



# 群落系统发育数据分析研讨班：第二天

## 群落数据处理和进化树的整合分析

张金龙

西双版纳

2014年1月16日

# 目录

- 1 R中的对象及操作
- 2 R与Phylogenetics
- 3 群落的alpha多样性, beta多样性和解释
- 4 空间分析
- 5 群落系统发育

# 目录

- 1 R中的对象及操作
- 2 R与Phylogenetics
- 3 群落的alpha多样性, beta多样性和解释
- 4 空间分析
- 5 群落系统发育

# R作为计算器

R可以作为计算器使用，输入数字，进行加减乘除运算

R code

```
2 + 2  
a <- 2
```

赋值

<-, =, or even ->

R code

```
b <- 2  
c <- a+b  
c
```

# 注释

# R中的对象

- 基本数据类型

Numeric数值型 100, 0, -4.335

Character字符串型 "China"

Logical逻辑型 TRUE, FALSE

Factor因子型 Different levels

Complex复数型  $2 + 3i$

- 基本对象类型

Vector向量 `c(1,2,3); c("a","a","b","b","c")`

Matrix矩阵 `x <- 1:12 ; dim(x) <- c(3,4)`

Dataframe数据框 由等长但是类型不同的向量，组成的数据结构

List列表 列表类似容器，可以装如其他类型的对象

Array数组 3维的数组，称为高维数组

# 变量命名

(1)变量名不能以数字开头，必须用英文字母开头，但是可以用数字结尾。

(2)R区分大小写，因此Data和data两个对象名是不同的。

(3)在为对象起名的时候，不能使用保留字，保留字主要包括：

`if, else, repeat`

`while, function, for`

`in, next, break`

`TRUE, FALSE, NULL`

`Inf, NaN, NA`

保留字主要涉及程序流程控制和特殊的变量类型等。

# 运算符Operators

## 数学运算

加：+，减：-，乘：\*，除：/，幂运算： $\wedge$

## 判断或比较

大于：>，小于：<，小于等于： $\leq$ ，大于等于： $\geq$ ，等于：==，  
不等于：!=

## 逻辑运算：与或非

否：!，与：&，与：&&，或：|，或：||

# 函数调用

函数的调用方法:函数名(对象,参数1,参数2, ...)  
调用函数时, 要注意函数中参数的默认值

## R code

```
x <- c(23, 14, 4, 5, 8, 2, 10)
sort(x)
sort(x, decreasing = TRUE)
```

一般在调用函数前, 一定要查看函数的帮助文件, 了解R函数的参数类型, 返回结果的类型, 并运行Example



# 向量的创建

## 字符型

```
character<-c("China", "Korea", "Japan", "UK", "USA")
```

## 数值型

```
numeric<-c(1, 3, 6, 7, 3, 8, 6, 4)
```

## 逻辑型

```
logical<-c(TRUE, FALSE, TRUE, FALSE)
```

## 创建向量的函数

`c(2,5,6,9)` 将两个以上向量或向量元素连接在一起

`rep(2,times=4)` 将某个数值或字符串重复若干次

`seq(from=3, to=21, by=3)` 给定步长, 生成从一个值到另一个值的序列

`1:15` 以1为步长, 生成一个向量

# 向量的常用函数

<code>sum()</code>	求和
<code>cumsum()</code>	累计求和
<code>min()</code>	最小值
<code>mean()</code>	平均值
<code>max()</code>	最大值
<code>range()</code>	极差
<code>median()</code>	中位数
<code>sd()</code>	标准差
<code>cor()</code>	两个向量的相关性
<code>length()</code>	向量的元素个数
<code>intersect()</code>	向量的交集
<code>setdiff()</code>	向量的相异元素
<code>%in%</code>	前者的各元素是否在后者中。
<code>names()</code>	向量的各元素名称。

# 创建矩阵

生成矩阵的函数`dim()`和`matrix()`  
`dim()` 定义矩阵的行列数，例如：

```
x <- 1:12
```

```
dim(x) <- c(3,4)
```

	[,1]	[,2]	[,3]	[,4]
[1,]	1	4	7	10
[2,]	2	5	8	11
[3,]	3	6	9	12

```
matrix.x <- matrix(1:12, nrow=3, byrow=TRUE)
```

# 数据框

数据框类似于数据表，但是各列的类型不同。

	plotname	species	abundance
1	plot1	sp1	3
2	plot1	sp2	6
3	plot1	sp3	1
4	plot1	sp4	2
5	plot1	sp5	1
6	plot2	sp1	8
7	plot2	sp3	30
8	plot3	sp4	2
9	plot3	sp2	1
10	plot3	sp6	1
11	plot3	sp7	3

# 数据框的创建和操作

创建数据框的函数:

<code>cbind()</code>	按列组合成数据框
<code>rbind()</code>	按行组合成数据框
<code>data.frame()</code>	生成数据框
<code>as.data.frame()</code>	转换成dataframe
<code>head()</code>	默认访问数据的前6行
<code>rownames()</code>	取行名或命名
<code>colnames()</code>	取列名或命名

# list

`list(character,numeric,logical,matrix.x)`

list类似百货公司，可以存放任何类型的对象，甚至包括List本身。

创建list()，将对象放入括号内即可。

list()若不放任何对象，则创建一个空白的list。

# 下标和索引

<code>[]</code>	引用向量内的元素
<code>[m,n]</code>	引用数据框或矩阵的第m行，第n列
<code>which()</code>	查询向量内哪个元素
<code>match()</code>	匹配两个向量
<code>unique()</code>	去重复

# 下标和索引

对于dataframe,

如果提取某一行或者列, 也可以用方括号[,]用法和矩阵相同。

\$ 可以用来提取特定名称的列。

List对象的下标和提取

\$ 访问list内某一个名称的对象

[[]] 访问list内第几个对象



# 数据读取

`read.table()` 读取tab间隔的txt文件

`read.csv()` 读取csv文件

`readLines()` 读取文本文件，

并将每一行作为一个字符串，返回值为多个字符串组成的向量

注意：不推荐用R直接读取Excel文件

读取Excel数据()

`library(RODBC)`

如果是Excel 2003

```
temp <- odbcConnectExcel ("test.xls")
```

如果是2007

```
temp <- odbcConnectExcel2007 ("test.xlsx")
```

```
dat <- sqlFetch(temp, "Sheet1")
```

```
close( temp )
```

## R Formula: 向量或者矩阵之间的基本关系

- + 各变量独立作用，不考虑交互
- : 只考虑变量间的交互作用
- \* 考虑变量的独立作用和交互作用
- 去掉某一部分
- I() 括号中可放数学表达式

# R公式举例

<b>a+b</b>	a 和b独立的作用
<b>a:b</b>	a 和b 的交互作用
<b>a*b</b>	相加和交互作用(等价于a+b+a:b)
<b>poly(a, n)</b>	a的n价多项式
<b>y~x-1</b>	表示过原点的线性回归
<b>y~1</b>	拟合截距

注意：不要在公式内进行数据下标操作，更不要用中文为变量命名，而让中文出现在公式中。这些都将导致不能预知的错误。

# 目录

- 1 R中的对象及操作
- 2 R与Phylogenetics**
- 3 群落的alpha多样性, beta多样性和解释
- 4 空间分析
- 5 群落系统发育

# ape程序包

ape是R的一个程序包(<http://ape-package.ird.fr/>), 是Analysis of Phylogenetics and Evolution首字母的缩写, 作者是法国进化生物学家E. Paradis博士。

ape程序包的主要功能

- 进化树的读取, 绘制, 进化树内部数据的调用和修改。
- DNA序列的读取, 以及距离计算。
- 建立进化树
- 进行分子钟校正并估计进化速率
- 模拟生成随机进化树

详情参考: Paradis, E. (2012) Analysis of Phylogenetics and Evolution with R (Second Edition). New York: Springer.

## 进化树的格式：Newick进化树转换成List:

```
owls(((Strix_aluco:4.2,Asio_otus:4.2):3.1,Athene_noctua:7.3):6.3,  
Tyto_alba:13.5)
```

```
tree.owls <- read.tree("ex.tre")
```

Phylogenetic tree with 4 tips and 3 internal nodes.

Tip labels:

```
[1] "Strix_aluco" "Asio_otus" "Athene_noctua" "Tyto_alba"
```

Rooted; includes branch lengths.

## 进化树的格式：Newick进化树转换成List

```
str(tree.owls)
```

```
List of 4
```

```
$ edge : int [1:6, 1:2] 5 6 7 7 6 5 6 7 1 2 ...
```

```
$ Nnode : int 3
```

```
$ tip.label : chr [1:4] "Strix_aluco" "Asio_otus" "Athene_noctua"  
"Tyto_alba"
```

```
$ edge.length: num [1:6] 6.3 3.1 4.2 4.2 7.3 13.5
```

```
- attr(*, "class")= chr "phylo"
```

```
- attr(*, "order")= chr "cladewise"
```

# 序列之间的遗传距离dist.dna

两两之间的距离矩阵

a matrix of pairwise distances from DNA sequences using a model of DNA evolution. Eleven substitution models (and the raw distance) are currently available.

哪11个模型？

"raw", "N", "TS", "TV", "JC69",  
"K80" (the default), "F81", "K81", "F84",  
"BH87", "T92", "TN93", "GG95", "logdet",  
"paralin", "indel", or "indelblock".



## 制作Neighbour Joining Tree

Neighbour Joining方法制作一个进化树

woodmouse数据

15 DNA sequences in binary format stored in a matrix.

All sequences of same length: 965

Labels: No305 No304 No306 No0906S No0908S No0909S ...

Base composition:

a	c	g	t
0.307	0.261	0.126	0.306

```
data(woodmouse)
trw <- nj(dist.dna(woodmouse))
plot(trw)
```

## R不适宜用来建立进化树

由于R语言本身是脚本语言，输入的命令随时可以通过解释器转换为机器语言，随时可以得到结果。在这个过程中，涉及大量的内存空间的安全方面的处理。不可避免要耗费运算时间。

完成同样的任务，R花费的时间是C语言或者C++程序的一百倍或者更多。

进化树的构建中，涉及大量的优化算法，包括，迭代，随机数生成，启发式搜索，蒙特卡罗马尔科夫链等等，运行时间太久。因此用R开发构建进化树的软件并不明智。

解决办法：调用现有的C或者C++的函数库。

# Phylomatic根据现有的类群关系构建进化树

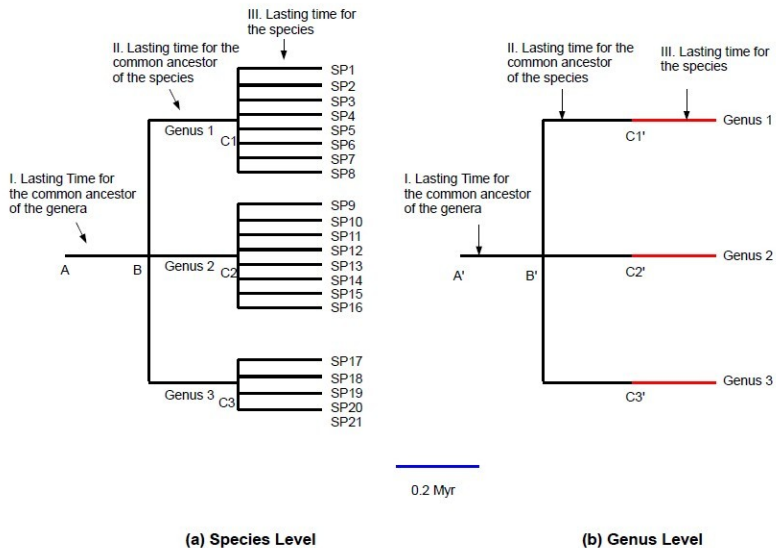
在DNA条形码无法获得时，人们希望通过植物名录，以及现有的各大类群的进化树，建立一个粗略的进化树。

2004年，美国哈佛大学阿诺德树木园的Campbell O. Webb博士发布了Phylomatic软件。用户只需要按照APG分类系统准备出科/属/种名录，Phylomatic软件就可以按照APG分类系统的科之间的关系，建立一个newick进化树。

进化时间，更是可以通过Webb博士提出的BLADJ算法，并根据Wikstroem对被子植物科的分化时间，以及更久远节点的时间进行校正。

在科以下，BLADJ算法假设，科的节点形成后，科属种，各自占据三分之一。

# BLADJ算法示意图



(a) Species Level

(b) Genus Level

## 查询被子植物的APG科apglst

```
library(apglst)
species2222 <- c(
  "Ranunculus japonicus",
  "Anemone udensis",
  "Ranunculus repens",
  "Ranunculus chinensis",
  "Solanum nigrum",
  "Punica sp."
)
col <- make.apglst(species2222)
taxa.table(col)
```

# 分子钟校正

chronompl()	用Mean Path Lengths校正进化树
chronopl()	用Penalized Likelihood校正进化树
chronos()	用Penalised Likelihood校正进化树
ACE	祖先状态重建
PIC	Phylogenetic independent contrast

# 目录

- 1 R中的对象及操作
- 2 R与Phylogenetics
- 3 群落的alpha多样性, beta多样性和解释
- 4 空间分析
- 5 群落系统发育

# 数据的准备和转换

## 野外记录的格式

	plotname	species	abundance
1	plot1	sp1	3
2	plot1	sp2	6
3	plot1	sp3	1
4	plot1	sp4	2
5	plot1	sp5	1
6	plot2	sp1	8
7	plot2	sp3	30
8	plot3	sp4	2
9	plot3	sp2	1
10	plot3	sp6	1
11	plot3	sp7	3

```
spaa::data2mat(testdata)
```



## 转换为物种矩阵

行表示样方，第一列作为行名

列表示物种，第一行作为物种名

行列交叉处表示物种的个体数，若没有出现，则用0来表示。

R中物种多样性指数计算，排序，系统发育多样性分析，种间联结计算和生态位重叠分析，都使用该矩阵。

	sp1	sp2	sp3	sp4	sp5	sp6	sp7
plot1	3	6	1	2	1	0	0
plot2	8	0	30	0	0	0	0
plot3	0	1	0	2	0	1	3

## spaa程序包计算种间联结和生态位重叠

### spaa代码

```
#数据转换
(spmatrix <- data2mat(testdata))
#计算种间联结
sp.assoc(spmatrix)
#所有种两两之间的联结
(result <- sp.pair(spmatrix))
#网络图
plotnetwork(result$Pearson)
title("Pearson Correlation Network")
#生态位宽度和生态位重叠
data(datasample)
niche.overlap.boot(datasample[,1:3], method = "levins")
niche.overlap(datasample, method = "levins")
niche.width(datasample[,1:3], method = "shannon")
```

# vegan程序包：群落数据分析

## Community Ecology Package: Ordination, Diversity and Dissimilarities

作者是芬兰生态学家Jari Oksanen.

- 多样性计算Shannon, Simpson, Pielou, EstimateR
- 排序CCA, DCA, CA, NMDS
- 种面积曲线
- 物种多度分布曲线
- 方差分解
- beta多样性的计算，以及分解

# vegan生物多样性指数

vegan程序包下计算多样性相关的几个指数:

<b>diversity</b>	多样性计算
<b>rarefy</b>	面积不等的两个群落比较时, 进行随机抽样
<b>fisher.alpha</b>	计算多样性
<b>specnumber</b>	物种累计计算。

## 多样性指数计算举例

```
data(BCI)
H <- diversity(BCI)
simp <- diversity(BCI, "simpson")
invsimp <- diversity(BCI, "inv")
S <- specnumber(BCI)
J <- H/log(S)
```

# vegan中的SAR和SAD

## 种面积曲线Species Area Relationship

```
SSarrhenius()
```

```
SSgitay()
```

```
SSgleason()
```

```
SSlomolino()
```

## 物种多度分布曲线Species abundance distributions

```
rad.null()
```

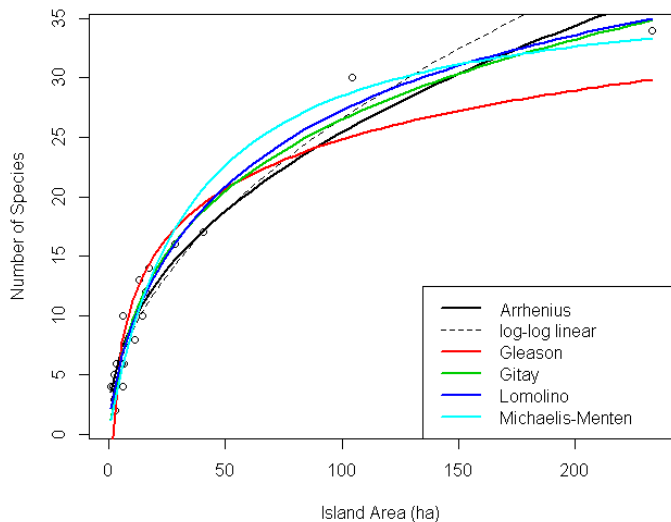
```
rad.preempt()
```

```
rad.lognormal()
```

```
rad.zipf()
```

```
rad.zipfbrot()
```

# 种面积曲线模型



# PCNM和方差分解

问题: 在物种丰富度格局中, 空间能解释多少变异?

PCNM是Principal Coordinates of Neighbourhood Matrix的缩写。是将距离矩阵转换成类似普通解释变量的一种方法。

将PCNM矩阵作为解释因子, 对物种丰富度等响应变量进行方差分解, 可以了解空间因素, 以及其他环境因素, 在格局形成中的相对重要性。PCNM将空间的距离关系, 转变为某一个尺度下, 不同点之间的逐渐变化。

示例:

```
library(phylotools)
library(vegan)
## 模拟一套数据
X <- seq(10, 590, by = 20)
Y <- seq(10, 390, by = 20)
```

```
## Label
XY <- expand.grid(X, Y)
names <- paste("X", (XY[,1] + 10)/20, "Y", (XY[,2] + 10)/20, sep =
"")
rownames(XY) <- names
## 计算样方两两之间的距离
distXY <- dist(XY)
## 进行pcnm
gtsPCNM <- pcnm(distXY)
head(gtsPCNM$vectors)
## 绘制pcnm图
for(i in 1:10){
  imagevect(gtsPCNM$vectors[,i], labels = names,
  col = topo.colors(100))
  Sys.sleep(0.3)
}
```



# 方差分解

PCNM

**vegan::pcnm()**: PCNM: the principal coordinate analysis of a truncated distance matrix

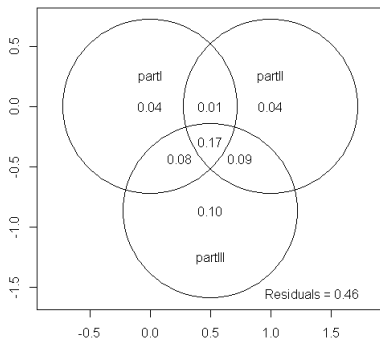
前向筛选

**forward.sel()**: forward selection by Dr. Stephene Dray

方差分解

**varpart**: variation partitioning:

# 方差分解varpart



partI: SubsDens + WatrCont

partII: Substrate + Shrub + Topo

partIII: mite.pcnm

# 计算群落之间的相似性或相异性

**vegan::vegdist**

```
vegdist(x, method="bray", binary=FALSE, diag=FALSE,  
upper=FALSE, na.rm = FALSE, ...)
```

相似性指数的计算，一般都转换为disimmilarity 才能进行后续计算。

**Jaccard**指数：适用于01数据，在计算beta多样性中，

**Bray-Curtis**指数：适用于物种多度数据

备选：物种相似性simba程序包

## Mantel 检验：距离矩阵之间的相关性

**mantel test:** 两个距离矩阵的相关性不能用 $r$ 来表示，需要用Mantel's test

```
ape::mantel.test(m1, m2, nperm = 999, graph = FALSE,  
alternative = "two.sided", ...)
```

```
## Is vegetation related to environment? data(varespec)  
data(varechem)  
veg.dist <- vegdist(varespec) # Bray-Curtis  
env.dist <- vegdist(scale(varechem), "euclid")  
mantel(veg.dist, env.dist)
```

## 环境距离的计算：欧几里得距离

设想二维平面直角坐标系, 有A,B两点, A坐标为 $(x_1, y_1)$ , B坐标为 $(x_2, y_2)$   
用`dist()`用来求欧几里得距离

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

推广到高维:

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

## 欧几里得距离的计算

```
x <- matrix(rnorm(100), nrow = 5)
dist(x)
dist(x, diag = TRUE)
dist(x, upper = TRUE)
m <- as.matrix(dist(x))
d <- as.dist(m)
```

获得的距离矩阵可以进一步用于进行聚类分析等。

# 地理距离的计算

若平面直角坐标系

`dist()`用来求欧几里得距离下三角矩阵

但是地球是椭球，经纬度亦为球面坐标

给定经纬度，如何计算地理距离？

`fields::rdist.earth`

```
data(ozone2)
```

```
out <- rdist.earth ( ozone2$lon.lat)
```

```
#距离矩阵
```

```
upper <- col(out) > row( out)
```

```
# histogram of all pairwise distances.
```

```
hist( out[upper])
```

## beta多样性的分解:地理距离

**adonis**: Analysis of variance using distance matrices

—for partitioning distance matrices among sources of variation and fitting linear models (e.g., factors, polynomial regression) to distance matrices; uses a permutation test with pseudo-F ratios.

**mrpp**: Multi Response Permutation Procedure and Mean Dissimilarity Matrix

**anosim**: Analysis of Similarities

**MRM**:基于Mantel test对beta多样性进行方差分解



# 目录

- 1 R中的对象及操作
- 2 R与Phylogenetics
- 3 群落的alpha多样性, beta多样性和解释
- 4 空间分析
- 5 群落系统发育

# 数据分析中考虑空间因素

## 1 空间自相关以及模型的应用

不一定用R, SAM - Spatial Analysis in Macroecology

[www.ecoevol.ufg.br/sam/](http://www.ecoevol.ufg.br/sam/)?

## 2 空间点格局分析整合群落生态学

为什么讲这些？

这方面的分析，是进行群落系统发育结构进一步研究的基础，是当前的重要方向。

生态学家已经不仅仅满足于计算群落系统发育多样性指数。还需要用空间分析的方法，了解成因。

# 空间自相关 Spatial Autocorrelation

## 什么是空间自相关？

空间自相关是指所研究的变量在空间上的不独立性，空间距离近的点，往往某一值更为接近。

## 为什么要考虑空间自相关？

空间自相关能够增加统计学上第一类错误的概率。

即在生态学分析中，没有考虑空间自相关的格局分析，往往会高估某些解释因子的影响，原本并没有显著关系的环境因子误认为具有显著影响，同时筛选出更多的解释变量。

# 空间自相关计算

空间自相关可以用Moran's I和Gary's C来表示  
Moran's I 的公式如下：

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

spdep程序包进行空间分析的核心程序包之一。  
其中既包括了计算Moran's I 的函数，也包括计算Spatial Auto  
Regressive Model 的函数

# spdep计算Moran's I 和SACSAR模型

## spdep

```
library(spdep)
## 为了检验ntaxa是否具有空间自相关
test0 <- read.csv("test.csv", header = TRUE, row.names = 1)
ntaxa <- test0$ntaxa
## 将test数据集转换成Spatial格式
test <- test0[,c(1,2)]
sptest <- SpatialPoints(test, proj4string = CRS("+proj=longlat
+datum=WGS84"))
## 计算每个点最近的几个neighbour (这里k = 1,表示只计算一个的)
nbk1 <- knn2nb(knearneigh(sptest, k = 5, longlat = TRUE))
## 将nbk1转换成spatial weight linkage object 对象
snbk1 <- make.sym.nb(nbk1)
#### n.comp.nb() finds the number of disjoint connected subgraphs
#### in the graph depicted by nb.obj - a spatial neighbours list object.
#### 查看每个点不相接的相邻点数量
```

## spdep计算Moran's I 和SACSAR模型II

### spdep

查看各点链接情况

```
plot(nb2listw(snbk1), cbind(test$longitude, test$latitude))  
#Moran's Test检验该数据集是否存在显著的空间自相关  
moran.test(ntaxa, nb2listw(snbk1))  
par(mfrow = c(1, 3))  
ntaxa.Moran.I <- sp.correlogram(snbk1, ntaxa, order=6, method="I",  
zero.policy = TRUE)  
lm.morantest.sad(lm(ntaxa ~ 1),nb2listw(snbk1))  
## sacsarlms  
COL.sacW.eig <- sacsarlms(ntaxa ~ Pre + Elev + factor(Time) +  
factor(Geology), data = test0, nb2listw(snbk1, style="W"))  
summary(COL.sacW.eig, correlation=TRUE)
```

# 点格局分析的程序包spatstat

## spatstat程序包的主要作用

- 创建点格局分析的数据类型，绘图
- 分析现有的点格局，并拟合到模型中，得到相应的系数
- 提供参数，生成模拟的点格局
- 进行模型相关的统计检验

# spatstat程序包中的ppp类

ppm: Fit Point Process Model to Data 那些模型?

Poisson, AreaInter, BadGey, Concom, DiggleGatesStibbard, DiggleGratton, Fiksel, Geyer, Hardcore, LennardJones, MultiStrauss, MultiStraussHard, OrdThresh, Ord, Pairwise, PairPiece, Saturated, SatPiece, Softcore, Strauss StraussHard.



# 点格局零模型的应用

1. 从样地中，实际的物种分布坐标，估计出每个种，拟定模型后的参数。
2. 用每个种的参数，分别按照随机过程的模型，生成满足某种指定模型的点格局。
3. 将2重复多次，生成n个满足某种指定模型的随机点格局。计算某统计量，获得统计量的分布。
4. 用真实数据，计算某统计量，再和某种模型下，随机过程的统计量分布进行比较，即可得到真实格局受到哪种随机过程的影响。

# spatstat进行参数拟合

```
data(redwood) ## California Redwoods Point Pattern  
u <- thomas.estK(redwood, c(kappa=10, sigma2=0.1))  
plot(u)
```

# 传播限制和生境异质对群落结构的影响

## Dispersal Limitation VS habitat heterogeneity

- Homogenous Poisson 过程rpoispp Generate Poisson Point Pattern
- Heterogenous Poisson过程
- Homogenous Thomas 过程rThomas Generate the Thomas cluster Point Pattern.
- Heterogenous Thomas过程

G Shen et al. (2009) Species-area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. Ecology 90 (11), 3033-3041 (Supplement)

考虑到系统发育结构与群落点格局整合，是重要研究方向

# 目录

- 1 R中的对象及操作
- 2 R与Phylogenetics
- 3 群落的alpha多样性, beta多样性和解释
- 4 空间分析
- 5 群落系统发育

# 群落系统发育结构Phylomatic, Phylocom, picante

NRI和NTI是最早提出的群落系统发育指数，表示群落内物种系统发育的距离高于或者低于零模型给出的距离。

计算群落内物种系统发育的距离常用指数有MPD和MNTD：

MPD是计算群落内物种两两之间的系统发育距离的平均值；

MNTD是寻找群落内每个物种系统发育关系最近的物种的系统发育距离，并计算所有最近物种的系统发育平均距离。

MPD和MNTD和与零模型进行比较之后，分别得到NRI和NTI指数。

$$NRI_{sample} = -1 \times \frac{MPD_{sample} - MPD_{rndsample}}{sd(MPD_{rndsample})}$$

$$NTI_{sample} = -1 \times \frac{MNTD_{sample} - MNTD_{rndsample}}{sd(MNTD_{rndsample})}$$

ses.mntd() ses.mpd()

# 系统发育多样性和信号

## 系统发育多样性PD : Phylogenetic Diversity

Faith(Faith, 1992)提出的系统发育多样性(Phylogenetic diversity, PD)给出的是某一地点所有物种在进化树上的枝长之和。

picante::pd()

**系统发育信号(Phylogenetic signal)**能够检验系统发育关系相近的物种是否具有相似性状。Blomberg's K(Blomberg and Garland 2002, Blomberg et al.

2003, Butler and King 2004)是最常用的系统发育信号指数, K值是类群间性状差异偏离于随机过程的度量。

picante::phylosignal()

# 系统发育beta多样性

系统发育beta多样性: Phylogenetic Beta diversity

系统发育beta多样性是群落或地点之间系统发育距离的度量(Fine and Kembel 2010)

<b>comdist() &amp; comdistnn()</b>	:MPD和MNTD (Webb 2000)
<b>phylosor()</b>	:Phylosor (Bryant et al. 2008)
<b>unifrac():</b>	:Unifrac(Lozupone et al. 2006)
<b>rao()</b>	:Rao 1982, Jost 2007, Webb et al. 2008
<b>pcd()</b>	:PCD (Ives and Helmus 2010)

# 物种组成结构研究中的零模型

群落系统发育分析中零模型，是将群落内物种组成的关系进行随机化的一系列方法。按照中性理论，物种与物种之间是等同的，因此，群落中物种应该是随机组合的。

群落系统发育分析中，一般通过以下四种方式进行随机化。

- **Null 0** 群落数据不变，但是物种在进化树末端随机排列。
- **Null 1** 进化树不变，物种在样方中随机排列，物种从所有样方中随机选取。
- **Null 2** 进化树不变，物种在样方中随机排列，物种从指定的物种库中选取。
- **Null 3** 进化树不变，与此同时，物种在样方中成对的关系保持不变。这种随机化的方法称为独立交换法(Independent swap)。



# 群落系统发育研究展望

- 1 地理尺度系统发育beta多样性的成因探讨，如纬度变化，解释宏生态学的一些假说；
- 2 群落尺度内，系统发育结构，种面积曲线，系统发育多样性面积曲线和点格局分析进一步整合，以探讨群落组成的理论问题。
- 3 群落尺度的数据，和物种分布信息和适应性进一步整合，了解物种在局域尺度的适应性。
- 4 物种的功能性状数据和物种在不同尺度的分布信息，以了解物种分布和适应性的一些机理。
- 5 用功能基因组学和转录组的手段，探讨物种的适应性和分布。
- 6 借助群体遗传学，亲本分析等手段进行物种群落分布格局的研究，进一步分析群落系统发育结构等的成因。

谢谢！  
敬请批评指正！

张金龙

jinlongzhang@gmail.com

香港嘉道理农场暨植物园植物保育部

香港 新界 大埔