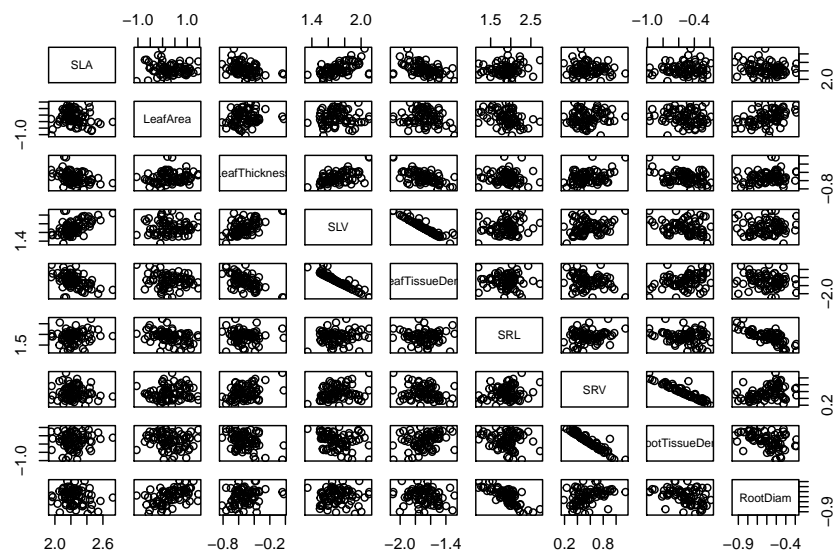


```
# 由于一些变量显著偏向一侧，因此进行对数转换
```

```
traits <- log10(traits)
```

```
# 查看转换后的数据
```

```
pairs(traits)
```



## 样地的总体情况 Metadata 主要包括样地的生境，采集地点，环境数据，如坡度、干湿度等

```
metadata <- read.csv("plot.metadata.csv", header = TRUE, row.names = 1)
```

```
# 查看前 6 行
head(metadata)
```

```
##           habitat    site slope aspect slope.position rel.moisture
## mix-0-1 Mixedgrass Onefour      0   270           3.0           1
## mix-0-2 Mixedgrass Onefour     20   130           1.5           2
## mix-0-3 Mixedgrass Onefour      5    90           1.0           2
## mix-0-4 Mixedgrass Onefour      5    40           2.0           1
## mix-0-5 Mixedgrass Onefour      5   130           2.0           1
## mix-0-6 Mixedgrass Onefour      1    90           3.0           1
```

### 3.3 进化树（或称为系统树）

newick 格式的进化树，用 read.tree 读取。Nexus 格式的进化树，用 read.nexus 读取。



```
phy <- read.tree("grassland.phylogeny.newick")
class(phy)
```

```
## [1] "phylo"
```

```
phy
```

```
##
```

```
## Phylogenetic tree with 76 tips and 68 internal nodes.
```

```
##
```

```
## Tip labels:
```

```
## Antennaria_neglecta, Antennaria_parvifolia, Erigeron_glabellus, Erigeron_pumilus, H
```

```
## Node labels:
```

```
## , , , , , ...
```

```
##
```

```
## Rooted; includes branch lengths.
```

读取到 R 中的进化树为 phylo 为格式。phylo 格式的详细说明参见 ape 的主页 (<http://ape.mpl.ird.fr/>)。phylo 对象的本质是 list，包括进化树末端分类单元名称 (tip label)，枝长 (edge length) 等等，ape 内置的相关函数可以针对 phylo 这种数据格式打印，汇总，绘图等。

```
# 列出 phy 对象的所有名称， 这些名称是 list 内部对象的名
names(phy)
```

```
## [1] "edge" "edge.length" "Nnode" "node.label" "tip.label"
```

```
## [6] "root.edge"
```

```
# 查看末端分类单元名称， 取前 5 个
```

```
phy$tip.label[1:5]
```

```
## [1] "Antennaria_neglecta" "Antennaria_parvifolia" "Erigeron_glabellus"
```

```
## [4] "Erigeron_pumilus" "Heterotheca_villosa"
```



```
# 检查物种是否匹配
combined <- match.phylo.comm(phy, comm)
# 返回结果是一个包含 phy 和 comm 的 list,
# 未在两个数据集中都出现的物种都已经去除,
# 并且顺序都排好了。
# 这里用 combined 中的数据, 替换原始数据。

phy <- combined$phy
comm <- combined$comm
# 性状数据也需要进行同样的处理
combined <- match.phylo.data(phy, traits)

# 用结果中检查过的数据, 替换掉原来的数据。
# 由于之前的返回结果是 list, 这里用 $ 调取 phy 和 data
phy <- combined$phy
traits <- combined$data

# 再检查一下 metadata 与群落数据出现的顺序是否一致?
all.equal(rownames(comm), rownames(metadata))

## [1] TRUE

# 若顺序不同, metadata 数据中样方出现的顺序就应该按照群落顺序排序
metadata <- metadata[rownames(comm), ]

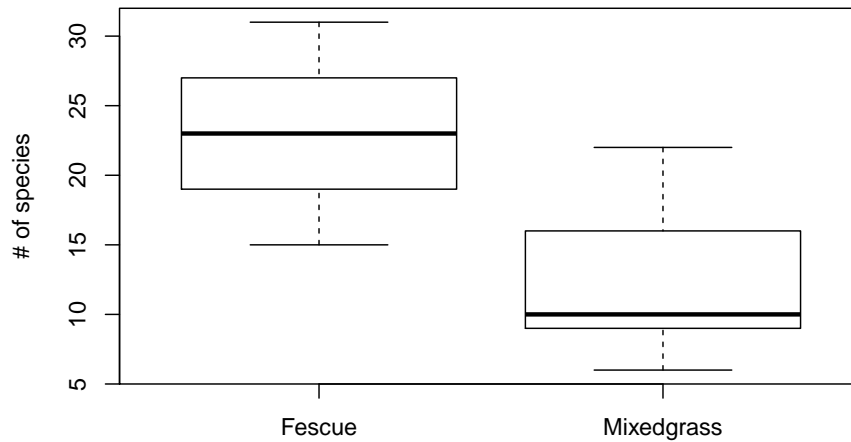
# 这样, 数据预处理就完成了。 后面即可开始各种生物多样性分析。
```

## 4 生物多样性数据的可视化及汇总

### 4.1 物种丰富度和多样性

每种生境上的物种数是否相同?

```
# 比较 fescue 和 mixedgrass 生境中物种数的差异
boxplot(specnumber(comm) ~ metadata$habitat, ylab = "# of species")
```

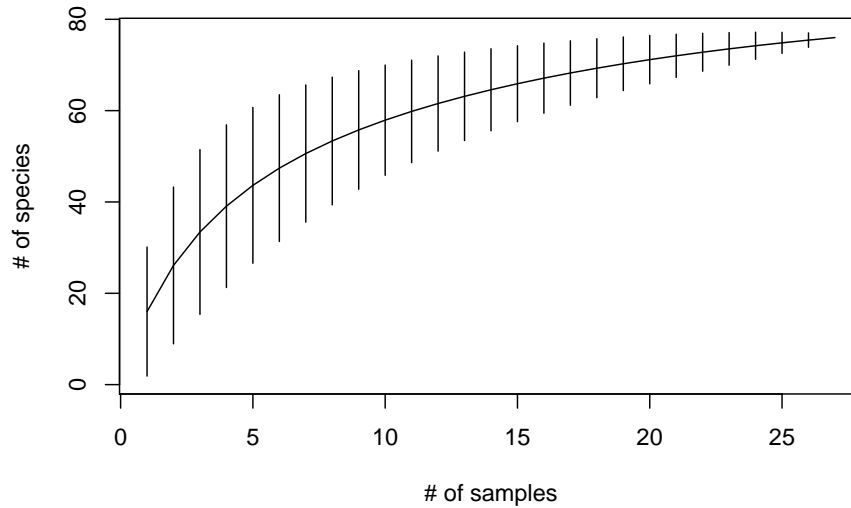


```
# 用 t 检验比较不同生境中物种数是否存在差异
t.test(specnumber(comm) ~ metadata$habitat)

##
## Welch Two Sample t-test
##
## data: specnumber(comm) by metadata$habitat
## t = 5.1371, df = 17.179, p-value = 7.972e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  6.274287 15.008066
## sample estimates:
##      mean in group Fescue mean in group Mixedgrass
##           22.70000           12.05882
```

```
# 物种累计曲线，用于查看样方总面积是否足够大
```

```
plot(specaccum(comm), xlab = "# of samples", ylab = "# of species")
```



```
# 群落多元分析
```

不同样地植物群落物种组成是怎样变化的？生境类型和环境变量与群落植物组成有什么关系？

要研究这些问题，需要用到多元排序方法。相关函数在 `vegan` 程序包中，相关的函数的帮助文件和使用指南也非常详尽。也可以参考 Borcard 等人编写的 *Numerical Ecology in R*，这本书在 2018 年出的第二版。

## 4.2 等级聚类

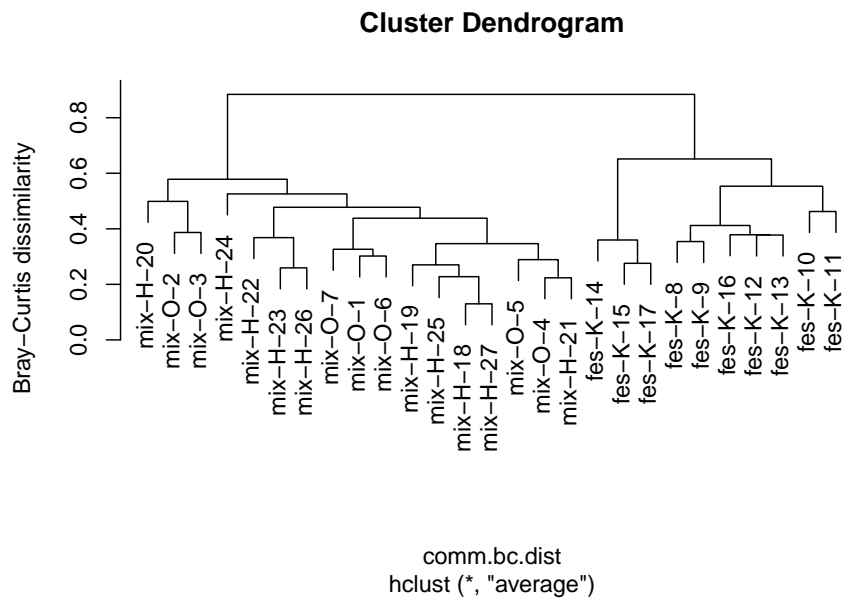
通过聚类，物种组成相近的群落能聚合到一起。先计算样方之间的 Bray-Curtis 距离，Bray-Curtis 不但考虑了物种组成的差异，同时考虑了每个种的多度。之后，用聚合等级聚类 (agglomerative hierarchical clustering algorithm) 算法即可实现聚类。

```
# 计算样地之间的 Bray-Curtis 距离
```

```
comm.bc.dist <- vegdist(comm, method = "bray")
```

```
# 用 UPGMA 方法聚类
comm.bc.clust <- hclust(comm.bc.dist, method = "average")

# 绘制聚类图
plot(comm.bc.clust, ylab = "Bray-Curtis dissimilarity")
```



两种类型的草地，由于有不同的物种组成，分别聚成两类

### 4.3 排序

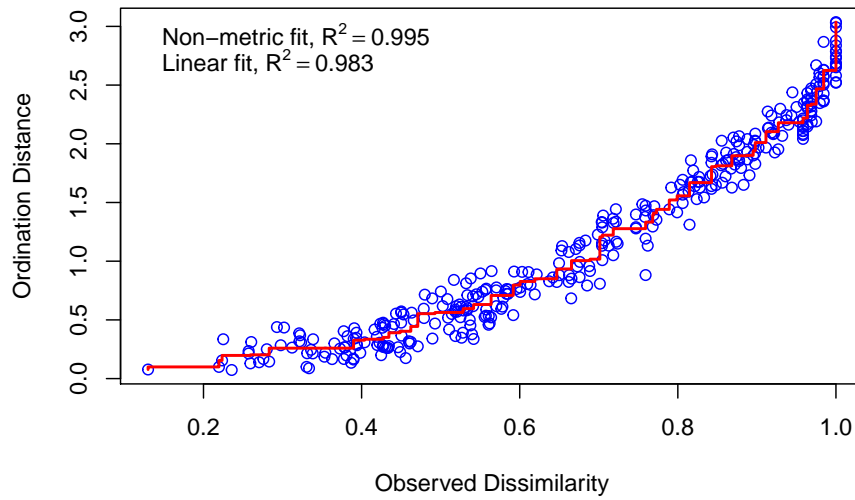
R 中的排序方法比较多，这里我们采用 NMDS（non-metric multidimensional scaling）方法。NMDS 的方法是基于距离矩阵的。

```
# metaMDS 函数会自动进行数据转换，并检查结果是否已经收敛。
comm.bc.mds <- metaMDS(comm, dist = "bray")
```

```
## Run 0 stress 0.07174232
## Run 1 stress 0.07174011
## ... New best solution
```

```
## ... Procrustes: rmse 0.0009381762  max resid 0.004274062
## ... Similar to previous best
## Run 2 stress 0.07344512
## Run 3 stress 0.089212
## Run 4 stress 0.07909591
## Run 5 stress 0.08520145
## Run 6 stress 0.07480965
## Run 7 stress 0.08502289
## Run 8 stress 0.08647953
## Run 9 stress 0.07886575
## Run 10 stress 0.07886618
## Run 11 stress 0.07480941
## Run 12 stress 0.07914226
## Run 13 stress 0.07892741
## Run 14 stress 0.07941515
## Run 15 stress 0.08562535
## Run 16 stress 0.08386448
## Run 17 stress 0.07173981
## ... New best solution
## ... Procrustes: rmse 0.0001514004  max resid 0.0006720928
## ... Similar to previous best
## Run 18 stress 0.08006783
## Run 19 stress 0.08174297
## Run 20 stress 0.07344492
## *** Solution reached
```

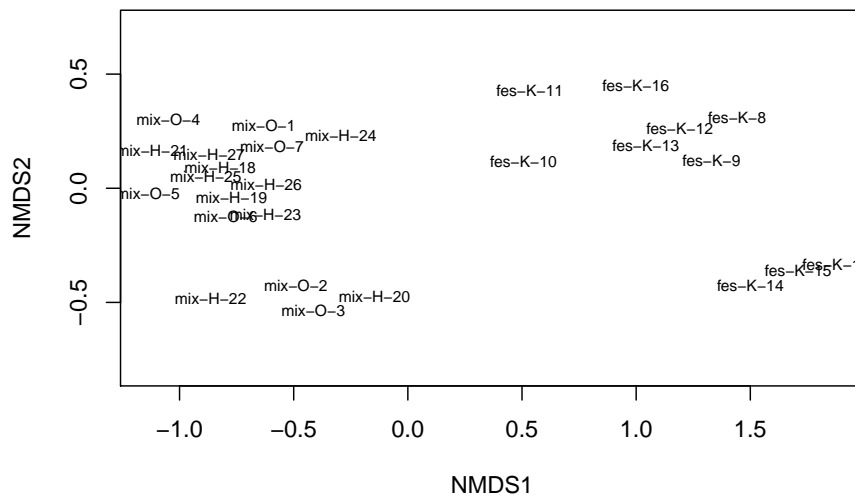
```
# 绘制 stressplot, 检查排序结果
stressplot(comm.bc.mds)
```



排序结果可通过多种方式展示

# 在 *Biplot* 中显示样地

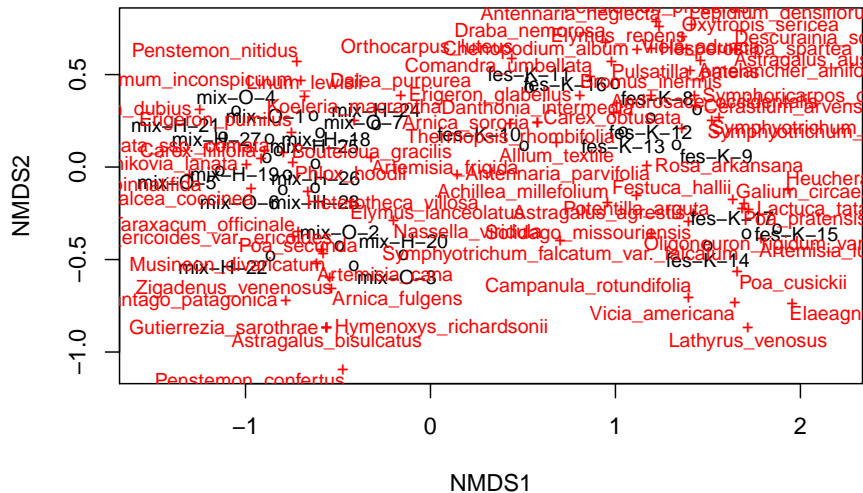
```
ordiplot(comm.bc.mds, display = "sites", type = "text")
```





```
# 添加物种名
```

```
ordipointlabel(comm.bc.mds)
```



```
# ordiplot 可以只打开绘图设备和画布，不绘制任何内容
```

```
mds.fig <- ordiplot(comm.bc.mds, type = "none")
```

```
# 只添加样地，颜色为样地类型，pch 这里设定的是点的形状。
```

```
# 试试 plot(1:30, pch = 1:30)
```

```
points(mds.fig, "sites", pch = 19, col = "green",
```

```
      select = metadata$habitat == "Fescue")
```

```
points(mds.fig, "sites", pch = 19, col = "blue",
```

```
      select = metadata$habitat == "Mixedgrass")
```

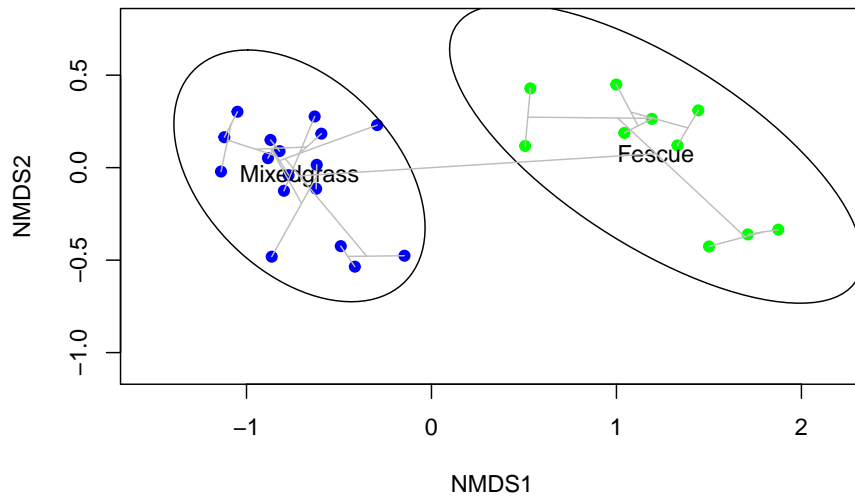
```
# 添加生境置信椭圆
```

```
ordiellipse(comm.bc.mds, metadata$habitat,
```

```
            conf = 0.95, label = TRUE)
```

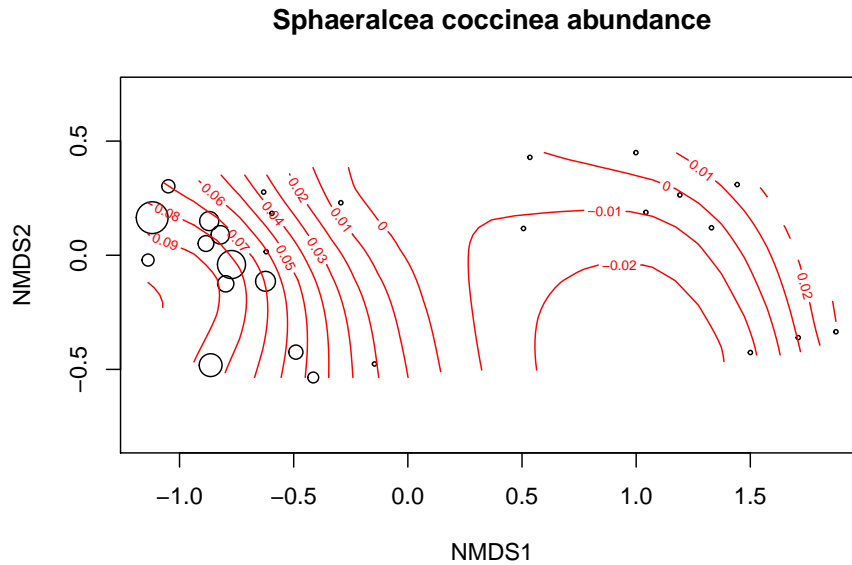
```
# 将聚类结果在 NMDS 排序图中显示
```

```
ordicluster(comm.bc.mds, comm.bc.clust, col = "gray")
```



也可以用 `ordisurf` 函数显示物种多度

```
# 显示球藻属 Sphaeralcea 的多度. cex 参数可以调整点的大小
ordisurf(comm.bc.mds, comm[, "Sphaeralcea_coccinea"],
         bubble = TRUE, main = "Sphaeralcea coccinea abundance",
         cex = 3)
```

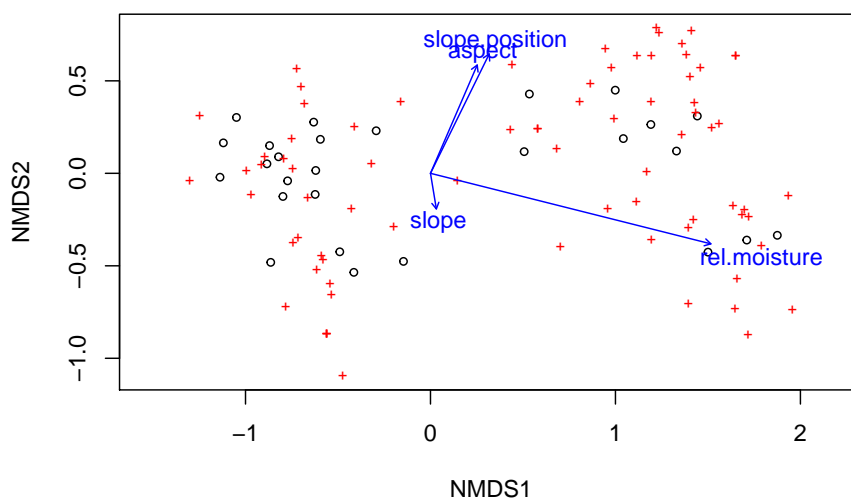


```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## y ~ s(x1, x2, k = 10, bs = "tp", fx = FALSE)
##
## Estimated degrees of freedom:
## 4.3 total = 5.3
##
## REML score: -43.60779
```

#### 4.4 添加环境信息和性状数据到排序中

环境变量与排序轴之间是怎样的关系？这里使用 `envfit` 函数拟合

```
ordiplot(comm.bc.mds)
plot(envfit(comm.bc.mds, metadata[, 3:6]))
```



用 `vegan` 中的 `cca` 或者 `rda` 等排序方法也是可以的。`cca` 或者 `rda` 直接考虑了物种组成和多度，并不用先计算样方之间的距离。

## 5 性状与进化树

### 5.1 系统发育信号

最近，系统发育保守性受到很多人关注。群落系统发育结构分析就假设物种的性状是系统发育保守的。

系统发育信号是度量不同物种的性状在进化树上相似程度的指数。Blomberg K 统计量就是比较性状的观察值与布朗运动模型预测值的统计量 (Blomberg et al. 2003)。K 值接近 1 表明性状进化过程接近布朗运动，表明有一定程度的系统发育信号或者呈现一定的保守性。K 接近于 0 表明性状进化倾向于随机， $K > 1$  表明性状保守。

K 的显著性检验可以用以下方法：将进化树上的物种名随机打乱，计算每次打乱物种名后，性状的系统发育独立差 (Phylogenetic independent contrast) 的方差，在多次打乱物种名生成零分布后，真实进化树的系统发育独立差方差与之进行比较。相关检验在 `Kcalc`, `phylosignal` 和 `multiPhylosignal` 函数中可实现。

下面用练习数据计算系统发育信号

```
# 可以用 Kcalc 依次计算 traits 数据框的所有列
apply(traits, 2, Kcalc, phy)
```

```
##           SLA      LeafArea LeafThickness      SLV LeafTissueDens
## 0.2563107 0.4231067 0.2418525 0.3310724 0.3298808
##           SRL      SRV RootTissueDens      RootDiam
## 0.2290093 0.2698531 0.2544904 0.3150881
```

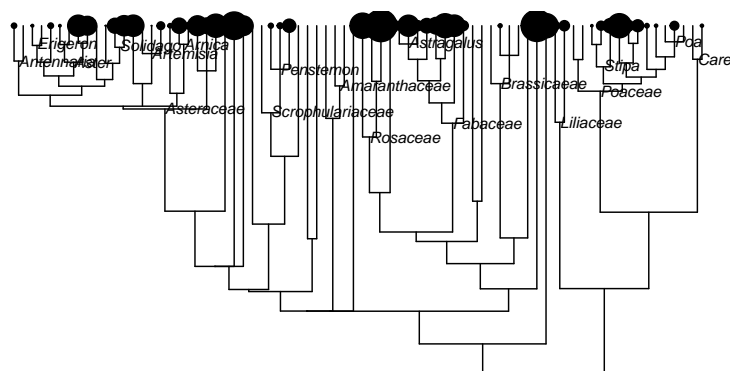
```
# multiPhylosignal 函数可以检验多列的 P 值,
# 但该函数需要的进化树为严格的二分叉树
multiPhylosignal(traits, multi2di(phy))
```

```
##           K PIC.variance.obs PIC.variance.rnd.mean
## SLA      0.2563107 0.0006068179 0.0008294584
## LeafArea 0.4231067 0.0055507994 0.0131253188
## LeafThickness 0.2418525 0.0006669402 0.0008842234
## SLV      0.3310724 0.0005728618 0.0010652977
## LeafTissueDens 0.3298808 0.0006688721 0.0012448133
## SRL      0.2290093 0.0028408725 0.0036614003
## SRV      0.2698531 0.0011064783 0.0016719717
## RootTissueDens 0.2544904 0.0010609020 0.0015168198
## RootDiam 0.3150881 0.0005510416 0.0009755421
##           PIC.variance.P PIC.variance.Z
## SLA      0.042 -1.614821
## LeafArea 0.001 -4.079502
## LeafThickness 0.084 -1.324053
## SLV      0.001 -3.270624
## LeafTissueDens 0.001 -2.969950
## SRL      0.074 -1.332157
## SRV      0.009 -2.210198
## RootTissueDens 0.019 -1.958894
## RootDiam 0.001 -2.993733
```

结果中包括 K 和 PIC.variance.P, 后者表示性状的非随机程度。大部分变量都完全随机表现出更强的系统发育信号。

## 5.2 性状进化的可视化

```
plot(phy, direction = "up", show.tip.label = FALSE,
     show.node.label = TRUE, cex = 0.7)
tiplabels(pch = 19, col = "black",
          cex = 3 * (traits[, "LeafArea"] / max(traits[, "LeafArea"])))
```



## 性状相关性的系统发育分析若功能性状表现出较强的系统发育信号，就违背了数据完全独立的假设。此时可以在 gls 中设定 corBrownian 关联矩阵，考虑物种之间的系统发育关系。（注：在分析过程中，可以使用 caper 的 pgls 回归。）

其中 gls 是一般最小二乘 (Generalised least squares) 法的简称，方法类似 ANOVA 或线性模型。无序类别变量以及连续变量都可以通过这种方式检验。

下面检验检测一下，考虑物种之间系统发育关系时，specific root length (SRL) 和 root tissue density 之间的关系。

```
# 普通的 gls 模型, 不考虑系统发育关系
```

```
root.gls <- gls(RootTissueDens ~ SRL, data = traits)
anova(root.gls)
```

```
## Denom. DF: 74
```

```
##           numDF  F-value p-value
## (Intercept)      1 611.0773 <.0001
## SRL              1   3.0592 0.0844
```

```
# 考虑系统发育关系
```

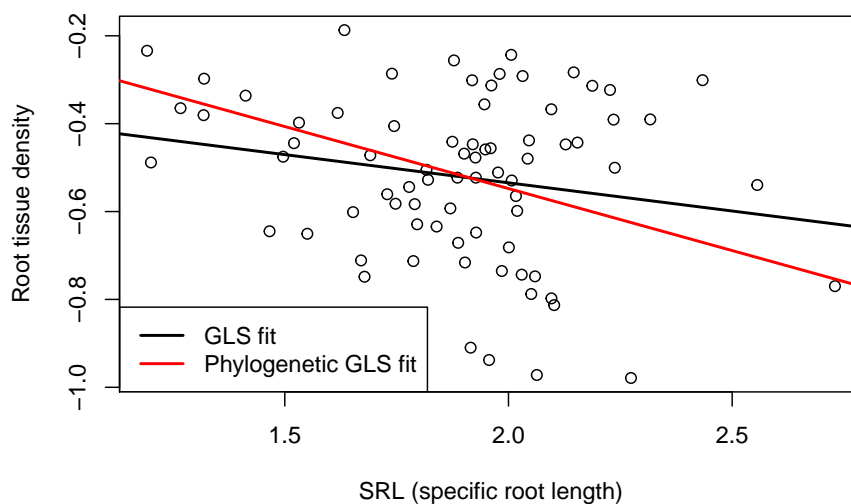
```
root.pgls <- gls(RootTissueDens ~ SRL,
                 correlation = corBrownian(value = 1, phy),
                 data = traits)
anova(root.pgls)
```

```
## Denom. DF: 74
```

```
##           numDF  F-value p-value
## (Intercept)      1 13.42163 5e-04
## SRL              1 20.07297 <.0001
```

```
# 绘图
```

```
plot(RootTissueDens ~ SRL, data = traits,
     xlab = "SRL (specific root length)",
     ylab = "Root tissue density")
abline(coef(root.gls), lwd = 2, col = "black")
abline(coef(root.pgls), lwd = 2, col = "red")
legend("bottomleft",
      legend = c("GLS fit", "Phylogenetic GLS fit"),
      lwd = 2, col = c("black", "red"))
```



当不考虑系统发育关系时，SRL 和 root tissue density 的关系很弱。考虑系统发育信号后，两者的相关性就明显了。

## 6 系统发育与性状多样性

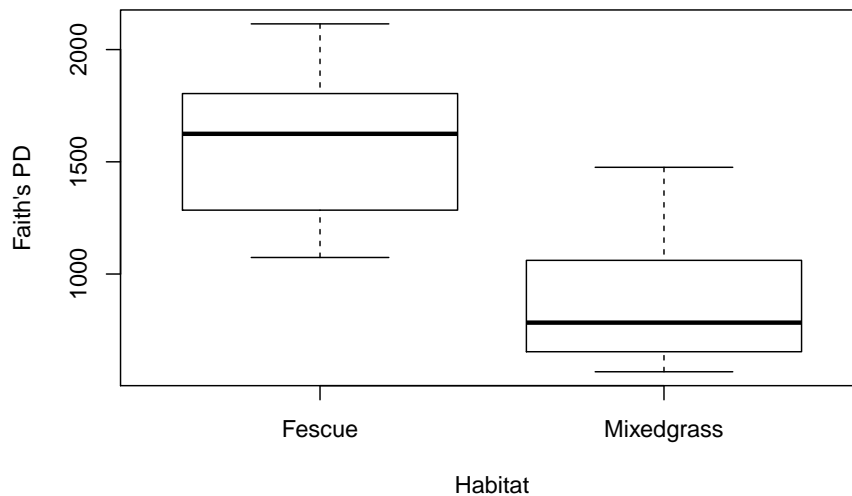
### 6.1 系统发育多样性

```
# 计算 Faith's PD
comm.pd <- pd(comm, phy)
head(comm.pd)
```

```
##              PD SR
## mix-0-1 1072.3697 16
## mix-0-2 1475.4767 22
## mix-0-3 1406.1708 21
## mix-0-4  564.5899  6
## mix-0-5  783.4028 10
## mix-0-6 1028.5796 13
```



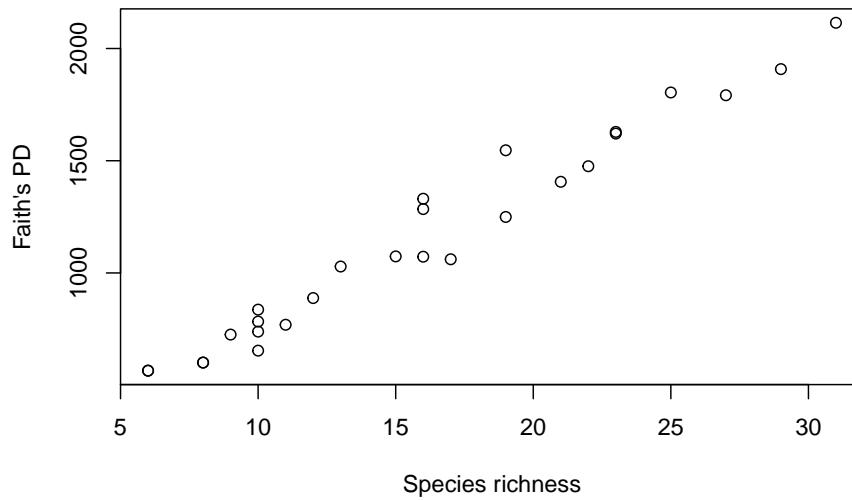
```
# 不同生境的 Faith's PD
boxplot(comm.pd$PD ~ metadata$habitat,
        xlab = "Habitat", ylab = "Faith's PD")
```



```
# 不同生境 PD 是否相同
t.test(comm.pd$PD ~ metadata$habitat)

##
## Welch Two Sample t-test
##
## data: comm.pd$PD by metadata$habitat
## t = 5.7161, df = 17.627, p-value = 2.195e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 451.1886 976.8551
## sample estimates:
## mean in group Fescue mean in group Mixedgrass
## 1602.2371 888.2153
```

```
# 查看 PD 和物种丰富度的关系
plot(comm.pd$PD ~ comm.pd$SR,
      xlab = "Species richness", ylab = "Faith's PD")
```



picante 可以用来计算  $MPD$ ,  $MNTD$ ,  $SES_{MPD}$  and  $SES_{MNTD}$  具体定义, 请参见 Phylocom 的说明书。

群落系统发育最重要指数之一就是标准效应值。公式如下  $SES_{metric} = \frac{Metric_{observed} - \text{mean}(Metric_{null})}{sd(Metric_{null})}$

一般包括:  $SES_{MPD}$  等于 -NRI,  $SES_{MNTD}$  等于 -NTI

picante 中提供了不同的零模型, 用来进行随机化, 以计算 SES。

```
# 系统发育距离矩阵
phy.dist <- cophenetic(phy)

# ses.mpd
comm.sesmpd <- ses.mpd(comm, phy.dist, null.model = "richness",
                       abundance.weighted = FALSE,
                       runs = 999)

head(comm.sesmpd)
```

```
##          ntaxa  mpd.obs mpd.rand.mean mpd.rand.sd mpd.obs.rank  mpd.obs.z
## mix-0-1      16 231.3054      237.7664   11.990722      255 -0.53883026
## mix-0-2      22 239.5479      237.8838    9.325874      517  0.17843002
## mix-0-3      21 236.5260      237.4709    9.375299      403 -0.10077982
## mix-0-4       6 222.5255      239.9238   25.108889      184 -0.69291631
## mix-0-5      10 234.2013      239.5197   15.717315      294 -0.33837344
## mix-0-6      13 239.4120      238.6373   14.049114      447  0.05513748
##          mpd.obs.p runs
## mix-0-1      0.255  999
## mix-0-2      0.517  999
## mix-0-3      0.403  999
## mix-0-4      0.184  999
## mix-0-5      0.294  999
## mix-0-6      0.447  999
```

```
# ses.mntd
comm.sesmntd <- ses.mntd(comm, phy.dist, null.model = "richness",
                        abundance.weighted = FALSE,
                        runs = 999)
head(comm.sesmntd)
```

```
##          ntaxa  mntd.obs mntd.rand.mean mntd.rand.sd mntd.obs.rank
## mix-0-1      16  94.98812      103.91986    19.07922      316
## mix-0-2      22  97.41972      93.34697    14.50653      615
## mix-0-3      21  98.71519      94.68340    15.56810      603
## mix-0-4       6 136.86094      154.65702    42.63607      340
## mix-0-5      10 107.36711      124.66227    28.94193      296
## mix-0-6      13 118.94915      112.83012    23.37778      609
##          mntd.obs.z mntd.obs.p runs
## mix-0-1 -0.4681396      0.316  999
## mix-0-2  0.2807528      0.615  999
## mix-0-3  0.2589780      0.603  999
## mix-0-4 -0.4173950      0.340  999
## mix-0-5 -0.5975813      0.296  999
```

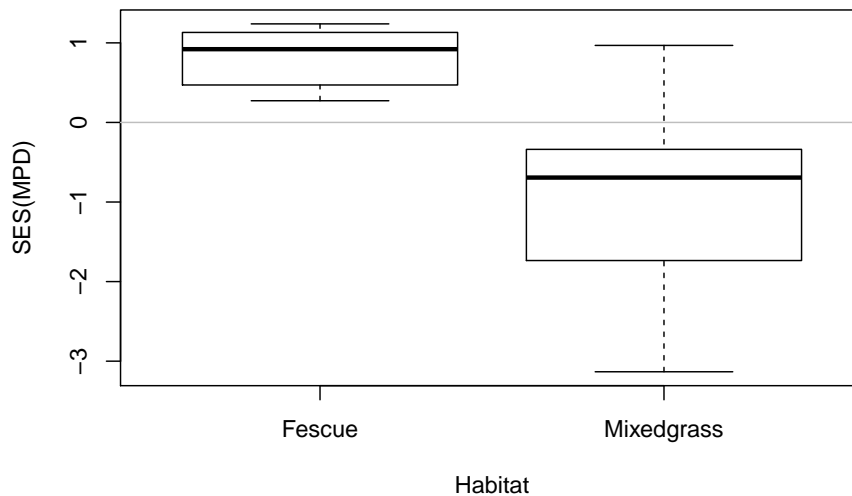
```
## mix-0-6 0.2617455 0.609 999
```

输出结果 `ntaxa` - 物种丰富度 `mpd.obs` - 群落中 `mpd` 观察值  
`mpd.rand.mean` - 零群落 `mpd` 平均值 `mpd.rand.sd` - 零模型 `mpd` 的标准差  
`mpd.obs.rank` - 观察值在零模型中的顺序 `rank` (秩) `mpd.obs.z` - `mpd` 标准效应指数 (等于  $-NRI$ ) `mpd.obs.p` - P-value (quantile) of observed `mpd` vs. null communities ( $= \text{mpd.obs.rank} / \text{runs} + 1$ )  
`runs` - 随机化次数

$SES.mpd > 0$ , 且  $mpd.obs.p > 0.95$  时, 表示群落系统发育均匀。  
 $SES.mpd < 0$ , 且  $mpd.obs.p < 0.05$  时, 表明与零模型相比系统发育聚集。

一般认为,MPD 对对系统发育整体的聚集性或均匀性较为敏感,MNTD 对靠近进化树末端的均匀性和聚集性更为敏感。

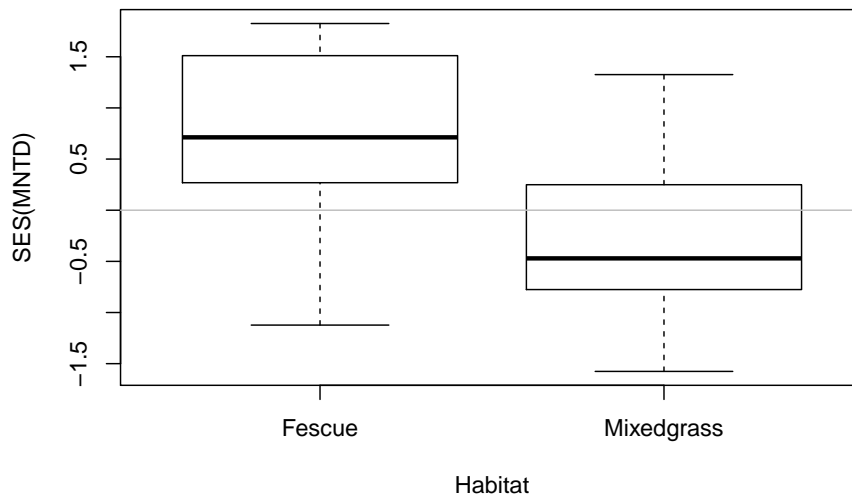
```
# 不同生境的 ses.mpd
plot(comm.sesmpd$mpd.obs.z ~ metadata$habitat,
      xlab = "Habitat", ylab = "SES(MPD)")
abline(h = 0, col = "gray")
```



```
t.test(comm.sesmpd$mpd.obs.z ~ metadata$habitat)

##
## Welch Two Sample t-test
##
## data:  comm.sesmpd$mpd.obs.z by metadata$habitat
## t = 5.8672, df = 20.206, p-value = 9.317e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.175249 2.470609
## sample estimates:
##      mean in group Fescue mean in group Mixedgrass
##                0.8187546                -1.0041742

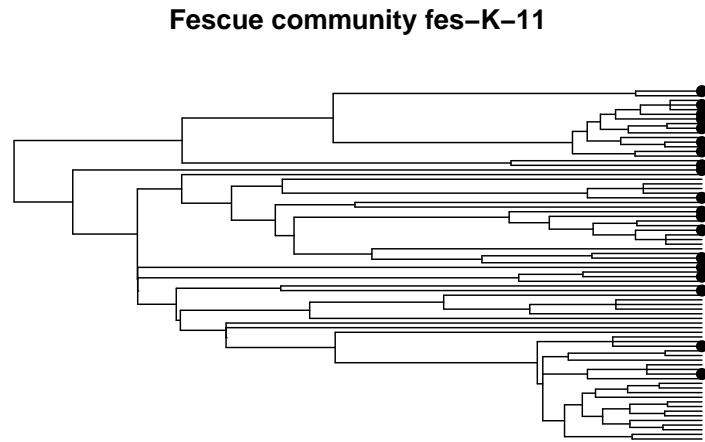
# 不同生境的 ses.mntd
plot(comm.sesmntd$mntd.obs.z ~ metadata$habitat,
      xlab = "Habitat", ylab = "SES(MNTD)")
abline(h = 0, col = "gray")
```



```
t.test(comm.sesmntd$mntd.obs.z ~ metadata$habitat)

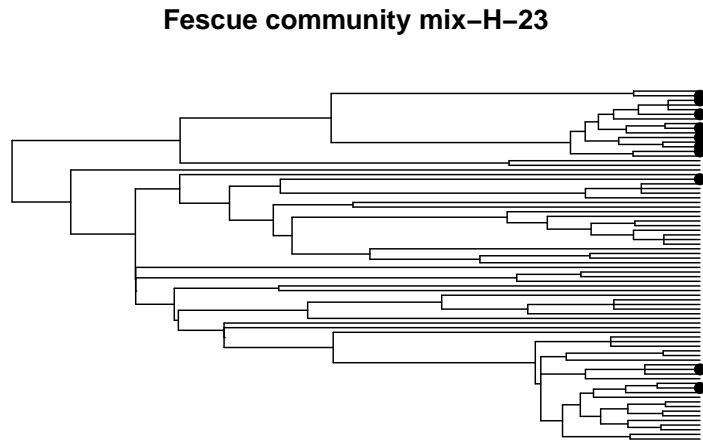
##
## Welch Two Sample t-test
##
## data:  comm.sesmntd$mntd.obs.z by metadata$habitat
## t = 2.7327, df = 14.223, p-value = 0.016
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2091817 1.7251809
## sample estimates:
##      mean in group Fescue mean in group Mixedgrass
##              0.5972522              -0.3699292
```

```
# 某 fescue 群落中的物种
plot(phy, show.tip.label = FALSE,
     main = "Fescue community fes-K-11")
tiplabels(tip = which(phy$tip.label %in%
                      colnames(comm)[comm["fes-K-11", ] > 0]),
          pch = 19)
```



‘mix-H-23’ Mixedgrass 群落中包含了一系列系统发育聚集的种。

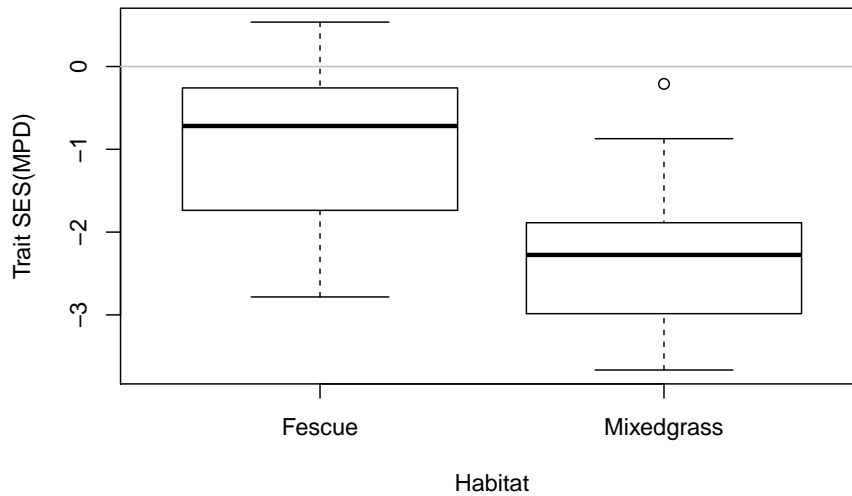
```
# 某 mixedgrass 群落中出现的个体
plot(phy, show.tip.label = FALSE,
     main = "Fescue community mix-H-23")
tiplabels(tip = which(phy$tip.label %in%
                      colnames(comm)[comm["mix-H-23", ] >
                      0]), pch = 19)
```



## 性状多样性性状的分析方法与进化树分析的方法类似，一般都是先进行标准化，再计算欧式距离，再计算功能性状相似性，计算 MPD 或者 MNTD

```
# 性状先用 scale 标准化，再用 dist 计算欧式距离计算
trait.dist <- as.matrix(dist(scale(trait), method = "euclidean"))
comm.sesmpd.traits <- ses.mpd(comm, trait.dist, null.model = "richness",
                             abundance.weighted = FALSE, runs = 999)
plot(comm.sesmpd.traits$mpd.obs.z ~ metadata$habitat,
     xlab = "Habitat", ylab = "Trait SES(MPD)")
abline(h = 0, col = "gray")
```





vegan 中的 `treedive` 函数与 `picante` 中的 `pd` 类似。

## 6.2 系统发育 beta 多样性

`unifrac` 函数和 `phylosor` 函数, 都是计算样方之间的 Faith's PD。`comdist` 和 `comdistnt` 的计算类似于 MPD 和 MNTD, 但是是针对样方之间的距离。计算过程中可以考虑多度, 由于之前我们计算的 Bray-Curtis 距离已经考虑了多度。为了方便比较, 后续的计算也考虑多度, 即 `abundance.weighted = TRUE`

```
# 系统发育 beta 多样性
comm.mntd.dist <- comdistnt(comm, phy.dist,
                             abundance.weighted = TRUE)

# 功能性状 beta 多样性
comm.mntd.traits.dist <- comdistnt(comm, trait.dist,
                                    abundance.weighted = TRUE)

# 计算两个距离矩阵的相关性, 用 mantel 检验
# 此处计算 Bray-Curtis 距离和 mntd 距离的相关性
```

```
mantel(comm.bc.dist, comm.mntd.dist)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = comm.bc.dist, ydis = comm.mntd.dist)
##
## Mantel statistic r: 0.8597
##      Significance: 0.001
##
## Upper quantiles of permutations (null model):
##   90%   95%  97.5%   99%
## 0.0749 0.1099 0.1363 0.1665
## Permutation: free
## Number of permutations: 999
```

```
# 此处计算 Bray-Curtis 距离和功能性状距离的相关性
```

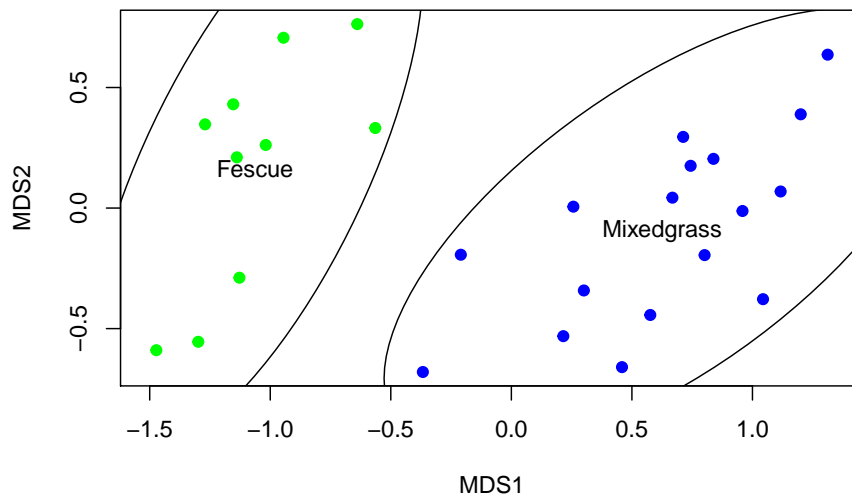
```
mantel(comm.bc.dist, comm.mntd.traits.dist)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = comm.bc.dist, ydis = comm.mntd.traits.dist)
##
## Mantel statistic r: 0.9524
##      Significance: 0.001
##
## Upper quantiles of permutations (null model):
##   90%   95%  97.5%   99%
## 0.0773 0.1081 0.1462 0.1969
## Permutation: free
## Number of permutations: 999
```

### 6.3 系统发育和功能性状排序

之前，我们用 Bray-Curtis 进行了 NDMS 排序，实际上，系统发育 MNTD 以及功能性状 MNTD 都可以用类似的方法排序，以展示样地之间的关系

```
# 由于只有样地之间的距离，这里用 monoMDS.  
# 具体请认真阅读 monoMDS 的说明书。  
  
# 系统发育距离  
comm.mntd.mds <- monoMDS(comm.mntd.dist)  
mds.fig <- ordiplot(comm.mntd.mds, type = "none")  
  
## species scores not available  
  
points(mds.fig, "sites", pch = 19, col = "green",  
       select = metadata$habitat == "Fescue")  
points(mds.fig, "sites", pch = 19, col = "blue",  
       select = metadata$habitat == "Mixedgrass")  
ordiellipse(comm.mntd.mds, metadata$habitat,  
             conf = 0.95, label = TRUE)
```

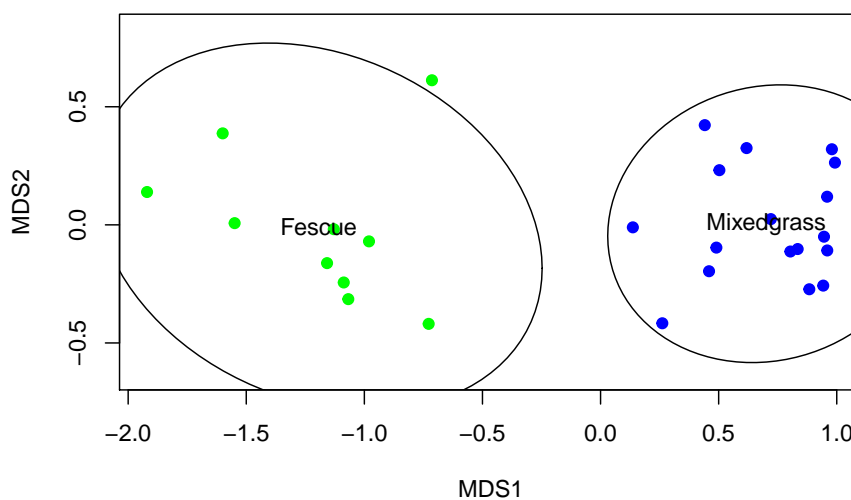


```
# 功能性状
```

```
comm.mntd.traits.mds <- monoMDS(comm.mntd.traits.dist)
mds.fig <- ordiplot(comm.mntd.traits.mds, type = "none")
```

```
## species scores not available
```

```
points(mds.fig, "sites", pch = 19, col = "green",
       select = metadata$habitat == "Fescue")
points(mds.fig, "sites", pch = 19, col = "blue",
       select = metadata$habitat == "Mixedgrass")
ordiellipse(comm.mntd.traits.mds, metadata$habitat,
            conf = 0.95, label = TRUE)
```



无论是对物种 beta 多样性，系统发育 beta 多样性还是功能性状 beta 多样性进行排序方式，两种生境类型的差别都比较大。

## 解释 beta 多样性

为了解释哪些因子对样方之间的系统发育关系有影响，可以用 permutational MANOVA (adonis) 计算，R 代码如下。输出结果类似 ANOVA 方差分析。

以下分别检验生境对于 Bray-Curtis 距离、系统发育距离、功能性状距离的解释能力。

```
# Bray-Curtis 距离
```

```
adonis(comm.bc.dist ~ habitat, data = metadata)
```

```
##
```

```
## Call:
```

```
## adonis(formula = comm.bc.dist ~ habitat, data = metadata)
```

```
##
```

```
## Permutation: free
```

```
## Number of permutations: 999
```

```
##
```

```
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## habitat     1     3.3656   3.3656  25.124 0.50123  0.001 ***
## Residuals  25     3.3490   0.1340          0.49877
## Total      26     6.7146          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 系统发育多样性距离
```

```
adonis(comm.mntd.dist ~ habitat, data = metadata)
```

```
##
## Call:
## adonis(formula = comm.mntd.dist ~ habitat, data = metadata)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## habitat     1    22814 22813.7  34.904 0.58267  0.001 ***
## Residuals  25    16340   653.6          0.41733
## Total      26    39154          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# 功能性状距离
```

```
adonis(comm.mntd.traits.dist ~ habitat, data = metadata)
```

```
##
## Call:
## adonis(formula = comm.mntd.traits.dist ~ habitat, data = metadata)
```

```
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##           Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## habitat     1   10.9154 10.9154  63.959 0.71897  0.001 ***
## Residuals  25    4.2666  0.1707          0.28103
## Total      26   15.1820          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```