

分子系统发育分析的生物信息学方法

张树波^{1,2,4} 赖剑煌^{3,4}

(广州航海高等专科学校计算机系 广州 510725)¹ (中山大学数学与计算科学学院 广州 510275)²
(中山大学信息技术与科学学院 广州 510275)³ (广东省信息安全技术重点实验室 广州 510725)⁴

摘 要 物种之间的系统发育分析研究物种的形成或进化历史,以及物种之间的进化关系,它是当代生命科学研究的一个重要分支。分子生物学和计算机科学的发展,为人类采用计算的方法从分子水平研究生物的进化历史以及探索生命的起源提供了新的思路。在过去的几十年中,基于分子水平的系统发育学研究取得很大的进展,各种各样基于分子水平的系统发育树推断方法被提出来。现从物种进化距离刻画和系统发育树重构算法两个方面,总结和评述了在过去十几年里,人类利用生物信息学方法在分子水平上研究物种进化关系的进展和取得的成果,分析了物种进化研究存在的问题和面临的挑战,指出了物种进化研究今后的主要方向。

关键词 生物信息学,系统发育分析,基因进化,最大简约法,最大似然法,邻接法

中图法分类号 TP3-05 **文献标识码** A

Bioinformatics Approach for Molecular Evolution Research

ZHANG Shu-bo^{1,2,4} LAI Jian-huang^{3,4}

(Department of Computer Science, Guangzhou Marine College, Guangzhou 510725, China)¹
(School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou 510275, China)²
(School of Information Science and Technology, Sun Yat-sen University, Guangzhou 510275, China)³
(Province Key Laboratory for Technology of Security Information, Guangzhou 510725, China)⁴

Abstract The phylogeny inference among different species aims to reveal the evolutionary history of certain species, as well as the evolutionary relationship among them, it is an essential branch in the today's research of science life. The progress in molecular biology supplies a large volume of data to the field of phylogeny inference, and the research of evolutionary relationship among different species at molecular level is receiving considerable attention. In the past decade, much progress in this area has been achieved by means of bioinformatics, and a variety of algorithms have been devised to reconstruct the phylogenetic tree of species. In this paper, the development of measuring the evolutionary distance and reconstructing algorithm for phylogeny were reviewed and commented, and the challenge of researchers in this field faced was then pointed out, in the end, perspectives in this realm were also proposed.

Keywords Bioinformatics, Phylogenetic analysis, Evolution of gene, Parsimony approach, Maximum likelihood approach, Neighbor joining approach

1 引言

系统发育学也称系统发生学(phylogeny),它研究的是物种的形成或进化历史,以及物种之间的进化关系,它是生物信息学的一个重要分支,通过系统发育研究,人们可以了解物种之间的进化历史,探索生命的起源,开展流行病学研究,探索未知基因的功能,进行物种差异性研究和微生物生态学的研究。

生命的进化是一个非常漫长的过程,如何找到物种之间进化的证据一直是摆在科学家面前的一大难题,长期以来人们一直认为生物界中不同物种的进化历史可以从化石中找到证据,但是由于我们能够得到的化石数量是有限的、零散

的^[1],采用这种方法很难全面把握生物之间的进化历史和进化关系。在分子生物学出现以前,利用生物的形态学特征、解剖学特征和发育学特征来研究物种之间的进化关系是一种最普遍的方法,但是这些方法对特征的选择依赖性很大。由于在进化过程中,相似的结构可能独立地进化,如脊椎动物和章鱼的眼睛虽然功能相同,但却是由完全不同的途径进化而来的;又如蝙蝠具有可以飞行的翅膀,从这个角度看,它应该属于鸟类,但是实际上,蝙蝠是一种哺乳动物,因此很难找到稳定的特征来分析物种之间的进化关系。在 20 世纪中叶,随着分子生物学的发展,进化研究开始进入分子水平,尤其是在过去的 30 多年时间里,基因测序技术快速发展,多种物种的生物序列数据(比如 DNA、RNA 和蛋白质序列)的积累已经越

来越多,这为人类从分子水平研究物种的进化关系提供了可能。目前人们普遍认为生物分子中蕴含着物种进化的痕迹,而且生物分子的获取远比化石的获取容易,因此,在分子水平上研究物种的进化关系逐渐引起人们的兴趣;另一方面,随着计算机科学的发展,计算机处理大规模数据的能力大大提高,而成本却明显降低,这为人类利用大量生物分子数据进行物种进化的系统发育分析提供了可能。从大量关于这方面的文献中可以看出,基于生物信息学方法研究物种的系统发育已经越来越引起人们的重视。尤其是在过去的几十年中,经过生物学家、数学家、统计学家和计算机科学家们的共同努力,系统发育学研究取得很大的进展,各种各样基于分子水平的系统发育树推断方法被提出来,表 1 列出了一些常用的系统发育树分析软件。

表 1 常用的一些系统发育分析软件

软件名称	网址	方法	开发机构
PHYLIP	http://evolution.genetics.washington.edu/phylip/	NJ, MP, ML	美国华盛顿大学
PAUP	ftp://onyx.si.edu/paup	MP	美国 Simthsonian 研究所
MEGA	http://bioinfo.weizmann.ac.il/databases/info/	NJ, MP	美国宾西法尼亚州立大学
MOLPHY	ftp://sunmh.ism.ac.jp/pub/molphy	ML	日本国立统计数理研究所
PAML	http://abacus.gene.ucl.ac.uk/software/	ML	英国伦敦大学
PUZZLE	http://www.tree-puzzle.de	ML	德国柏林马普学会分子基因研究所
PHYML	http://atgc.lirmm.fr/phyml/	ML	法国蒙彼利埃市生物信息学和微电子学实验室
MrBayes	http://mrbayes.csit.fsu.edu/	Bayesian	美国加州大学圣地亚哥分校
MAC5	http://www.agapow.net/software/	Bayesian	英国雷丁大学
SplitsTree	http://www-ab.informatik.uni-tuebingen.de/	NJ	德国蒂宾根大学

注: NJ 指邻接法, MP 指最大简约法, ML 指最大似然法, Bayesian 指贝叶斯统计推断方法

利用生物信息学方法在分子水平进行系统发育分析重点需要解决两个方面的问题: (1) 分子水平上的物种进化信息提取; (2) 系统发育树重构算法。本文围绕这两个方面进行综述, 总结基于分子水平的系统发育树分析研究领域所取得的成果, 探索该领域面临的问题和今后的研究方向。第 2 节主要讨论在分子水平上刻画两个物种之间进化关系的方法; 第 3 节评述重构系统发育树的一些主要算法; 最后指出了系统发育研究的发展方向。

2 分子水平上的进化信息

当代分子生物学研究认为物种进化的最根本原因在于生物分子的进化, 生物分子主要包括 DNA、RNA 和蛋白质, 这些生物序列中隐藏着物种进化的痕迹, 包含着物种进化的信息, 因此, 物种之间的进化关系可以用生物分子之间的相似性来表示, 而生物分子相似性可以从多方面来刻画, 包括核酸序列或氨基酸序列之间的相似性, 也包括这些分子结构上的相似性, 或者在它们之间功能上的相似性。所以系统发育关系

研究可以在序列水平上, 也可以在结构上或者在功能上利用相关的信息。

生物分子的结构信息是指二级结构、三级结构或者更加复杂的结构信息, 其中二级结构主要包括分子的空间折叠结构, 可以利用这种折叠信息分析物种的进化关系^[2]。由于生物分子的结构非常复杂, 目前可用的分子结构信息不是很充分, 利用生物分子的结构信息研究物种进化关系不是目前的主要方法。分子的功能信息主要包括一些对生物分子的功能进行描述和解释的信息, 由于这方面提供的数据不是非常充分, 因此基于分子功能方面的信息研究物种之间的系统发育关系的工作较少。但是随着生物数据的不断积累, 相信在不久的将来, 分子的结构和功能信息也能够被用于系统发育分析。

在序列水平上, 生物序列的进化体现在序列之间相应位点上的碱基或者氨基酸残基之间的变异, 即从一条生物序列变异产生另一条生物序列, 就是核酸或者蛋白质序列之间的进化。就核酸序列而言, 目前用于系统发育分析的信息主要是来自编码蛋白质的那些序列, 即不同物种之间同源基因的进化信息。与分子的结构和功能信息相比, 目前研究系统发育可用的序列信息比较多, 本文主要讨论基于分子序列的系统发育分析问题。下面根据提取进化信息所采用的同源基因或同源蛋白质的数量不同, 对分子水平上系统发育分析研究中所利用的信息进行总结和评述。

2.1 单条生物序列中的进化信息

在高通量测序技术出现以前, 由于基因测序技术和测序成本的原因, 在分子水平上研究物种之间的系统发育关系主要是利用单个基因或蛋白质序列, 即利用不同物种之间的单个同源基因或同源蛋白质。通过计算由一条序列进化到另外一条序列的概率或者需要变异的操作次数来度量它们之间的进化距离。基于单条序列的分子进化信息刻画的是局部位点上碱基或者氨基酸残基的变异情况, 包括它们的插入、删除和替代等进化事件。

单基因水平的进化信息提取就是从不同物种的同源基因中分析提取能够反映物种之间进化关系的信息, 这种信息体现在两条基因序列之间的差异, 差异越大说明物种之间的进化距离越远, 反之, 认为进化距离越近。

基于单序列的进化信息可以用序列之间的比对得分值作为它们之间的进化距离度量, Wu 等人用蛋白质序列的比对信息构建分析果蝇的系统发育关系^[3]; 或者可以用统计技术来计算两条序列之间的进化距离, 如 Ranwez 等人利用最大似然法从序列中计算序列之间的最大似然距离^[4]; 也可以用其它的方法来定义两条序列的相似程度, 如 Otu 等人定义了一种称为 LZ complexity 的距离度量, 用于描述从一条序列开始产生另外一条序列所需要的操作次数^[5]。Qi 等人采用了基于多肽序列频率分析的方法进行系统发育推断^[6]。最近, Hoppenrath 等人提出用 rDNA 序列中的信息来分析藻类的系统发育关系^[7]。

由于在物种的进化过程中, 不同基因的进化速度实际上是不同的^[8], 因此当选择不同的基因来重构系统发育树时, 难免得到相互冲突的结果, 出现通常所说的基因树和物种树不一致的现象。同时, 在生物的进化过程中可能存在基因水平(横向)移动的现象^[9-11], 这对于寻找保守基因造成一定的干

扰^[12]。另外,直系同源基因和旁系同源基因的存在,也会对分析结果产生不良的影响^[8]。然而,单调序列所包含的信息量是有限的^[13],难以全面地刻画物种的进化信息。

2.2 多条生物序列中的进化信息

由于利用单条序列进行系统发育分析存在不稳定性,生物体内的生物分子数量巨大,不同的生物分子中都包含着一定的进化信息,因此利用多条序列来研究物种之间的进化关系是一种自然的想法。而随着生命科学技术的发展,提供了越来越多的生物序列,这使得科研人员利用多条序列进化信息进行系统发育分析成为可能。为了利用多条生物序列中的进化信息,人们在研究中采用了两种主要策略:(1)对每个同源序列分别利用上面提到的单序列信息构建系统发育树,然后根据这些树构造一致树来表示物种系统发育树,文献[14, 15]采用的就是这种策略。(2)先利用比对算法分别对每个同源基因进行多序列比对,然后将所有的比对结果串联起来,作为多条序列比对的结果,再根据这个比对结果重构系统发育树,文献[12, 16]采用了这种方法。最近 James 等人从 200 种菌类中找出 6 个同源基因,然后利用这些基因的信息对它们进行了有效的分类^[17]。

从本质上讲,基于多条序列的分子进化信息刻画的仍然是序列上局部位点碱基或者氨基酸残基的变异情况,包括它们的插入、删除和替代等进化事件,由于这种信息综合了不同序列中的进化信息,这类方法总体上能够得到比利用单条序列信息更加合理的结果,但是其代价就是计算的复杂程度提高了。

2.3 全基因组序列中的进化信息

随着高通量测序技术的发展,基因测序成本大幅度下降,很多物种的全基因组序列已经被测序出来,全基因组中包含着更加全面的进化信息,这为人类在分子水平上寻找物种进化的标记物提供了大量的数据资料。另外,随着计算机技术的发展,计算机的运算速度和运算能力不断提高,以及统计技术在生物科学某些研究领域的成功应用,使系统发育分析已经进入到基因组时代,越来越多的学者将注意力转移到全基因组水平上。目前,这方面的研究已经成为计算机科学家和生物学家普遍关注的问题。基于全基因组的系统发育推断方法主要可以归纳为三大类。

(1)全基因组中的 gene order(gene rearrangement)信息。这类方法利用了不同基因组之间共有的基因排列顺序变化信息来度量它们之间的进化距离,并以此推断物种之间的系统进化关系。全基因组中基因排列顺序变化是生物进化中的一类重要进化事件,包括基因颠倒、基因转座和基因反转座三种类型。基因排列顺序变化信息计算的是不同基因组中共有基因排列顺序的差异程度,而这种差异程度通常是由从一个基因组转换为另一个基因组所需要的最小变化次数来定义的,即编辑距离。早在 1938 年, Dobzhansky 和 Sturtevant 就发现不同种类的果蝇的基因组中存在基因顺序重排的现象,并且首先开始了这方面的研究^[18]。之后在很长一段时间内,由于受基因组测序技术发展水平的限制,人们在基因排列顺序方面的研究进展相对缓慢。直到最近十几年,这方面的研究才取得明显的进展,目前 gene order 信息已经被用于研究植物^[19]、细菌^[20]、藻类^[21]、哺乳动物^[22]和病毒^[23]等物种的进化关系。

基于基因顺序重排的物种进化信息是通过两条基因组中基因的编辑距离刻画的,计算两个基因组之间与基因排列顺序相关的进化距离通常考虑了基因颠倒、基因转座和基因反转座三种进化事件,然后通过计算两个排列之间的编辑距离来表示两个基因组之间的进化距离,即从一条全基因组序列利用基因顺序重排算子,将一条全基因组序列变化为另外一条全基因组序列所需要的最少编辑操作次数。计算这类编辑距离已经被证明是 NP-难题^[24],关于如何有效计算这种进化距离的问题,目前仍然没有得到很好的解决,大多数算法采用的是近似搜索策略。

(2)全基因组中 gene content 的信息。这类方法利用不同基因组中同源基因(也称共有基因)数量占全基因组中基因数量的比例来刻画两条全基因组序列之间的相似程度,进而根据这个相似性度量刻画它们之间的进化距离。gene content 的概念由是 Snel 等人^[25]在 1999 年研究物种的系统发育关系时首先提出的,其计算方法如下:

$$Dist_content_{A,B} = 1 - \frac{|C|}{\min(|A|, |B|)} \quad (1)$$

式中, $|A|$ 和 $|B|$ 分别表示基因组 A 和 B 中基因的数量, $|C|$ 表示基因组 A 和 B 中同源基因的数量。根据 gene content 的概念,如果两条全基因组序列中同源基因越多,表示它们越相似,进化距离越短,反之,进化距离越远。目前, gene content 信息已经被成功应用于分析病毒^[16, 26-28]、古细菌^[29, 30]等的系统进化关系。

后来 Gu 和 Zhang 为了刻画基因组中基因家族的出现或者缺失情况,引进了 extended gene content 的概念,构造了一种基于随机框架的可加性进化距离度量^[31]。Huson 等人根据基因发生和基因丢失进化事件提出了一种基于最大似然方法的进化距离度量^[32]。

(3)全基因组中的高分匹配对信息。这种信息是通过全基因组比对结果来提取的,首先对全基因组序列进行比对运算,在比对的结果中,按照一定的 E-值或得分找出两条序列中的高的分匹配对,然后记录匹配对的长度、在基因组中的位置和 E-值或得分,然后利用这些高的分匹配对的信息计算出两条全基因组序列的相似性,其中一种可以采用的方法就是像计算 gene order 信息的 breakpoint 距离那样。Alexander 等人利用这种方法计算质体基因组和线粒体基因组的进化距离信息^[13], Henz 等人则用这种方法分析原核生物的全基因组序列之间的进化关系^[33]; Deng 等人在哺乳动物线粒体全基因组上采用这种信息进行系统发育推断。

上述三种基于全基因组序列的信息中, gene content 反映的是不同基因组中共有基因数量的变化情况,刻画的是基因的丢失、基因的分裂、基因的合并或基因的复制等进化事件; gene order 反映的是基因排列顺序的变化情况,包括基因移动、基因方向颠倒等事件;而高分匹配对信息反映的是局部基因组序列在局部位置上的变异情况,包括插入、删除和替代等进化事件。

3 系统发育树重构算法

系统发育树是用一个树状结构图描述物种之间的进化关系,树的叶节点表示分类单元,节点分为内节点和外节点两类,外节点表示生物分子,内节点表示生物分子之间的进化位

置。节点之间的连线表示生物分子之间的进化关系。按照是否有根节点,系统发育树可以分为有根树和无根树,有根树中存在唯一的根节点,表示它是所有其它节点的共同祖先,根节点下面的各层节点表示了树的进化方向;无根树没有层次结构,无法表示物种之间的进化方向,只是刻画了它们之间的进化关系。基于分子水平的系统发育推断方法可以分为两大类,即基于字符序列的方法和基于距离的方法。

3.1 基于字符序列的系统发育树重构算法

基于字符序列的系统发育树重构算法通过搜索各种可能的树,从中选出最能够解释给定序列的那棵树,用于刻画研究的物种之间的系统发育关系。这类方法利用统计技术定义一个最优化标准,对树的优劣进行评价,包括如下3种方法。

(1)基于最大简约法的方法。这种方法首先由 Camin 和 Sokal 提出^[34],后来 Hein 等人^[35,36]对该方法进行了推广。最大简约法的理论基础是 Ockham 的哲学原则,其思想是希望用最小的改变来解释所要研究的分类群体之间观察到的差异,基本做法是检查所有可能的发育树,从中选择需要最小进化次数的那棵树来表示研究物种之间的进化关系,它在过去的30年时间里,被成功用于研究一些物种的进化关系^[37,38],但是这种方法已经被证明是 NP-难题^[39],这就导致了在物种数量较大时,它变得不实用。另外,最大简约法没有考虑树中分支的长度,这样就无法反映出不同物种之间进化距离的长短。

(2)基于最大似然法的方法。这种方法最早由 Felsenstein 提出^[40],该方法采用显式的进化模型,使得用模型产生的数据与真实数据之间的相似程度最高。它具有良好的统计性质,因此在选取合适的模型时,往往能够得到理想的结果。与最大简约法相比,最大似然法考虑了不同物种之间进化距离的问题,能够更加有效地刻画它们之间的进化距离。但是基于最大似然法的系统发育树重构方法需要估计分支的长度和搜索似然率最大的树拓扑结构,计算量非常大,也是一个 NP-难题^[39],对于物种数量较大的情况是不适应的。另外,这种方法对模型的依赖程度较高,导致其在很多情况下的应用存在困难。

(3)基于贝叶斯推断的方法。这种方法的理论基础是最大后验概率原理,根据先验知识求出后验分布,用于刻画所推断发育树为真的可能性,选择具有最大后验概率的那棵树作为系统发育树的最佳估计。尽管这种方法已经成功地应用于多种场合^[42,43],但是由于需要估计分布的先验概率,并且需要集成各种参数,同时这种方法的计算也很耗费时间,因此基于贝叶斯推断方法的应用也受到限制。

基于字符序列的系统发育树推断方法往往需要事先假定特定的进化模型,因此结果的有效性受模型选择的影响很大。

3.2 基于距离的系统发育树重构算法

基于距离的系统发育树重构算法是目前应用最广的一类系统发育树推断方法。其理论基础是最小进化的原理(minimum evolution, 简称为 ME)^[41],这类方法首先构造一个距离矩阵来表示每两个物种之间的进化距离,然后基于这个距离矩阵,采用聚类算法对研究的物种进行分类。通常的做法是根据物种之间的进化距离数据,采用合并策略来重构系统发育树。第一个基于距离数据的系统发育树重构算法由 Kidd 和 Sgaramella-Zonta 提出^[42],该方法的目标是寻找一棵具有

最小进化分支长度总和的发育树。标准的基于距离数据的方法是检查所有可能的进化树,从中选择需要最小变化次数的那棵树,对于含有较多物种的种群来说,这种方法也是不现实的。为了解决这个问题,人们提出了一些启发式的搜索策略,从一个距离矩阵开始,采用一定的准则,递归地合并矩阵中距离最短的节点,并重构新的距离矩阵,一直到只有3个分类单元为止。

基于距离度量重构系统发育树的两个经典算法是 UPGMA 法(unweighted pair-group method using arithmetic averages)^[43]和邻接法(Neighbour Joining, NJ)^[44]。邻接法是一种应用最广的合并算法,由 Saitou 和 Nei 首先提出,尽管 NJ 算法通常无法找到精确的最小进化树,只能找到近似的最小进化树,但是它的计算速度非常快,而且准确率较高,因此被广泛应用于系统发育分析。而且也出现了它的多个改进版本,如 Studier 和 Keppler 在 NJ 算法提出的第二年就对该方法进行了改进^[45],Bruno 等人提出了 Weighted NJ 算法^[46],Gasuel 提出了 BIONJ 算法^[47],Desper 等人提出了 FASTME 算法^[48],最近 Criscuolo 等人提出了一种解决距离矩阵中存在缺失数据的快速 NJ 算法^[49]。

基于距离的方法在原理上比基于字符的方法更加直观,更加容易理解;同时在计算的速度上通常也比基于字符序列的方法更加有优势,一旦能够计算出物种之间的进化距离矩阵,就可以很快地利用聚类算法构造出它们之间的系统发育树。

结束语 在过去的十几年里,基于分子水平的系统发育研究已经取得了明显的进步,这一方面是得益于生命科学的进步,另一方面是得益于计算科学的发展。在生物分子进化信息提取方面,从简单的序列比对,到基于全基因组序列的 gene order 和 gene content 信息,深深地烙着分子生物学技术的发展,尤其是大规模基因测序技术发展的痕迹,同时也由于计算机技术的发展,才使得在基因组水平上的大规模计算成为可能;在系统发育树重构算法方面,从早期的基于字符序列的算法,到后来的 NJ 算法,以及它的各种改进版本,也是因研究对象不断复杂化而相应地发展的。

在分子水平上研究物种的进化关系,开辟了人类探索生命的新途径,但是由于生物系统的复杂性和生命进化过程的复杂性,在某些时候,在分子水平上分析物种之间的进化关系并不能得到合理的结果。原因可能是多方面的,例如不同的生命分子的进化速度并不一致,它们所提供的进化信息存在差异;又如不同的进化事件产生不一样的进化信息,这些信息从不同的角度刻画物种之间的进化关系,如 gene content 和 gene order 反映的是基因组进化过程中的不同类型的进化事件,用它们来分析物种之间的系统发育关系,很可能得到不同的结果。有效地提取和利用物种之间的进化信息,给出物种之间的进化关系的合理描述和解释,将是我们面临的重要研究课题;在系统发育树重构算法研究方面,尽管合并算法是目前应用最广的一类系统发育树推断方法,但是由于计算过程中的每一步都需要合并和估计新的距离矩阵,这是比较费时的的工作,而且距离矩阵的估计误差可能随着计算的进行被逐渐放大,这就导致这类算法无法保证找到最准确的发育树,同时,合并算法采用的是一种局部的观点,通常只能得到局部最优解,而不是全局最优解,这就要求我们继续寻找更加有效的

系统发育树重构算法。

另外,根据目前已有的知识推断的物种进化关系存在一些冲突的现象,或者存在一些不确定的分支途径,而现有的树状模型不能很好地刻画这种现象,因此需要进一步研究更加合理的工具来描述物种之间的进化关系,比如网络结构^[50]在这方面的应用已经开始了很好的尝试。

随着生命科学研究的发展,人类在分子水平上对物种之间的进化关系的了解将更加深入和全面,对物种之间进化关系的刻画将更加合理和接近真实的生物现象。同时,随着机器学习理论的发展,新的机器学习方法和技术也将被用于分析物种之间的系统发育关系,物种进化研究将取得一些新的突破。

参 考 文 献

- [1] Nei M, Kumar S. *Molecular Evolution and Phylogenetics*[M]. New York: Oxford University, 2000
- [2] Lin J, Gerstein M. Whole-genome trees based on the occurrence of folds and orthologs; implications for comparing genomes on different levels[J]. *Genome Research*, 2000, 10: 808-818
- [3] Wu J Y, Xu X, J X L X. FlyPhy: a phylogenomic analysis platform for *Drosophila* genes and gene families[J]. *BMC Bioinformatics*, 2009, 10: 123
- [4] Ranwez V, Gascuel O. Improvement of distance-based phylogenetic methods by a local maximum likelihood approach using triplets[J]. *Mol Biol Evol*, 2002, 19: 1952-1963
- [5] Otu H H, Sayood K. A new sequence distance measure for phylogenetic tree construction[J]. *Bioinformatics*, 2003, 19 (16): 2122-2130
- [6] Qi J, Wang B, Hao B I. Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach [J]. *Journal of molecular evolution*, 2004, 58(1): 1-11
- [7] Hoppenrath M, Bachvaroff T R, Handy S M, et al. Molecular phylogeny of ocelloid-bearing dinoflagellates (Warnowiaceae) as inferred from SSU and LSU rDNA sequences[J]. *BMC Evolutionary Biology*, 2009, 9(1): 116
- [8] Huynen M A, Bork P. Measuring genome evolution[J]. *Proc. Natl Acad Sci USA*, 1998, 95: 5849-5856
- [9] Bapteste E, Boucher Y, Leigh J, et al. Phylogenetic reconstruction and lateral gene transfer[J]. *Trends Microbiol*, 2004, 12: 406-411
- [10] Choi I G, Kim S H. Global extent of horizontal gene transfer [J]. *Proc. Natl Acad Sci USA*, 2007, 104(11): 4489-4494
- [11] Bratke K A, McLysaght A. Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach[J]. *BMC Evolutionary Biology*, 2008, 8: 67
- [12] Rokas A, Williams A L, King N, et al. Genome-scale approaches to resolving incongruence in molecular phylogenies[J]. *Nature*, 2003, 425: 798-804
- [13] Alexander F A, Stefan R H, Barbara R H, et al. Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences[J]. *BMC Bioinformatics*, 2006, 7: 350
- [14] Bininda-Emonds O R P, Gittleman J L, Steel M A. The (super) tree of life: procedures, problems, and prospects[J]. *Annu Rev Ecol Syst*, 2002, 33: 265-289
- [15] Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life[J]. *Nat Rev Genet*, 2005, 6: 361-375
- [16] McLysaght A, Baldi P F, Gaut B S. Extensive gene gain associated with adaptive evolution of poxviruses[J]. *Proc. Natl Acad Sci USA*, 2003, 100: 15655-15660
- [17] James T Y, Kauff F, Schoch C L, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny[J]. *Nature*, 2006, 443(7113): 818-822
- [18] Dobzhansky T, Sturtevant A. Inversions in the chromosomes of *Drosophila pseudoobscura*[J]. *Genetics*, 1938, 23: 28-64
- [19] Bafna V, Pevzner P. Sorting by reversals: Genome rearrangement in plant organelles and evolutionary history of X chromosome[J]. *Mol Biol Evol*, 1995, 12: 239-246
- [20] Belda E, Moya A, Silva F J. Genome Rearrangement Distances and Gene Order Phylogeny in $\{\gamma\}$ -Proteobacteria[J]. *Mol Biol Evol*, 2005, 22: 1456-1467
- [21] Luo H, Shi J, Arndt W, et al. Gene Order Phylogeny of the Genus *Prochlorococcus*[J]. *PLoS ONE*, 2008, 3(12): e3837
- [22] Murphy W J, Larkin D M, Wind A E D, et al. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps[J]. *Science*, 2005, 309(5734): 613-617
- [23] Goodman D, Ollikainen N, Sholley C. Baculovirus Phylogeny Based on Genome Rearrangements[J]. *LNBI*, 2007, 4751: 69-82
- [24] Caprara A. Sorting by reversal is difficult[C]// *Proc. 1st Ann Int Conf on Research in Computational Molecular Biology*, 1997. 1997: 75-83
- [25] Snel B, Bork P, Huynen M A. Genome phylogeny based on gene content[J]. *Nat Genet*, 1999, 21: 108-110
- [26] Hughes A L, Friedman R. Poxvirus genome evolution by gene gain and loss[J]. *Mol Phylogenet Evol*, 2005, 35(1): 186-195
- [27] Gubser C, Hue S, Kellam P, et al. Poxvirus genomes: a phylogenetic analysis[J]. *J Gen Virol*, 2004, 85: 105-117
- [28] Iwasaki W, Takagi T. Reconstruction of highly heterogeneous gene-content evolution across the three domains of life [J]. *Bioinformatics*, 2007, 23: i230-i239
- [29] Snel B, Bork P, Huynen M A. Genomes in flux: the evolution of archaeal and proteobacterial gene content [J]. *Genome Res*, 2002, 12(1): 17-25
- [30] Kunin V, Ouzounis C A. GeneTRACE-reconstruction of gene content of ancestral species[J]. *Bioinformatics*, 2003, 19: 1412-1416
- [31] Gu X, Zhang H M. Genome phylogenetic analysis based on extended gene contents[J]. *Mol Biol Evol*, 2004, 21: 1401-1408
- [32] Huson D H, Steel M. Phylogenetic trees based on gene content [J]. *Bioinformatics*, 2004, 20(13): 2044-2049
- [33] Henz S R, Huson D H, Auch A F, et al. Whole Genome-based Prokaryotic Phylogeny[J]. *Bioinformatics*, 2005, 21: 2329-2335
- [34] Camin J, Sokal R. A method for deducing branching sequences in phylogeny[J]. *Evolution*, 1965, 19: 311-326
- [35] Hein J. A heuristic method to reconstruct the history of sequences subject to recombination[J]. *J Mol Evol*, 1993, 36: 396-405

- Piscataway; IEEE Computer Society, 2007; 1-5
- [6] Won-Ik Kim, Bong-Ju Lee, Jae-Su Song, et al. Ping-Pong avoidance algorithm for vertical handover in wireless overlay networks [C]//IEEE VTC'07. Piscataway; IEEE Computer Society, 2007; 1509-1512
- [7] Stevens-Navarro E, Lin Yuxia, Wong V W S. An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks [J]. IEEE Transactions on Vehicular Technology, 2008; 1243-1254
- [8] Chen L-J, Sun T, Chen B, et al. A smart decision model for vertical handoff [C]//ANWIRE'04. Piscataway; IEEE Inc., 2004; 1-5
- [9] Hassawa A, Nasser N, Hassanein H. Tramcar: a context-aware cross-layer architecture for next generation heterogeneous wireless networks [C]//ICC'06. Piscataway; IEEE Inc., 2006; 240-245
- [10] Song Qing-yang, Jamalipour A. A quality of service negotiation-based vertical handoff decision scheme in heterogeneous wireless systems [J]. European Journal of Operational Research, 2008, 191; 1059-1074
- [11] 时岩, 陈山枝. 一种基于 QoS 满意度评价的 mSCTP 垂直切换决策方法 [J]. 微电子学与计算机, 2008, 25(2); 9-13
- [12] 刘敏, 李忠诚, 过晓冰. 一种基于速度感知的垂直切换算法 [J]. 电子学报, 2008, 30(6); 1198-1201
- [13] 刘侠, 蒋铃鸽, 何晨. 一种基于模糊逻辑的预判决辅助垂直切换算法 [J]. 电子学报, 2007, 30(10); 1989-1993
- [14] 周永华, 毛宗源. 一种新的全局优化搜索算法—人口迁移算法 [J]. 华南理工大学学报: 自然科学版, 2003, 31(3); 1-5
- [15] 宋承先. 现代西方经济学(微观经济学) [M]. 上海: 复旦大学出版社, 2004; 515-521
- [16] 王兴伟, 郭磊, 秦培玉, 等. 一种 ABC 支持型接入网络选择机制 [OL]. 中国科技论文在线. http://www.paper.edu.cn/paper.php?serial_number=200909-109
- [17] Blake S. An architecture for differentiated services. RFC 2475, 1998. [OL]. <http://www.ietf.org/rfc/rfc2475.txt>
- [18] Odlyzko A. Paris metro pricing: The minimalist differentiate services solution [C]//NJ, USA. IEEE/IFIP IWQoS'99. Piscataway; IEEE Inc., 1999. 559-561
- [19] Maillé P, Tuffin B. Pricing the Internet with multibid auctions [J]. IEEE/ACM Transactions on Networking, 2006, 14(5); 992-1004
- [20] 杨纶标, 高英仪. 模糊数学原理及应用(第三版) [M]. 广州: 华南理工大学出版社, 2001; 76-79
- [21] Cao X R, Shen H X, Mילו R, et al. Internet pricing with a game theoretical approach: concepts and examples [J]. IEEE/ACM Transactions on Networking, 2002, 10(2); 208-216
- [22] 徐雷鸣, 庞博, 赵耀. NS 与网络模拟 [M]. 北京: 人民邮电出版社, 2003; 1-9
- [23] 王晓东. 计算机算法设计与分析(第二版) [M]. 北京: 电子工业出版社, 2005; 86-113
- [24] Zhu Fang, McNair J. Multiservice vertical handoff decision algorithms [J]. EURASIP Journal on Wireless Communications and Networking, 2006; 1-13

(上接第 51 页)

- [36] Hein J. Reconstructing evolution of sequences subject to recombination using parsimony [J]. Math Biosci, 1990, 98; 185-200
- [37] Jin G H, Nakhleh L, Snir S, et al. Efficient parsimony-based methods for phylogenetic network reconstruction [J]. Bioinformatics, 2006, 23; e123-e128
- [38] Sridhar S, Lam F, Billelloch G E, et al. Direct maximum parsimony phylogeny reconstruction from genotype data [J]. BMC Bioinformatics, 2007, 8; 472
- [39] Foulds L R, Graham R L. The steiner tree problem in phylogeny is np-complete [J]. Adv Appl Math, 1982, 3; 4-49
- [40] Felsenstein J. Evolutionary trees from DNA sequences; a maximum likelihood approach [J]. J Mol Evol, 1981, 17; 368-376
- [41] Saitou N, Nei M. The number of nucleotides required to determine the branching order of three species with special reference to the human-chimpanzee-gorilla divergence [J]. J Mol Evol, 1986, 24; 189-204
- [42] Kidd K K, Sgaramella-Zonta L A. Phylogenetic analysis: concepts and methods [J]. Am J Hum Genet, 1971, 23; 235-252
- [43] Sokal R R, Sneath P H A. Numerical Taxonomy [M]. San Francisco, CA; W. H. Freeman and Co, 1963
- [44] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. Mol Biol Evol, 1987, 4; 406-425
- [45] Studier J A, Keppler K J, et al. A note on the neighbor-joining algorithm of Saitou and Nei The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. Mol Biol Evol, 1988, 5; 729-731
- [46] Bruno W J, Succi N D, Halpern A L. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction [J]. Mol Biol Evol, 2000, 17; 189-197
- [47] Gascuel O. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data [J]. Mol Biol Evol, 1997, 14; 685-695
- [48] Desper R, Gascuel O. Fast and accurate phylogeny reconstruction algorithms based on the Minimum-Evolution principle [J]. J Compt Biol, 2002, 19; 687-705
- [49] Criscuolo A, Gascuel O. Fast NJ-like algorithms to deal with incomplete distance matrices [J]. BMC Bioinformatics, 2008, 9; 166
- [50] Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks [J]. Mol Biol Evol, 2004, 21(2); 255-256