

<<分子系统发生学>>

图书基本信息

书名：<<分子系统发生学>>

13位ISBN编号：9787030330260

10位ISBN编号：7030330269

出版时间：2012-6

出版时间：科学出版社

作者：黄原

页数：557

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<分子系统发生学>>

### 内容概要

分子系统发生学是应用分子数据重建系统发生关系的学科。

《分子系统发生学》全面系统地论述了分子系统发生学的基础、原理、方法及应用。

《分子系统发生学》由18章组成，可以归纳为五大部分：第一部分包括第1~3章，分别介绍了系统发生和系统树的基本知识；第二部分包括第4~7章，是分子系统发生分析的基础，其中第4章和第5章是分子系统发生学的信息学基础，第6章是数据集系统发生信号评估，第7章讨论了分子进化模型及模型选择原理与方法；第三部分中的第8~12章是各种系统发生分析方法，分别就目前主要的系统发生分析方法（距离矩阵法、简约法、最大似然法、贝叶斯推论法和系统发生网络法等）从原理、软件操作、应用及局限性等方面进行了详细的介绍，第13章讨论了系统发生假设检验的原理和方法，第14章讨论了系统发生分析可靠性与影响因素；第四部分主要涉及各类数据集分析策略，其中第15章总结了不同类型数据的分析策略，第16章对复杂数据系统发生的分析策略与方法进行了详细地介绍，第17章是多基因数据分析策略和方法；最后一部分即第18章是系统树的可视化、注释与应用方面的内容。

《分子系统发生学》可作为生物学、生物技术、生态学和生物信息学专业的本科生、研究生及科研人员学习分子系统发生学的教材或参考资料。

## &lt;&lt;分子系统发生学&gt;&gt;

## 书籍目录

前言第1章 系统发生学概论1.1 系统发生与系统发生学1.2 系统发生关系的含义1.2.1 表征关系1.2.2 分支关系1.2.3 遗传关系1.2.4 系统发生关系1.2.5 年代关系1.2.6 地理分布关系1.3 分子系统发生分析的原理和假设1.3.1 分子系统发生分析的原理1.3.2 分子系统发生分析的假设1.3.3 分子数据的优点1.4 分子系统发生学的方法论1.5 分子系统发生学的发展历史1.6 系统发生分析的策略与步骤1.7 分子系统发生学的文献资源1.7.1 分子系统发生学期刊1.7.2 分子系统发生学领域主要专著和教科书1.8 分子系统发生学的成就和问题第2章 系统发生分析基础2.1 分子进化基础2.1.1 分子进化的动力2.1.2 分子进化的中性理论2.1.3 溯祖理论2.2 系统发生分析的分类学基础2.2.1 系统发生与分类学的关系2.2.2 分类阶元的系统发生意义2.3 性状和性状分析方法2.3.1 性状的分类2.3.2 关于性状的基本假设2.3.3 性状进化分析方法2.3.4 性状的加权2.3.5 性状的同源2.3.6 性状的同型2.4 系统发生分析的数学基础2.5 系统发生分析的统计学基础2.5.1 概率分布2.5.2 系统发生的统计学检验2.5.3 零假设与零模型2.5.4 常用检验方法2.5.5 随机数据及其在系统发生中的应用2.6 理论系统发生学2.7 模拟系统发生研究2.7.1 系统树的模拟2.7.2 序列的模拟2.7.3 系统发生模拟研究的优势2.8 系统发生分析的算法2.8.1 精确算法2.8.2 启发式算法第3章 系统树3.1 系统树的概念和含义3.2 系统树的要素3.2.1 系统树的拓扑结构3.2.2 系统树的节点3.2.3 系统树的分枝和分枝长度3.3 演化历史与系统树的完整性3.4 系统树表达的信息3.5 系统树概念和表达形式的发展3.6 系统树的类型3.6.1 树状图与网状图3.6.2 有根树和无根树3.6.3 标度树与未标度树3.6.4 基因树和物种树3.6.5 基础树和合一树、源树和超树3.6.6 期望树与实际树3.6.7 普适生命树与完全树3.6.8 二歧树和多歧树3.6.9 系统树的表示形式3.7 系统树的数学描述3.7.1 系统树各部位的名称3.7.2 二分树及其表示方式3.7.3 二歧树的性质3.8 系统树的赋根方法3.9 系统树的生物学描述和解释3.9.1 描述系统树的基本术语3.9.2 系统树的分类学解释3.9.3 系统树的进化解释第4章 系统发生信息学4.1 系统发生信息学概述4.2 系统发生信息学研究内容4.3 系统发生数据文件格式4.3.1 数据文件格式4.3.2 格式转换软件4.3.3 系统树文件格式4.4 系统发生分析软件4.4.1 系统发生分析软件概述4.4.2 系统发生分析软件的编程语言4.4.3 系统发生分析软件的使用4.5 PAUP\*软件及使用4.5.1 PAUP\*软件的历史和版本4.5.2 PAUP\*的安装4.5.3 PAUP\*的功能4.5.4 PAUP\*命令及操作4.5.5 PAUP\*使用的一般步骤4.5.6 ClustalX和PAUP\*连用4.5.7 PAUP\*4辅助软件4.6 MEGA 5软件包简介4.7 DAMBE软件包简介4.8 Sea View 4软件包简介4.9 PHYLIP软件包简介4.10 系统发生的自动化分析工具4.11 系统发生网络资源4.11.1 系统发生软件目录4.11.2 CIPRES4.11.3 分子进化和系统发生专题研讨会4.12 系统发生数据库介绍4.12.1 系统发生知识数据库4.12.2 生命之树数据库4.12.3 Species 2000数据库4.12.4 NCBI分类数据库4.13 系统发生信息学展望第5章 数据集准备与序列比对5.1 分子数据的获得5.1.1 自测数据5.1.2 序列拼接5.2 来源于公共数据库的分子数据5.2.1 查看分类单元中已知基因序列分布的方法5.2.2 查看一个分类单元被提交到GenBank中序列数量的方法5.2.3 查看一个分类单元有序列记录物种数量的方法5.2.4 数据库序列获取方法5.2.5 批量下载序列的方法5.2.6 比对序列数据库5.3 序列比对5.3.1 比对的分类和概念5.3.2 序列比对的原理5.3.3 序列比对算法5.3.4 比对方法的分类5.4 常用比对软件5.4.1 ClustalX5.4.2 T-Coffee5.4.3 DIALIGN5.4.4 MUSCLE和MAFFT5.4.5 ProAlign5.4.6 POA和ABA5.5 比对软件的选择5.6 不同类型的序列比对方法和策略5.6.1 DNA序列比对方法和策略5.6.2 RNA基因序列的比对方法与策略5.6.3 蛋白质序列比对5.7 比对结果的美化显示与格式转化5.7.1 比对结果的美化和位点信息显示5.7.2 比对结果的格式转化5.8 比对与系统发生分析5.9 数据集中空位、模糊区、多态位点和丢失数据的处理5.9.1 数据集中空位的处理5.9.2 模糊比对序列的处理5.9.3 多态性状的处理5.9.4 丢失数据的处理5.10 多源数据集组装5.10.1 公共数据库数据的组装5.10.2 多基因数据的连接5.11 序列管理与数据提交5.11.1 序列管理5.11.2 系统发生数据提交第6章 数据集系统发生信号评估6.1 系统发生数据信号描述6.2 数据集质量的评价6.2.1 数据集组成特征分析6.2.2 替换型式分析6.2.3 分子进化参数计算6.2.4 替换饱和作图6.3 系统发生信号与结构分析6.3.1 序列数据系统发生信号强弱的评价6.3.2 系统发生信号评估软件与方法6.3.3 系统发生信号组成结构分析6.4 系统发生数据探索与实验性分析6.4.1 数据特征的探索6.4.2 系统发生数据的实验性分析第7章 进化模型及其选择7.1 进化模型及其在系统发生分析中的作用7.2 系统发生模型7.3 形态性状进化模型7.4 DNA序列进化模型7.4.1 DNA序列上发生的进化改变7.4.2 同质性模型7.4.3 碱基组成异质性模型7.4.4 Indel模型7.5 RNA进化模型7.5.1 结构RNA序列的进化特征7.5.2 RNA替换模型7.6 蛋白质序列进化模型7.6.1 蛋白质序列进化及建模7.6.2 经验模型7.6.3 机理模型7.6.4 氨基酸频

## &lt;&lt;分子系统发生学&gt;&gt;

率变异和位点之间速率变异模型7.6.5 混合模型7.7 进化模型的选择7.7.1 进化模型选择原理7.7.2 LRT检验法7.7.3 AIC信息标准法7.7.4 贝叶斯信息标准法7.7.5 贝叶斯因子法7.7.6 决策论法7.7.7 进化模型选择注意事项7.8 DNA进化模型选择7.8.1 用PAUP\*选择模型的LRT检验7.8.2 DNA模型选择软件7.8.3 jModelTest的使用7.9 蛋白质进化模型的选择和使用7.9.1 蛋白质进化模型选择概述7.9.2 蛋白质进化模型选择软件ProtTest3.07.10 进化模型参数的准确估计7.11 混合模型和平均模型第8章 距离矩阵方法8.1 遗传距离的概念8.2 距离数据的数学特征和生物学意义8.3 将序列数据转化为距离的方法8.3.1 未校正的遗传距离8.3.2 校正距离的计算方法8.3.3 最大似然法估计的校正距离8.3.4 LogDet距离8.3.5 基因组距离8.3.6 蛋白质遗传距离8.3.7 计算遗传距离的软件8.3.8 校正距离的选择和使用注意事项8.4 距离矩阵方法概述8.5 聚类分析方法8.6 邻接法8.6.1 邻接法原理8.6.2 邻接法的算法8.7 最小进化法8.8 叠加树法8.8.1 原理8.8.2 平均距离法8.8.3 转换距离法8.8.4 最小平方方法8.8.5 其他叠加树方法8.9 距离树可靠性评价8.10 距离矩阵建树方法的比较及应用8.11 距离矩阵法建树软件8.11.1 PAUP\*4距离法建树8.11.2 MEGA5的距离法8.11.3 TREECON使用8.11.4 T-REX软件使用8.11.5 ProfDist使用方法第9章 简约法9.1 简约性方法原理9.2 简约法的分析过程9.2.1 性状分布模式9.2.2 性状优化9.2.3 多态性内部节点祖先状态的重建方法9.2.4 性状加权9.2.5 最简约树搜索9.2.6 简约树分枝长度和树长的计算9.2.7 最简约树的选择9.2.8 MP树分支支持度计算9.3 数据集中同型性状水平的分析和评价9.4 简约法分析结果9.5 简约性方法的优缺点9.6 简约法分析软件9.7 用PAUP\*进行MP法分析9.7.1 利用PAUP\*进行简单简约法分析9.7.2 加权简约法分析9.7.3 PAUP\*限制树搜索9.7.4 PAUP\*4简约法的脚本命令运行9.8 TNT软件9.9 WinClada和NOVA第10章 最大似然法10.1 最大似然法原理及其在系统发生分析上的应用10.2 最大似然法建树原理10.3 最大似然法建树过程10.3.1 进化模型的选择及参数计算10.3.2 系统树搜索方法10.3.3 分枝长度的优化10.3.4 似然值的计算10.3.5 分支支持度计算10.4 最大似然法建树结果的表示10.5 最大似然法的优缺点10.5.1 最大似然法的优点10.5.2 最大似然法的缺点10.6 最大似然法分析软件10.6.1 PAUP\*4的ML分析方法10.6.2 PAUP\*与ModelTest联合运行选择进化模型10.6.3 TREEFINDER软件使用方法10.6.4 TREE-PUZZLE软件使用方法10.6.5 RAxML10.6.6 PhyML10.6.7 MetaPIGA10.6.8 IQPNNI10.6.9 GARLI第11章 贝叶斯系统发生推论法11.1 贝叶斯系统发生分析原理11.1.1 贝叶斯统计原理11.1.2 贝叶斯系统发生推论法历史和现状11.1.3 贝叶斯系统发生推论原理11.2 贝叶斯分析过程11.2.1 贝叶斯方法选择模型11.2.2 先验概率的设置11.2.3 马尔可夫链运行设置11.2.4 提议、混合与接受11.2.5 贝叶斯推论法克服局部优化的方法11.2.6 评估和促进后验概率分布收敛的方法11.2.7 影响系统树后验概率计算的因素11.3 贝叶斯法运行结果汇总11.4 贝叶斯推论法结果的分析、判断与表示11.5 贝叶斯系统发生软件及使用11.5.1 贝叶斯系统发生软件11.5.2 MrBayes 3.2使用方法11.6 贝叶斯系统发生推论法优缺点11.7 贝叶斯法与最大似然法的联系及区别11.8 贝叶斯后验概率与自举支持度的关系第12章 系统发生网络、超树和无比对方法12.1 系统发生网络12.1.1 网状进化型式与机制12.1.2 系统发生网络的构建方法12.1.3 网状图的构建软件12.1.4 系统发生网络的应用12.2 系统树的整合方法——超树12.2.1 超树的概念12.2.2 超树构建方法12.2.3 超树方法的优缺点12.3 无比对方法12.3.1 比对和系统发生的联合估计方法12.3.2 完全无比对方法第13章 系统发生假设检验13.1 系统发生假设检验概述13.2 似然比检验13.3 数据随机化检验13.3.1 比较双树检验13.3.2 PTP检验和限制树T-PTP检验13.4 配对位点检验13.4.1 Templeton检验13.4.2 KH检验13.5 非参数自举法13.5.1 SH检验13.5.2 AU检验13.6 参数自举法13.7 贝叶斯统计检验法13.8 PAUP\*执行的系统发生假设检验方法13.9 CONSEL软件使用第14章 系统发生分析的可靠性与影响因素14.1 系统发生分析方法的可靠性14.1.1 方法可靠性的评价标准14.1.2 系统发生分析方法的比较研究14.1.3 不同构树方法的优缺点14.2 系统树的可靠性14.2.1 系统树的两类误差14.2.2 系统误差和随机误差14.2.3 检验系统树可靠性的统计学方法14.3 随机误差及统计分析14.3.1 评估分支支持度的方法14.3.2 自举法14.3.3 自减法14.3.4 贝叶斯后验概率法14.3.5 计算分支支持度的软件14.4 系统误差的消除方法14.4.1 系统误差的来源14.4.2 导致系统误差的条件14.4.3 系统误差的识别14.4.4 系统误差的消除方法14.5 系统发生分析疑难解答14.5.1 有异常分支的系统发生14.5.2 随机误差14.5.3 分类单元抽样14.5.4 序列长度与类型14.5.5 序列比对问题14.5.6 进化模型选择问题14.5.7 建树方法的选择14.5.8 搜索算法选择14.5.9 分子进化速率对系统发生的影响14.5.10 替换速率变异14.5.11 碱基组成偏向性的影响14.5.12 碱基组成异质性的影响14.5.13 外群选择与系统树的赋根问题14.5.14 谱系缺失的影响14.5.15 数据缺失对系统发生分析的影响14.5.16 基因水平转移14.5.17 序列和位点同源关系14.5.18 选择作用的影响14.5.19 重组的影响14.5.20 分支支持度低的问题14.5.21 计算时间太长

## &lt;&lt;分子系统发生学&gt;&gt;

的问题14.5.22 总结第15章 不同类型数据的分析策略15.1 不同类型数据的特点15.2 DNA序列分析策略和方法15.2.1 用DNA序列还是蛋白质序列15.2.2 编码蛋白质DNA序列的分析15.2.3 DNA序列的加权简约法分析15.2.4 DNA序列的ML和贝叶斯法分析15.3 蛋白质序列分析策略和方法15.3.1 蛋白质序列数据的获得15.3.2 必须使用蛋白质序列的情况15.3.3 蛋白质序列的分析策略15.3.4 蛋白质立体结构分析15.4 RNA序列分析策略和方法15.4.1 RNA序列数据的特点15.4.2 rRNA基因序列系统发生分析策略15.4.3 rRNA基因序列分析软件第16章 复杂数据和困难系统发生的分析策略与方法16.1 早期适应辐射的系统发生16.2 近期发生过适应辐射的系统发生16.3 存在长枝吸引问题的系统发生16.3.1 长枝吸引现象16.3.2 产生长枝吸引现象的可能原因16.3.3 识别长枝吸引的方法16.3.4 消除长枝吸引现象的方法16.4 大数据集的系统发生16.4.1 大数据集系统发生及其面临的问题16.4.2 大数据集系统发生分析策略16.4.3 大数据集的系统发生分析需要的计算机和软件16.4.4 大数据集分析实例16.5 碱基组成异质性数据集的分析16.5.1 序列组成偏向性及其对系统发生分析的影响16.5.2 碱基组成异质性数据分析方法16.5.4 氨基酸组成异质性数据分析方法16.6 种上与种下数据的联合分析第17章 多源数据集分析策略和方法17.1 多源数据集概述17.2 数据集之间的不相合性及检验方法17.2.1 不相合性的类型17.2.2 数据集之间不相合性的原因17.2.3 数据集之间不相合性的检验方法17.3 多源数据集的分析策略17.3.1 联合方法17.3.2 分类学相合性分析17.3.3 数据划分方法17.4 多源数据集的划分分析实例17.5 谱系基因组学方法17.5.1 谱系基因组学17.5.2 谱系基因组学分析策略17.5.3 谱系基因组学分析方法第18章 系统树的可视化、注释与应用18.1 系统树的可视化18.1.1 TreeView18.1.2 Dendroscope18.1.3 Mesquite18.1.4 FigTree18.1.5 MrEnt18.1.6 2D和3D曲面表示方法18.1.7 iTOL18.2 系统树的注释18.2.1 分类学命名标注18.2.2 分歧年代和地质时代的标注18.2.3 重建祖先状态18.2.4 性状进化18.2.5 协同系统发生18.3 系统树表达的信息及其应用18.3.1 拓扑结构和分支长度18.3.2 系统树的树形及应用18.3.3 系统发生的不平衡性18.3.4 系统树用于分析分歧速度18.4 系统发生的应用参考文献

## &lt;&lt;分子系统发生学&gt;&gt;

## 章节摘录

第1章 系统发生学概论1.1 系统发生与系统发生学系统发生 (phylogeny, 由希腊词根phylon = stem、tribe、race 和genesis = origin构成) 是指任何生物实体 (基因、个体、种群、物种和种上阶元) 的起源和演化关系。

达尔文首次使用系统发生一词是在《物种起源》第5版提及Haeckel的著作Generelle Morphologie时, 并将系统发生等同为所有生物的传代线 (the lines of descent of all organic beings), 这与Haeckel的原意不同, Haeckel书中的系统发生是生命之树的传代线上生物在形态上发生的主要改变, 而非传代线本身 (Dayrat, 2003)。

但达尔文此处对系统发生概念的使用与我们现在的定义基本上一致。

分子系统发生 (molecular phylogeny) 是利用各种分子性状构建的生物实体之间起源和演化关系, 采用的分子数据主要是DNA和蛋白质序列, 也包括其他类型的分子数据。

系统发生学 (phylogenetics) 是研究利用各种性状构建基因、个体、种群、物种和种上单元之间系统树或网络的原理和方法的学科。

系统发生学重建进化历史依赖于对取样物种的性状分布进行数学推论, 这种重建涉及不同类群共享的同源性状, 并通过这些性状推断系统树。

这种数学推断的准确性完全依赖于对性状进化的假设和模型。

20世纪50年代以来, 蛋白质和DNA测序技术为系统发生重建带来了曙光。

DNA和蛋白质序列数据作为生物信息分子具有线性数字编码特征, 并且能够建立位点之间的同源关系, 逐渐成为系统发生分析的主要数据来源。

分子系统发生学 (molecular phylogenetics) 就是在这种背景下诞生的, 是研究利用各种分子性状构建基因、个体、种群和物种之间系统树或进化网络的原理和方法的学科。

分子数据的使用导致了系统发生研究的革命。

在20世纪80年代后期, 由于保守引物的PCR扩增和DNA测序技术的应用, 使得系统发生分析可利用的同源位点 (即性状) 数量达到500个, 有的甚至超过数千个, 与此前几十个、最多上百个形态特征相比, 大大地增强了解决系统发生推论的数据力度。

此时, 一些线粒体基因和rDNA成为最广泛应用的标记, 其中编码SSU rRNA的基因识别出了作为生命树的第三分支的古细菌 (Archaea)。

随着更多基因标记, 尤其是大量单拷贝核基因的使用, 基于单个基因推论的系统发生关系之间的冲突逐渐显露。

而且, 来自单个基因的信息经常不足以对系统发生的节点提供坚实的统计学支持。

所以, 自20世纪90年代以来, 多基因数据逐渐成为分子系统发生研究的主流。

目前, 成千上万个物种的全基因组序列信息已经通过新一代的高通量测序技术产生, 并由此产生了一个新的分支学科——谱系基因组学 (phylogenomics), 就是在基因组水平上进行系统发生研究。

谱系基因组学将基因座位的进化作为一种随机过程看待, 将分子水平的基因座位和序列位点进化模型及群体历史过程整合在一起, 分析基因树和物种树之间的关系, 引发分子系统发生学思想的又一次革命。

基因组学数据增加了用于系统发生学分析的性状数量和类型, 期望能够减少先前由于序列或基因取样偏差造成的系统发生推论误差。

分子系统发生学已经成为当前生物学的核心领域。

根据SCI Web of Science引文数据库统计, 到2009年底已经有30 000多篇关于系统发生分析的论文, 并且每年以3000篇的速度增加 (Pagel and Meade, 2008)。

Rokas和Carroll (2006) 估计世界范围平均每天发表15棵系统树。

最近发起的重建生命之树计划和DNA条形码计划是生物学历史上能够与基因组计划媲美的生物学大科学项目, 加之廉价而快速的新一代高通量测序技术引发的全基因组测序的普及, 如人类千人基因组计划、宏基因组学 (metagenomics)、脊椎动物基因组10K计划和昆虫基因组5K计划等, 将极大地推进分子系统发生学的研究。

## &lt;&lt;分子系统发生学&gt;&gt;

分子系统发生学数据的增加速度很快，目前NCBI 核苷酸数据库有序列记录的物种数超过30 万种。过去5 年GenBank 的物种数以每年约1.7 万种的速度增加，也就是170 万种已描述物种中，每年约有1% 的物种被进行至少一个基因的测序。

即便如此，至少含有一条分子序列的生物体只占全部已知物种的17% 左右。

而在系统发生信息数据库TreeBASE 中，目前只录入了2000 多项研究的5000 多棵系统树，包括100 000 个类群 (<http://www.treebase.org/>)。

因此，实现重建生命之树的宏伟计划还有漫长的路要走。

1.2 系统发生关系的含义不同生物学家对系统发生概念的认识和理解有所不同。

生物之间在系统发生学上的相关性称为系统发生关系 (phylogenetic relationship)。

生物之间存在着各种各样的相互关系，系统发生关系只是其中最重要的关系之一，其他的还包括表征的 (phenetic)、分支的 (cladistic)、时序的 (chronistic)、遗传的或亲缘的 (patristic) 和相互作用 (interaction) 关系等，这些复杂的关系从不同的角度反映了生物之间的相关性。

1.2.1 表征关系表征关系是不考虑进化关系，仅以所有可利用性状为基础的全面相似性程度排列的关系。

Sneath 和Sokal (1972) 将表征关系定义为“在所研究的机体表型特征基础上的相似性”。

以表征关系为基础的分类学研究称为表征分类学 (phenetics)，根据生物表征总体相似性为依据获得的有机体之间的关系图解称为表征图 (phenogram)。

表征分类学认为有机体之间的演化关系是无法弄清楚的，因而表征图不需要代表机体之间的演化关系。

根据表征图显示的类群之间的聚类关系就可以直接转化为分类体系。

1.2.2 分支关系分支关系指物种或类群之间与共同祖先相对近度 (relative recency) 的关系。

以分支关系为基础的系统学研究称为支序系统学 (cladistics) 或系统发生系统学 (phylogenetic systematics) (Hennig, 1966)。

支序系统学派认为，判别系统发生关系远近的唯一标准是共同祖先的近度 (recency of common ancestry)，共同祖先关系可以通过性状的分布分析来发现，支序系统学派将性状分为祖征 (plesiomorphy)、共享祖征 (sym-plesiomorphy)、衍征 (apomorphy)、共享衍征 (synapomorphy) 和自裔衍征 (auto-pomorphy)，认为只有共享衍征才是共同祖先的证据，共享祖征及由趋同进化和平行进化形成的相似性 (同型性状) 均不能作为共同祖先的证据。

通过共享衍征推论的有机体分支关系的树状图称为支序图 (cladogram)。

支序图的纵轴仅表示分支发生的相对时间，图上的二叉分支节点代表一次物种形成事件。

支序图只是关于共享衍征分布的陈述 (图1-1)，而不是系统发生关系的陈述，要将支序图转化成系统发生关系还需要进一步对进化过程作出假设。

支序图上的分类单元 (无论是现存种还是化石种) 总是在末端分枝，而在系统树上必须明确分类单元的祖裔关系。

图1-2 中左框的分支图 ((A, B), C) 就可以解释为右框中6 种不同的系统树。

支序系统学认为系统发生关系是生物之间最核心的关系，所有的分类学体系都必须建立在系统发生关系的基础上，也就是要求所有的分类单元必须是单系性的。

已经建立了基于系统发生的分类学命名法规 Phylocode ([www.ohio.edu/phylocode/](http://www.ohio.edu/phylocode/))。

1.2.3 遗传关系遗传关系是生物在遗传组成方面的关系，在群体遗传学中采用遗传相关性系数

(coefficient of genetic relatedness) 来度量，在种上阶元之间采用亲缘距离 (patristic distance) 来度量。亲缘距离是指在传代线内发生的遗传变异数量，表现在标度系统树上两个物种经过其共同祖先节点的所有通径的分枝长度之和。

在分子系统树上，亲缘距离实际上等价于它们从共同祖先分歧以来在两个支系上发生的遗传改变，如果以基因组序列来度量的话，就等价于分支之间的遗传组成差异。

遗传关系起源于遗传物质的继承与传递，包括两种不同的遗传方式：垂直遗传和水平遗传。

垂直传递是通过繁殖方式进行的，在有性生殖群体内个体之间的遗传关系是一种网状关系 (特称为tokogeny)。

## &lt;&lt;分子系统发生学&gt;&gt;

垂直遗传包括双亲遗传（如常染色体遗传）、父系遗传（如Y染色体遗传）和母系遗传（如线粒体基因组遗传）三种不同的方式。

双亲遗传标记是生物之间的主要遗传标记，可以用于研究生物主要遗传组成的演化历史；父系遗传的标记可以推论父本谱系的历史；母系遗传标记可以推论母本谱系的历史。

三类垂直遗传标记都可用于类群系统发生关系的重建。

水平遗传的主要方式是基因水平转移（horizontal gene transfer, HGT），也称为侧向转移（lateral gene transfer, LGT），是指在不同物种之间进行的遗传物质的交流。

LGT类似于物种内部的重组，但种内不同染色体/DNA分子的重组是共享基因库分子之间的混合，虽然也产生了不同进化历史的DNA分子的嵌合体，但重组分子对推论物种之间的系统发生关系影响不大，因为它们的遗传传递方式与分支发生方式一致。

而LGT是跨越生殖隔离的DNA分子之间的混合，是与分支发生关系毫无关联的遗传传递，因而会对系统发生关系产生误导。

水平基因转移事件作为推动物种进化的重要动力，在生命起源和进化的早期发挥了十分重要的作用，后来也对原核生物基因组的进化产生了深刻的影响（图1-3），但相对来说在真核生物中发生的规模不大。

垂直遗传和水平遗传的概念在细胞形态的生物之间是很容易区分的，因为垂直遗传是通过细胞膜体系和遗传系统的双重复制及分裂过程完成的，而水平遗传仅仅是部分遗传物质的整合。

因此，也有人将这种以细胞传承为基础的垂直遗传系统发生称为细胞之树（tree of cell）。

以遗传关系为基础的系统学研究即分子系统学，从带遗传信息的分子数据建立的树状图称为分子树（molecular tree）或基因树（gene tree）。

基因树可以是群体内部取样的等位基因之间的系统发生关系，特称为基因谱系（gene genealogy），反映的是等位基因的起源和演化关系；也可以是基因组内部一个基因家族成员之间的系统发生关系，反映的是基因重复事件；还可以是不同物种的直系或并系同源基因之间的系统发生关系，反映的是基因重复与物种形成双重进化事件。

基于垂直遗传分子标记构建的基因树可以转化为物种树，而根据水平遗传分子标记构建的基因树就不能直接转化为物种树。

所以，只有垂直遗传关系的基因才能表达物种之间的系统发生关系。

1.2.4 系统发生关系广义的系统发生（phylogenetic）或种系发生（phyletic）关系包括任何生物实体（基因、个体、群体、物种和种上阶元）的起源和演化关系，而狭义的系统发生关系仅指物种和种上单元的起源和演化关系。

对狭义的系统发生关系的含义有不同的看法，有些人认为系统发生关系应是包括以上三种关系的总和，有些人则将系统发生关系仅看成是分支关系或遗传关系。

在此，我们将狭义的系统发生关系定义为由分支发生（cladogenesis）产生的存在于任何支系（lineage）之间的祖裔关系和姐妹群关系。

因此，系统发生关系实际上是由垂直遗传构成的分支关系，而分支发生实际上是连续的物种形成过程。

在这个连续的过程中，永恒存在的仅仅是作为复制模板的细胞膜系统和遗传系统，个体只是作为这两个复制模板的中间载体而短暂存在。

除这种分支关系外，系统发生关系还包括少量的由物种杂交形成产生的网络关系。

自达尔文以来，系统发生关系被认为是树状分支的，表示有机体之间系统发生关系的树状图解称为系统树（phylogenetic tree或phylogram）或进化树（evolutionary tree）。

传统上，系统发生关系通过寻找共同祖先及祖裔关系来重建。

由于现存的所有物种都是由共同祖先进化形成的，因此，系统发生关系的远近程度是一个相对概念。

如果两个谱系享有比其他谱系更近的共同祖先，则这两个谱系相互之间的系统发生关系更接近且远离其他谱系。

由于灭绝和分类单元取样不完整等原因，系统发生关系一般无法重建完整的生物进化历史，而是简



## <<分子系统发生学>>

化的进化历史，是一种对进化历史的假设，但随着研究的深入可无限逼近进化历史。以上四种关系是生物之间最核心的关系，它们之间在概念上的区别可以通过熟悉的爬行类和鸟类进化关系的树状图解来说明（图1-4）。

图1-4中显示出表征关系、分支关系和亲缘距离三种关系度量方法之间的矛盾。

蜥蜴和鳄鱼的形态特征最接近，表征关系最近；鳄鱼和恐龙之间在系统树上的通径长度最小，故亲缘距离最近；恐龙和鸟类是拥有最近共同祖先的类群，它们之间的分支关系最近。

· · · · · ·

<<分子系统发生学>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>