遗传 HEREDITAS (Beijing) 2009 年 7 月, 31(7): 689－697
ISSN 0253-9772 www.chinagene.cn

综 述

DOI: 10.3724/SP.J.1005.2009.00689

整合系统发育信息: 概念、方法与挑战

吴良[1, 2], 宋明华[1], 欧阳华[1, 3]

1. 中国科学院地理科学与资源研究所, 北京 100101;
2. 中国科学院研究生院, 北京 100049;
3. 国际山地综合发展中心, 加德满都, 尼泊尔

摘要: DNA 序列、形态和其他同源性状可以用于推断物种的起源和历史。整合所有可利用的系统发育信息可以大大拓展所覆盖类群的范围, 推进我们对现存生物的认识, 而且使得生物学家提出和验证的假说尺度更广, 更有统计说服力。文章综述了整合系统发育信息的概念及其与传统分析的异同, 重点讨论了整合系统发育信息中应用最广的超级树(Supertree)和超级矩阵(Supermatrix)方法; 在比较分析了这两个方法的优缺点之后, 介绍了近些年提出的新的方法。文章详细分析了整合系统发育信息的发展所面临的来自数据和理论方面的挑战, 认为尽管整合分析的发展困难较多, 它仍然是到目前为止构建完整生命之树(网)的唯一方法; 它的完善必将拓展我们对于生物进化过程的认识, 并对进化生物学相关学科产生积极影响。

关键词: 系统发育学; 整合分析; supertree; supermatrix

Combining phylogenetic information: concept, methodology, and challenges

WU Liang[1, 2], SONG Ming-Hua[1], OUYANG Hua[1, 3]

1. *Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing* 100101, *China*;
2. *Graduate School of Chinese Academy of Sciences, Beijing* 100049, *China*;
3. *International Centre for Integrated Mountain Development, G.P.O. Box* 3226, *Khumaltar, Kathmandu, Nepal*

Abstract: The DNA sequences, morphological and other homologous characters can be used to infer the origins and histories of biological taxa. Combining all the phylogenetic information available can produce more inclusive phylogenies, improve our understanding of living organisms, and enable biologists to prompt and test hypotheses on a larger scale and with stronger statistical power. In this article, the concept of combining phylogenetic information and its comparison with traditional analysis were reviewed. The most popular approaches of supertree and supermatrix were discussed in detail, and novel ways were presented. Although the combining analysis is facing rigid challenges from data and foundation, it is currently the only approach for realization of the Tree(Net) of Life, and its development will definitely expand our knowledge of evolution on the earth and contribute to the progress of evolutionary related disciplines.

Keywords: phylogenetics; combining analysis; supertree; supermatrix

生物界的各类群由于一个共同的进化历史而有着不同程度的关联[1], 这种关联在达尔文的进化论出现以后, 才开始为人们所认识。系统发育学(Phylogenetics)通过物种同源性状的相似性比较,

收稿日期: 2008–12–08; 修回日期: 2009–02–22
基金项目: 国家重点基础研究发展计划项目(编号：2005CB422005)和国家自然科学基金项目(编号：30600070)资助
作者简介: 吴良(1984–), 男, 博士研究生, 研究方向：系统发育生物地理学。Tel: 010-64889809; E-mail: wul1984@163.com
通讯作者: 欧阳华(1958–), 男, 博士, 研究员, 研究方向：生态系统格局与过程。E-mail: ohua@igsnrr.ac.cn

分析物种之间的系统发育关系，探究各类群的起源和历史，最终构建全部生物的生命之树(Tree of Life, ToL)[2]。生物进化过程的多时空特征决定了能够用于系统发育分析的同源性状多种多样。早期探讨生物类群之间进化关系的研究主要局限于形态学或者是对超微结构的分析，比如化石和现存生物的比较解剖研究[3]。直到上个世纪 70 年代DNA测序技术发明之后，研究人员才开始从序列水平探讨生物界各类群的进化历史和系统发育关系[4]。实际上，除了化石和DNA序列，基于合理的进化假设的性状，比如细胞学、生理学、分子生物学乃至地理学和地层学的资料都可用于系统发育分析。因为它们从不同侧面反映出生物在进化过程中面对的各种不同尺度的选择作用以及做出的响应[5, 6]，而研究进化的目的也在于能够从不同的尺度和侧面对生物界的各个类群进行阐述。

过去的 30 多年内，生物信息学、系统发育学的发展，DNA测序技术的提高以及大规模基因组测序工作的深入，给系统发育研究带来了海量的分子序列和同源基因；而古生物学、进化生物学等学科的发展，为系统发育学研究积累了大量的性状数据。利用这些数据进行的单个的系统发育学研究使我们对于生物界各个类群有了一些基本的认识。然而，由于生物系统及其进化过程的复杂性，往往使得采用单个基因或者单个性状的传统系统发育学分析由于缺乏统计支撑而难以推断某些关键节点的系统发育关系[3]。从而有必要将单个的性状或者系统发育关系联合起来；同时，某一类群的完整的系统发育关系对于研究生物的适应性、验证宏观进化假说，以及指导物种保护的实践都具有重要价值[7]。因而，合并零散的系统发育资料用于构建大的系统发育树，从而获得覆盖更广、可信度更高的系统发育关系不仅是获取新的系统发育关系的一种简单、便宜而且有效的方法，而且已经逐渐成为一种需要。然而，传统的分析方法不能直接处理不同类型的数据；而合并不同来源的数据，甚至不同基因位点的序列都很可能产生不一致的系统发育关系[8, 9]，或者说信息冲突，而且这种冲突会随着数据的增多而更加突出[10]。因而，如何分析大量的系统发育信息、对产生的系统发育关系进行评价，以及处理数据分析过程中的冲突是决定系统发育分析可靠性的重要因素，也成为了构建生命之树的关键。

1 整合系统发育信息的概念

整合系统发育信息是指合并所有能够反映类群(Taxa)进化历 史的信息，从而获得许多在单独的数据中不能获得的新的系统发育关系，而且数据的规模和完整性也可以使生物学家们在一个前所未有的更大的尺度上、更有统计说服力地验证假说[11]。整合系统发育信息是相对于传统分析而言的，其数据基础和理论方法相互关联。传统的系统发育分析主要包括两个步骤：一是辨认在不同生物体之间共享的同源性状(Homologous characters)；二是通过比较这些性状，推断系统发育关系[3](图 1 中箭头 Ⅰ)。这两个步骤在整合分析中得以保留，也存在一些差异。

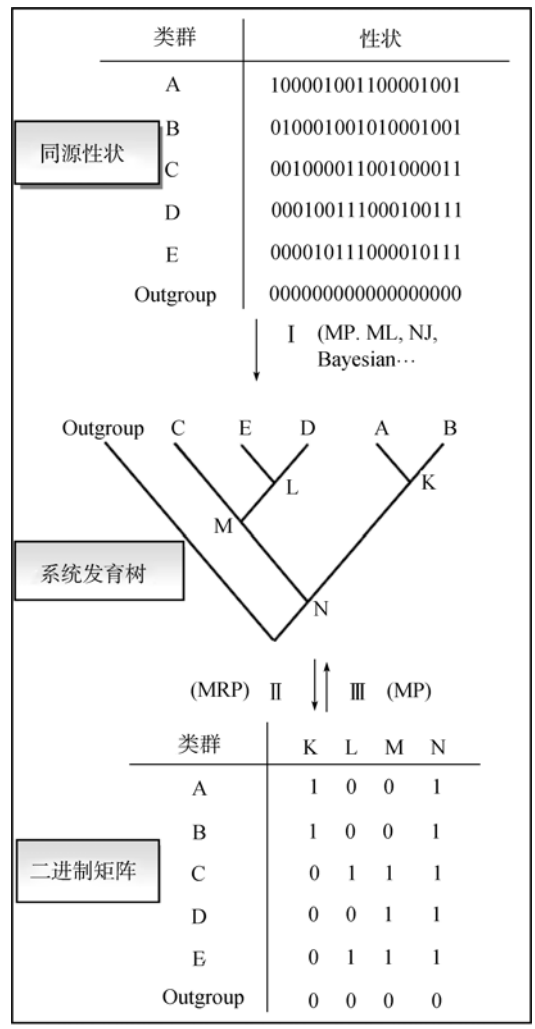

图 1 同源性状、系统发育树与矩阵表达及其相互关系

箭头Ⅰ、Ⅱ和Ⅲ分别表示从同源性状构建系统发育树，对系统发育树进行加性二进制编码，以及编码矩阵还原为原来的树的过程。同源性状的编码数据选自文献[12]。

首先，与传统的分析相比，整合分析的数据有了很大的拓展。传统系统发育分析利用的同源性状主要包括分子序列和形态学、细胞学、生物化学、生理学等方面的性状数据，类型比较单一，而且来源通常都相同；整合分析不仅囊括了传统分析的数据，还包括从传统分析得到的系统发育信息，主要是各种系统发育树，如最简约（MP)树、最大似然(ML)树、邻接（NJ)树等，导致整合分析的数据类型多样，来源也各不相同。整合系统发育的信息大致可以分成三类：一类是直接反映生物变异的DNA或者蛋白质序列；第二类是间接反映生物变异的离散特征性状，比如有性器官的演化特征或者动物的比较解剖证据等；第三类是已经发表的合理的系统发育树。分子序列虽然也是离散的性状数据，但由于其反映生物最本质的变异，而且不能同其他的性状数据合并处理，所以将其与性状数据分列。相同类型的数据之间相互兼容，反之则不兼容。

其次，数据的拓展使得系统发育关系的推断方法也有所区别。传统的系统发育分析通常是由多个类群的性状构成“性状-类群”矩阵，然后用传统的系统发育软件进行分析(比如广泛应用的PHYLIP软件包[13])。整合分析除了要处理性状数据，还要处理从性状数据分析得到的系统发育信息，甚至是这两部分数据的集合。整合分析的研究内容主要包括以下两个方面：一是相互兼容的数据的整合，比如合并不同位点的DNA序列数据，从而获得目标类群的更多基因序列或者覆盖更多的类群；二是相互不兼容数据的合并分析，比如生物形态性状矩阵同DNA序列的合并分析。这两方面的内容并非完全独立，而是可以同时存在。比如，Lee[14]在分析蛇的起源的时候，加入了一个化石类群*Pachyhachis*，对这个类群和现存相关类群的DNA序列和形态数据的分析都支持蛇的海洋起源，而非以前研究认为的陆地起源。这其中既包括化石与现存类群多个基因树的整合，也包括DNA序列与形态性状的综合。虽然在数据和内容方面存在差异，但整合分析的推断方法在很大程度上还是借鉴了传统分析的理论和技术，从而构建了较为完整的分析框架。

2　整合系统发育信息的方法

2.1　Supermatrix 途径

在探索整合分析各类系统发育信息，构建大的系统发育树的过程中，性状一致性(“Character Congruence”[15])和类群一致性(“Taxonomic Congruence”[16])成为解决整合系统发育信息最基本的策略[17]。性状一致性通常被称为合并方法，是基于Total Evidence (即全部证据)的哲学原则产生的[18]。在系统发育分析中，“Total Evidence”认为利用所有可以获得的系统发育信息才可以产生最好的科学假设[11, 18]。将同一类型的数据(比如DNA序列)串联(Concatenation)，组成一个类群-序列矩阵，缺失的部分用“？”代替，就可以把矩阵当作一个“Supergene”[19]，在常见的系统发育分析软件如PAUP*[20]、PHYLIP[13]、MEGA[21]中，运用MP、ML、MrBayes等算法构建系统发育树[22]。这个方法也就是后来为多数人所知的Supermatrix途径[22](图2)。Supermatrix途径的一个很好的特点在于它保存了所有的原始数据，这样数据可以被充分挖掘；数据之间的相互关系也得到保留。由于分子序列数据不断丰富，加上Supermatrix操作简便，而且往往能产生比较一致的结果，应用Supermatrix方法构建系统发育树的研究也不断增多。例如，Mcmahon和Sande- rson[23]分析了从GenBank中获取的Papilionoid Legume的2 228条DNA序列，利用Supermatrix方法从深浅两个尺度构建系统发育树，尽管面临很多困难，而且缺失数据高达96.6%，分析结果仍然可以作为Papilionoid Legume的较合理系统发育关系的集成；Pirie等[24]利用“自上而下”(Top down)和“自下而上”(Bottom up)的Supermatrix方法，获得了含79%的Danthonioid物种的系统发育关系，而且发现这个方法可以提高系统发育分析的分辨能力。不过，Supermatrix要求合并的数据的类型兼容，使得其在实际的研究中，只能处理性状矩阵与性状矩阵或者分子序列和分子序列整合的情况。在分子系统发育的研究中，由于只有少数几个基因(如*rbcL*)的覆盖范围广，数据缺失在相当长的时期内仍将是阻碍Supermatrix应用的主要因素。

2.2　Supertree 途径

与性状一致性相对应的是类群一致性。类群一致性方法也被称为单独分析方法，是指先单独构建各个数据的系统发育树，然后将这些树整合得到最终的树的方法[10, 25]，因而，它也被称为超级树(Supertree)途径(图2)。Supertree途径的算法很多，最早的Supertree采用的是非正式(Informal)的方法，将不

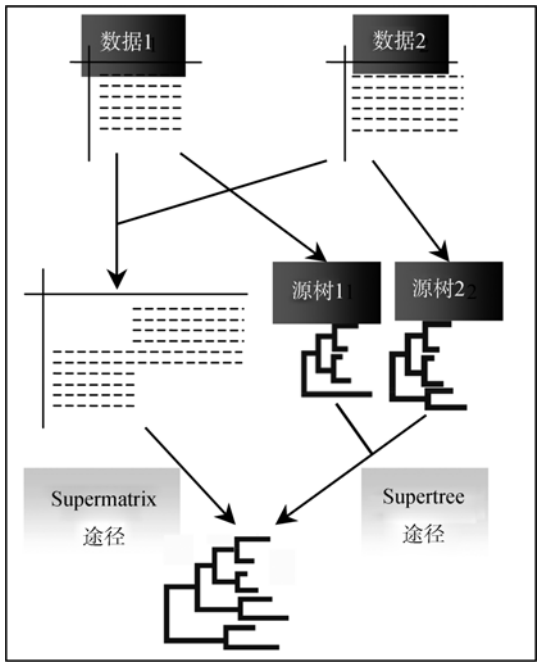

图 2 **supermatrix** 和 **supertree** 途径构建系统发育树的模式图

同的树依据其分类地位叠加到一起，形成一个大的系统树。这实际上就是将一棵大的系统树从某些节点分组形成若干小的树的逆操作，相当于在这些节点建立索引。后来的方法主要可以分成两类，即一致性(Agreement)和优化(Optimization)的方法。一致性方法以合意(Consensus)的算法为基础，比如最为保守的严格超级树(Strict supertree)算法[26]利用重叠的类群将树直接合并，保留一致的系统发育关系，而不一致关系的往往被瓦解(Collapse)，这个算法生成的系统发育树不能同源树的任意树相冲突[27]。除了严格超级树，一致性的算法还包括Mincut-supertree[28]、Semi-strict[29]等，但是同优化的方法相比，这些算法的应用非常有限。

优化的方法通常都包含一个优化标准(Optimization criterion)，即构树方法，也有很多包括了矩阵表征(Matrix representation)，其中由Baum[30]和Ragan[31]分别单独提出的MRP的算法最受关注，应用得也最多[11]。MRP的含义可以直接从它的名字中得出：即Matrix Representation(即矩阵表示) 和Parsimony(即简约性分析)。源树通过加性二进制编码(Additive binary coding)[32]转换成等同的性状矩阵(图 1 中箭头II)。编码规则如下：从结点(Node)衍生的类群编码为1，不从该结点衍生但是仍然在这个源树中的类群编为 0；所有其它的类群编为缺失，用“？”表示。源树之间只要有两个以上的类群重叠就可以合并其矩阵，空的位点用“？”填充，然后利用最大简约法构建系统发育树[30, 31]，并将所有的结点性状都编码为“0”的类群作为外类群(Outgroup) [27,30, 31]。MRP的两个主要过程目前都可以利用PAUP*[20]，SuperTree[33]，r8s[34]等软件实现自动处理。

Baum和Ragan的方法被称为标准MRP方法，后来有许多学者对编码的过程和赋值形式进行了修改[35~37]。比如可以对具有不同支持率的分支(依据Bootstrap值的大小)甚至对源树(依据源树的可信度)做加权处理。除了MRP，还有Matrix Representation using Distance(MRD)、Matrix Representation using Flipping(MRF)[38]等，但由于兼容性好、计算简便、易于操作等优点，MRP及其变型在哺乳动物、高等植物、细菌等类群的系统树构建中已经得到很好的应用[11]，并逐渐取代其他算法而成为构建supertree最流行的方法，因而后来所指的supertree方法通常都是指MRP。

2.3 Supermatrix 还是 supertree？

Supermatrix 和 Supertree 分别代表合并和独立的分析方法，是目前应用最广泛的分析方法。这两种方法虽然都可以有效合并系统发育信息，但由于它们的数据源以及理论依据不同，所采用的算法也大不一致，因而对于选用何种方法一直都存在争论。

Supermatrix较大程度保存了原始数据，信息丢失很少，而且往往能够产生单一的、具有较强自展支持率的系统进化树[39]，因而有人认为它较全面地反映了系统发育关系[24, 40]。但是，Supermatrix面临许多问题亟需解决：首先，Supermatrix在搜索树的时候只能使用一个核苷酸替换模型，因而用这个方法同时分析不同进化历史的性状的做法是有问题 的[11, 41]。Supermatrix方法通常假定所有的性状都是独立的，每个性状都形成同一棵物种树[22]。然而，实际的情况是，一些性状，比如许多DNA位点由于连锁(Genetic linkage)、基因重组(Gene recombination)和渗入(Introgression)等原因而并不独立，串联多个性状有可能混合代表不同进化历史的系统发育信号[42]。因此需要仔细追溯这些性状的来源以确定是否适合同其他的性状进行整合分析。其次，Supermatrix只能处理同一类型的数据，其他的数据，如距

离矩阵，DNA-DNA杂交数据等，都只能通过Supertree的方法合并进去。因而，严格从这个意义上考虑，Supermatrix甚至都不能成为一个完整的分析方法。第三，随着数据的不断加入，数据矩阵越来越大，给对位排列处理带来很大的压力[24, 39]；而且一旦加入新的数据，所有的分析工作就得重新进行，效率大大降低。第四，缺失数据会随着类群的增多而越来越多，导致分辨率的丢失，并有可能带来错误的系统发育关系[42]。尽管许多研究认为一定比例(通常在 12.5%~25%之间[43~45]) 的数据缺失是可以容忍的，甚至利用含 99.6%的缺失数据的supermatrix分析仍然得到了Crocodylians的较完整的系统发育关系[46]。但是，大量的缺失数据还是会让人产生不安，分析结果也未必会令人信服[45~47]。

同Supermatrix相比，Supertree一个很大的优势在于可以有效处理各种不同类型的数据，而且由于拓扑结构得以保留，因而受缺失数据的影响小。但是，它脱离原始数据的分析方法还是引起了强烈的质疑[46], Springer和de Jong[48]甚至认为Supertree只是一种有用的源树的归纳法，而非精确的系统发育重建方法。其次, Supertree方法不能直接获得各分支的长度。虽然Lapoints和Cucumel[37]利用平均合意(Average cunsensus)计算Supertree同各源树之间的距离的方法可以得到，但他们的方法建立在多个假定条件之上，使得直接应用较为困难。此外, Supertree的应用还存在一个数据重复的问题，即同样的一套数据出现在不同的谱系中，导致重复的类群的比重显著提高，分析结果产生偏向。这往往是由于相同的研究成果被许多人用于构建系统发育关系，而这些关系又被不加甄别的使用导致的。Gatesy[49]发现Liu及其同事[50]利用MRP方法构建的胎盘类哺乳动物的Supertree在个别的分枝关系上与通过Supermatrix方法得到的系统树有争议。通过重新分析他们所使用的数据，发现数据中存在大量的重复，比如β-血红蛋白序列在总共 30 棵源树中就出现过 6 次。不过，这个问题在很大程度上还是可以解决的。最好的办法就是仔细检查数据来源，剔除重复的数据[49]。

2.4　新的方法

从上面的分析可以看出，Supermatrix和Supertree这两种方法优劣并存，而且没有哪一个方法能解决所有的问题[3, 19]。因而，选择何种方法的问题实质上还是方法完善的问题。实际上，目前构建大的系统发育树的研究仍然在很大程度上受限于数据，以至于采取的策略也倾向于尽可能多地囊括类群或数据。许多人讨论将Supertree和Supermatrix相结合，比如，Sanderson和Driskellkr47提出将整合系统发育信息的策略分成三类: 单个基因、Supermatrix和Supertree，认为应该针对数据分布的实际情况采取不同的策略。更多人倾向的实际上是由Supermatrix和Supertree组合成的“ divide-and-conquer ”策略[42]，即“ 分治法 ”，也就是先利用同源性状构建Supermatrix矩阵，并生成单个的树，然后依据重叠的类群，利用Supertree将各个树组装起来，达到构建大的系统发育树的目的。这个策略甚至被认为是整合系统发育信息的最终出路[11]，因为Supermatrix充分利用了数据，而Supertree则最大程度的整合了数据，所以是对计算精度、计算量和类群覆盖度折中处理的结果，可以最大限度地利用已有的数据。

除此之外，Supermatrix和Supertree处理系统发育信息的思路也引发了人们探索新的方法。Steel和Rodrige[51]提出最大似然超级树(Maximum likelihood supertrees)的理论，将整合系统发育分析置于一个似然模型的框架中予以分析。这个模型在整合不同来源或者进化模式不一致的数据的时候非常有意义，因为它依据源树之间距离的分布状态来构建相应的模型，从而允许在同一个框架下采用不同的进化模型，并且可以进行统计检验[52]。这个模型具有很好的统计一致性，在实际应用中，随着数据的不断增多，模型的结果都必然趋向一致[52]。不过这个模型目前仍然只停留于理论阶段，具体的搜索树的算法和应用软件还在开发中。

此外，基于Supermatrix的新方法也已被提出。Smith等[53]提出了Mega-phylogeny的策略，借助序列数据库和分类等级(Taxonomic hierarchy)方法，利用高级程序语言和管道(Pipeline) 技术，使得处理过程很少人工干预，从而实现自动完成。利用这个方法，他们构建出目前最大的两棵巨型系统发育树: 菊目(Asterales)包括近 5 000 个物种、含有 5 个非编码位点，而植物界(Virdiplantae)则包括*rbcL*序列的 13 000 多物种。这个方法实际上也可以看作“ divide-and-conquer ”方法的策略，但由于这个方法依据同

源测试(Orthology test)和饱和度测试(Saturation test)对性状数据进行分组，然后依次加入预先设置好的指导树的末端，因而分析过程可操作性强，结果也更加可靠。

3 整合分析面临的挑战

3.1 数据缺乏

截止 2008 年 10 月, GenBank中序列数据总量达到 2.33×10^{11}碱基(除去基因组和contig的序列数为9.7×10^{10}碱基)，而序列数目达到 96.4×10^{6}(GenBank release 168)。即便如此，在GenBank中，至少含有一条分子序列的生物体(约 310 000 物种)只占到全部为人类所知的生物体(约 1 800 000 物种)的 17%左右(http://ncbi.nlm.nih.gov/)。而在系统发育信息数据库TreeBASE中，目前只录入了 2 000 多项研究的 5 000 多系统发育树，包括 100 000 个类群(http://www.treebase.org/)。尽管人们在测序方面的投入不断增加，今后的采样工作仍将侧重那些引人瞩目的、极具经济或保护价值的物种[54]。其结果必将是少数的物种具有较多的DNA序列(如模式生物)，也有少数的序列分布于较多的类群(如*rbcL*序列)，大多数的物种在将来相当长的一段时间内还将缺乏可以代表的分子序列[17]。因而，数据缺乏，以及由此而产生的长枝吸引(Long branch attraction, LBA)[55]和数据缺失将会对整合系统发育信息的分辨率和准确度产生严重影响。对于属下水平的系统发育学或者大尺度的生物地理学研究而言，这些问题将显得尤为突出。因而，原始数据的积累依然是制约着构建完整系统发育关系的最主要的因素[17]。

与化石和其他类型的系统发育数据相比较，分子序列数据算得上是相当丰富的了。分子序列以外的同源性状由于取样困难和分析相对复杂等因素，在系统发育分析中所占的比重越来越小，许多基于传统分析得出的系统发育关系正受到分子系统发育的挑战。如果没有新的化石或者发育证据出现，许多传统的观点将被推翻。因而，新的关键证据的发掘不光是满足分析数据的需要，也是保证系统发育学正确发展的重要因素。

3.2 数据冲突

同分析数据缺乏相对应的是数据的持续增加引起的数据冲突的不断增多。系统发育学的主要任务是构建物种树而不是性状树[56]，但是，即使是在数据和方法都无误的情况下，性状树依然会表现出同物种树不一致的系统发育关系，即性状树-物种树冲突[8, 9, 57, 58]。特别是在利用多基因片段进行系统发育重建的过程中，这种情况更加普遍。性状树-物种树冲突的原因往往是多方面的，比如DNA不同位点和不同类群进化速率的差异，一致进化(Convergent evolution)，横向基因迁移(Horizontal gene transfer, HGT)等等。邹新惠和葛颂[57]认为目前讨论的基因树冲突的实质是“探讨基因树不能正确反映物种树的原因”，并且将导致基因树冲突的原因归纳为 3 个大方面，即基因片段选择导致的随机误差(Stochastic error)、系统发育重建过程中的系统误差(Systematic error)和由复杂的生物学过程所产生的生物学因素(Biological factor)[59]。实际上，由于生物体进化的高度复杂性，或者由于我们目前的知识水平所限，在进行系统发育分析的时候这些原因都只能作为我们推断的依据，真实的系统进化情况是不可能充分了解的。因而，Maddison[60]认为，简单地把一些基因树看成支持，而把另外一些看成冲突的做法是错误的。他认为每个基因树都是物种树的一部分，所有的基因树形成基因历史的“云团”(Cloud)[60]，即物种树应该看作所有性状树的统计分布。

为了更全面地反映系统发育树的全貌，降低由少数的基因树反映物种树的片面性，一个直接的办法就是增加分析的基因片段的数目[19]。Rokas等[40]的研究表明，要获得较强的自展支持，利用多个基因的整合分析需要最少 20 个基因。不过近些年大规模的基因组测序工作和系统发育基因组学(Phylogenomics)[61]的飞速发展为系统发育分析提供了几乎无限的基因数目，因而这方面的限制正在逐步减少。

3.3 从生命之树到生命之网

随着人们对一系列遗传学事件，比如杂交、横向基因迁移(Horizontal gene transfer, HGT)和基因重组的认识，传统的系统发育学的理论和方法不断受到冲击，生命之树的宏伟设想开始变得不切实际，许多人倾向于生命之网(Net of life)[62]的表述—— 网的纵向代表进化，横向代表基因(群)的传递；网状进化取代了简单的分支过程。值得注意的是，生物

的网状进化都是在不一致的系统发育关系中发现的，而当前的许多(整合)系统发育分析方法都把这些不一致看成是“麻烦”，而非有用的信息。因而，在最终的系统发育树中这些不一致的信号都被过滤了。这样的结果导致我们不可能真正认识进化的历史[62]。在整合系统发育分析的过程中这种情况更为突出，因为整合分析的数据来源多种多样，类似MRP和Supermatrix的方法本身就可以减弱甚至消除这些不一致的信息。而且由于整合系统发育分析往往覆盖的类群范围较广，它们对杂交或者HGT事件的鉴别就显得更为薄弱了。即便如此，整合系统发育信息的方法仍然是当前建立完整生命之树(网)的唯一途径。而且，生物的网状进化通常只是在低等生物中表现得特别突出，高等生物中虽然也有发现，但主要还是集中于叶绿体和线粒体基因组。系统发育的网状进化的整合研究到目前为止还是空白，但传统系统发育的网状分析方法已经引起了很多关注[63~65]，相信能够检测、区分信息冲突并进行整合的分析方法在不久后也将成为可能。

4　展　望

随着生物学各领域研究的深入，考查生物体的进化历史和类群之间的亲缘关系已经成为几乎所有生物学研究的前提[3]。对于某一类群而言，完整的系统发育关系对于研究生物的适应性、验证宏观进化假说，以及指导物种保护的实践都具有重要价值[7]。系统发育重建是了解生物体进化历史最直接的途径，它的发展一方面得益于生物学其他学科的进步，另一方面也通过联系生物界各类群，促进生物学系统科学的发展。整合系统发育信息虽然在数据、理论和应用方面都面临一些问题，但它兼容处理各类数据的能力以及分析的简便性使其不断受到关注。整合系统发育信息是构建生命之树(网)的唯一途径，它的完善无疑会极大推动系统发育学的发展，扩展人们对于生物界各个类群的认识，并对保护生物学、古生物学及进化生物学做出积极贡献。

参考文献(References):

[1] 于黎, 张亚平. 系统进化基因组学—重建生命之树的一条迷人途径. 遗传, 2006, 28(11): 1445–1450.

[2] Maddison DR, Schulz KS, Maddison WP. The tree of life web project. *Zootaxa*, 2007, 1668: 19–40.

[3] Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*, 2005, 6(5): 361–375.[DOI]

[4] Avise JC. Phylogeography: Retrospect and prospect. *J Biogeogr*, 2009, 36(1): 3–15.[DOI]

[5] Doyle JJ. Trees within trees: Genes and species, molecules and morphology. *Syst Biol*, 1997, 46(3): 537–553.

[6] Lazarus DB, Prothero DR. The role of stratigraphic and morphologic data in phylogeny. *J Paleontol*, 1984, 58(1): 163–172.

[7] Bininda-Emonds ORP, Gittleman JL, Purvis A. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). *Biol Rev*, 1999, 74(2): 143–175.[DOI]

[8] Palmer JD, Zamir D. Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. *Proc Natl Acad Sci USA*, 1982, 79(16): 5006–5010.[DOI]

[9] Zou XH, Zhang FM, Zhang JG, Zang LL, Tang L, Wang J, Sang T, Ge S. Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biol*, 2008, 9(3): R49.[DOI]

[10] de Queiroz A, Donoghue MJ, Kim J. Separate versus combined analysis of phylogenetic evidence. *Annu Rev Ecol Syst*, 1995, 26: 657–681.[DOI]

[11] Bininda-Emonds ORP. The evolution of supertrees. *Trends Ecol Evol*, 2004, 19(6): 315–322.[DOI]

[12] Brooks DR, van Veller MGP, McLennan DA. How to do BPA, really. *J Biogeogr*, 2001, 28(3): 345-358.[DOI]

[13] Felsenstein J. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics*, 1989, 5: 164–166.

[14] Lee MSY. Molecular evidence and marine snake origins. *Biol Lett*, 2005, 1(2): 227–230.[DOI]

[15] Michevich MF. Taxonomic congruence. *Syst Zool*, 1978, 27: 143–158.[DOI]

[16] Kluge AG, Wolf AJ. Cladistics: What's in a word? *Cladistics*, 1993, 9(2): 183–199.[DOI]

[17] Bininda-Emonds ORP, Gittleman JL, Steel MA. The (Super)tree of life: Procedures, problems, and prospects. *Annu Rev Ecol Syst*, 2002, 33: 265–289.[DOI]

[18] Kluge AG. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes) *Syst Zool*, 1989, 38(1): 7–25.[DOI]

[19] Rannala B, Yang ZH. Phylogenetic inference using whole genomes. *Annu Rev Genom Hum Genet*, 2008, 9:

217–231.[DIO]

[20] Swofford DL, PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4. Sinauer Associates, Sunderland, Massachusetts, 2003.

[21] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol*, 2007, 24(8): 1596–1599.[DOI]

[22] de Queiroz A, Gatesy J. The supermatrix approach to systematics. *Trends Ecol Evol*, 2007, 22(1): 34–41.[DOI]

[23] McMahon MM, Sanderson MJ. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Syst Biol*, 2006, 55(5): 818–836.[DOI]

[24] Pirie MD, Humphreys AM, Galley C, Barker NP, Verboom GA, Orlovich D, Draffin SJ, Lloyd K, Baeza CM, Negritto M, Ruiz E, Sanchez JHC, Reimer E, Linder HP. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Mol Phylogenet Evol*, 2008, 48(3): 1106–119.[DOI]

[25] Eernisse DJ, Kluge AG. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Mol Biol Evol*, 1993, 10(6): 1170–1195.

[26] Gordon AD. Consensus supertrees: The synthesis of rooted trees containing overlapping sets of labeled leaves. *J Classif*, 1986, 3(2): 335–348.[DOI]

[27] Sanderson MJ, Purvis A, Henze C. Phylogenetic supertrees: Assembling the trees of life. *Trends Ecol Evol*, 1998, 13(3): 105–109.[DOI]

[28] Semple C, Steel M. A supertree method for rooted trees. *Discrete Appl Math*, 2000, 105(1-3): 147–158.[DOI]

[29] Goloboff PA, Pol D. Semi-strict supertrees. *Cladistics*, 2002, 18(5): 514–525.

[30] Baum BR. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon*, 1992, 41(1): 3–10.[DOI]

[31] Ragan MA. Phylogenetic inference based on matrix representation of trees. *Mol Phylogenet Evol*, 1992, 1(1): 53–58.[DOI]

[32] Farris JS, Kluge AG, Eckardt MJ. A numerical approach to phylogenetic systematics. *Syst Biol*, 1972, 19: 172–191.

[33] Salamin N, Hodkinson TR, Savolainen V. Building supertrees: An empirical assessment using the grass family (Poaceae). *Syst Biol*, 2002, 51(1): 136–150.[DOI]

[34] Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, 2003, 19(2): 301–302.[DOI]

[35] Bininda-Emonds ORP, Bryant HN. Properties of matrix representation with parsimony analyses. *Syst Biol*, 1998, 47(3): 497–508.

[36] Rodrigo AG. A comment on Baum's method for combining phylogenetic trees. *Taxon*, 1993, 42(3): 631–66.[DOI]

[37] Lapointe FJ, Cucumel G. The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa. *Syst Biol*, 1997, 46(2): 306–312.

[38] Chen D, Eulenstein O, Fernández-Baca D, Sanderson M, Supertrees by Flipping, in Computing and Combinatorics. Springer Berlin: Heidelberg, 2002, 128–137.

[39] Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol*, 2007, 56(1): 17–24.[DOI]

[40] Rokas A, Williams BL, King N, Carroll SB. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 2003, 425(6960): 798–804.[DOI]

[41] Daubin V, Moran NA, Ochman H. Phylogenetics and the cohesion of bacterial genomes. *Science*, 2003, 301(5634): 829–832.[DOI]

[42] Sanderson MJ, Driskell AC. The challenge of constructing large phylogenetic trees. *Trends Plant Sci*, 2003, 8(8): 374–379.[DOI]

[43] Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature*, 1999, 402 (6760): 404–407.[DIO]

[44] Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. *Nature*, 2001, 409(6820): 614–618.[DOI]

[45] Bapteste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Durufle L, Gaasterland T, Lopez P, Muller M, Philippe H. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc Natl Acad Sci USA*, 2002, 99(3): 1414–1419.[DOI]

[46] Gatesy J, Baker RH, Hayashi C. Inconsistencies in arguments for the supertree approach: Supermatrices versus supertrees of Crocodylia. *Syst Biol*, 2004, 53(2): 342–355.[DIO]

[47] Kearney M. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. *Syst Biol*, 2002, 51(2): 369–381.[DOI]

[48] Springer MS, de Jong WW. Phylogenetics-Which mammalian supertree to bark up? *Science*, 2001, 291(5509):

1709–1711.[DOI]

[49] Gatesy J, Matthee C, DeSalle R, Hayashi C. Resolution of a supertree/supermatrix paradox. *Syst Biol*, 2002, 51(4): 652–664.[DOI]

[50] Liu F-GR, Miyamoto MM, Freire NP, Ong PQ, Tennant MR, Young TS, Gugel KF. Molecular and morphological supertrees for Eutherian (Placental) mammals. *Science*, 2001, 291(5509): 1786–1789.[DOI]

[51] Steel M, Rodrigo A. Maximum likelihood supertrees. *Syst Biol*, 2008, 57(2): 243–250.[DOI]

[52] Cotton JA, Wilkinson M. Supertrees join the mainstream of phylogenetics. *Trends Ecol Evol*, 2008, 24(1): 1–3.

[53] Smith S, Beaulieu J, Donoghue M. Mega-phylogeny approach for comparative biology: An alternative to supertree and supermatrix approaches. *BMC Evol Biol*, 2009, 9(1): 37.[DOI]

[54] O'Brien SJ, Eizirik E, Murphy WJ. GENOMICS: On choosing mammalian genomes for sequencing. *Science*, 2001, 292(5525): 2264–2266.

[55] Bergsten J. A review of long-branch attraction. *Cladistics*, 2005, 21(2): 163–193.[DOI]

[56] Tateno Y, Takezaki N, Nei M. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol Biol Evol*, 1994, 11(2): 261–277.

[57] McClean PE, Hanson MR. Mitochondrial DNA sequence divergence among *Lycopersicon* and related solanum species. *Genetics*, 1986, 112(3): 649–667.

[58] Kim ST, Donoghue MJ. Molecular phylogeny of *Persicaria* (Persicarieae, Polygonaceae). *Syst Bot*, 2008, 33(10): 77–86.[DOI]

[59] 邹新惠，葛颂．基因树冲突与系统发育基因组学研究. 植物分类学报, 2008, 46(6): 795–807.

[60] Maddison WP. Gene trees in species trees. *Syst Biol*, 1997, 46(3): 523–536.

[61] Eisen JA, Fraser CM. Phylogenomics: Intersection of evolution and genomics. *Science*, 2003, 300(5626): 1706–1707.[DOI]

[62] Doolittle WF. Phylogenetic classification and the universal tree. *Science*, 1999, 284(5423): 2124–2128.[DOI]

[63] Linder CR, Rieseberg LH. Reconstructing patterns of reticulate evolution in plants. *Am J Bot*, 2004, 91(10): 1700–1708.[DOI]

[64] Makarenkov V, Legendre P. From a phylogenetic tree to a reticulated network. *J Comput Biol*, 2004, 11(1): 195–212.[DOI]

[65] McBreen K, Lockhart PJ. Reconstructing reticulate evolutionary histories of plants. *Trends Plant Sci*, 2006, 11(8): 398–404.[DOI]

•综合信息•

《遗传》被评为“RCCSE 中国核心学术期刊”

本刊讯：由中国学术期刊评价委员会、武汉大学中国科学评价研究中心(Research Center for Chinese Science Evaluation, RCCSE)发布的《中国学术期刊评价研究报告》(2009–2010)中,《遗传学报》被评为“RCCSE 中国权威学术期刊”; 同时《遗传》被评为“RCCSE 中国核心学术期刊”。2009 年 6 月，《遗传学报》和《遗传》编委会收到了上述证书。

据悉，《中国学术期刊评价研究报告》将中国的学术期刊评为 5 个等级: A+等为权威期刊(占 5%); A 等为核心期刊(占 15%); B+等为准核心期刊(占 30%); B 等为一般期刊(占 30%); C 等为较差期刊(占 20%)。

另外,《中国学术期刊评价研究报告 —— RCCSE 权威、核心期刊排行榜与指南》(书号: 03-024128-3)已于 2009 年 4 月由科学出版社出版。这是国内外期刊评价中第一种分类分级排行榜和权威与核心期刊指南。作者采用定量评价与定性分析相结合的方法，构建了科学、合理的多指标评价体系，得出了 65 个学术期刊排行榜，包括分学科的 61 个排行榜和分类型的 4 个高校学报排行榜。这次共有 6170 种中国学术期刊参与评价，计 1324 种学术期刊进入核心区，其中权威期刊 311 种、核心期刊 1013 种，约占总数的 21.46%；并分析了核心期刊的学科分布、地区分布，自然科学类核心期刊被国外重要数据库收录，综合性核心期刊的核心效应，中国英文学术期刊的国际学术影响力等状况。本书还有 1324 种权威期刊与核心期刊的基本信息与投稿指南。附录中汇集了 SCI、EI 收录的中国期刊和中国出版的其他英文学术期刊及缩略语表等，便于广大读者阅读和投稿时查阅使用。

该《指南》既可为各级各类的科学评价和科研管理工作提供重要基础和定量依据，又能为各个图书馆及文献情报单位选购期刊、优化馆藏提供必不可少的有效工具，还可供广大读者、作者、期刊编辑部、政府管理部门、图书情报人员、信息工作者、广大知识分子以及社会各界人士阅读和使用。