



种间联结和生态位重叠的计算:spaa 程序包

张金龙^{1*} 马克平²

(1 香港嘉道理农场暨植物园植物保育部, 香港大埔;

2. 中国科学院植物研究所植被与环境变化国家重点实验室, 北京 100093)

摘要 群落的构建机制是生态学研究的核心问题之一。生态位理论在解释群落构建机制中占据着重要地位。研究物种的种间联结、生态位宽度以及生态位重叠,可以进一步检验生态位理论,从而进行更深入的研究。但是相应的分析计算一直缺乏统一的平台,在繁琐的数据处理中很容易引入人为错误。为此,作者用 R 语言编写了 spaa 程序包,用来计算群落内物种的种间联结、物种生态位宽度和生态位重叠、绘制半矩阵图和网络图等。该程序包代码公开,数据处理方式标准,简单易用,降低了计算的难度。

关键词 群落,种间联结,生态位,R 语言

群落的物种构成机制是生物多样性研究的核心问题之一(Ackerly, 2003; Chase, 2003; Cornwell & Ackerly, 2009; Pavoine & Bonsall, 2011)。通过样地调查,人们得以了解某地区典型地段的物种数目以及多样性等指标(马克平等, 1995; 方精云等, 2004)。为了解释群落中的物种共存机制,生态学家提出了多种假说,如生态位理论和中性理论,并研究各种机制在群落物种多样性形成中的作用,如同种的密度制约、不同种的竞争排斥、种子传播限制作用等等(杨利民等, 2001; 牛克昌等, 2009)。生态学家认为生物单位(即个体、种群或物种生态位)即为该生物单位适应性的总和。群落尺度上,物种的生态位是指其种在群落中利用资源的能力(张金屯, 2004)。

生态位在种群大小、群落多样性维持等多方面具有重要的理论意义,研究结论对于物种多样性保育、物种恢复等有很重要的指导意义(杨利民等, 2001)。生态位的宽度、物种间生态位重叠以及种间联接等是物种生态位研究的主要内容。近年来,对于种间联接、种间生态位重叠等,国内已有较多研究(如孙中伟和赵士洞, 1996; 郭志华等, 1997; 史作民等, 2001; 邢福和郭继勋, 2001; 张思玉和郑世群, 2002; 张光明等, 2003; 张志勇等, 2003; 朱圣潮, 2006; 韩文衡等, 2009; 李军玲和张金屯, 2010; 王乃江等, 2010)。但是在数据分析过程中多缺乏统一的平台,很容易人为引进一些错误。由于缺少相应的软件包,处理群落数据、绘制图形等十分不便。为此,我们用 R 语言编写了程序包 spaa。spaa 是 SPecies Association Analysis 的缩写,该程序包可以对群落数据进行简单的处理,计算物种的生态位宽度、生态位重叠,绘制相应的半矩

* 通讯作者:张金龙,嘉道理农场暨植物园植物保育部,香港新界大埔林锦公路,植物鉴定和记录主任。

E-mail: jinlongzhang01@gmail.com。

阵图,并对生态位宽度进行 bootstrap 等显著性检验。

本文介绍如何用 spaa 进行数据处理以及种间关联、生态位宽度和重叠等分析。

1 spaa 的安装

spaa 程序包由开源的 R 语言写成,源代码完全公开,目前已经上传到 R 的镜像网站 CRAN 上,用户可以在任何一个 CRAN 镜像下载。

推荐的安装方法为,在 R 界面输入命令 `install.packages("spaa")`,R 软件会自动提示选择一个镜像,自动下载该程序包,并进行安装。

2 spaa 主要功能介绍

利用 spaa 进行群落数据的前期处理。

2.1 群落数据转换为矩阵

在野外调查中,原始数据常常为以下格式(表 1)

表 1 野外原始数据记录

Table 1 Raw data from field record

样地名 Plotname	物种 Species	多度 Abundance
plot1	sp1	3
plot1	sp2	6
plot1	sp3	1
plot1	sp4	2
plot1	sp5	1
plot2	sp1	8
plot2	sp3	30
plot3	sp4	2
plot3	sp2	1
plot3	sp6	1
plot3	sp7	3

spaa 提供了函数 `data2mat()` 将数据转换成矩阵(表 2)。转换成矩阵后,即可方便地进行多种计算。

表 2 物种分布矩阵

Table 2 Species distribution matrix

	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

实际情况下,建议将原始记录输入到 MS Excel 中,并另存为 csv 文件,通过 `read.csv()` 函数读取为数据框(`data.frame`),并将对应的各列分别命名为“plotname”,“species”,“abun-

dance”, 经过这样标准化处理后, 即可通过 `data2mat()` 转换为矩阵。

调用方式为 `data2mat(testdata)`

2.2 计算种间链接的函数

种间链接的计算一般包括总体链接以及种与种之间的关联(周先叶等, 2000; 史作民等, 2001; 张思玉和郑世群, 2002; 张志勇等, 2003; 康冰等, 2005; 王文进等, 2007; 王乃江等, 2010)。

2.2.1 总体的链接计算

物种相对多度的方差:

$$\delta_T^2 = \sum_{i=1}^s P_i(1 - P_i) \quad (1)$$

物种数的方差:

$$S_T^2 = 1/N \sum_{j=1}^N (T_j - t)^2 \quad (2)$$

相对多度频率:

$$P_i = n_i/N \quad (3)$$

方差比率:

$$VR = S_T^2/\delta_T^2 \quad (4)$$

如果 $VR > 1$, 呈现正相关; 如果 $VR < 1$, 呈现负相关。

用统计量 W 检验方差比率的显著程度(张倩媚等, 2006), 95% 置信区间为 $\text{Chi}_{0.95, N^2}^2 < W < \text{Chi}_{0.05, N^2}^2$

$$W = VR \times N \quad (5)$$

公式(1)–(5)中, N 为样方数, S 为总物种数, n_i 为一个物种所占据的样方数, T_j 为每个样方的物种数, t 为所有样方的物种数平均值。

2.2.2 种之间的联结

种之间的联结一般采用 2×2 列联表。假设有物种 A 和物种 B 分别出现于各样方中, a, b, c, d 分别表示两者之间共同出现和不出现的情况: a : 物种 A 和物种 B 共同占据的样方数; b : 只出现物种 A 的样方数; c : 只出现物种 B 的样方数; d : 物种 A 和物种 B 都不出现的样方数, n 为所有样方数, $n = a + b + c + d$ (张金屯, 2004)。

在判断两个种之间是否存在显著的关联时, 常常应用经过 Yates 校正过的卡方; 而 V 值以及 Jaccard 指数、Ochiai 指数、Dice 指数、点相关指数 PCC、Pearson 相关系数以及 Spearman 相关系数则体现两个种关联性的紧密程度(郭志华等, 1997; 邢福和郭继勋, 2001; 张光明等, 2003; 张志勇等, 2003; 张金屯, 2004; 朱圣潮, 2006), 其公式分别为:

经过 Yates 矫正的卡方:

$$\chi_{Yates}^2 = \frac{(|ad - bc| - 0.5n)^2 n}{(a+b)(a+c)(b+d)(c+d)} \quad (6)$$

V 值:

$$V = \frac{(a+d) - (b+c)}{a+b+c+d} \quad (7)$$

Jaccard 指数:

$$JD = a/(a + b + c) \quad (8)$$

Ochiai 指数:

$$OI = \frac{a}{\sqrt{(a + b)(a + c)}} \quad (9)$$

Dice 指数:

$$DI = \frac{2a}{2a + b + c} \quad (10)$$

PCC 点相关系数:

$$PCC = \frac{ad - bc}{\sqrt{(a + b)(a + c)(c + d)(b + d)}} \quad (11)$$

连接系数 AC:

如果 $ad \geq bc$

$$AC = \frac{ad - bc}{(a + b)(b + d)} \quad (12)$$

如果 $bc > ad$, 且 $d \geq a$, 则

$$AC = \frac{ad - bc}{(a + b)(a + c)} \quad (13)$$

如果 $bc > ad$ 且 $d < a$, 则

$$AC = \frac{ad - bc}{(b + d)(d + c)} \quad (14)$$

Spearman 秩相关系数

$$r_{ik} = 1 - \frac{6 \sum_{j=1}^N (x_{ij} - x_{kj})_j^2}{N^3 - N} \quad (15)$$

Pearson 相关系数

$$r_{ik} = \frac{\sum_{j=1}^N (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{\sqrt{\sum_{j=1}^N (x_{ij} - \bar{x}_i)^2 \sum_{j=1}^N (x_{kj} - \bar{x}_k)^2}} \quad (16)$$

这些指数的计算结果通过 spaa 的 sp.assoc() 函数即可得到。结果整合在该函数返回的 list 中, 如果需要调出相应的结果, 用 \$ 符号即可。

2.3 计算生态位宽度的函数

常用 Levins 方法或 Shannon 方法来度量生态位宽度(吴大荣, 2001; 向悟生等, 2002; 张桂莲和张金屯, 2002; 胡正华和于明坚, 2005; 何小娟等, 2008; 钱逸凡等, 2012)。

Levins 生态位宽度的公式为:

$$B_i = \frac{1}{\sum_{j=1}^r (P_{ij})^2} \quad (17)$$

Shannon 生态位宽度的公式为

$$B_i = - \sum_{j=1}^r (P_{ij} \ln P_{ij}) \quad (18)$$

其中 B_i 为第 i 种的生态位宽度, j 表示样方, r 表示样方的数量。

spaa 程序包提供了函数 `niche.width()` 来计算每个种的生态位宽度。调用方法为 `niche.width(mat, method = c("shannon", "levins"))`。其中 `mat` 为物种分布矩阵, `method` 选择其一即可。

2.4 计算生态位重叠的函数

物种的生态位重叠是人们特别关注的问题之一。spaa 提供了 `niche.overlap()` 函数来计算矩阵内所有物种对之间的生态位重叠系数。这些系数包括: `levins`, `schoener`, `petraitis`, `pianka`, `morisita`, 返回结果为距离矩阵(`dist`), 以清晰地表示种与种两两之间的生态位重叠关系。

Levins 生态位重叠指数:

$$O_{ik} = \frac{\sum_{j=1}^r P_{ij} P_{kj}}{\sum_{j=1}^r (P_{ij})^2} \quad (19)$$

Schoener 生态位重叠指数:

$$O_{ik} = 1 - \frac{1}{2} \sum_{j=1}^r |P_{ij} - P_{kj}| \quad (20)$$

Petraitis 特定重叠指数:

$$O_{ik} = e^{\sum_{j=1}^r (P_{ij} \ln P_{kj})} - \sum_{j=1}^r (P_{ij} \ln P_{ij}) \quad (21)$$

Pianka 重叠指数:

$$O_{ik} = \frac{\sum_{j=1}^r P_{ij} P_{kj}}{\sqrt{(\sum_{j=1}^r P_{ij})^2 (\sum_{j=1}^r P_{kj})^2}} \quad (22)$$

Czechanowski 指数:

$$O_{12} = O_{21} = 1 - 0.5 \sum_{i=1}^n |p_{1i} - p_{2i}| \quad (23)$$

简化的 Morisita 指数:

$$O_{ik} = \frac{2 \sum_{j=1}^r P_{ij} P_{kj}}{\sum_{j=1}^r P_{ij}^2 + \sum_{j=1}^r P_{kj}^2} \quad (24)$$

其中 O_{ik} 为种 i 和种 k 的生态位重叠系数, P_{ij} 和 P_{kj} 分别为种 i 和种 k 在第 j 个样方的多度, r 为样方的总数, e 为自然对数的底。

调用方法举例:

```
data(datasample)
```

```
niche.overlap(mat, method = c("levins", "schoener", "petraitis", "pianka", "czech", "morisita"))
```

其中 `mat` 为群落物种分布矩阵, `method` 选择 "levins", "schoener", "petraitis", "pianka", "czech", "morisita" 等其中任意一个指数。

若只想得到两个物种之间的生态位重叠系数, 则需要运行 `niche.overlap.pair()` 函数。调用方式为:

```
niche.overlap.pair(vectA, vectB, method = c("pianka", "schoener", "petraitis",
```

"czech", "morisita", "levins"))

其中 vectA 和 vectB 分别为两个向量,表示在对应的群落中,物种 A 和物种 B 出现的个体数, method 则需要选取 pianka, schoener, petraitis, czech, morisita, levins 中的任意一个。

2.5 生态位宽度置信区间的自展分析

为了估计两个种之间的生态位重叠的置信区间, spaa 提供了生态位重叠的自展分析 bootstrap 函数 niche.overlap.boot()。该函数各参数如下:

```
niche.overlap.boot(mat, method = c("pianka", "schoener", "petraitis", "czech",
"morisita", "levins"), times = 999, quant = c(0.025, 0.975))
```

其中 mat 为输入的物种分布矩阵, method 是要选择的方法, times 为 bootstrap 进行的次数, quant 为生态位重叠指数的分位数, 默认为 0.025 和 0.975, 即 95% 置信区间。

在计算过程中, niche.overlap.boot() 会调用 niche.overlap.boot.pair(), 先计算物种两两之间的生态位重叠置信区间。一般情况下, 用户均无需调用 niche.overlap.boot.pair()。niche.overlap.boot() 会给出群落内所有种两两之间的生态位重叠指数, 以及相应的 bootstrap 结果(图 1)。

```
> data(datasample)
> niche.overlap.boot(datasample[,1:4], method = "pianka")
Observed Boot mean Boot std
Castanopsis.eyrei-Schima.superba 0.887 0.894 0.044
Castanopsis.eyrei-Rhododendron.ovatum 0.820 0.836 0.084
Castanopsis.eyrei-Cyclobalanopsis.glauca 0.551 NaN NA
Schima.superba-Rhododendron.ovatum 0.693 0.722 0.151
Schima.superba-Cyclobalanopsis.glauca 0.491 0.520 0.122
Rhododendron.ovatum-Cyclobalanopsis.glauca 0.253 0.305 0.138
Boot CI1 Boot CI2 times
Castanopsis.eyrei-Schima.superba 0.791 0.963 999
Castanopsis.eyrei-Rhododendron.ovatum 0.634 0.977 999
Castanopsis.eyrei-Cyclobalanopsis.glauca 0.251 0.868 999
Schima.superba-Rhododendron.ovatum 0.421 0.948 999
Schima.superba-Cyclobalanopsis.glauca 0.310 0.772 999
Rhododendron.ovatum-Cyclobalanopsis.glauca 0.096 0.593 999
```

图 1 物种生态位重叠的 bootstrap 结果

Fig. 1 Bootstrap output of niche overlap between each pair of species

第一、二行为输入的 R 命令, 其余为返回的结果
Row 1 and 2: R commands, others: the outputs

各列的含义分别如下:

第 1 列, 两个列的种名 *Castanopsis.eyrei-Schima.superba*, 表示对应的种对。

第 2 列, Observed 表示实际的物种生态位重叠指数;

第 3 列, Boot mean 表示 bootstrap 结果的算数平均值;

第 4 列, Boot std 表示生态位重叠指数的标准差(standard deviation);

第 5 列, Boot CI1 表示物种生态位重叠指数的下分位数, 默认为 0.025;

第 6 列, Boot CI2 表示物种生态位重叠指数的上分位数, 默认为 0.975

当某个样方中缺少一个物种时, 在 bootstrap 的过程中会出现 NaN, 这表明在根据公式计算生态位重叠的过程中, 出现了除数为 0 的情况。在随机有放回抽样中, 这种物种出现组合是可能发生的, 此时的 bootstrap 结果以及 CI 的结构都会变得不准确, 建议此时要审慎考虑和解

释。

2.6 其他函数

spaa 程序包还提供了一些函数用来绘制物种关联半矩阵图等(图 2)。

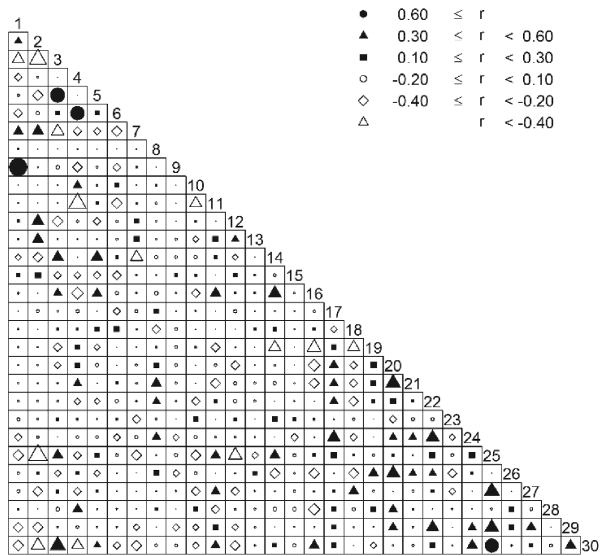


图 2 spaa plotlowertri() 函数绘制的半矩阵图

Fig. 2 Semi-matrix plot generated by function plotlowertri() in spaa

其函数调用方式为:

plotlowertri(input, valuname = "r", pchlist = c(19, 17, 15, 1, 5, 2, 7), interval = 6, cex = 1, ncex = 1, int = 1.2, add.number = TRUE, size = FALSE, add.text = FALSE, show.legend = TRUE, digits = 2), plotlowertri 提供了众多的参数。

用户可以对要显示的相关系数矩阵或者距离矩阵进行灵活调整。

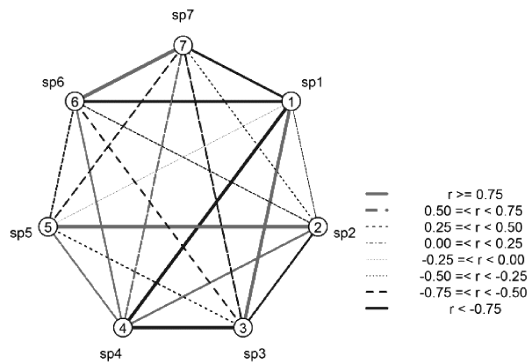


图 3 物种对之间相互关联的网络图

Fig. 3 Network plot showing the association between each pair of species

spaa 还可以用网络图来显示物种之间的相互关联,即用不同颜色、不同粗细以及类型的线条来连接各物种,表示种与种之间的关联,或者地点之间物种的相似性。plotnetwork()函数即可以用来绘制网络图(图 3)。但是网络图不宜用来表示物种较多的情况,如超过了 10 个物种,则所绘制的线过多,难以清晰地表达物种之间或者地点之间的联系。此时推荐使用半矩阵图(图 1)或进行聚类后,用树状图来表示。

除此之外,spaa 还提供了用于物种筛选的函数,sub.sp.matrix(spmatrix,freq = 0.5,common = NULL),可以选出频度大于某一数值的所有物种,common 参数则是选出最常见的多少个物种的数据,常见物种经过筛选之后,可以方便进行生态位宽度、生态位重叠等计算。

spaa 程序包还包括计算地理距离的函数 lgeodist()和 geodist(),给定两点的经纬度即可计算两地点间的地理距离,前者为低精度,后者为高精度;deg2dec()可以用来进行度、分、秒的转换;数据转换函数 dist2list()、turnover()、lab.mat()等在 beta 多样性的计算以及大样地数据的处理中可以提供帮助。

3 展望

spaa 程序包已经提交到 CRAN,并且受到了部分用户的反馈和好评。我们一直在致力于定期更新 spaa 程序包,并及时发现和解决程序的问题,使程序包更加友好。spaa 主要涉及物种关联性相关的分析,对于栅格数据的支持还十分有限,我们希望将来 spaa 能够对栅格数据有良好的支持。

致谢:感谢丁琼、黄继红、李宗善博士在程序包开发中给予的帮助,感谢马进泽、张雪妮、姜俊、高梅香、方晓峰、游泳、Wei Li、Alfredo H. Zúñiga Á.、Ísis Arantes、Patricia Martínez、Wilson Martins da Silva、Jessica L. Sabo、Kurt Hornik 博士以及 Brian Ripley 教授对程序包提出的意见和建议。

参考文献

- 方精云,沈泽昊,唐志尧,王志恒. 2004. 中国山地植物物种多样性调查及若干技术规范. 生物多样性,12:5-9.
- 郭志华,卓正大,陈洁,吴梅凤. 1997. 庐山常绿阔叶、落叶阔叶混交林乔木种群间联结性研究. 植物生态学报,21:424-432.
- 韩文衡,李先琨,叶铎,吕仕洪,向悟生,宋同清,曹洪麟. 2009. 桂西北喀斯特区常绿落叶阔叶混交林种群间联结性与相关性. 山地学报,27:719-726.
- 何小娟,洪滔,何东进,刘勇生,卞莉莉,陈笑玲,苏炳霖. 2008. 武夷山风景名胜区天然林主要种群生态位特征研究. 中国生态农业学报,16:285-291.
- 胡正华,于明坚. 2005. 古田山青冈林优势种群生态位特征. 生态学杂志,24:1159-1162.
- 康冰,刘世荣,蔡道雄,温远光,史作民,郭文福,朱宏光,张广军,刘磊. 2005. 南亚热带人工杉木林灌木层物种组成及主要木本种间联结性. 生态学报,25:2173-2179.
- 李军玲,张金屯. 2010. 太行山中段植物群落草本植物优势种群间联结性分析. 草业科学,27(9):119-123.
- 马克平,刘灿然,刘玉明. 1995. 生物群落多样性的测度方法. II. β 多样性的测度方法. 生物多样性,3:38-43.
- 牛克昌,刘恽宁,沈泽昊,何芳良,方精云. 2009. 群落构建的中性理论和生态位理论. 生物多样性,17:579-593.
- 钱逸凡,伊力塔,胡军飞,张超,余树全,沈露,彭东琴. 2012. 普陀山主要植物种生态位特征. 生态学杂志,31:

561-568.

- 史作民,刘世荣,程瑞梅,蒋有绪. 2001. 宝天曼落叶阔叶林种间联结性研究. 林业科学, **37**(2):29-35.
- 孙中伟,赵士洞. 1996. 长白山北坡椴树阔叶红松林群落木本植物种间联结性与相关性研究. 应用生态学报, **7**(1):1-5.
- 王乃江,张文辉,陆元昌,范少辉,王勇. 2010. 陕西子午岭森林植物群落种间联结性. 生态学报, **30**:67-78.
- 王文进,张明,刘福德,郑建伟,王中生,张世挺,杨文杰,安树青. 2007. 海南岛吊罗山热带山地雨林两个演替阶段的种间联结性. 生物多样性, **15**:257-263.
- 吴大荣. 2001. 福建罗卜岩闽楠(*Phoebe bournei*) 林中优势树种生态位研究. 生态学报, **21**:851-855.
- 向悟生,李先琨,苏宗明,欧祖兰,宁世江,唐润琴,李瑞棠. 2002. 元宝山冷杉群落主要树木种群生态位的初步研究. 武汉植物学研究, **20**:105-112.
- 邢福,郭继勋. 2001. 糙隐子草草原 3 个放牧演替阶段的种间联结对比分析. 植物生态学报, **25**:693-698.
- 杨利民,周广胜,王国宏. 2001. 草地群落物种多样性维持机制的研究 II 物种实现生态位. 植物生态学报, **25**:634-638.
- 张光明,杨大荣,徐磊,彭艳琼,卢耀. 2003. 西双版纳聚果榕榕果小蜂种间联结性研究. 生态学杂志, **22**(4):20-26.
- 张桂莲,张金屯. 2002. 关帝山神尾沟优势种生态位分析. 武汉植物学研究, **20**:203-208.
- 张金屯. 2004. 数量生态学. 北京:科学出版社.
- 张倩媚,陈北光,周国逸. 2006. 鼎湖山主要林型优势树种种间联结性的计算方法研究. 华南农业大学学报, **27**(1):79-83.
- 张思玉,郑世群. 2002. 福建永定桫欏群落内主要灌木种群的种间联结性研究. 云南植物研究, **24**(1):17-22.
- 张志勇,陶德定,李德铎. 2003. 五针白皮松在群落演替过程中的种间联结性分析. 生物多样性, **11**:125-131.
- 周先叶,王伯荪,李鸣光,管启杰. 2000. 广东黑石顶自然保护区森林次生演替过程中群落的种间联结性分析. 植物生态学报, **24**:332-339.
- 朱圣潮. 2006. 中华水韭松阳居群的群落结构与种间联结性研究. 生物多样性, **14**:258-264.
- Ackerly D D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences*, **164**:S165-S184.
- Chase J M. 2003. Community assembly: when should history matter? *Oecologia*, **136**:489-498.
- Cornwell W K and Ackerly D D. 2009. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. *Ecological Monographs*, **79**:109-126.
- Pavoine S and Bonsall M. 2011. Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews*, **86**:792-812.

spaa: An R Package for Computing Species Association and Niche Overlap

ZHANG Jinlong¹, MA Keping²

(1. Flora Conservation Department, Kadoorie Farm and Botanic Garden, Hong Kong SAR;

2. State Key Laboratory of Vegetation and Environmental Change, Institute of Botany,
Chinese Academy of Sciences, Beijing, 100093)

Abstract Species assembly in communities is a central issue in ecological studies. The niche theory played important roles in explaining the community assembly processes. Testing the niche theory by investigating the association between species, niche width and overlap between species would improve our understanding in the mechanisms underlying the diversity patterns. Although there have been a number of reports focused on this issue, an integrated and user friendly software would be helpful to the researchers to lower the risk in introducing errors. We therefore developed a package spaa using R, an open source statistical language. spaa allows the users to compute species association, niche width and overlap with much ease. The distance/similarity matrices can also be displayed in certain kinds of figures. The source code of spaa is open to the public. Computing the indices has been made easy because of the standardized data structures in spaa.

Key words community, species association, niche, R language