

联合使用 PAUP 和 MacClade 程序 进行支序分析*

孟凯巴依尔^{①②} 李枢强^{①**} 傅金钟^③

(①中国科学院动物研究所 北京 100080; ②新疆大学生命科学与生物技术学院 乌鲁木齐 830046;

③Department of Zoology, University of Guelph, Guelph, Canada)

摘要: 针对一组数据,同时利用 MacClade 和 PAUP 两个软件进行系统发育分析,是支序系统学前沿学者广泛使用的研究方法。本文主要介绍了 PAUP* 4.0、MacClade 4.0 以及两者联合使用的优势,同时介绍了具体的操作方法。

关键词: PAUP; MacClade; 支序分析; 系统发育学

中图分类号: Q-39 **文献标识码:** A **文章编号:** 0250-3263(2002)06-59-04

Phygenetic Analysis Using Both PAUP and MacClade

MENG Kaibaryer^{①②} LI Shu-Qiang^① FU Jin-Zhong^③

(① *Institute of Zoology, Chinese Academy of Sciences Beijing 100080;*

② *College of Life Sciences & Technology, Xinjiang University Urumqi 830046, China;*

③ *Department of Zoology, University of Guelph Guelph Canada)*

Abstract: MacClade and PAUP are useful and popular programs for phylogenetic analysis. This paper describes major features of PAUP* 4.0 and MacClade 4.0 and introduces basic manipulation for using them together.

Key words: PAUP; MacClade; Cladistic analysis; Phylogeny

目前流行的系统发育分析软件各有特色。PAUP* (phylogenetic analysis using parsimony* and other methods) 和 Hennig 86 便于搜索优化树; MacClade 和 CAIC (comparative analysis of independent contrasts) 便于性状进化分析; TreeMap 和 TreeView 便于分支树研究。其它软件如 Clustal V、PUZZLE、COMPONENT 以及 CLADOS 等也有不少优点。但在实际操作时,包括数据录入、分支树搜索和结果分析等,单独使用这些软件常有不能令人满意的方面。因此对同一组数据,同时使用两个或更多软件进行分析是不少支序系统学前沿学者的选择。

若干年来, MacClade 和 PAUP 的组合成为一种重要

的系统发育分析软件组合方式^[1-5]。特别是 MacClade 4.0 和 PAUP* 4.0 版本出现以后,由于两者的兼容性进

* 中国科学院知识创新工程生物领域重要方向项目“中国及其邻近区域生物区系及生物地理学研究”(No. KSCX2-1-06A), 国家自然科学基金资助项目(No. 39970102), 部分得到中国科学院生命科学与生物技术特别支持费(No. STZ-00-19)资助;

** 通讯作者;

第一作者介绍 孟凯巴依尔, 男, 26岁, 硕士研究生; 研究方向: 资源动物学; E-mail: lisq@panda.ioz.ac.cn.

收稿日期: 2002-05-10, 修回日期: 2002-09-23

一步得到加强,受到更多学者的欢迎和广泛使用^[6,7]。本文拟就 PAUP、MacClade 以及两者的合并操作予以介绍,以期引起国内支序系统学者更为广泛的注意。

本文介绍的 MacClade 和 PAUP 是指 Macintosh 版本。作者对这两个软件的评论是基于 Macintosh Power Mac G4466 和 iBook/500 机型, OS9.2 操作系统。使用其它机型或操作系统结论可能有所不同。

1 PAUP* 4.0 及其特点

PAUP* 4.0^[8] 版本是继 PAUP 3.1 后的升级产品。它囊括了 PAUP 以前版本的所有优点,包括把组分分析和 Farris 的定量简约法结合在一起、方便的用户对话系统、完善的文件汇编特点、具有自显菜单和对话框、内置的正文编辑程序、多种简约模式、能进行分支交替分析以及能与 MacClade 相互衔接等。

PAUP* 4.0 同时还增加了许多新的特点。如多种多样的输入输出格式。用于分析的文件可以用 PHYLIP、Hennig86、CGC MSF、MEGA、NBRF-PIR、Text 及 FREQPARS 等格式输入,分析完成后可以以 PHYLIP、Hennig86、PICT、Text 或 NEXUS 等格式输出。另外,PAUP* 4.0 除了包括多种优化标准外,如最大简约(maximum parsimony)、最大似然(maximum likelihood)、距离(distance)等,还囊括了多种搜索系统树算法,如穷尽搜索(exhaustive search)、启发式搜索(heuristic search)、分支限界搜索(branch-and-bound search)、邻居连接法(neighbor-joining)和使用算术平均的不加权对群法(UPGMA, unweighted pair group method with arithmetic mean)等。

PAUP* 4.0 提供数学方法对模拟出的进化树进行评估,减少支序分析过程中性状选取和算法选择的人为影响。不同的算法有不同的适用检测方法,PAUP* 4.0 包括十几种检测程序,如用于检测系统发育信号检测的 PTP(permutation tail probability)和 gI(穷尽搜索的统计结果之一),用于数据联合性检测的 Partition Homogeneity, 确定分支可靠性的 Bootstrap, Jackknife, Decay Index 和 T-PTP 以及检测数据在不同树中效应的 Compare-2, Winning Sites, Templeton Test 和 Kishino-Hasegawa 等。

PAUP* 4.0 根据分子生物学、形态学及行为数据得出的结论不仅仅局限在解释进化生物学领域的系统演化关系,在结合一定的程序或分析方法之后还可以用于保护生物学、生态学和生物地理学等研究领域。

2 MacClade 4.0 及其特点

MacClade 4.0^[9] 是目前 MacClade 系列的最高版本,是针对 Macintosh 设计的系统发育软件。经过 20 多年

的努力,这个程序在数据编辑和性状进化分析方面有明显的优势。

MacClade 4.0 包括数据编辑、树图描述分析和表格三种操作窗口。前两种操作窗口主要用于数据输入、编辑、系统分析以及形成树的图和表,后一种操作窗口主要用于分类分析。

MacClade 4.0 同时是一个优秀的编辑工具,其数据编辑操作窗口可以支持多种标准格式,因此输入的数据类型广泛,包括从网络下载的分子序列和手动设置的行为和形态数据。这些数据可以是离散的或是连续的。数据编辑过程允许操作者进行加权、包括或排除分类单元或性状、定义外群以及挑选编码格式等。同时,为适应系统发育分析中越来越多的分子序列方面的数据,包括 DNA 序列、RNA 序列、蛋白质和核酸序列等,4.0 版的 MacClade 还增加了许多针对性的编辑功能,如对序列进行对齐和移动操作、碱基间的替换、核苷酸序列与蛋白质序列的转换、调整蛋白质序列、对差额型和缺失型数据的处理等等。每种操作基本上都有相应的快捷方式。

MacClade 4.0 还有一个直观的树图分析界面。在这个界面可以观察所有搜索到的树,并分析性状进化格局,确定每一个性状状态最优化的进化过程。在 MacClade 4.0 的树界面中专门为操作树和描述性状信息提供了一系列工具,它们可以对树赋根、移动分支、取消分支、除去分类单元,或修剪全部的分支。这些功能对解释性状集中现象(平行进化或性状逆转)很有意义。MacClade 可以用多种形式显示拓扑树,如种系图(phylogram)通过分支长度的变化表示性状变化次数;分支图(cladogram)则着重显示类群间的关系。MacClade 还可以列出详细的统计结果,如一致性指数(consistency index, CI)、保留指数等(retention index, RI)、最大最小可能步长以及序列中核苷的百分比等。

MacClade 4.0 的性状演化分析主要是通过 Trace 和 Chart 菜单完成。所有性状的重建变化(reconstructed changes)可以通过 Trace 菜单中 Trace all changes 命令显示在树图上。通过这个命令可以在拓扑图上看到性状分布和分支长度的变化,而且 MacClade 4.0 对每个性状都给出一个简明的解释。Chart 菜单中的选项可用图或表的形式表示当前树或所有树中性状与步长、CI、RI 的关系、性状变化统计量和比率以及比较两个不同树间的差别。通过这些图和表所给出的信息,研究人员可以推测出性状的进化规律。

MacClade 4.0 所具有的伸缩式界面和拖拉式菜单,方便了与其它程序的交互使用,例如 PAUP、Mega、Nona、

TNT 和 Winclada 等。

3 联合 PAUP* 4.0 和 MacClade 4.0 的优势

MacClade 4.0 和 PAUP* 分别是目前上百种支序系统学软件中最有特色的程序之一。MacClade 侧重于性状进化的研究,便于输入数据和观察树图,但优化树的搜索功能较差。PAUP* 则侧重优化树的搜索、数据和树的检测,但数据输入和对树的分析比较繁琐,所得的结论仅限于各类群之间的关系,对性状间的关系不能做明确的解释。MacClade 4.0 与 PAUP* 的联合使用可以互补各自的不足,发挥各自的优势,完成一个完整的系统发育分析过程。PAUP 和 MacClade 的联合使用主要体现在 Nexus 文本的管理、强制树的建立和假设、衰减指数的推算等方面。

3.1 Nexus 文本的管理 MacClade 4.0 可以操作在数据文件中包括的各种 NEXUS 文本,包括 PAUP* 4.0 程序。PAUP* 4.0 执行这种文件时允许对其进行编辑命令^[10]。

3.2 强制树的建立和假设 强制树是在 MacClade 4.0 的树图分析界面中选取一个树进行调整后通过 PAUP 再次搜索产生的一组新的等简约树,是人为调整分析结果的一种方法。当系统发育分析结果与研究者的预期有较大的矛盾时,有些研究者常常根据自己掌握的资料进行这种操作。强制树的建立需要有充分的理论和事实依据后才能执行,否则分析的结果难以令人信服。

3.3 推算延迟指数 尽管 MacClade 4.0 本身不能进行衰减指数(decay index)计算(也称 bremer 支持值),但它可以建立一个包含文本命令的 PAUP 文件,在 MacClade 树界面中计算所有树的衰退指数。在 PAUP 中执行这个文件,PAUP 将执行所有的相关搜索用以计算衰退指数。

由于 PAUP* 4.0 与 MacClade 4.0 都是针对 Macintosh 设计的,两个软件的兼容性较强。尽管这两个程序在一定条件下可以在 Windows 运行,但只有在 Macintosh 下它们的操作界面最为便利,操作过程最为简单,而且只有在 Macintosh 操作系统下 MacClade 的功能才能完全体现出来。

PAUP* 还可以与其它程序联合使用,如 Treeview 和 MrBayes 等,但就性状进化分析功能来看,它们的结合都比不上与 MacClade 的结合。与 Treeview 结合的 PAUP* 仅能进行树的研究,与 MrBayes 的组合只是为 PAUP* 增

添了一种新的算法——贝叶斯判决规则法(Bayesian method)。

4 联合 PAUP* 4.0 和 MacClade 4.0 的操作方法

Macintosh 版本的 PAUP* 4.0 和 MacClade 4.0 的合并操作并不复杂。首先在 MacClade 4.0 数据处理界面下编辑文件、设定假设;然后在 PAUP* 4.0 界面中打开设定好的文件,检查设定后进行优化树搜索;分类单元间的关系可以在 PAUP* 界面下观察或在 MacClade 4.0 树图分析界面中观察;最后在 MacClade 4.0 树图分析界面下对产生的等简约树或合意树进行比较、分析,找出性状进化规律。详细过程如下所示。

(1) 打开 MacClade

- * 建立新文件,进入数据编辑界面
- * 输入数据(性状、性状状态、原始数据矩阵)
- * 定义假设(性状加权值、状态顺序)
- * 保存

(2) 打开 PAUP

- * 打开上述保存的文件
- * 定义外群、设置简约方法
- * 搜索优化树
- * 保存搜索到的树
- * 计算合意树并保存

(3) 在 MacClade 的树图分析界面下

- * 打开保存的树,观察后根据需要强制安排分类单元
- * 执行“Place PAUP Command”命令并保存强制安排,或在 PAUP 中使用“load constraint”命令

(4) 在 PAUP 界面下

- * 运行搜索命令,包含或排除上述定义的强制关系
- * 保存搜索的最短树(强制树)

(5) 在 MacClade 树图分析界面下

- * 打开合意树,运行“Decay Index PAUP”命令
- * 保存操作

(6) 在 PAUP 界面下

- * 打开并执行“PAUP Decay Commands”文件
- * 保存操作,分析分支的衰减指数

(7) 在 MacClade 树图分析界面下

- * 选择首选分析树,分析分类单元和性状进化关系
- * 保存结果
- * 编辑、打印

应该说明的是:第一,并不是每一次分析都必须遵从上述过程,例如研究者可以取消强制树搜索(第3、4步)或将衰减指数检验(第5、6步)替换为其它检测手段,如 Bootstrap, PTP 等;第二, PAUP* 4.0 和 MacClade 4.0 交互使用的文件必须首先进行存盘处理,因为无论是 PAUP* 4.0 还是 MacClade 4.0 读取数据和执行程序都是通过硬盘进行,而分析结果放在随机存储器(RAM, random access memory)内。PAUP* 4.0 不能从随机存储器中读取 MacClade 4.0 的数据,同时 MacClade 4.0 也不能从随机存储器中读取 PAUP* 4.0 的数据。只有当分析文件被保存后,变动的数据被记录到硬盘上, MacClade 4.0 和 PAUP* 4.0 的文件才能交互使用。

5 存在的问题

这两个程序由于在缺省方面的要求、计算方法以及数据文件的格式等方面有所不同,在一定程度上不完全兼容。例如 MacClade 4.0 不把缺省性状作为额外性状状态解释, PAUP* 4.0 不支持地层学的性状类型。MacClade 4.0 的说明材料对不兼容的内容提供了详细的表述^[6]。

参 考 文 献

- [1] Beutel R G, Haas F. Phylogenetic relationships of the suborders of Coleoptera (Insecta). *Cladistics*, 2000, 16: 103 ~ 141.
- [2] Blaxter M L, Ley P D, Garey J R *et al.* A molecular evolu-

tionary framework for the phylum Nematoda. *Nature*, 1998, 392: 71 ~ 75.

- [3] Eisen J A. The RecA protein as a model molecule for molecular systematic studies of Bacteria: comparison of trees of RecAs and 16S rRNAs from the same species previously. *Journal of Molecular Evolution*, 1995, 41(12): 1105 ~ 1123.
- [4] Koenemann S, Holsinger J R. Phylogenetic analysis of the amphipod Family Bogidellidae s. lat., and revision of taxa above the species level. *Crustaceana*, 1999, 72 (8): 781 ~ 816.
- [5] Schram F R, Hof C H J. Fossils and the interrelationships of major crustacean groups Arthropod Fossils and Phylogeny. In: Gregory D E ed. *Arthropod Fossils and Phylogeny*. Columbia University Press, 1998. Chapter 6: 233 ~ 302.
- [6] Edgecombe G D, Wilson G D F, Colgan D J *et al.* Arthropod cladistics: combined analysis of histone H3 and U2 snRNA sequences and morphology. *Cladistics*, 2000, 16: 155 ~ 203.
- [7] Hanley R. Phylogeny and higher classification of Hoplandriini (Coleoptera: Staphylinidae: Aleocharinae). *Systematic Entomology*, 2002, 27, 301 ~ 321.
- [8] Swofford D L. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Method). Version 4. Sinauer Associates, Sunderland, Massachusetts. 1998.
- [9] Maddison D R, Maddison W P. MacClade 4: analysis of phylogeny and charater evolution. Version 4.0. Sinauer Associates, Sunderland, Massachusetts. 2000.
- [10] Maddison D R, Swofford D L, Maddison W P. NEXUS: an extensible file format for systematic information. *Systematic Biology*, 1997, 46: 590 ~ 621.