

DOI: 10.1360/yc-007-0659

## 系统发育研究中“长枝吸引”假象概述

黎一苇<sup>1,2</sup>, 于黎<sup>1,2</sup>, 张亚平<sup>2,1</sup>

1. 云南大学生物资源保护与利用重点实验室, 昆明 650091;  
2. 中国科学院昆明动物研究所细胞与分子进化实验室, 昆明 650023

**摘要:** 系统发育研究(phylogeny)不仅有助于重建地球所有生物体的进化历史, 而且还可以揭示进化生物学领域中的一些基本问题。清晰了解各生物物种进化历程及不同物种之间的进化关系, 是进一步研究和探索生物学其他学科的基础。但是现今广泛应用的所有系统发育分析方法都存在一定的局限性, 在一定程度上不能有效消除各种误差, 从而不能客观地处理和分析数据, 也就不能成功重建生物进化历程, 真实反映物种进化关系。系统发育研究中, “长枝吸引”(Long-branch Attraction, LBA)假象是最为困扰研究者的问题。文章从“长枝吸引”问题的产生原由、检测方法以及消除策略等多个方面进行详尽概述, 并通过列举典型实例, 阐述了解决“长枝吸引”问题的途径。

**关键词:** 系统发育研究; 长枝吸引; 系统发育分析方法

## “Long-branch Attraction” artifact in phylogenetic reconstruction

LI Yi-Wei<sup>1,2</sup>, YU Li<sup>1,2</sup>, ZHANG Ya-Ping<sup>2,1</sup>

1. *Laboratory of Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China;*  
2. *Laboratory of Molecular Evolution and Genome Diversity, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming 650023, China*

**Abstract:** Phylogenetic reconstruction among various organisms not only helps understand their evolutionary history but also reveal several fundamental evolutionary questions. Understanding of the evolutionary relationships among organisms establishes the foundation for the investigations of other biological disciplines. However, almost all the widely used phylogenetic methods have limitations which fail to eliminate systematic errors effectively, preventing the reconstruction of true organismal relationships. “Long-branch Attraction” (LBA) artifact is one of the most disturbing factors in phylogenetic reconstruction. In this review, the conception and analytic method as well as the avoidance strategy of LBA were summarized. In addition, several typical examples were provided. The approach to avoid and resolve LBA artifact has been discussed.

**Keywords:** phylogenetic reconstruction; Long-branch Attraction; phylogenetic methods

收稿日期: 2006-12-26; 修回日期: 2007-02-09

基金项目: 云南省科技厅项目和国家自然科学基金项目资助[Supported by Scientific and Technical Fund Project of Yunnan and Chinese National Natural Science Foundation]

作者简介: 黎一苇(1981-), 女, 重庆人, 硕士, 专业方向: 分子进化。E-mail: azalee\_713@yahoo.com

通讯作者: 于黎(1976-), 女, 辽宁人, 博士, 研究员, 研究方向: 分子进化。Tel: 0871-5033362, Fax: 0871-5034838;

E-mail: yuli1220@yahoo.com.cn

张亚平(1965-), 男, 云南人, 博士, 研究员, 中国科学院院士, 研究方向: 分子进化。Tel: 0871-5190761, Fax: 0871-5195430;

E-mail: zhangyp@mail.kiz.ac.cn

追溯生物界不同生物类型的起源及进化关系,即重建生物类群的系统发育树是进化生物学领域中的一个十分重要的内容。系统发育研究不仅对重建地球生物进化历程,真实反映各物种间进化关系具有重大意义,而且也是进一步研究和探索生物学其他学科的基础。就目前而言,推断物种间系统发育关系已成为分析所有进化问题的必经过程<sup>[1]</sup>。遗憾的是,目前系统发育研究中使用的所有分析方法都有一定的局限性,不能有效消除各种误差。因此,得到的系统发育结果往往不能正确反映所研究物种的真实进化历程。在系统发育分析中,“长枝吸引”(Long-branch Attraction, LBA)假象是最为困扰研究者的问题。“长枝吸引”是指在用系统发育分析方法分析一个有限数据集时,由于高频率的相似变化(如趋同、平行进化)和加速的进化速率等因素的存在使序列达到相同状态而人为地将这些不是来自于共同祖先的序列的代表分类元聚在一起,使这些分类元之间相互“吸引”。因此,在进行系统发育分析时,应尽可能避免“长枝吸引”假象的产生,从而构建出可靠的系统发育树。

## 1 “长枝吸引”(LBA)的概念

“长枝吸引”假象由Felsenstein(1978)首次提出。1978年,Joseph Felsenstein在对由4个分类元组成的数据集进行最大简约分析(Maximum Parsimony, MP)时发现,当4个分类元突变速率显著不同时,进化速率快的分类元会因为形成长枝而互相发生吸引,将原本不是姐妹群关系的分类元错误地聚在一起。如图1所示,假设A, B, C, D 4个分类元具有显著不同的突变速率,其中A和B之间, C和D之间是真实的姐妹群关系。 $p$ 和 $q$ 分别代表各分类元的突变速率,当 $p < q^2$ 时,由于A和C分类元突变速率相对较快,MP分析会将A与C或B与D人为地聚在一起,呈现姐妹群关系。Felsenstein将这种现象称为“长枝吸引”,而受到“长枝吸引”影响的区域称为Felsenstein zone。1989年, Hendy和Penny在Felsenstein的基础上对产生长枝吸引的条件加以扩充,他们认为在分析数据时,即使各个分类元进化速率相等,只要产生的拓扑结构中各分类元枝长不等也同样可以产生“长枝吸引”假象<sup>[2]</sup>。除了进化速率差异以外,非对称的拓扑结构(主要由不同的物种消亡速率和形成速率以及不完全采样形成)也会导致

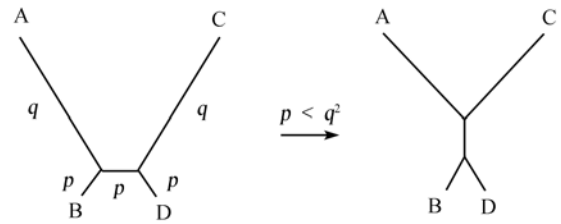


图1 “长枝吸引”假象<sup>[2]</sup>

Fig. 1 The “Long-branch Attraction” artifact<sup>[2]</sup>

不等枝长的产生。

目前的研究不仅对“长枝吸引”假象进行了大量的理论探索,而且还使用真实数据对其进行了详尽探讨<sup>[2,3]</sup>。无论是在种内水平的系统发育研究,如水蚤<sup>[4]</sup>,蜜蜂<sup>[5]</sup>以及甲虫<sup>[6]</sup>,还是种上水平的研究如硬骨鱼属<sup>[7]</sup>,硬骨鱼科<sup>[8]</sup>,昆虫目<sup>[9-11]</sup>,哺乳动物<sup>[12-16]</sup>以及鸟类<sup>[17]</sup>,甚至基于门阶元和纲阶元水平如后生动物<sup>[18,19]</sup>,被子植物<sup>[20]</sup>,种子植物<sup>[21]</sup>,红藻<sup>[22]</sup>以及真核生物<sup>[23-26]</sup>,细菌<sup>[27]</sup>和生命之树<sup>[28-31]</sup>等研究中,都发现了“长枝吸引”假象<sup>[2]</sup>,这在一定程度上干扰了物种系统发育树的正确构建。

## 2 “长枝吸引”假象的产生原因

在系统发育分析过程中,有多种原因可造成“长枝吸引”假象,常见的主要包括以下几个方面:(1)使用的数据集中包含进化速率显著加快的分类元,这些分类元序列在拓扑结构中表现为“长枝”;(2)使用不恰当的或亲缘关系过远的分类元作为外群,外群分类元很有可能与内群中的长枝分类元发生相互吸引从而改变内群内部拓扑结构;(3)使用的数据集中存在趋同、平行等进化特征,系统发育分析会将这些进化特征认为是同源性状(Synapomorphies),将非真实姐妹群关系的分类元人为地聚在一起,造成“长枝吸引”假象;(4)分析的数据集与系统发育分析中所假设的核苷酸替换模型之间不匹配;(5)系统发育分析结果得到非对称拓扑结构。

## 3 “长枝吸引”假象的避免方法

为了尽量避免受到“长枝吸引”假象的影响,系统发育研究者提出了一系列有效的方法<sup>[2]</sup>。最早是将简约法加以改良来分析数据<sup>[32,33]</sup>,随后的研究普遍认为相对于简约法,最大似然法不太容易产生“长枝吸引”假象,因此最大似然法被广泛应用<sup>[35-37]</sup>。但是,也有研究表明最大似然法对于“长枝吸引”来说并不具有绝对免疫力,即使使用似然率检验

(Likelihood ratio test)方法也不能确保不会造成“长枝吸引”<sup>[38]</sup>。因此,就目前来看,还没有一种方法可以完全避免“长枝吸引”假象,一般建议将几种不同的方法结合使用。目前避免“长枝吸引”的方法主要包括:

### 3.1 模型优化法

分析数据时,尽可能考虑位点间替换速率的异质性,通过设定 gamma 分布参数,优化核苷酸替换模型,避免形成“长枝吸引”假象。

### 3.2 排除法

#### 3.2.1 去除序列中第三密码子位点

相对于第一和第二密码子,第三密码子进化速率更快,将其去除可以大大降低由于不同位点进化速率不一致造成的“长枝吸引”假象。但是同时,由于第三密码子所包含的有用信息量往往又是最丰富的,所以第三密码子的去除在一定程度上会影响系统发育结果,降低数据对物种进化关系的解决能力。

#### 3.2.2 去除分类群中进化速率较快的长枝分类元

这种方法已在很多研究中证明可以较好的避免“长枝吸引”,但是如果这些长枝分类元是研究中不可缺少的部分,那么该方法的可行性也有一定的限制。鉴于此,有些系统发育学家建议在进行系统发育分析时应尽可能选用进化速率较慢的分类元,以避免形成“长枝”<sup>[18,28,37]</sup>。

### 3.3 打断长枝法

增加与长枝分类元关系较近的分类元进行系统发育分析,以打断长枝。多数情况下,这种方法能够避免形成“长枝吸引”。但是,如果需要增加的分类元是已消亡物种时,这种方法则无法应用。此外,我们还要意识到即使增加了分类元,打断了长枝,新增加的分类元数据同时也会给系统发育分析带来新的问题<sup>[39,40]</sup>。

### 3.4 无关联数据集整合分析法

同时使用两个或多个无关联的独立数据集进行整合分析,比如形态学数据和分子数据的联合分析<sup>[41-43]</sup>。这种方法的优点是分析结果不再单纯地来源于一个数据集,一个数据集中无法提供的系统发育信息可以通过另一个数据集弥补,使分析结果更为可靠,同时避免了诸如“长枝吸引”等进化噪音的形成。

## 4 “长枝吸引”假象的检测策略

在系统发育分析时,如果怀疑所得结果可能受到“长枝吸引”影响而呈现错误的系统发育关系(拓扑结构),那么就应当对系统发育结果进行检验。目前,检测“长枝吸引”的方法主要包括以下几种<sup>[2]</sup>:

### 4.1 单独比较分析法(Separate partition analysis)

这种方法主要包括4个层面的内容:(1)相对于分子数据,形态学数据不易形成“长枝吸引”假象,因此可以首先将形态学数据和分子数据进行单独分析,如果得到不一致的结果,那么分子数据很有可能受到了“长枝吸引”的影响。当然,这种方法只能作为形成“长枝吸引”假象的一个佐证,因为不一致的研究结果也可以来源于物种树和基因树之间的拓扑结构差异,并不能说明一定受到了“长枝吸引”假象的影响。(2)通过分析比较不同基因的进化速率,检测进化速率较快的基因是否将长枝分类元聚在一起<sup>[44]</sup>。(3)将进化速率较快的序列位点单独形成矩阵与进化速率较慢序列位点形成的矩阵进行比较分析,如果形成相冲突的拓扑结构,那么进化速率较快的序列位点很有可能造成了“长枝吸引”假象。同时也可以去除进化速率快的位点进行重新分析,观察所得结果是否有差。(4)分析编码蛋白的序列时,可以将进化速率明显不同的第三密码子与第一、二密码子分开分析,或者是将第三密码子去除以后重新分析,检验结果是否一致。由于同一密码子中处于不同位置的核苷酸不可能有着不同的进化历史,因此可推断不一致结果可能是分析方法的不恰当以及随机误差或是人为错误引起的。

总之,当研究者使用其他方法检测到可能形成了“长枝吸引”的前提下,单独分析方法可以通过启发式的分析方法较好地指引研究者确定那些受到“长枝吸引”影响的数据,提供一些有价值的研究信息。而且在分类元采集以及如何选取各种进化特征进行系统发育分析方面,这种方法也有一定的指导意义。

### 4.2 长枝抽提法(Long-branch extraction)

长枝抽提法是基于“长枝吸引”假象产生的一个必要条件而提出的一种有效的检测方法。由于只有在树图结构中存在两个或两个以上的长枝分类元时“长枝吸引”才有可能发生,因此,可以先保留一个

长枝分类元, 同时将剩余长枝分类元去除, 进行分析得到一个结果, 然后再将另一长枝分类元保留, 去除其余的长枝单元, 进行分析得到另一个结果, 以此重复分析, 最后将各个长枝分类元单独分析得到的结果进行比较。

长枝抽提法简单易操作, 而且与去除第三密码子或去除进化速率较快位点的方法比较, 这种方法尽可能地保留了有效信息, 丢失的信息也可以通过增加分类元以弥补。已有很多研究证实此方法可以有效检测到“长枝吸引”假象<sup>[6,45]</sup>。

#### 4.3 随机生成序列检测法(Adding artificial LB sequences)

在系统发育分析中, 当使用亲缘关系较远的分类元作为外群进行有根树分析时, 外群枝长越长其系统发育关系推断的准确度越低<sup>[46]</sup>。基于此, 我们可以通过随机生成与外群等长序列若干条(如 100 条), 分别将每条随机序列与真实外群调换后进行相同方法分析, 并将结果与使用真实外群进行相同分析得到的结果进行比较, 如果随机生成序列的分析结果与真实外群结果相同的概率较高, 那么可推断真实数据集中很可能受到了“长枝吸引”的影响。已有多项研究通过这种方法成功检测到“长枝吸引”假象<sup>[31,47-51]</sup>。

随机生成序列法一个明显的不足之处表现在当长枝分类元确实位于拓扑结构中较基部的位置的情形中。因此在进行分析时我们必须尽可能涵盖外群所含的分类元, 以打断外群过长的枝长, 避免与内部长枝分类元发生吸引, 同时尽量选用进化速率慢的分类元进行分析, 避免长枝分类元的出现。

#### 4.4 参数模拟检测法(Parametric simulation)

这种方法用于检测两个分类元的枝长是否足够长以致发生相互吸引。首先假设一个核苷酸替换模型和一个模型树, 树中长枝分类元为非姐妹群关系。基于真实数据估计模型参数(比如枝长)后, 模拟一系列与真实数据集同等大小的重复数据集, 最后使用与真实数据集相同的分析方法对重复数据集进行分析, 将得到的结果与真实数据集得到的结果进行比较, 如果非姐妹群关系的分类元在重复数据集中高频率的以姐妹群关系出现, 那么可推断所分析的数据集中很可能形成了“长枝吸引”假象。

参数模拟法被普遍用于检测“长枝吸引”, 已有

多项研究通过这种方法检测到数据中确实存在“长枝吸引”假象<sup>[4,7,21,52-55]</sup>。

#### 4.5 相关显著共源性状分析法(Relative apparent synapomorphy analysis, RASA)

RASA方法是基于统计学原理检测特征矩阵中的系统发育信号<sup>[56]</sup>, 随后用于检测长枝分类元<sup>[48,57-63]</sup>。但是, 由于在许多实际研究中RASA都无法检测到“长枝吸引”假象的存在<sup>[64,65]</sup>, 因此, 越来越多的研究者对这种方法提出了质疑<sup>[64-67]</sup>。Grant和Kluge<sup>[41]</sup>还针对RASA方法的局限性作了详细地阐述<sup>[2]</sup>。

#### 4.6 谱分析(Split decomposition and spectral analysis)

该方法可用于检测系统发育数据中是否存在冲突信号。方法原理主要基于: 由于趋同进化作用的存在可能会导致“长枝吸引”假象的产生, 扰乱真实的系统发育信号, 所以在研究的数据集中很有可能存在着冲突信号。如果在由这种方法得到的图表中发现box-like结构时则表明数据中有冲突信号<sup>[68,69]</sup>。Grant和Kluge<sup>[41]</sup>对谱分析方法的检测优势进行了详尽的讨论<sup>[2]</sup>。到目前为止, 已有很多研究通过使用这种方法检测到“长枝吸引”假象<sup>[8,36,70-72]</sup>。

#### 4.7 不同方法的不一致分析(Methodological discordance)

对一个数据集, 如果怀疑有“长枝吸引”假象的存在, 那么可以通过使用不同的分析方法对这一数据集进行分析, 将所得到的各种结果加以比较, 如果出现不一致现象, 那么数据集中很有可能存在扰乱真实系统发育的信号, 以致形成“长枝吸引”假象。综上所述, 现在常用的“长枝吸引”假象检测方法都有各自的优缺点, 没有一种方法可以适用所有的数据。因此在实际检测过程中, 要尽可能同时使用多种方法检测“长枝吸引”。检测结果之间表现出来的一致程度, 也可作为结果可靠性的标准之一。

### 5 “长枝吸引”典型研究实例

为了更直观, 更系统地说明“长枝吸引”问题, 以下列举了在“长枝吸引”方面较为典型的研究实例并加以讨论说明。

#### 5.1 原生生物小孢子虫(Microsporidia)系统发育研究<sup>[73]</sup>

小孢子虫是一类真菌寄生虫。由于它们缺少有“发电站”作用的有氧呼吸细胞器—线粒体, 因此一

直被认为是进化过程中比较原始的生物。基于核糖体 RNA 的研究结果也显示小孢子虫位于系统发育树中较基部的位 置(图 2), 甚至比后生植物更早出现。在拓扑结构中明显发现, 3 个分类元, 即毛滴虫 (Trichomonodida)、眼虫(Euglenozoa)和小孢子虫的枝长相对其他分类元显著增长, 小孢子虫序列的枝长最长位于最基部, 成为最先分化的物种。直到 Germot A(1997)使用线粒体 HSP70 蛋白(一种分子伴侣, 在细菌、真核生物胞质中, 内质网、叶绿体中和线粒体中普遍存在)作为分子标记重新对这些分类元进行系统发育分析, 结果表明(图 2)小孢子虫虽然仍有显著增长的枝长(因其较快的进化速率), 但并不是最先分化的物种, 而是与真菌聚在一起, 晚于 Metazoa 出现。很明显, 基于核糖体 RNA 的分析结果受到了“长枝吸引”的影响, 枝长较长的 3 个分类元与外群发生相互吸引而位于系统发育树拓扑结构中较基部的位 置, 得到了错误的系统发育关系。

在一些情况下, 由于突变饱和效应(Mutational saturation)<sup>[75~79]</sup>, 无根树分析中的长枝分类元在有根树分析时与其他分类元相比枝长并没有显著增长。突变饱和和“长枝吸引”之间的关系可以用图 3 说明。图 3: A 中假设分类元 A、B 有相同的突变速率, 即它们与外群 O 之间的遗传距离相等。在另外一种情况下, A 和 C 的突变速率有显著差异, C 相对于 A 有更快的突变速率。但 A 和 C 的突变速率刚好位于突变饱和和平台期(当真实进化距离难以估算, 再加上多次替换以及不恰当地修正使得图 3 中 A、C 的突变率刚好位于曲线的平台期), 那么同时分析 A、B、C 以及外群 O 时, 多数检测方法都会检测到分子钟 (Molecular-clock)的存在, 认为 A、B 和 C 有相同的进化速率(图 3: B)。这种情况在使用亲缘关系较远的外群时更为严重。

在真核生物(Eukaryotes)的系统发育研究中就出现了这样的情况, 如图 4 所示<sup>[73]</sup>。当选取亲缘关系较近的  $\alpha$ - 蛋白菌 ( $\alpha$ -Proteobacteria) 作为外群时, *Nosema locustae*(属于小孢子门中的一种物种)表现出明显增长的枝长, 在内群分类元中与酵母 (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*)聚为一枝, 分歧时间较近(图 4: A)。但当选取亲缘关系较远的胞质溶胶同分异构体(Cytosolic isoform)作为外群分析时, 小孢子虫就处于内群分类元

类元中最基部的位 置, 成为仅次于外群而最先分化的分类元, 而且其枝长相对其他分类元来说并未显著增长(图 4: B)。很明显, 使用亲缘关系较远的外群, 会使外群与内群中进化速率较快的小孢子虫发生“相互吸引”而将小孢子虫分类元位于系统发育树拓扑结构的基部位 置, 错误的显示为最早分化的内群分类元。因此研究者在进行类似研究时往往会忽略“长枝吸引”假象的存在, 误认为所得到的结果是可靠的。

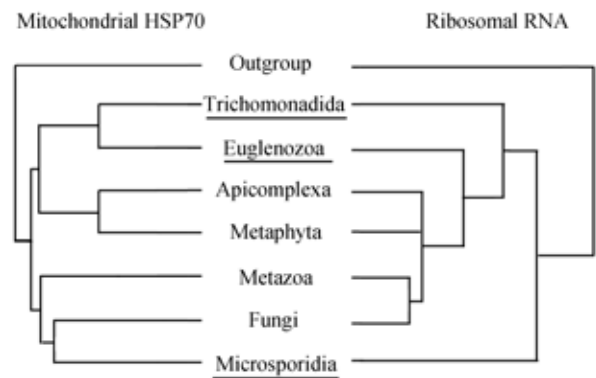


图 2 小孢子虫系统发育研究  
图中划横线部分表示枝长显著增长 的分类元。  
Fig. 2 *Microsporidia* phylogeny  
Taxa underlined represent long branch.

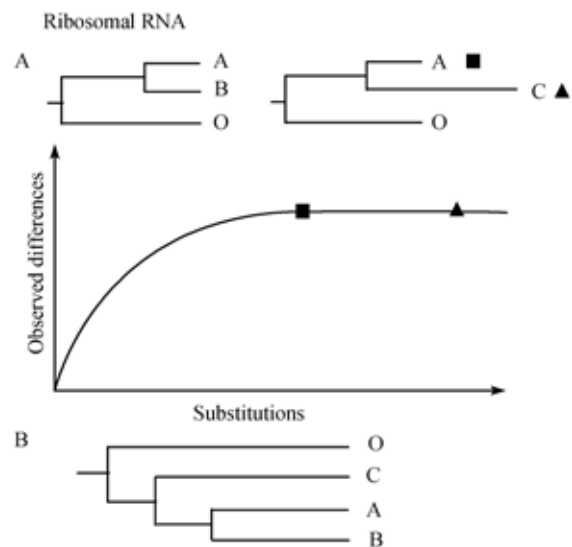


图 3 突变饱和与“长枝吸引”的关系  
图中  $\blacksquare$  表示 A 的突变率;  $\blacktriangle$  表示 C 的突变率。  
Fig. 3 The relationship between mutational saturation and “Long-branch Attraction”  
Black square and triangle represent mutational rates of A and C , respectively.

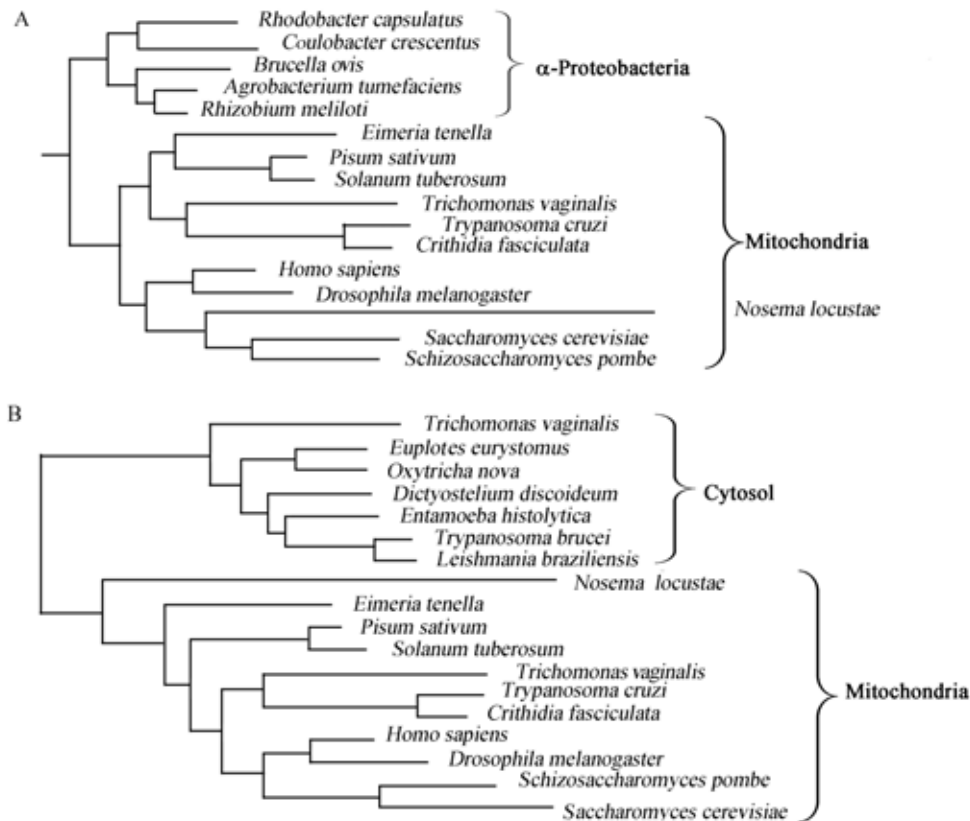


图 4 真核生物系统发育研究<sup>[73]</sup>

Fig.4 Phylogeny of eukaryotes<sup>[73]</sup>

## 5.2 胎盘哺乳类动物系统发育研究

### 5.2.1 天竺鼠(guinea-pig)是否属于啮齿类动物？啮齿类动物是否是单系起源？

在胎盘哺乳类动物的系统发育研究中，其中一个一直悬而未决，长期存有争议的问题是天竺鼠的系统发育位置。自D'Erchia等<sup>[80]</sup>在1996年《自然》杂志上发表了一篇通过分析天竺鼠全线粒体基因组数据得到不支持天竺鼠属于啮齿类动物的报道后，天竺鼠的进化地位便引起了学术界的普遍关注。直到Sullivan和Swofford等<sup>[81]</sup>对这一问题进行重新分析的结果发现，在D'Erchia等人的研究中，由于鼠科物种(包括大、小鼠)进化速率很快，加上其特有的核苷酸组成偏好，使得在系统发育分析时，鼠科进化枝和外群分类元发生了相互吸引而使其位于系统发育树拓扑结构的基部位置，天竺鼠也就排除在啮齿类动物进化枝之外，得到了天竺鼠不属于啮齿类动物或者啮齿类动物不是单系起源的错误结果。Sullivan和Swofford通过增加分类元打断“长枝”以及

考虑位点替换速率异质和优化模型对数据进行了重新分析，得到的系统发育结果支持啮齿类动物单系起源<sup>[81]</sup>。

### 5.2.2 刺猬(hedgehog)是最先分化的胎盘哺乳类动物吗？

由于刺猬表现出显著加快的进化速率，因此，关于刺猬在哺乳类群中确切的系统发育位置一直难以解决。Krettek等<sup>[82]</sup>在1995年对刺猬线粒体全基因组序列分析认为刺猬是最先分化的胎盘哺乳类动物。后来，一些研究者<sup>[82-84]</sup>发现刺猬由于进化速率太快，以及其显著不同的核苷酸组成，使它在系统发育树上形成长枝并与有袋动物外群发生相互吸引从而得到刺猬是最先分化的胎盘哺乳类动物的错误结果。Murphy等<sup>[85]</sup>发表在2001年《科学》杂志的研究中通过增加分类元以及选取合适的核基因作为分子标记，优化似然模型后，支持刺猬与劳亚古大陆兽亚纲哺乳动物进化枝中的食虫动物聚在一起，而不是最先分化的胎盘哺乳类动物。

## 6 小结

“长枝吸引”假象在分子系统发育研究中广泛存在, 并日益受到关注。就目前所使用的系统发育分析方法, 包括简约法, 还是最大似然法, 贝叶斯法, 距离法等, 均有报道受到了“长枝吸引”的影响。因此, 在期待更加合理的分析方法出现的同时, 目前能够有效避免和消除“长枝吸引”假象的方法是通过增加分类元, 合并多个独立数据集以及使用相对不易产生“长枝吸引”的分析方法(如最大似然法), 并结合使用以上阐述的各种检测方法以尽可能地去除“长枝吸引”假象, 构建可靠的系统发育树。因此, 我们建议在系统发育研究过程中, 如果怀疑所得结果可能受到了“长枝吸引”的影响, 首先可将外群分类元去除, 进行无根树分析, 将得到的结构与有根树结构相比较, 如果得到不一致结果, 那么可以推断数据很有可能确实受到了“长枝吸引”的影响。如结果一致时, 可以通过去除可疑分类元重新进行分析, 检测拓扑结构是否发生改变, 寻找合理途径以解决“长枝吸引”问题, 使我们的结果能够更接近所研究物种间真实的系统发育关系。

## 参考文献(References):

- [1] PENG Yi-Xin, HUANG Shi-Jian. Evolutionary Biology. Wuhan: Wuhan University Press, 1997.  
彭奕欣, 黄诗笺主编. 进化生物学. 武汉: 武汉大学出版社, 1997.
- [2] Bergsten J. A review of long-branch attraction. *Cladistics*, 2005, 21(2): 163–193.
- [3] Andersson FE, Swofford DL. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. *Mol Phyl Evol*, 2004, 33(2): 440–451.
- [4] Omilian AR, Taylor DJ. Rate acceleration and long-branch attraction in a conserved gene of cryptic daphniid (Crustacea) species. *Mol Biol Evol*, 2001, 18(12): 2201–2212.
- [5] Schwarz MP, Tierney SM, Cooper SJB, Bull NJ. Molecular phylogenetics of the allodapine bee genus *Braunsapis*: A-T bias and heterogeneous substitution parameters. *Mol Phyl Evol*, 2004, 32(1): 110–122.
- [6] Bergsten J, Miller KB. *Acilius* phylogeny (Coleoptera: Dytiscidae), problems with long-branch attraction and morphological intersexual coevolution [Meeting Abstract]. *Cladistics*, 2004, 20: 76–77.
- [7] Tang KL, Berendzen PB, Wiley EO, Morrissey JF, Winterbottom R, Johnson GD. The phylogenetic relationships of the suborder Acanthuroidei (Teleostei: Perciformes) based on molecular and morphological evidence. *Mol Phyl Evol*, 1999, 11(3): 415–425.
- [8] Clements KD, Gray RD, Choat JH. Rapid evolutionary divergences in reef fishes of the family Acanthuridae (Perciformes: Teleostei). *Mol Phyl Evol*, 2003, 26(2): 190–201.
- [9] Carmean D, Crespi BJ. Do long branches attract flies? *Nature*, 1995, 373(6516): 666.
- [10] Huelsenbeck JP. Is the Felsenstein zone a fly trap? *Syst Biol*, 1997, 46(1): 69–74.
- [11] Steel M, Huson D, Lockhart PJ. Invariable sites models and their use in phylogeny reconstruction. *Syst Biol*, 2000, 49(2): 225–232.
- [12] Philippe H. Rodent monophyly: pitfalls of molecular phylogenies. *J Mol Evol*, 1997, 45(6): 712–715.
- [13] Sullivan J, Swofford DL. Are Guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J Mamm Evol*, 1997, 4(2): 77–86.
- [14] Waddell PJ, Kishino H, Ota R. A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics*, 2001, 12: 141–154.
- [15] Lin YH, McLenachan PA, Gore AR, Phillips MJ, Ota R, Hendy MD, Penny D. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Mol Biol Evol*, 2000, 19(2): 2060–2070.
- [16] Lin YH, Waddell PJ, Penny D. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. *Gene*, 2002, 294(1-2): 119–129.
- [17] Garcia-Moreno J, Sorenson MD, Mindell DP. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. *J Mol Evol*, 2003, 57(1): 27–37.
- [18] Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*, 1997, 387(6632): 489–493.
- [19] Kim JH, Kim W, Cunningham CW. A new perspective on lower metazoan relationships from 18S rDNA sequences. *Mol Biol Evol*, 1999, 16(3): 423–427.
- [20] Soltis DE, Soltis PS. Amborella not a basal angiosperm? *Am J Bot*, 2004, 91: 1199.
- [21] Sanderson MJ, Wojciechowski MF, Hu JM, Khan TS, Brady SG. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol Biol Evol*, 2000, 17(5): 782–797.
- [22] Moreira D, Le Guyader H, Philippe H. The origin of red algae and the evolution of chloroplasts. *Nature*, 2000, 405(6782): 69–72.
- [23] Moreira D, Le Guyader H, Philippe H. Unusually high evolutionary rate of the elongation factor *lar* genes from the ciliophora and its impact on the phylogeny of eukaryotes. *Mol Biol Evol*, 1999, 16(2): 234–245.
- [24] Archibald JM, O’Kelly CJ, Doolittle WF. The chaperonin genes of jakobid and jakobid-like flagellates: implications for eukaryotic evolution. *Mol Biol Evol*, 2002, 19(4): 422–431.

- [25] Dacks JB, Marinets A, Doolittle WF, Cavalier-Smith T, Logsdon JM. Analyses of RNA polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. *Mol Biol Evol*, 2002, 19(6): 830–840.
- [26] Inagaki Y, Susko E, Fast NM, Roger AJ. Covarion shifts cause a long-branch attraction artefact that unites microsporidia and archaeobacteria in EF-1 alpha phylogenies. *Mol Biol Evol*, 2004, 21(7): 1340–1349.
- [27] Bocchetta M, Grimaldo S, Sanangelantoni A, Cammarano P. Phylogenetic depth of the bacterial genera Aquifex and Thermotoga inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. *J Mol Evol*, 2000, 50(4): 366–380.
- [28] Brinkmann H, Philippe H. Archaea sister group of bacteria? Indications from tree reconstruction artefacts in ancient phylogenies. *Mol Biol Evol*, 1999, 16(6): 817–825.
- [29] Grimaldo S, Philippe H. Ancient phylogenetic relationships. *Theor Pop Biol*, 2002, 61(4): 391–408.
- [30] Lopez P, Forterre P, Philippe H. The root of the tree of life in the light of the covarion model. *J Mol Evol*, 1999, 49(4): 496–508.
- [31] Philippe H, Forterre P. The rooting of the universal tree of life is not reliable. *J Mol Evol*, 1999, 49(4): 509–523.
- [32] Lake JA. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol Biol Evol*, 1987, 4(2): 167–191.
- [33] Willson SJ. A higher order parsimony method to reduce longbranch attraction. *Mol Biol Evol*, 1999, 16(5): 694–705.
- [34] Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (Eds.). *Phylogenetic Inference*. Sinauer Associates, Sunderland, MA, USA, 1996, pp. 407–514.
- [35] Huelsenbeck JP. Is the Felsenstein zone a fly trap? *Syst Biol*, 1997, 46(1): 69–74.
- [36] Kennedy M, Paterson AM, Morales JC, Parsons S, Winington AP, Spencer HG. The long and short of it: branch lengths and the problem of placing the New Zealand short-tailed bat, *Mystacina*. *Mol Phyl Evol*, 1999, 13(2): 405–416.
- [37] Kim JH, Kim W, Cunningham CW. A new perspective on lower metazoan relationships from 18S rDNA sequences. *Mol Biol Evol*, 1999, 16(3): 423–427.
- [38] Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 1998, 14(9): 817–818.
- [39] Kim JH. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Syst Biol*, 1996, 45: 363–374.
- [40] Poe S, Swofford DL. Taxon sampling revisited. *Nature*, 1999, 398(6725): 299–300.
- [41] Grant T, Kluge AG. Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics*, 2003, 19(5): 379–418.
- [42] Baker RH, Yu X, DeSalle R. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Mol Phyl Evol*, 1998, 9(3): 427–436.
- [43] Jenner RA. Accepting partnership by submission? Morphological phylogenetics in a molecular millennium. *Syst Biol*, 2004, 53(2): 333–342.
- [44] Moreira D, Kervestin S, Jean-Jean O, Philippe H. Evolution of eukaryotic translation elongation and termination factors: variations of evolutionary rate and genetic code deviations. *Mol Biol Evol*, 2002, 19(2): 189–200.
- [45] Hampl V, Cepicka I, Flegr J, Tachezy J, Kulda J. Critical analysis of the topology and rooting of the parabasalian 16S rRNA tree. *Mol Phyl Evol*, 2004, 32(3): 711–723.
- [46] Wheeler WC. Nucleic-acid sequence phylogeny and random outgroups. *Cladistics*, 1990, 6: 363–367.
- [47] Sullivan J, Swofford L. Are Guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J Mamm Evol*, 1997, 4: 77–86.
- [48] Stiller JW, Hall BD. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol Biol Evol*, 1999, 16(9): 1270–1279.
- [49] Qiu YL, Lee J, Whitlock BA, Bernasconi-Quadroni F, Dombrowska O. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Mol Biol Evol*, 2001, 18(9): 1745–1753.
- [50] Stiller JW, Riley J, Hall BD. Are red algae plants? A critical evaluation of three key molecular data sets. *J Mol Evol*, 2001, 52(6): 527–539.
- [51] Graham SW, Olmstead RG, Barrett SCH. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Mol Biol Evol*, 2002, 19(10): 1769–1781.
- [52] Maddison DR, Baker MD, Ober KA. Phylogeny of carabid beetles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). *Syst Entomol*, 1999, 24(2): 103–138.
- [53] Tourasse NJ, Gouy M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol Phyl Evol*, 1999, 13(1): 159–168.
- [54] Wiens JJ, Hollingsworth BD. War of the iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in iguanid lizards. *Syst Biol*, 2000, 49(1): 143–159.
- [55] Wilcox TP, Garcia de Leon FJ, Hendrickson DA, Hillis DM. Convergence among cave catfishes: long-branch attraction and a Bayesian relative rates test. *Mol Phyl Evol*, 2004, 31(3): 1101–1113.
- [56] Lyons-Weiler J, Hoelzer GA, Tausch RJ. Relative apparent synapomorphy analysis (RASA) I: the statistical measurement of phylogenetic signal. *Mol Biol Evol*, 1996, 13(6): 749–757.
- [57] Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, El-



- lisens WJ, Moore G, Wolfe AD, dePamphilis CW. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc Natl Acad Sci USA*, 2000, 97(24): 13166–13171.
- [58] Belshaw R, Downton M, Quicke DLJ, Austin AD. Estimating ancestral geographical distributions: a Gondwanan origin for aphid parasitoids? *Proc. R. Soc. London (B), Biol Sci*, 2000, 267(1442): 491–496.
- [59] Bowe LM, Coat G, dePamphilis CW. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc Natl Acad Sci USA*, 2000, 97(8): 4092–4097.
- [60] Culligan KM, Meyer-Gauen G, Lyons-Weiler J, Hays JB. Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. *Nucl Acids Res*, 2000, 28(2): 463–471.
- [61] Reyes A, Pesole G, Saccone C. Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. *Gene*, 2000, 259(1-2): 177–187.
- [62] Teeling EC, Scally M, Kao DJ, Romagnoli ML, Springer MS, Stanhope MJ. Molecular evidence regarding the origin of echolocation and flight in bats. *Nature*, 2000, 403(6799): 188–192.
- [63] Stiller JW, Riley J, Hall BD. Are red algae plants? A critical evaluation of three key molecular data sets. *J Mol Evol*, 2001, 52(6): 527–539.
- [64] Simmons MP, Randle CP, Freudenstein JV, Wenzel JW. Limitations of relative apparent synapomorphy analysis (RASA) for measuring phylogenetic signal. *Mol Biol Evol*, 2002, 19(1): 14–23.
- [65] Xiang QY, Moody ML, Soltis DE, Fan CZ, Soltis PS. Relationships within Cornales and circumscription of Cornaceae –matK and rbcL sequence data and effects of outgroups and long branches. *Mol Phyl Evol*, 2002, 24(1): 35–57.
- [66] Faivovich J. On RASA. *Cladistics*, 2002, 18(3): 324–333.
- [67] Farris JS. RASA attributes highly significant structure to randomized data. *Cladistics*, 2002, 18(3): 334–353.
- [68] Bandelt HJ, Dress AWM. Split decomposition: a new unambiguous approach to phylogenetic analysis of distance data. *Mol Phyl Evol*, 1992, 1(3): 242–252.
- [69] Hendy MD, Penny D. Spectral analysis of phylogenetic data. *J Classification*, 1993, 10: 5–24.
- [70] Flook PK, Rowell CHF. The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Mol Phyl Evol*, 1997, 8(2): 177–192.
- [71] Waddell PJ, Cao Y, Hauf J, Hasegawa M. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid invariant sites LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. *Syst Biol*, 1999, 48(1): 31–53.
- [72] Lockhart PJ, Cameron SA. Trees for bees. *TREE*, 2001, 16(2): 84–88.
- [73] Urban & Ficher Verlag. Opinion: Long branch attraction and protist phylogeny. *Protist*, 2000, 151(4): 307 - 316.
- [74] Germot A, Philippe H, Le Guyader H. Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol Biochem Parasitol*, 1997, 87(2): 159 - 168.
- [75] Philippe H, Adoutte A. What can Phylogenetic Patterns Tell us about the Evolutionary Processes Generating Biodiversity? In Hochberg M, Clobert J, Barbault R (eds) *Aspects of the genesis and maintenance of Biological Diversity*. Oxford University Press, 1996, 41–59.
- [76] Philippe H, Adoutte A. The Molecular Phylogeny of Eukaryota: Solid Facts and Uncertainties. In Coombs G, Vickerman K, Sleigh M, Warren A (eds). *Evolutionary Relationships Among Protozoa*. London: Chapman & Hall, , 1998, 25–56.
- [77] Philippe H, Laurent J. How good are deep phylogenetic trees? *Curr Opin Genet Dev*, 1998, 8: 616–623.
- [78] Philippe H, Chenail A, Adoutte A. Can the cambrian explosion be inferred through molecular phylogeny? *Development*, 1994, 120(Suppl.): 15–25.
- [79] Philippe H, Sörhannus U, Baroin A, Perasso R, Gasse F, Adoutte A. Comparison of molecular and paleontological data in diatoms suggests a major gap in the fossil record. *J Evolution Biol*, 1994, 7: 247–265.
- [80] D'Erchia AM, Gissi C, Pesole G, Saccone C, Arnason U. The guinea-pig is not a rodent. *Nature*, 1996, 381(6583): 597–600.
- [81] Sullivan J, Swofford DL. Are Guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J Mamm Evol*, 1997, 4(2): 77–86.
- [82] Krettek A, Gullberg A, Arnason U. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. *J Mol Evol*, 1995, 41(6): 952–957.
- [83] Nikaido M, Cao Y, Harada M, Okada N, Hasegawa M. Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. *Mol Phyl Evol*, 2003, 28(2): 276–284.
- [84] Nikaido M, Kawai K, Cao Y, Harada M, Tomita S, Okada N, Hasegawa M. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a re-evaluation of the phylogeny of bats and insectivores. *J Mol Evol*, 2001, 53(4-5): 508–516.
- [85] Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, 2001, 294(5550): 2348–2351.