

利用 DNA 序列构建系统树的方法

李 涛¹, 赖旭龙¹, 钟 扬²

(1. 中国地质大学地球科学学院, 武汉 430074; 2. 复旦大学生命科学学院, 上海 200433)

摘要:利用 DNA 序列进行系统发生分析是分子进化研究的必要手段。构建系统树的方法有距离法、简约法、最大似然法以及贝叶斯推断法等。要解决特定的系统发生问题, 首先要挑选合理的分类群及序列, 尽量减少数据的偏倚, 然后选择构树方法, 最后还要对结果进行评价并给出进化学上的解释。通过讨论挑选数据的原则及存在的问题, 介绍了几种构树方法的基本原理及步骤, 并列出了它们的优缺点。

关键词:系统树; 距离法; 简约法; 最大似然法; 贝叶斯推断法

中图分类号: Q349

文献标识码: A

文章编号: 0253-9772(2004)02-0205-06

The Methods of Constructing Phylogenetic Trees with DNA sequences

LI Tao¹, LAI Xu-Long¹, ZHONG Yang²

(1. Faculty of Earth Sciences, China University of Geosciences, Wuhan 430074, China;

2. School of Life Sciences, Fudan University, Shanghai 200433, China)

Abstract: Construction of phylogenetic trees is a key means in molecular evolutionary studies. The methods of constructing phylogenetic trees include the distance-based methods, parsimony, maximum likelihood, and Bayesian inference methods. To resolve a special problem about phylogeny, several notices are necessary: first, to select the reasonable data at less bias as possible; second, to choose the proper method to reconstruct phylogenetic tree; third, to evaluate the conclusions and explain them on the field of evolution. The present paper provides a brief introduction of the principles of data selection and tree-construction methods, and discusses about their advantage and disadvantage points.

Key words: Phylogenetic Tree; Distance-based Method; Parsimony; Maximum Likelihood; Bayesian Inference

自达尔文以来, 重建生物进化史成为进化生物学家们的梦想。由于化石保存的不完备性使得由化石记录推导出的谱系树缺乏中间环节, 利用现存物种的比较形态和比较生理学的研究大致填补了化石谱系树的空缺, 但对分类单元何时与最近祖先分歧等细节性问题含糊不清。随着分子生物学的发展, 人们发现生命的密码蕴涵在 DNA 链中, 四种核苷酸的排列变化反映了进化信息。利用 DNA 序列进行发育分析就是要推断并评价分子水平的进化关系, 并用分支图的形式表现出来, 这种图就是系统发生

树, 简称系统树。系统树可分有根树和无根树, 前者以外类群作为树根, 后者无外类群为树根。有根树的根节点为全部分类单元最近共同祖先, 它反映了分类单元间的进化关系, 而无根树仅反映出分类单元间的分类关系。无根树可通过加入外类群或利用分子钟理论, DNA 不可逆取代模型推导的方法转化为有根树。

1 数据的选取

目前, 分子数据的大量涌现为系统发生分析提

收稿日期: 2002-12-30; 修回日期: 2004-01-29

基金项目: 国家自然科学基金项目(40172005)资助[Supported by National Nature Science Foundation of China (No. 40172005)]

作者简介: 李 涛(1980-), 男, 硕士研究生, 研究方向: 分子古生物学。Tel: 027-62008478, E-mail: giantboyli@163.com

通讯作者: 赖旭龙(1964-), 男, 教授, 博士生导师, 研究方向: 分子古生物学, 微体古生物学。Tel: 027-62867063, E-mail: xllai@cug.edu.cn

供了丰富的素材,但并非所有的数据都适合对特定问题的分析,在构建系统树时要求事先作出一个优先决定:哪些数据是合理的,哪些是不合理的。由于事先并不知道真实的进化关系,又如何选取合理的数据呢?

若选取分类群过少,则有可能遗漏一些关键的分类群,分析结果将十分粗糙,甚至是错误的。因而在原有基础上增加一些额外的分类群,这样可以消除“长枝吸引”,提高分子系统树的可靠性^[1],并抵消非同源相似或误解带来的影响;然而分枝长度估计偏倚也会随分类群的增加而放大,当选取的进化模型能无偏倚地调节进化力(如突变)时能减弱这种影响^[2]。不过,每增加一个分类群,就会导致内部分枝的增加,对内部分枝不一致估计的可能性也增大,如果对进化问题的研究不需要大量的数据时,最好选取少些^[3],而并非是分类群越多则越接近“真实树”。在甄选分类群时还依赖于比较解剖、个体发育以及化石记录的背景知识。

对于分子性状而言,序列长度过短则不能全面地反映一个基因或基因组的进化情况,但序列越长,也越可能得出错误的拓扑结构^[4]。假定分析的是一个进化速率快的基因,加入另一个进化速率快的基因时可以得到一致的结论,但如果加入的是一个或一些进化慢的基因呢?于是有人提出能否假想出一个 Felsenstein 域,增加该域中的“好”基因时能得到正确的结果^[5]?然而基因并不能简单的划分“好”“坏”,从 Felsenstein 域中挑选性状并不意味着可以两者选一,因而应谨慎挑选分子性状。

2 构树方法

构建系统树的常用方法有 3 大类^[6]:距离法、简约法、似然法。一种称为贝叶斯推断的统计学方法也开始使用。

2.1 距离法

不同 DNA 域进化模式不一致,通常用于系统发生分析的 DNA 序列均来自编码区,若不考虑密码子的简并,并假定所有位点的替代速率均相同,就能根据核苷酸替代模型估算出进化距离(如 JC 距离、Kimura 距离等)。若替代速率服从 Γ 分布,则为相应的 Γ 距离。距离法首先估算序列两两间的进

化距离,然后根据不同的算法,从进化距离最短的开始依次聚类,利用距离方阵计算出最优树,或将分枝长度之和最小化,获得优化树。

依据不同的聚类算法,距离法可分为 4 类:UPGMA 法、Fitch-Margoliash 法、最小进化法、邻接法。

2.1.1 UPGMA 法

UPGMA 法采用最小距离法聚类:首先计算各分类群两两间的进化距离,将进化距离最小的两个分类群(如 A、B)最先聚合,将 A、B 合并成一个聚合群 C,用 C 代替 A 和 B,计算 C 与剩余分类群两两间的进化距离,再次将距离最小的两个分类群聚合,合并为一个聚合群 D,依此循环计算,直至所有分类群都包含在一棵树中。

2.1.2 Fitch-Margoliash 法

Fitch-Margoliash(FM)法首先估计树中各长度:如果仅有 3 个分类群,先假定分类群 A 和 B 之间的进化距离最小,然后与 C 聚合;如果多于 3 个分类群,则将其他分类群合并为 C,AB 距离不变,A(或 B)与 C 的距离 d_{AB} (或 d_{BC})为 A(或 B)与 C 中所有分类群的距离算术平均值。

根据图 1 的拓扑结构,AB 间的距离为分枝 x 和 y 之和,即 $d_{AB} = x + y$,同样有 $d_{BC} = y + z$; $d_{AC} = x + z$,解出 x, y, z ;将 AB 合并成一个新的分类群,用类似方法计算出新的 x, y, z 值,依此可计算出所有分支长度的估计值。考虑所有可能的进化树的残差平方和^[7]:

$$R_s = \sum_{i=1}^n \sum_{j=1}^n \frac{(d_{ij} - \hat{d}_{ij})^2}{d_{ij}^2}$$

其中 n 为分类群个数(以下同), d_{ij}, \hat{d}_{ij} 分别为分类群 i, j 间的观察距离和估计距离¹,取 R_s 值最小的为最优树。

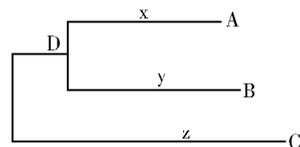


图 1 Fitch-Margoliash 法示意图

Fig. 1 Sketch map of Fitch-Margoliash method

2.1.3 最小进化法

最小进化法(ME)的理论基础源于 Rzhetsky 和 Nei 的数学证明^[8]:当距离使用无偏估计时,树的真

1 连接两分类群的所有分支长度估计值之和。

实拓扑结构的分支长度和的期望值最小。对分枝长度的估计除 Fitch-Margoliash 法外,还可使用最小二乘估计, Rzhetsky 和 Nei 给出了一种快速算法^[8,9];分类群两两间的距离、抽样误差和分支长度分别构成列向量 $\vec{d}, \vec{\epsilon}, \vec{L}$, 用矩阵 A 代表树的拓扑结构矩阵, 行对应于分类群的两两配对, 列对应于分支序号; 连接分类群 i, j 间的路径包含第 k 个分支时, 矩阵元素 $a_{(ij)k}$ 为 1, 否则为 0, 有:

$$\vec{d} = AL + \epsilon$$

抽样误差服从平均值 0 和方差 $V(d_{ij})$ 分布, 令 $T = (A^T A)^{-1} A^T$, 向量 T 的第 i 行为向量 \vec{T}_i , 则第 i 个分枝长度 (\hat{L}_i) 的估计为:

$$\hat{L}_i = T_i \vec{d}$$

比较所有分支长度估计的和 S , 值最小的为最优树, 或对任意两个树 A、B 分支长度和的差 $S_B - S_A$ 进行零假设检验^[9]; 若差值大于 0, 则 A 比 B 优越; 若差值小于 0, 则 B 比 A 优越; 差值等于 0 时则不能判断 A、B 树的好坏, 对所有的树两两对比, 最终获得最优树。

2.1.4 邻接法

邻接法(NJ)从星形树出发, 假定 X 为共同祖先(图 2), 可计算出星形树的分支长度和:

$$S_0 = \sum_{i=1}^n L_{iX} = \frac{1}{n-1} \sum_{i < j}^n d_{ij}$$

定义由单一的内节点连接的分类群为一对邻居, 假定分类群 1、2 为一对邻居(图 3), 利用最小二乘法估算树的分支长度和 S_{12} ^[10]:

$$S_{12} = L_{XY} + L_{1X} + L_{2X} + L_{XY} =$$

$$\frac{1}{2(n-2)} \sum_{i=3}^n (d_{1i} - d_{2i}) + \frac{1}{2} d_{12} + \frac{1}{n-2} \sum_{3 < i < j}^n d_{ij}$$

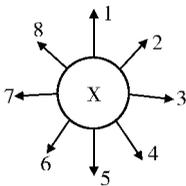


图 2 NJ 的星形树

Fig. 2 A starlike tree in NJ method

由于事先不知哪两个分类群为一对邻居, 因而首先应分别计算任意两个分类群为第一对邻居时树

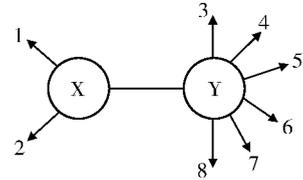


图 3 分类群 1 和 2 相聚的树

Fig. 3 A tree with OTUs 1 and 2 are clustered

的分支长度和, 选取值最小时的两个分类群作为邻居, 连接它们的内节点为第一个内节点, 然后在剩余分类群中选取任意两个分类群为第二对邻居, 类似地找到第二个内节点, 依此, 直至找到所有内节点。确定拓扑结构后采用 Fitch-Margoliash 法估算树的分支长度^[10]。

2.2 最大简约法

最大简约法(MP)理论基础来自 Occam 剪刀哲学原理: 对一现象的解释, 应选择最简单(假设最少)的解释而不是更复杂的。对系统发生分析而言, 该法则有两层含义^[11]: 一是用最少的进化事件(如突变)去解释观察到的数据; 二是在任何进化模型或机制下, 假设尽可能少。具体来说最优树就是在解释整个进化过程时, 全部位点的最小核苷酸替代数之和最小的树。

假定有四条序列 i : AGGGTAACTG; j : AC-GATTATTA; k : ATAATTGTCT; l : AATGTT-GTCG。可构建出 3 种无根树(图 4):

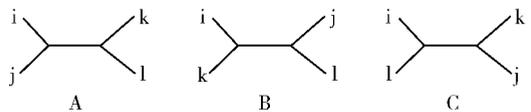


图 4 3 种可能的四分类群无根树

Fig. 4 Three Possible Unrooted Trees with Four OTUs

将四条序列相对应的各位点核苷酸依次代入树的相应位置, 遍历所有可能进化路径, 分别计算出其中最小的核苷酸替代数: 树 A(0 3 2 2 0 1 1 1 1 3); 树 B(0 3 2 2 0 1 2 1 2 3); 树 C(0 3 2 1 0 1 2 1 2 3), 树 A 中全部位点的最小核苷酸替代数之和最小, 为最优树。上述的计算仅位点 4、7、9 对构建 MP 树提供了有用的信息, 其他位点在 3 种拓扑结构中都有一致的最小核苷酸替代数, 为简化计算, 要求选取的位点为信息位点^[11]。当分类群较少时, 可穷尽式

2 对构建 MP 树能提供信息的位点, 该位点至少有两种不同的核苷酸, 每种至少出现两次。

搜索所有可能的树从而找到最优树;当分类群较多时,则只搜索那些可能正确的树,可以采用分支限界法和启发式搜索两种搜索方法。与距离法不同,MP法利用每个可变位点的所有进化路径的分支(包括内部或外部)的核苷酸平均替代数估算分支长度^[8]。考虑到转换比颠换更常见,对转换赋予相对高的权重,即加权 MP 法。Lake 提出进化简约法(EP)与 MP 法相关,不过进化简约法的原则是寻找与树的外部分支一致的最小核苷酸替代数^[12],四种核苷酸的替代情况分别用 1 到 4 之间的数赋值^[13],然后,计算四条序列的 3 种可能树型净支持率,最后进行统计显著性检验^[11]。

2.3 最大似然法

最大似然法(ML)基于两条基本假设:不同的性状进化是独立的;物种发生分歧后进化独立。单个位点的似然值是指在核苷酸替代模型中该位点每个可能被取代或再现的概率之和,将所有位点似然值相乘就得到进化树的似然值。

下面以 Felsenstein 替代模型为例计算位点的核苷酸替代概率,假定某位点的核苷酸为 i ,分枝 m 上的核苷酸替代数为 V_m 时,核苷酸 i 在整个序列中的相对频率为 g_i ,该位点变为核苷酸 j 的概率 $P_{ij}(v)$ ^[14]:

$$P_{ij}(v) = \begin{cases} 1/4 + (1 - g_i) e^{-v} & (j = i) \\ 1/4(1 - e^{-v}) & (j \neq i) \end{cases}$$

若有四条序列在某位点上的核苷酸分别为 i, j, k, l ,节点 x, y, z 为四种核苷酸中的任一种,在图 5 的有根树中(无根树也采用类似的计算方法),假定节点 x 为共同祖先,该位点的似然值为:

$$f(i, j, k, l) = \sum_x \sum_y \sum_z g_x P_{xy}(v_5) P_{yi}(v_1) \cdot P_{yj}(v_2) P_{xz}(v_6) P_{zk}(v_3) P_{zl}(v_4)$$

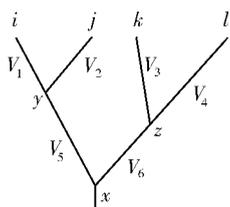


图 5 4 分类群 ML 有根树

Fig. 5 A Rooted ML tree with Four OTUs

将 4 条序列其他对应位点的核苷酸依次代入图 5 相应位置,可计算出各位点的似然值,对所有位点

似然值相乘就得此树的似然值。

ML 法对最优树的搜索有两种策略:一是对所有可能树的似然值对比,找出最大值;二是对树的似然函数最大化,估算分枝长度、核苷酸频率、替代速率等未知参数,从而确定树的拓扑结构。

2.4 贝叶斯推断法

贝叶斯推断法首先涉及两个基本概念:树的先验概率(prior probability)和后验概率(posterior probability)。树的先验概率是指对进化树未进行任何观测时的概率,具体来说就是认为所有进化树都相同的可能性;树的后验概率是指通过观测,进化树的条件概率,即在给定的序列数据条件下,某进化树正确的概率^[15]。因而后验概率最大的进化树为最优树。

假设有 n 条序列,序列比对矩阵为 X , τ_i 代表第 i 棵系统树的参数(包括分枝长度、拓扑结构以及核苷酸替代模型参数),则树的先验概率为 $P(\tau_i)$,树的似然值为 $P(X | \tau_i)$,利用贝叶斯公式可计算该树的后验概率:

$$P(\tau_i | X) = \frac{P(X | \tau_i) P(\tau_i)}{\sum_{j=1}^{B(n)} P(X | \tau_j) P(\tau_j)}$$

其中, $B(n)$ 为所有可能进化树的个数,有根树的 $B(n)$ 值为 $(2n-3)! / [2^{n-2} (n-2)!]$,无根树的 $B(n)$ 值为 $(2n-5)! / [2^{n-3} (n-3)!]$ ^[16]。

要得到最优树,常规的方法是计算出所有可能的进化树的后验概率,对每个进化树还要考虑分支长度与模型参数之间的组合,而这种分析几乎是不可能的^[15]。因此只能采用后验概率的估算方法,最常用的是 MCMC (Markov Chain Monte Carlo) 法,其基本思想是构造出一条马尔柯夫链,该链的状态空间为统计模型参数和不变后验分布参数。链的构造由多步完成,每步状态空间的状态都被推荐为链的下一个连接点。首先在状态空间中随机挑选一个状态作为链的当前态,随机扰动当前态各参数,从状态空间中推荐一个新态,计算推荐态的相对后验概率密度,若推荐态的后验概率密度高于当前态,则链的移动被接受,推荐态则作为下一循环的当前态。若推荐态的后验概率密度低于当前态,则计算由 Metropolis 等和 Hastings 提出的推荐态与当前态后验概率的比率,该值接近 1 时接受推荐态,接近 0 时则拒绝推荐态,此时,当前态作为自身的下一个连接点。对上述过程重复若干次,最终马尔柯夫链将停

留在后验概率高的状态,某态的后验概率就是马尔柯夫链停留该态的时间分值^[15,16]。

3 不同构树方法的优缺点

当序列间的分歧度不高且序列较多时,距离法、简约法、似然法所构建的系统树往往具有相似的拓扑结构^[17],然而,一般应用这些方法所构建的进化树常有拓扑差异,虽然利用自展法³可以大致判断分支是否合理,但自展值低的分支不一定就该舍弃,因此有必要比较几种构树方法,从而对特定的序列选择合适的构树方法。

UPGMA 法在不同谱系间进化速率有较大差异或有同源序列的平行进化时常得出错误的拓扑结构^[18],而且当进化树的状态空间较大时,UPGMA 法的可操作性极差,因而该建树方法的使用极为有限。NJ 法的运算速度最快,但该算法每迭代运算一次均只搜索最近邻居配对,对其他可能的配对不加考虑,最终只生成单一的最优树,可能会遗漏一些拓扑结构更合理的次优树为弥补缺陷,William 等提出遍及邻接法(Generalized NJ),将一些最接近的树也包含在搜索算法中^[19],逐步迭加,期望找到更好的 NJ 树。ME 与 FM 法从所有可能的进化树中挑选树长最短的作为最优树,是距离法建树中两种相对较好的方法。不过,值得注意的是距离法在将原始数据转换成距离矩阵时难免会丢失一些进化信息^[20]。

简约法是一种不依赖任何进化模型的无噪声统计方法^[21],能快速地分析出大量序列之间的系统发生关系,所构建的树中的短分支更接近真实。但简约树的分值完全决定于所有重建祖先序列中的最小突变数,而突变是否按照事先约定的核苷酸最少替代的途径进行是不得而知的,单一的突变图谱可能会得出是似而非的结论。再者,所有分支的突变数不可能相同,由于没有考虑核苷酸的突变过程,使得长分支末端的序列由于趋同进化而显示较好的相似性趋同现象违背了简约法则,导致的结果是对“长枝吸引”的敏感^[22]。因此,当序列单位位点上核苷酸替代数相对较大时,MP 法则极可能得出错误拓扑结构的树^[23]。

最大似然法似乎是几种常用方法中最为近似的方法,它考虑了所有可能的突变路径,能完全利用数据的系统发生信息。然而,最大似然法构建的系统树在很大程度上依赖于对核苷酸替代模型的选择^[24],不同的位点核苷酸替代速率不一致,在核苷酸替代一般模型中包含了反映进化过程的参数,如颠换和置换的偏倚以及替代率变异的分布模型,但并非替代模型越复杂,结果就越理想。似然法运算强度极大,对于分类群较多时十分费时,这也是似然法应用的最大障碍,而且似然法并没有评估拓扑结构的优劣,而是假定分支长度估计最精确的拓扑结构为最优树,实际上,系统发生所关心的是树的拓扑结构,分支长度反而成为干扰参数,忽略分支长度似乎更合理些^[25]。

贝叶斯推断法和最大似然法紧密相关,不过它衡量分支的支持率所用的 MCMC 法要比似然法的自展法要快得多。似然法的目的是寻找参数空间中的最高点,因而最终结果仅取决于某个点的数值,贝叶斯推断法的分析结果则取决于所有可能的参数值,寻找对参数的最广泛的支持,因而能得到比似然法更好的结果,尤其相对参数的数据量较少时,似然法常得出不合理的结论^[22]。当然,应用 MCMC 法估算后验概率时,对于何时才能使得马尔柯夫链运行足够充分,从而得到合适的样本,我们并不知晓。此外,后验概率对先验概率十分敏感,而先验概率分布与进化模型参数(分支长度、拓扑结构、核苷酸替代模型)息息相关,其值完全依赖于数学上的计算,和最大似然法一样都没有考虑生物学背景知识^[16],这一点值得注意。

总的来说,随着分子生物学研究的深入,越来越多序列的进化意义最终会被揭示,对各种构树方法也提出了新的挑战,如何增加算法对复杂数据的处理能力将成为今后系统发生分析的重点。

参 考 文 献 (References):

- [1] Naylor G J, Brown W M. Amphioxus mitochondrial DNA, chromosome phylogeny and the limits of inference based on comparisons of sequences. *Syst Biol*, 1998, 47(1): 61~76.
- [2] Cummings M P, Otto S P, Wakeley J. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol Biol Evol*,

3 自展法对原始数据矩阵随机重抽样形成同长度的假数据集,利用同样的算法对重抽样数据构建系统发生树(自展树)。原始树与自展树相同时,分支赋值 1,否则为 0,分支的自展支持率为赋值 1 的次数百分数。

- 1995,12;814~822.
- [3] Kim J. General inconsistency condition for maximum parsimony; effects of branch lengths and increasing number of taxa. *Syst Biol*,1996,45;363~374.
- [4] Felsenstein J. Cases in which parsimony and compatibility methods will be positively misleading. *Syst Zool*,1978,27;401~410.
- [5] Huelsenbeck J P, Kirkpatrick M. Do phylogenetic methods produce trees with biased shape? *Evolution*,1996, 50;1418~1424.
- [6] LÜ Bao-Zhong. Constructing molecular evolutionary trees. *Zool Research*,1993,14(2);186~193.
吕宝忠. 分子进化树的构建. *动物学研究*, 1993, 14(2): 186~193.
- [7] Fitch W M, Margoliash E. Construction of phylogenetic trees. *Science*,1967,155;279~284.
- [8] [America]Nei M, Kumar S. Molecular Evolution and Phylogenetics. LÜ Bao-Zhong, ZHONG Yang, GAO Li-Ping translate, ZHAO Shou-Yuan, ZHANG Jian-Zhi collect. Higher Education Press, Beijing, 2002,76~163.
[美]根井正利,库马著(吕宝忠,钟扬,高莉萍译,赵寿元,张建之校). 分子进化与系统发生. 北京:高等教育出版社,2002, 76~163.
- [9] Rzhetsky A, Nei M. A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol*,1992,9;945~967.
- [10] Saitou N, Nei M. The neighbor-joining method; a new method for reconstructing phylogenetic trees. *Mol Biol Evol*,1987,4; 406~425.
- [11] ZHONG Yang, ZHANG Liang, ZHAO Qiong. An Introduction to Bioinformatics. Higher Education Press, Beijing, 2001, 116~149.
钟扬,张亮,赵琼著. 简明生物信息学. 北京:高等教育出版社,2001, 116~149.
- [12] Lake J A. A rate-independent technique for analysis of nucleic acid sequence; evolution parsimony. *Mol Biol Evol*, 1987, 4(2);167~191.
- [13] Jin L, Nei M. Limitation of the evolution parsimony method of phylogenetic analysis. *Mol Biol Evol*,1990,7(1);82~102.
- [14] Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol*,1981,17(6);368~376.
- [15] Huelsenbeck J P, Ronquist F, Nielsen R, Bollback J P. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*,2001,294(5550);2310~2314.
- [16] Huelsenbeck J P, Larget B, Miller R E, Ronquist F. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst Biol*,2002,51(5);673~688.
- [17] Saitou N, Imanishi T. Relative Efficiencies of the Fitch-Margoliash, Maximum-Parsimony, Maximum-Likelihood, Minimum-Evolution, and Neighbor-joining Methods of Phylogenetic Tree Construction in Obtaining the Correct Tree. *Mol Biol Evol*,1989,6;514~525.
- [18] LI Jing-Yan. The pdaric method for constructing molecular evolutionary trees from sequences data. *Zool Research*, 1992, 13(4);387~396.
李靖炎. 削偶合今法的提出. *动物学研究*, 1992, 13(4): 387~396.
- [19] Pearson W R, Robins G, Zhang T. Generalized neighbor-joining; more reliable phylogenetic tree reconstruction. *Mol Biol Evol*,1999,16(6);806~816.
- [20] Penny D. Towards a basis for classification; the incompleteness of distance measures incompatibility analysis, and phenetic classification. *J Theor Biol*,1982,96(2);129~142.
- [21] Sourdiss J, Nei M. Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Mol Biol Evol*,1988;5(3);298~311.
- [22] Holder M, Lewis P O. Phylogeny estimation; traditional and Bayesian approaches. *Nature*,2003,4;275~284.
- [23] Li W H. Evolutionary change of restriction cleavage sites and phylogenetic inference. *Genetics*,1986,113; 187~213.
- [24] Mike S, David P. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol Biol Evol*,2000,17;839~850.
- [25] Goloboff P A. Parsimony, likelihood, and simplicity. *Cladistics*,2003,19;91~103.