

# 历史生物地理学中的 BPA 分析法

## II: 二级 BPA \*

孟凯巴依尔<sup>①</sup> 李枢强<sup>①\*\*</sup> 傅金钟<sup>②</sup>

(<sup>①</sup>中国科学院动物研究所 北京 100080; <sup>②</sup>Department of Zoology, University of Guelph, Canada)

**摘要:** BPA(brooks parsimony analysis)是历史生物地理学研究的重要分析方法之一,包括初级 BPA 和二级 BPA。前者主要用于寻找共同格局,后者是对网状进化格局的解释。本文详细介绍了二级 BPA 的分析方法,并就 BPA 与其它历史生物地理学/支序学派的研究方法进行了区别。

**关键词:** BPA;历史生物地理;支序学派

**中图分类号:** Q332 **文献标识码:** A **文章编号:** 0250-3263(2004)01-52-08

## The Application of BPA in Historical Biogeography

### Part Two: Secondary BPA

MENG Kaibayier<sup>①</sup> LI Shu-Qiang<sup>①</sup> FU Jin-Zhong<sup>②</sup>

(<sup>①</sup>*Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, China;*

*<sup>②</sup>Department of Zoology, University of Guelph, Canada)*

**Abstract:** BPA(brooks parsimony analysis) is one of the most popular methods in the study of historical biogeography. A brief introduction on the secondary BPA is given in the present paper. A comparison to other methods of historical biogeography is presented.

**Key words:** BPA; Historical biogeography; Cladistics

BPA 初级分析(primary BPA)可以非常有效的推断出研究类群的共同格局<sup>[1]</sup>。但大多数情况下,一个区域内的类群常常具有网状进化格局,这种情况下不能用初级 BPA 分析进行解释。因此,BPA 进行历史生物地理学研究时还需要进行二级 BPA 分析(secondary BPA)。二级 BPA 可以对初级 BPA 分析结果中出现的平行演化和网状区域做出明确的解释。

### 1 二级 BPA 分析

在对初级 BPA 分析方法介绍<sup>[1]</sup>,提到初级 BPA 分析是在先验地假设“被分析的分布区内不具有网状历史”的条件下进行,它忽略了网状进化与物种现有分布区的关系。这与 BPA 的假设“每个分布区内所有被分析的种仅有单一的形成过程”相矛盾,同时 inclusive ORing 编码方式也不能有效地避免网状进化的结果。为此,Brooks 等人<sup>[2,3]</sup>针对造成网状进化的原因提出了

利用重复区域惯例(duplicated areas convention)进一步分析的解决方案。这就是 BPA 进行历史生物地理学深入分析的第二步——二级 BPA 分析。

导致网状进化格局类群主要的表现形式是:多余分类单元(redundant taxa)和广布分类单元(widespread taxa)。下面分别对这两种情况进行分析和解释。

**1.1 多余分类单元的分析** 多余分类单元是指在初级

\* 国家自然科学基金资助项目(NSFC-30270183, 30370263),同时部分得到国家基础科学人才培养基金(NSFC-J0030092),中国科学院知识创新工程(KSCX2-1-06A, KSCX3-IOZ-01)和中国科学院生物科学与技术研究特别支持费(STZ-00-19)的资助;

\*\* 通讯作者, E-mail: lisq@panda.ioz.ac.cn;

**第一作者介绍** 孟凯巴依尔, 26 岁, 硕士, 实习研究员; 研究方向: 动物系统进化。

**收稿日期:** 2003-04-18, **修回日期:** 2003-09-20

BPA 分析产生的区域支序图中,一个给定的支序图的多个分类单元出现在同一分布区,或不同支序图的多个成员对同一个分布区表现出不同的区域关系。

1.1.1 多个分类单元表现相同区域关系 根据某生物类群种 1~6 的系统发育关系(图 1),通过初级 BPA 分析的数据矩阵(表 1),可以得到区域支序图(图 2),其一致性指数(consistence index, CI)为 100%。由于采用 inclusive ORing 编码方式,分布区 B 的分支上出现 4 个种(种 3、5、6 和种 8)。

根据系统发育树(图 1)可以看出种 5 和种 6 互为姐妹种,都是种 8 的后裔。它们与种 3 有着不同的进化历史,同时也不能确定形成种 5 和种 6 的时间是否相同。因此,这种分布格局与 BPA“每个分布区内所有被分析的种仅有单一的形成机制和演化过程”的假设矛盾。依据假设,应该把分布区 B 分为 3 个区: B<sub>1</sub> 包含种 3, B<sub>2</sub> 包含种 5, B<sub>3</sub> 包含种 6,在此基础上通过编制新的数据矩阵(表 2),可以产生新的区域支序图(图 3)。

图 3 中 B<sub>1</sub> 出现在相对底层位置, B<sub>2</sub> 和 B<sub>3</sub> 形成姐妹分布区位于相对衍生的位置。根据最终的区域支序图可以推断相对衍生的种 5 和种 6 之所以出现在分布区 B,是因为它们的祖先种(种 8)由分布区 D 扩散到分布区 B,进而在分布区 B 进化为种 5 和种 6。分布区 B<sub>2</sub> 和 B<sub>3</sub> 可以被简约的认为属于同一分布区。

这种分析过程被 Brooks 等人<sup>[2,3]</sup>称为重复区域惯例。同时说明, CI 值不能准确地佐证区域支序图。

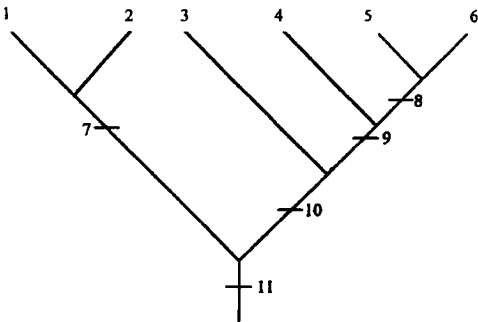


图 1 种 1~6 系统发育树  
数字为现生种和假设祖先种

表 1 种 1~6 的初级 BPA 分析的数据矩阵		
分布区	分类单元	特征序列
A	1	10000010001
B	3, 5, 6	00101101111
C	2	01000010001
D	4	00010000111

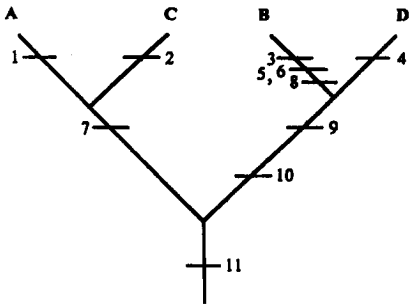


图 2 种 1~6 的初级 BPA 分析的区域支序图  
数字表示特征序号

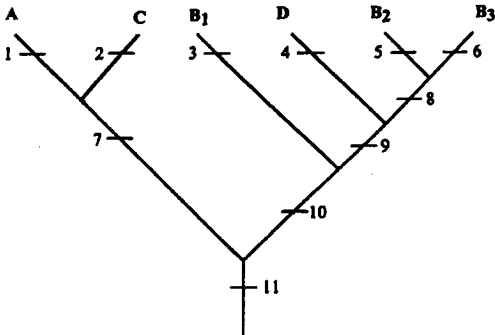


图 3 细化区域 B 后的二级 BPA 分析的区域支序图  
数字表示特征序号

表 2 细化区域 B 后二级 BPA 分析的数据矩阵		
分布区	分类单元	特征序列
A	1	10000010001
B <sub>1</sub>	3	00100000011
B <sub>2</sub>	5	00001001111
B <sub>3</sub>	6	00000101111
C	2	01000010001
D	4	00010000111

BPA 认为一个分布区内的多个种可以是多个历史事件形成的。这种多重性造成分布区内的进化历史呈网状。二级分析通过重复区域惯例解决了这种分布格局,即细化地理分布区,之后依据种的系统发育关系,重建立区域支序图用来解释网状进化格局。

需要注意的是,图 2 能否代表正确的区域共同格局还需要进一步验证,如与地质区域支序图比较,或与更多类群的系统发育关系联合分析。

1.1.2 分类单元表现不同区域关系 多余分类单元的另一种表现形式是,不同支序图的多个成员对同一个分布区表现出不同的区域关系(图 4)。这种情况经常使进行重复区域惯例后的分区与共同格局出现不同。

例如,根据种 1~4、种 8~11、种 15~18 和种 22~25 系统发育关系,可以得出 4 组系统发育树(图 4),每个现生种生活在其中的一个分布区内。根据 BPA 初级分析数据矩阵(表 3),可以建立区域支序图(图 5),其中种 1~4、种 8~11 和种 15~18 的系统发育关系与该区域支序图完全匹配,这说明这个区域支序图表现的就是分布区 A~D 的共同格局((A (B (C, D)))。但种 22~25 的分布格局与 BPA 的假设矛盾,因为祖先种 26 和 27 表现出平行演化现象,说明分布区 A 上的种有多个物种形成事件,所以这种共同格局不被种 22~25 认可。实际上,通过简单的逻辑分析也可以看出分布区 A 具有网状进化历史。就是说种 1~4、8~11、15~18 在分布区 A 的分布原因与种 22~25 在分布区 A 的分布原因不同。既然如此,如何通过分析得出这一结论呢?

通过观察发现祖先种 26 和祖先种 27 分别出现两次(图 5),说明祖先种 26 和种 27 曾同时存在于分布区 A 和 D,其中祖先种 26 的后裔为种 24 和种 25,它们分别出现在分布区 A 和分布区 D。但 BPA 认为一个种的形成方式只有一种,上述分析结果与此矛盾。为使分析结果既能满足假设又能符合共同分布格局,可应用重复区域惯例细化分布区 A 或 D。分布区 A 可分为 2 个区: A<sub>1</sub> 包括现生种 1,8 和 15,因为它们支持共同格局; A<sub>2</sub> 包括现生种 25,因为它的祖先种与假设矛盾。重

复后的区域结构重新依据系统发育关系编制数据矩阵(表 4),然后根据简约法得出新的区域支序图(图 6)。

由此可以推测,祖先种 26 的一些种由分布区 D 扩散到分布区 A,然后形成种 25。因此,分布区 A 中种 1,8,15 的存在是由于共同历史,而分布区 A 中种 25 的存在是由于种的扩散造成。

分布区 D 细化后的结果与图 5 相近,它的解释仍然与假设矛盾。

以上两点(1.1.1 和 1.1.2)主要分析了扩散造成的物种形成对共同格局的影响。当这种情况存在时, inclusive ORing 编码方式使种聚集在一个特定分布区中,如 1.1.1 中的分布区 B 和 1.1.2 中的分布区 A。重复区域惯例使扩散格局不再与共同格局重叠,并有了相应的解释。

表 3 种 1~4、8~11、15~18 和 22~25 的  
初级 BPA 分析的数据矩阵  
(仿 Brooks *et al.*, 2001)

分布区	分类单元	特征序列
A	1, 8, 15, 25	1000001100000110000010001111
B	2, 9, 16, 22	0100011010001101000111000001
C	3, 10, 17, 23	0010111001011100101110100011
D	4, 11, 18, 24	0001111000111100011110010111

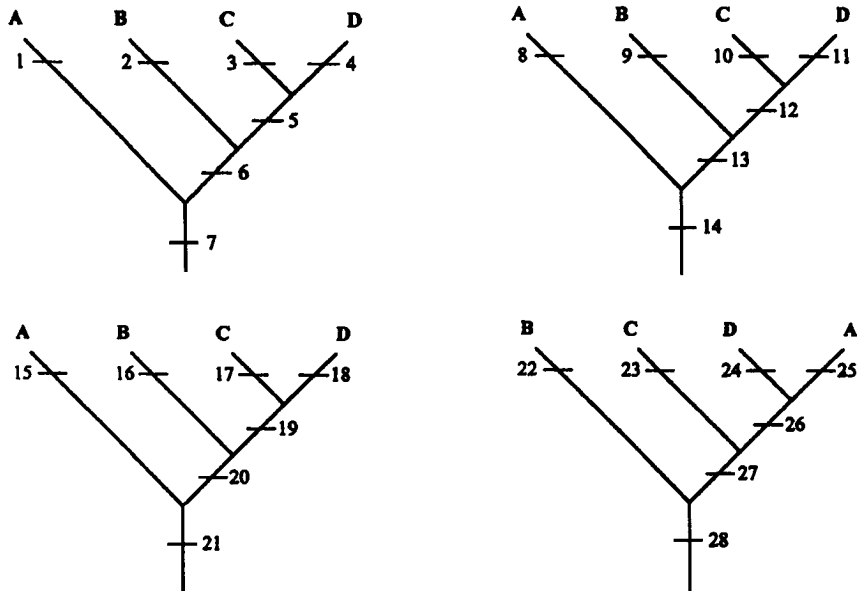


图 4 种 1~4、8~11、15~18 和 22~25 系统发育关系树及现生种在分布区 A~D 的分布  
数字代表现生种和假设祖先种(仿 Brooks *et al.*, 2001)

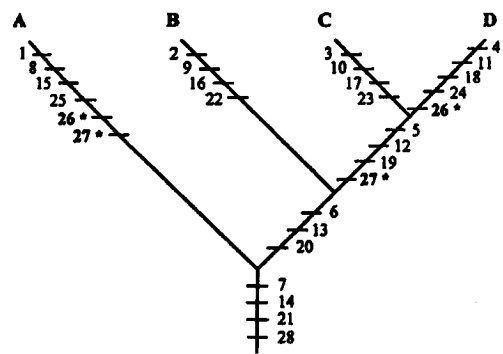


图 5 种 1~4、8~11、15~18 和 22~25 的  
共同格局区域支序图  
带星号种的系统发育关系与 BPA 假设矛盾  
(仿 Brooks *et al.*, 2001)

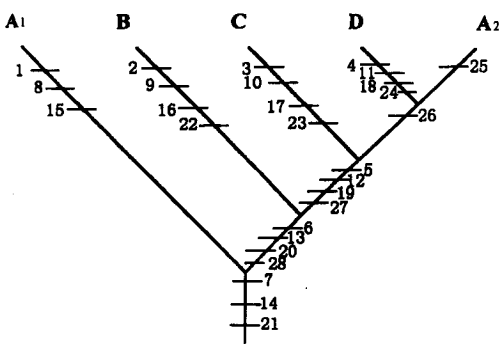


图 6 细化区域 A 后的二级 BPA 分析的区域支序图  
数字表示特征序号

1.2 对广布分类单元的分析 广布分类单元是指相同的分类单元出现在一个以上的分布区内。它同样包括两种情况:一是广布种在单个分支上,二是广布种在多个分支上。广布分类单元的存在可能是由于种从原始分布区扩散到其它分布区,或者由于相应的替代事件对它们没有产生影响。它的存在也会造成网状进化历史并影响 BPA 分析结果。

表 4 细化区域 A 后的二级 BPA 分析的  
数据矩阵(仿 Brooks *et al.*, 2001)

分布区	分类单元	特征序列
A <sub>1</sub>	1, 8, 15	100000110000011000001???????
B	2, 9, 16, 22	0100011010001101000111000001
C	3, 10, 17, 23	0010111001011100101110100011
D	4, 11, 18, 24	0001111000111100011111001011
A <sub>1</sub>	25	????????????????????0001111

“?”表示分布区内数据缺失

1.2.1 广布种在一个分支上 根据 4 个假设种的系统发育树(图 7),以及由此得出的 BPA 初级分析的数据矩阵(表 5),可以得到区域支序图(图 8),其 CI 值为 100%。这个由生物学数据得出的区域支序图(图 8)与由地质数据得出的区域支序图相同(图 9),即图 8 中表现的区域关系就是分布区 A~D 的共同格局(A(B(C, D)))。尽管共同格局(图 8)的 CI 值为 1,但它在多处表现出与假设或系统发育关系的矛盾。如分布区 C 中种 1 的缺失可以解释为逆转,或是灭绝事件。如果是逆转现象,则 C 分布区内的种 1 应该位于种 2、3 和 4 的祖先位置,这与实际情况中(图 7)种 1 与其它种形成姐妹分支的事实不符。如果解释为种 1 在分布区 D 灭绝,则系统发育学方面的证据不足。但后一种是以简约原则为基础的解释,相对来说更加符合规律。因此, Brooks 等人<sup>[2]</sup>认为这种情况也可应用重复区域惯例。

根据分布区 A~D 的共同格局和种间的系统发育关系所提供的信息,种 1 与其它种不具有相同的进化历史。通过对分布区 B 和 D 细化,建立新的数据矩阵(表 6),并由此得到新的区域支序图(图 10)。

图 10 中分布区 A、B<sub>2</sub> 和 D<sub>2</sub> 形成多分支并且种 1 位于其根部。说明种 1 是由分布区 A 扩散到分布区 B 和 D,但种 1 扩散到分布区 B 和 D 的顺序不能确定。

表 5 种 1~4 的初级 BPA 分析的数据矩阵

分布区	分类单元	特征序列
A	1	1000001
B	1, 2	1100011
C	3	0010111
D	1, 4	1001111

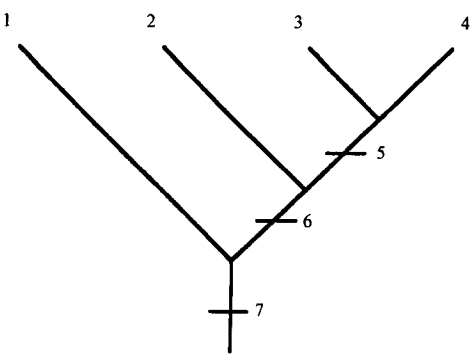


图 7 种 1~4 的系统发育树  
数字代表现生种和假设祖先种

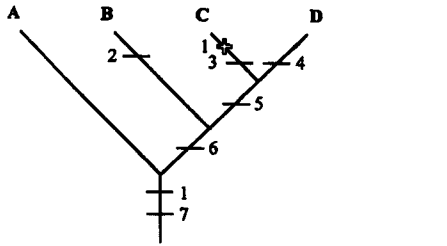


图 8 种 1~4 的共同格局区域支序图  
种 1 与图 7 的描述相矛盾

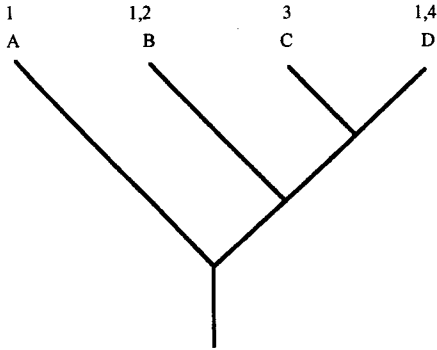


图 9 种 1~4 的地质区域支序图  
其中种 1 在分布区 B 和 D 均有分布

表 6 细化区域 B 和 D 后的二级 BPA 分析  
的数据矩阵

分布区	分类单元	特征序列
A	1	1000001
B <sub>1</sub>	2	0100011
B <sub>2</sub>	1	1000001
D <sub>1</sub>	4	0001111
D <sub>2</sub>	1	1000001
C	3	0010111

### 1.2.2 广布种在多个分支上 初级 BPA 分析结果中有

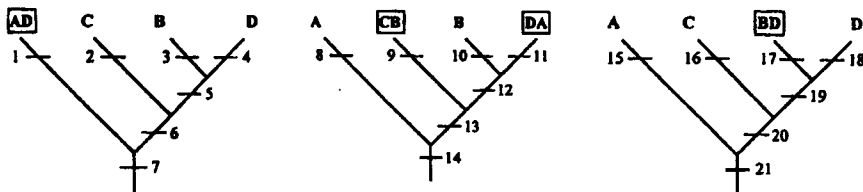


图 11 种 1~4、8~11、和 15~18 的系统发育树及相应的分布状况  
数字表示现生种和假设祖先种(仿 Brooks *et al.*, 2001)

时出现多个分支上广布种的镶嵌现象。BPA 认为这种格局的形成是由于每个种都是一个独立进化系统。Brooks 等人<sup>[4]</sup>认为通过重复区域惯例对广布种的分布区进行反复迭代操作,同样可以选出最简约的区域支序图。

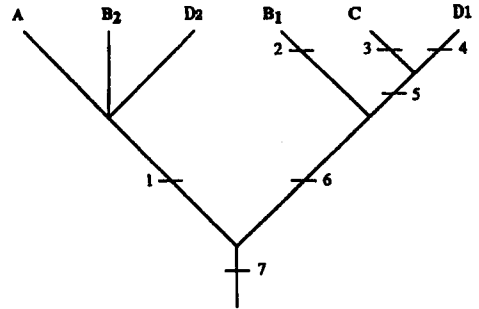


图 10 细化区域 B 和 D 后的二级 BPA 分析  
的区域支序图  
其中分布区 B 和 D 重复

举例来说:假设有 3 个类群各有 4 个种分布在 4 个分布区(图 11),在每个分支图上至少有一个种发生在多个分布区。根据初级 BPA 分析的数据矩阵(表 7)可以得到区域支序图(图 12)。这个共同格局被大部分种支持,但仍有广布分类单元产生的平行演化现象,如种 1、种 9 以及种 11 与它的祖先种(种 12),与 BPA 的假设矛盾。另外,种 13 和种 14 以及种 17 和种 18 在共同格局区域支序图(图 12)中表现的关系与它们在系统发育树(图 11)中的系统发育关系矛盾。

根据共同格局和系统发育关系可以确定分布区 A、D 和 B 需要重复,其中 A 和 B 区分别重复一次,D 区重复两次。基于这种分析建立数据矩阵(表 8),通过简约分析得到新的区域支序图(图 13)。

分析认为这 3 个类群在分布区 A~D 的共同格局为(A(C(B, D))),广布种可能是由于物种形成后 4 次扩散事件造成:种 1 从 A 区扩散到 D 区,种 9 从 C 区扩

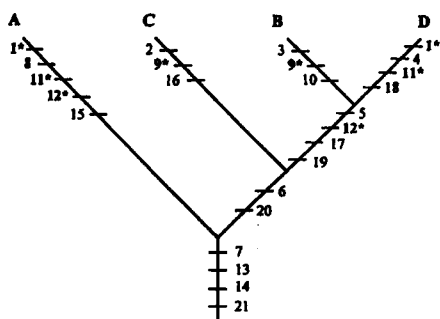


图 12 种 1~4、8~11、和 15~18 的共同格局  
区域支序图

带星号种与 BPA 假设矛盾(仿 Brooks *et al.*, 2001)

表 7 种 1~4、8~11、和 15~18 的初级 BPA 分析的数据矩阵(仿 Brooks *et al.*, 2001)

分布区	分类单元	特征序列
A	1,8, 11, 15	100000110011111000001
B	3,9, 10,17	00101110101110010111
C	2,9, 16	010001101000110100011
D	1,4, 11,17, 18	1001111000111100111111

表 8 细化区域 A、B 和 D 后的二级 BPA 分析的数据矩阵(仿 Brooks *et al.*, 2001)

分布区	分类单元	特征序列
A <sub>1</sub>	1, 8, 15	100000110000011000001
A <sub>2</sub>	11	??????? 0001111??????
B <sub>1</sub>	3, 10, 17	001011100101110010111
B <sub>2</sub>	9	??????? 0100011??????
C	2, 9, 16	010001101000110100011
D <sub>1</sub>	4, 11, 18	0000111100011110001111
D <sub>2</sub>	1	1000001??????????????
D <sub>3</sub>	17	000000000000000010111

“?”表示分布区内数据缺失

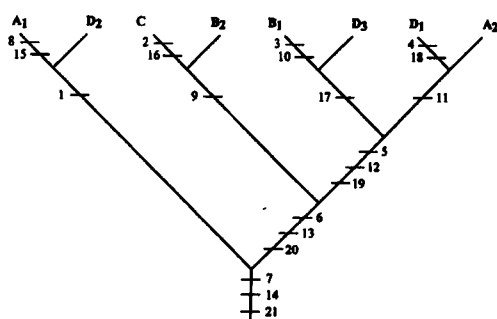


图 13 细化区域 A、B 和 D 后的二级 BPA 分析  
的区域支序图

分布区 A 和 B 分别重复 2 次,分布区 D 重复 3 次  
(仿 Brooks *et al.*, 2001)

散到 B 区,种 11 从 D 区扩散到 A 区,种 17 从 B 区扩散到 D 区。它们都是在种形成以后才扩散的。

## 2 综合实例

任何复杂的历史过程都可以应用 BPA 进行最简约的分析和解释。大多数情况下,用于分析的数据常同时包括分布区缺失、广布种和多余种现象,这使分析变得复杂。假如有 5 个支序图(图 14)。根据初级 BPA 分析的数据矩阵(表 9),得到惟一的共同格局区域支序图(图 15)。从共同格局中可以推断出:种 33 在分布区 D 的缺失可能是由于灭绝事件;种 1、种 9 和种 11 及其祖先种(种 12)发生平行演化现象;种 13 和种 14、种 17 和种 18 以及种 28 和种 29 表现出的关系与它们的系统发育关系不一致;种 23 和种 24 的关系与 BPA 的假设矛盾。通过重复区域惯例分析,获得最终的区域支序图(图 16,表 10 为二级 BPA 分析的数据矩阵),其中 A 分布区重复一次, B 分布区和 D 分布区分别重复三次。

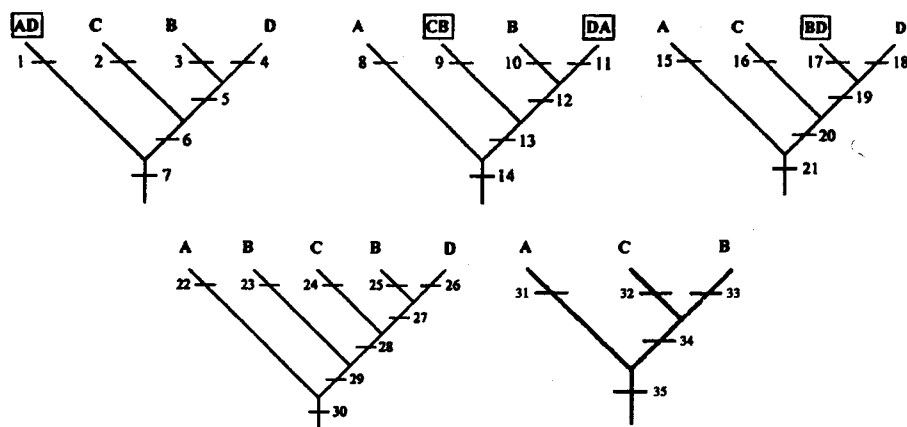


图 14 种 1~4、8~11、15~18、22~26 和 31~33 的系统发育树及分布现状

数字表示现生种和假设祖先种(仿 Brooks *et al.*, 2001)

最后可以得出结论,这 5 个类群的共同格局为 (A (C(B,D))),包括一个灭绝种(D 分布区的种 33),一个通过扩散形成种(种 23 由分布区 A 扩散到分布区 B),4 个成种后扩散现象(种 1 和种 17 分别从分布区 A 和分布区 B 扩散到分布区 D,种 9 由分布区 C 扩散到分布区 B,种 11 由分布区 D 扩散到分布区 A)。

BPA 二级分析尽可能的从所掌握的数据得到一个明确的结果,包括规律性的和非规律性的。但这不一定获得惟一的最简约答案。例如,如果认为上例中 B 区的种 25 假设为不知道或不存在,那么它们的初级 BPA 分析结果相同,但进行二级 BPA 分析后产生两个

等简约的区域支序图。其中一个与图 16 相同,认为种 23 是扩散形成的种;另一个区域支序图认为种 24 是扩散形成的种。

表 9 种 1~4、8~11、15~18、22~26 和 31~33 的初级 BPA 分析的数据矩阵  
(仿 Brooks *et al.*, 2001)

分布区	分类单元	特征序列
A	1,8,11,15,22,31	1000001100111110000011000000110001
B	3,9,10,17,23,25,33	0010111011011100101110101011100111
C	2,9,16,24,32	01000110100011010001100100011101011
D	1,4,11,17,18,26	100111100011110011111000011111?????

“?”表示分布区内数据缺失

表 10 细化区域 A、B 和 D 后的二级 BPA 分析的数据矩阵  
(仿 Brooks *et al.*, 2001)

分布区	分类单元	特征序列
A <sub>1</sub>	1,8,15,22,31	10000011000001100000110000000110001
A <sub>2</sub>	11	???????0001111????????????????????
B <sub>1</sub>	3,10,17,31,33,35	00101110010111001011100010111100111
B <sub>2</sub>	9	???????0100011????????????????????
B <sub>3</sub>	23	?????????????????????010000011?????
C	2,9,16,24,32	01000110100011010001100100011101011
D <sub>1</sub>	4,11,18,25	000111100011110001111000011111?????
D <sub>2</sub>	1	1000001????????????????????????????
D <sub>3</sub>	17	???????????????0010111????????????

“?”表示分布区内数据缺失

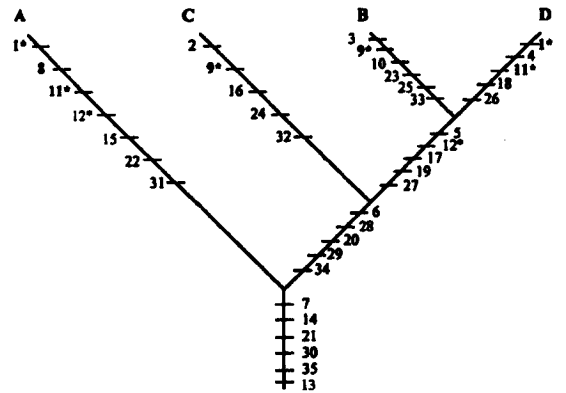


图 15 种 1~4、8~11、15~18、22~26 和 31~33 的共同格局区域支序图  
带星号种与 BPA 的假设矛盾(仿 Brooks *et al.*, 2001)

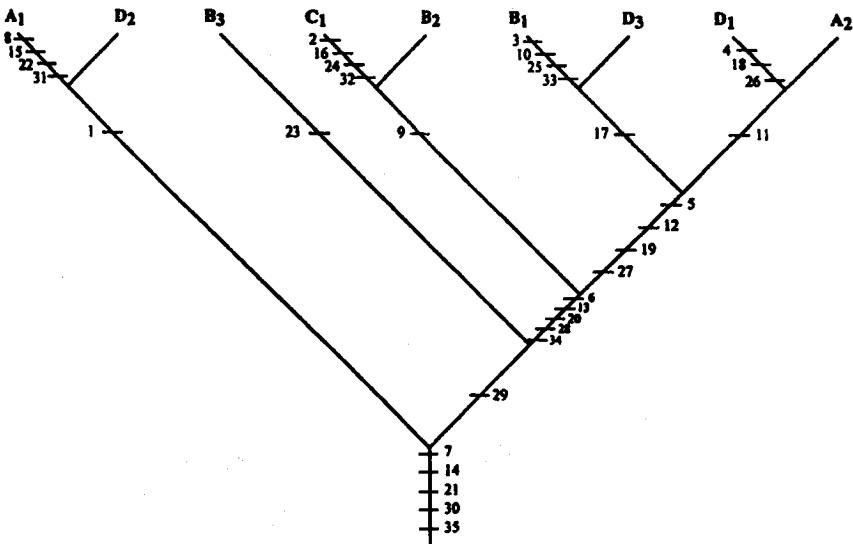


图 16 细化区域 A、B 和 D 后的二级 BPA 分析的区域支序图  
分布区 A 重复 2 次,分布区 B 和 D 分别重复 3 次(仿 Brooks *et al.*, 2001)

### 3 BPA 与其它历史生物地理学方法的比较

如前所述,历史生物地理学中的支序学派,除了 BPA 外,还有组分相容分析(CCA)、组分分析(CA)、协调树分析(RTA)和三域综合分析(TAS)等几种方法。它们分别选择假设 0、1 或 2 进行分析。Morrone 和 Carpenter<sup>[5]</sup>曾以误差程度或一致性为基础对由 CA, RTA, TAS 和 BPA 的分析结果做了比较,认为它们分析得出的区域支序图各有优点,但也有各自的不足之处。

除了在选择假设上的区别外,CA、RTA、TAS 和 BPA 在系统发育信息与生物地理信息发生冲突时解决的方式也不同。CCA 和 BPA 主要以分析后的修正来处理系统发育信息与生物地理信息间的冲突。它们不允许改变分类单元-区域支序图(taxon-area cladogram)的历史关系。

当出现多余分类单元时,CCA 和 BPA 不是简单的把多余类群产生的数据增加到原数据中,而是以类元历史约定<sup>[6]</sup>(taxon history protocol)和类元关系途径<sup>[7]</sup>(taxon relationship approach)解释。CA、RTA 和 TAS 以先验的方法处理这种冲突。它们依据假设 1、2 修剪和添加分类单元与分类单元关系。换言之,这些方法是以假设种的形成过程为基础,寻求合适的解释(如扩散、灭绝),然后再通过相应的地质演化历史事件来验证/推断这种解释的正确性。

致谢 本文写作过程中,曾与加拿大多伦多大学教授, BPA 创始人 D.R. Brooks 博士多次商榷。对他的无私帮助,在此表示诚挚的感谢。

### 参 考 文 献

- [1] 孟凯巴依尔,李枢强,傅金钟.历史生物地理学中的 BPA 分析法 I:初级 BPA. 动物学杂志, 2003, 38(5): 64 ~ 68.
- [2] Brooks D R. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Systematic Zoology*, 1990, 39: 14 ~ 30.
- [3] Brooks D R, Veller M Van, McLennan D A. How to do BPA, really. *Journal of Biogeography*, 2001, 28: 345 ~ 358.
- [4] Brooks D R, McLennan D A. Phylogeny, Ecology and Behavior: A Research Program in Comparative Biology. Chicago: Univ Chicago Press, 1991.
- [5] Morrone J J, Carpenter J M. In search of a method for cladistic biogeography: an empirical comparison of component analysis, brooks parsimony analysis, and three-area statements. *Cladistics*, 1994, 10: 99 ~ 153.
- [6] Hovenkamp P. Vicariance events, not areas, should be used in biogeographical analysis. *Cladistics*, 1997, 13: 67 ~ 79.
- [7] Van Veller M G P, Zandee M, Kornet D J. Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography. *Cladistics*, 1999, 15: 393 ~ 406.