

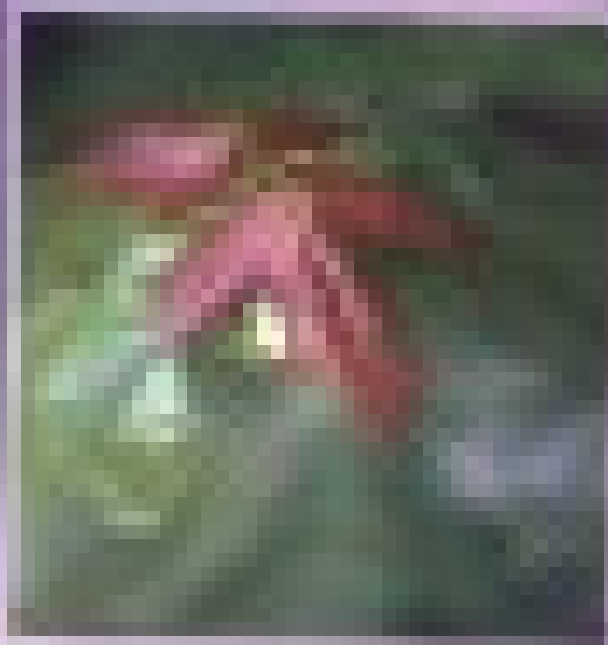
植物系统学进展

陈之端 冯 旻 编译



科学出版社

植物系统学进展



— — — — —

權

受

401514

科学出版社

1998

内 容 简 介

本书根据最近几年来植物系统学上出现的热点而精选出7篇论文编译而成。作者都是目前活跃在该学科领域前沿的植物学家。论文内容包括分子系统学、分支系统学、生殖结构化石三方面,反映了当前植物系统学研究的进展和趋势。其中有叶绿体 DNA 系统学的实验方法和数据处理综述;根据 *rbcL* 基因序列比较完成的迄今最大规模的分子系统学分析;从功能、进化、技术和系统学应用方面阐述 ITS 区可在较低分类阶元上成为系统学家有价值的证据来源;18S rRNA 和26S rRNA 数据在较高分类阶元上对于理解被子植物起源的价值;分支系统学的进展和前景;被子植物起源和早期分化的研究进展。

本书适合于大学生物系植物学、古植物学各专业及农、林院校相关专业师生、研究生和科研单位研究人员参考。

图书在版编目(CIP)数据

植物系统学进展/陈之端,冯旻编译.-北京:科学出版社,1998

ISBN 7-03-006292-2

I. 植… II. ①陈… ②冯… III. 植物分类学-进展 IV. Q949

中国版本图书馆 CIP 数据核字(97)第21949号

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

北京科地亚印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1998年6月第 1 版 开本:787×1092 1/16

1998年6月第一次印刷 印张:11 3/4

印数:1—1 800 字数:262 000

定价:24.00 元

译校者序

80年代初,现代最流行的几个被子植物分类系统相继修订(如 Dahlgren 1980, Takhtajan 1980, Cronquist 1981, Thorne 1983)。从那时起到现在的15年间,现代植物系统和进化生物学研究日新月异,取得了一些激动人心的进展。Hennig 的分支分析方法为近年来剧增的分子数据的系统发育处理提供了一条有效途径,被大多数系统学家所采用。DNA 序列资料对于理解被子植物起源、分化和现存各大类群之间的关系作用很大,为评价各个分类阶元上的系统发育关系提供了大量新的假设。分子系统学的蓬勃开展很大程度上得益于全球范围内的计算机网络技术,快捷和极大丰富的网上信息为各类资料的综合分析提供了方便。古植物学的新发现,尤其是白垩纪花化石的研究,揭示了被子植物主要类群间的关系,并提供了一个大概的有关它们在地层上分化和出现的时间尺度。系统发育方法应用于传粉生物学领域,发现了一些动态的协同演化系统。生物地理学由描述发展到分析,为生物区系和演化历史提供了更多的精确模型。植物系统和进化生物学与植物生理学、发育生物学和分子遗传学间的交叉渗透可从根本上研究功能与结构之间的关系。新兴的分子生物学技术是改良作物品种的有效手段,同时为作物的起源提供了新证据。全球生态学研究使人们感到空前的环境危机,尤其是不可再生的生物多样性迅速丧失带来的恐惧,于是保护生物学应运而生。以上是和植物系统学发展有关的一些方面,本书主要从分子系统学、分支系统学和生殖结构化石三个角度展示该领域的进展和趋势。

在第一篇,Olmstead 和 Palmer 从实验方法和数据处理方面讨论了如何利用叶绿体基因组来研究植物的系统发育,客观论述了限制性内切酶位点研究和序列研究的优缺点与适用范围,讨论比较了目前常用的几种数据处理方法及存在问题。第二篇就是 Mark Chase 等 42 位作者根据 *rbcL* 基因序列比较完成的迄今最大规模的分子系统学分析。这篇论文虽然引起很多争论,原因是由于数据矩阵太大,从大量的分析结果中选择的分支图可信度受到怀疑,但它仍有许多令人振奋的发现,说明分子序列变异包含了大量有关植物演化的信息,对于将来人们利用分子和分子以外的性状研究植物系统发育具有启发和指导作用。该文除了讨论 *rbcL* 数据的分析结果外,还对一系列的问题进行了评述,如平行核苷酸置换的影响,如何对总分支结构作内部支持分析,不同密码子位置的核苷酸置换和密码子使用率,分枝步长的不等性和不等率等等。接着我们选用 Soltis 夫妇在《密苏里植物园年报》第 82 卷中的导言并以此引出利用其他基因或 DNA 片段研究系统发育的两篇论文:Bruce Baldwin 等人以其工作为例,从功能上、进化上、技术上和系统学应用上详细阐述了 ITS 区可在较低分类阶元上成为系统学家有价值的证据来源;而 Doyle, Donoghue 和 Zimmer 则综述了 18S rRNA 和 26S rRNA 数据在较高分类阶元上对于理解被子植物起源的价值。这反映出当前的分子系统学研究不断从细胞器基因组和核基因组选择适用于在各分类阶元上进行系统发育研究的基因或 DNA 片段。此外,Doyle 等的文章从方法论的角度考虑价值也是不可低估的。该文提供了一种方法,即如何对分子数据和形态学数据单独或结合在一起作 PAUP 和 MacClade 分析,以及如何处理在分析过程中经常遇到

的一些问题,如性状编码和加权,特别是如何对复合分类单元的性状赋值。

当前植物系统学出现的另一个热点是分支系统学。分支系统学的发展与分子生物学资料的猛增有关,由于数据矩阵加大,两者结合在一起出现了许多新的概念、方法和争论,如分枝吸引和系统树岛屿等。Michael Donoghue 的论文对分支系统学的进展和前景作了评述。

最后,我们选用了 Peter Crane 等发表在 Nature 上的一篇综述,目的是对最近人们在被子植物起源与早期演化问题上所取得的进展作一总结。同时也反映出和本世纪中叶出现的热点(如数量分类学和细胞分类学)相比,人们对于分子系统学的到来保持了谨慎和冷静,并不认为 DNA 序列就能解决所有的系统学问题,因而不排斥其他性状资料或证据来源。在这些证据来源中,花的化石研究取得了令人鼓舞的进展。近年来,由于能够成功地得到大量保存完好的白垩纪的花,从而大大加深了我们对于早期被子植物系统发育和花演化的认识(见 Friis & Crepet 1987, Crane, Friis & Pedersen 1995)。利用 SEM 和 TEM 对现存原始被子植物花和花粉的微形态、演化式样和生殖生物学研究(Friis & Endress 1990, Endress 1987, 1989)加速了化石材料的鉴定和解释。在此基础上,利用分子遗传学手段对于花发育的直接分析预计将成为未来植物系统学研究的热点。因为通过花部器官发生的基因控制方面的证据可以帮助我们判别被子植物与其他种子植物,以及被子植物各大类群之间花的各个部分的同源性。

在过去的十几年里,虽然植物系统学面貌发生了根本性改变,但在国内还没有一本论著反映这一领域的进展。我们怀着满腔热情,希望本书能对我国从事植物系统学工作的同行有所帮助,同时也希望本书能对我们目前开展的国家自然科学基金委员会重点项目“原始被子植物的结构、分化与演化”起到参考和指导作用。恳请读者对书中译文欠妥之处提出宝贵意见。

最后,我们无比感谢张志耘、杨亲二、卢宝荣、张大明、邢树平、桑涛和文军博士,美国密苏里植物园主任 Peter Raven 博士,芝加哥 Field Museum 的 Peter Crane 博士,西雅图华盛顿大学的 Richard Olmstead 博士,以及加州大学戴维斯分校的 James Doyle 博士对本书翻译的关心和支持。感谢原文作者以及《密苏里植物园年报》、《自然》和《美国植物学报》同意我们无偿翻译这些论文并在科学出版社出版。黄长珍、孙海英承担本书的录入、打印工作,在此一并表示真诚的谢意。

参 考 文 献

- Chase M W et al. 1993. Phylogenetics of seed plants; An analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Missouri Bot. Gard. 80: 528—580
- Coen E S, Meyerowitz E M. 1991. The war of the whorls; Genetic interactions controlling flower development. Nature 353: 31—37
- Crane P R 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. Ann. Missouri Bot. Gard. 72: 716—793
- Crane P R, Friis E M, Pedersen K R. 1995. The origin and early diversification of angiosperms. Nature 374 (2): 27—33
- Cronquist A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York
- Dahlgren R T. 1980. A revised system of classification of angiosperms. Bot. J. Linn. Soc. 80: 91—124
- Doyle J A, Donoghue M J, Zimmer E A. 1994. Integration of morphological and ribosomal RNA data on the origin of

- angiosperms. *Ann. Missouri Bot. Gard.* 81: 419—450
- Endress P K. 1987. Floral phyllotaxis and floral evolution. *Bot. Jahrb. Syst.* 108: 417—438
- Endress P K. 1990. Evolution of reproductive structures and functions in primitive angiosperms (Magnoliidae). *Mem. New York Bot. Gard.* 55: 5—34
- Friis E M, Crepet W L. 1987. Time of appearance of floral features. In Friis, E. M., W. G. Chaloner & P. R. Crane (eds.): *The Origins of Angiosperms and Their Biological Consequences*. 145—179. Cambridge Univ. Press, Cambridge
- Friis E M, Endress P K. 1990. Origin and evolution of angiosperm flowers. *Adv. Bot. Res.* 17: 99—162
- Takhtajan A. 1980. Outline of the classification of flowering plants (Magnoliophyta). *Bot. Rev.* 46: 225—359
- Thorne R F. 1983. Proposed new realignments in the angiosperms. *Nordic J. Bot.* 3: 85—117

目 录

译校者序

I. 叶绿体 DNA 系统学: 方法与数据分析综述	
..... R. G. Olmstead & J. D. Palmer(1)	
II 种子植物的系统发育——来自质体基因 <i>rbcL</i> 的核苷酸序列分析.....	
..... M. W. Chase 等(23)	
III 用于植物系统发育重建的各种不同的基因(导言)	
..... P. S. Soltis & D. E. Soltis(75)	
IV. 核核糖体 DNA ITS 区: 被子植物系统发育研究中有价值的证据来源.....	
..... B. G. Baldwin 等(76)	
V. 形态学与核糖体 RNA 数据应用在被子植物起源问题的综论.....	
..... J. A. Doyle, M. J. Donoghue & E. A. Zimmer(102)	
VI. 重建植物系统发育研究的进展和前景	
..... M. J. Donoghue(135)	
VII. 被子植物的起源和早期分化	
..... P. R. Crane, E. M. Friis & K. R. Pedersen(146)	
参考文献.....	(158)

I. 叶绿体 DNA 系统学： 方法与数据分析综述*

摘要 随着植物分子系统学研究的迅猛发展,全新的与改进的方法日渐投入使用。本文综述了有关叶绿体基因应用中实验方法与数据分析的近期进展。叶绿体基因组限制性位点图谱分析已经被广泛应用,但它所适合的分层阶元有限。其适用上限(即最大趋异)是比较基因组中核苷酸置换率低的反向重复区域。而限制性位点研究中经得起检验的应用下限是使用具 4-碱基识别位点的限制性内切酶去分析快速进化的、由聚合酶链式反应(PCR)扩增得到的叶绿体基因组中的部分片段。DNA 序列比较是分子系统学所取得的最大进展。PCR 的出现及 PCR 产物直接测序方法使序列数据如雨后春笋般涌现。理论上,任何程度的变异都可对其进行序列比较研究。而在实践中,植物系统学家们则青睐于进化慢的两个序列(*rbcL* 和 *rRNA* 基因)。有时也包括一些进化较快的叶绿体基因、叶绿体内含子和基因间隔区以及核糖体 RNA 重复区的非编码部分。本文综述了有关限制性位点图谱比较与 DNA 测序各自的优点和缺点。对于这两种方法而言,数据分析必须在足够的分类单元和性状取样的基础上,以期对系统发育关系作尽可能全面的估计。由于一个系统树上会出现长分枝问题,致使简约性分析对分类单元取样非常敏感。然而,当分类单元足够多时,又使数值矩阵面临严重的计算困难,有可能造成无法获得最大简约的系统树或是找不到所有的最短系统树。

在过去十年中,分子生物学技术用于被子植物系统发育研究已呈现出繁盛景象。选用于系统发育研究的分子变异的两个主要来源是叶绿体基因组(Palmer 1987, Palmer et al. 1988, Olmstead et al. 1990, Clegg & Zurawski 1992, Downie & Palmer 1992b)与核糖体 DNA 重复区(Knaak et al. 1990, Baldwin 1992, Hamby & Zimmer 1992)。动物系统学中,线粒体基因组扮演着中心角色(Moritz, Dowling & Brown 1987, Avise 1991),而相比之下,植物中的线粒体基因组很少用来进行系统发育研究(Palmer 1992a)。

分子系统学研究所用的实验技术与其他所有分子生物学家使用的大致相同(例如 Sambrook, Fritsch & Maniatis 1989),但在不同植物种的大规模比较(Palmer 1986, Palmer et al. 1988, Hillis & Moritz 1990)和生化特性(biochemical idiosyncracies)研究中,这些方法需要经常修改。完整叶绿体基因组的限制性位点图谱(例如 Palmer et al. 1988, Jansen et al. 1991, Olmstead & Palmer 1992)和 *rbcL* 基因测序(Doebley et al. 1990, Soltis, Soltis & Bothel 1990, Giannasi et al. 1992, Olmstead et al. 1992)两种方法已为被子植物系统发育重建提供了大部分叶绿体 DNA (cpDNA)数据。另外,叶绿体基因组内独立的结构重排(structural rearrangements)[例如倒位(inversions)和内含子丢失(intron loss)]也可作为确定单系类群的标记,但是这类突变的发

* Richard G. Olmstead & Jeffrey D. Palmer 1994: Chloroplast DNA systematics: A review of methods and data analysis. Amer. J. Bot. 81(9): 1205-1224.

生频率太低,无法提供足够的数据来进行大多数类群的系统发育重建(Jansen & Palmer 1987a, Bruneau, Doyle & Palmer 1990, Lavin, Doyle & Palmer 1990, Olmstead et al. 1990, Downie et al. 1991, Raubeson & Jansen 1992b)。

本文的目的是综述和阐明 cpDNA 限制性位点和序列数据在植物系统发育研究上的方法和应用,并在任何可能的情况下对二者进行比较。本文并不打算对使用这两种方法已取得的成就进行总结,有关这方面的内容读者可参考其他综述(Palmer et al. 1988, Soltis, Soltis & Doyle 1992, Clegg 1993, Doyle 1993, Sytsma & Hahn 1994, Soltis & Soltis 待发表)。

细胞器基因组

在被子植物中,叶绿体基因组的大小、结构和基因组成是相当一致的。典型的被子植物叶绿体基因组长 135—160kb,具有一个大约 25kb 的反向重复序列,该序列将基因组的其余部分分隔为一大一小两个单拷贝区(Palmer 1985, Sugiura 1989)。曾报道过的较小的叶绿体基因组实际上是由一个反向重复区缺失造成的[例如豆科的一个类群(Palmer & Thompson 1982);所有的松柏类(Raubeson & Jansen 1992a)],而在不行光合作用的寄生植物 *Epifagus*(列当科)中,小基因组的存在是因为其叶绿体基因组中有大量缺失,其中包括许多基因丢失(dePamphilis & Palmer 1990)。在天竺葵属 *Pelargonium*(牻牛儿苗科)植物中曾发现过较大的叶绿体基因组(217kb),其大小的增加主要归因于它具有一个异常扩展的反向重复区,但这并不能代表基因组复杂程度的增加(Palmer, Nugent & Herbon 1987)。

4 种陆生植物的叶绿体 DNA (cpDNA)全序列确定(烟草 *Nicotiana tabacum*——Shinozaki et al. 1986, 地钱 *Marchantia polymorpha*——Ohya et al. 1986, 水稻 *Oryza sativa*——Hiratsuka et al. 1989, 寄生植物 *Epifagus virginiana*——Wolfe, Morden & Palmer 1992)为叶绿体基因组的结构和基因组成提供了丰富的可比数据。烟草代表了被子植物的典型基因组成,共有大约 113 个不同基因,其中 4 个为 rRNA 基因,30 个 tRNA 基因和 79 个推测的蛋白质编码基因(Shinozaki et al. 1986, Wolfe 1991)。叶绿体基因组的基因次序和组成很少变化,而变化经常是由基因组的一部分发生倒位或基因丢失造成。例如,非维管陆生植物地钱属 *Marchantia* 与有花植物烟草属 *Nicotiana* 之间的不同在于一个 30kb 的倒位和 5 个蛋白质基因及一个 tRNA 基因在一个或另外一个基因组中的缺失(Ohya et al. 1986, Wolfe & Sharp 1988)。同样,单子叶植物稻属 *Oryza* 和双子叶植物烟草属之间的差别包括 3 个倒位、3 个蛋白质基因丢失、一些与反向重复区末端移动有关的基因重复和基因重排。所有这些均代表了水稻谱系(lineage)的衍生变化。

植物线粒体基因组研究与叶绿体基因组形成强烈反差。对于植物线粒体基因组目前还了解得很少;从苔藓植物地钱中得到的唯一已完整测序的基因组(Oda et al. 1992)含有 90 个不同基因,包括 3 个 rRNA 基因、27 个 tRNA 基因和大约 60 个蛋白质基因。被子植物线粒体 DNAs (mtDNAs)的部分测序结果显示出它们具有相似的一套基因(A. Brennicke 个人通讯)。被子植物线粒体基因组在大小、结构和基因次序上变异很大(Newton 1988, Palmer 1990, 1992b),难以进行整个基因组限制性位点研究。然而,植

物 mtDNA 的沉默置换(silent substitution)率远小于植物的其他基因组或动物基因组(Sederof 1987, Wolfe, Li & Sharp 1987),因而表明植物线粒体基因组的比较测序可能在解释植物系统发育的深层分枝(deep branches)时会起作用。但是,植物线粒体基因的另一个复杂性在于它们具有 RNA 剪辑造成的不同程度的(在被子植物中高,在地钱中低或没有)转录后序列变化(Bonnard et al. 1992)。

与了解较为贫乏的植物线粒体基因组相比,动物线粒体基因组在所有生物类群中是研究得最多的。其大小(大多动物的为 15—19kb)、高度保守的结构和基因组成(Brown 1985)以及高的同义置换(synonymous substitution)率(Brown, George & Wilson 1979, Wolfe, Li & Sharp 1987, Avise 1991)使其适用于较低分类阶元上的居群遗传学和系统学研究(Avise et al. 1987, Moritz, Dowling & Brown 1987)。但是,其上的结构重排和若干进化慢的基因(例如, rRNA 基因和细胞色素氧化酶基因)也可应用于较高分类阶元上的比较研究。现已从 15 种后生动物中得到了可以进行比较的 mtDNA 全序列(Wolstenholme 1992)。大多数动物 mtDNA 含 37 个基因,分别是 2 个 rRNA 基因、22 个 tRNA 基因和 13 个蛋白质基因。

植物叶绿体基因组和动物线粒体基因组在其各自类群的系统发育研究中是自然的对应物。动物 mtDNA 比植物 cpDNA 更快速的沉默置换率,为有兴趣将分子技术用于居群遗传学研究的动物学家们提供了便利(Avise et al. 1987)。尽管如此,叶绿体基因组为某些物种提供了有用的种内差异,虽然对所有已被调查过的物种并非均是如此(Soltis, Soltis & Milligan 1992)。比较而言,叶绿体基因组的三个特点为在种级及以上阶元的系统发育研究提供了显著的优点。第一,约大出 10 倍的叶绿体基因组和约多出 6 倍的蛋白质基因为限制性位点研究提供了一个大得多的数据基础和序列比较的更多选择。第二, cpDNA 的沉默置换率比动物 mtDNA 的低 10 倍还多(Wolfe, Li & Sharp 1987),使得将 cpDNA 核苷酸序列直接比较用于更高阶元的系统发育研究比动物 mtDNA 更为可行。第三,尽管结构重排在 cpDNA 与动物 mtDNA 中均不常见,但相比之下在 cpDNA 中较普遍些,特别是在被子植物中发现有许多倒位和基因或内含子丢失(Jansen & Palmer 1987a, Downie & Palmer 1992b)。而在与被子植物进化时间相仿的哺乳类动物的 mtDNA 中仅发现一个重排,该重排把有袋类与真兽次亚纲的哺乳动物区分开来(Paabo et al. 1991)。

核基因组和叶绿体基因组及动物线粒体基因组相比均相差甚远。由于它太大和太复杂,且包括许多重复的和非编码的序列,使得对其不可能进行限制性位点图谱分析。限制性片段长度多态性分析(RFLPs)表明随机核探针应用于系统发育研究中有许多缺点(见下)。大部分核基因中存在种源(orthologous)(来源于物种形成)和基源(paralogous)(来源于基因重复)拷贝,使得核基因的应用复杂化。尽管对核基因组进行比较研究非常困难,但核核糖体 RNA 重复区(rDNA)却是一个可以进行比较研究的片段。rDNA 由含 rRNA 基因和非编码间隔区 DNA 的串联重复片段组成,并且表现出足够快的一致进化(concerted evolution),以至于在大多数情况下,可把它看成长度相当的 cpDNA 片段来分析。(Zimmer et al. 1980, Arnheim 1983, Hamby & Zimmer 1992)。

叶绿体 DNA 的限制性位点图谱

应用与局限 将 cpDNA 用于系统发育研究,基本方法是分析它的限制性片段变异。这一方法已从在一块经染色的琼脂糖凝胶上观察到的整个基因组的限制性片段式样的简单比较(Palmer & Zamir 1982, Clegg, Brown & Whitfeld 1984, Hosaka et al. 1984),发展到用 DNA 印迹杂交(Southern blot hybridization)和用覆盖了一个已有图谱的基因组的大部分或全部的克隆的探针来更清晰地比较限制性位点图谱(Sytsma & Gottlieb 1986, Coates & Cullis 1987, Jansen & Palmer 1988, Palmer et al. 1988)。

限制性位点图谱的某些优点使得将其应用于系统发育研究成为可能。1) 在分子生物学的诸多方法中,这一技术相对简单(即可使用“菜单”操作),并使研究者可以同时比较大量数据。2) 叶绿体基因组足够大,从而使每一种酶均可产生许多位点。对每个基因组可进行 100 次酶切的酶,容易获得图谱。3) 限制性内切酶所切位点代表一个对 cpDNA 近于随机的取样,因而接近于在系统发育分析中所有方法都需要的、重要但常被忽视的假定:性状的独立性。然而我们应牢记在心的是,整个叶绿体基因组是作为单个连锁群(single linkage group)被遗传的。因此,如果杂交和渗入(introgression)已经使 cpDNA 从一个谱系转移到另一个谱系,其变异状态可能就无法代表种的差异(Doyle 1992)。4) 限制性位点上许多有系统发育信息的变异存在于非编码区(Olmstead & Palmer 数据未发表),而非功能性状(nonfunctional characters)用于解释系统发育长期以来备受推崇(例如 Darwin 1859)。5) 利用限制性位点图谱研究同属的种或近缘属的问题,所得的数据矩阵显示出的同塑性(homoplasy)水平相当低(Palmer et al. 1988 及其中的文献;Schilling & Jansen 1989, Wallace & Jansen 1990, Lavin & Doyle 1991, Rieseberg et al. 1991)。6) 基因组不同区域的置换率相差很大,使该方法可用于虽有限但有相当大范围的分类阶元上。最近的研究表明,cpDNA 反向重复区内含有大多数被子植物 cpDNA 总基因组的约 20% 的复杂性,可以得到可靠的图谱来对被子植物的整个目,甚至于一个亚纲进行分析(Downie & Palmer 1992a, Manos, Nixon & Doyle 1993)。这要归功于在反向重复区内所观察到的明显较低的同义置换率(Wolfe, Li & Sharp 1987)。这一方法,结合小杂交探针(200—3000bp)和高频内切酶的应用,在相当程度上扩展了比较图谱的使用上限。而在另一个极端,若干研究(Arnold, Buchner & Robinson 1991, Liston 1992, Liston, Rieseberg & Hanson 1992, Rieseberg, Hanson & Philbrick 1992)用 PCR 扩增一个长 2—4kb、长度保守但快速进化(序列上)的 cpDNA 片段检测了近缘种的关系。该扩增产物用不同的限制性内切酶消化,再直接从凝胶上比较得到的片段。此方法很适合用于那些 DNA 量受限或许多重排使图谱不易得到的类群。7) 与 PCR 测序研究相比,由于其样品不易被污染,限制性位点分析得到的假象较少(见下面讨论)。

某些因素也制约着限制性位点图谱在系统发育研究中的应用。1) cpDNA 进化的保守性决定了系统发育分析中的实际下限。许多研究中近缘种对所使用的内切酶来说得到的图谱是相同的(Schilling & Jansen 1989, Wallace & Jansen 1990, Olmstead et al. 1990 及其中的文献,Rieseberg et al. 1991)。2) 在较大分子距离上也存在上限,此时限制性位点的同源性不再能被可靠确认。对整个基因组进行限制性位点图谱的比较工作已在

某些大的科内开展(茜草科 Rubiaceae——Bremer & Jansen 1991, 菊科 Asteraceae——Jansen et al. 1991, 茄科 Solanaceae——Olmstead & Palmer 1992, 柳叶菜科 Onagraceae——K. Sytsma 个人通讯),而且对反向重复区进行分析可以使这一技术扩展到更高分类阶元(如下述)。然而,对于那些古老科或 cpDNA 快速进化的科而言,整个基因组的分子差异可能会太大,以至于其图谱无法比较。菊亚纲中目间及目内的图谱比较已被普遍证明是不可靠的,至少对基因组的单拷贝区是如此(R. Jansen 个人通讯, Olmstead & Palmer 数据未发表)。3) 含有高度重排的 cpDNA 的植物类群,如在豆科 Leguminosae, 牻牛儿苗科 Geraniaceae 和 Orobanchaceae 中观察到的(可能还包括其他寄生的, 不行光合作用的植物类群),整个基因组的图谱研究并不能有效地反映类群关系,尽管重排本身有时也被用作系统发育分析(例如 Knox, Downie & Palmer 1993)。如前所述,这一问题可通过对 PCR 扩增出的基因组内非重排部分进行限制性位点分析而得以避免。同样地,那些缺失一个反向重复拷贝的植物类群也无法使用仅仅基于反向重复区的图谱比较研究,因为在这些分类群中,基因组内该区域的进化速率可能出奇地快。4) 进行整个基因组的图谱研究需要相当数量的 DNA($\sim 10-100\mu\text{g}$),以满足使用较多限制性内切酶对基因组 DNA 进行消化。提取足够 DNA 所需的组织会因为物种稀少或难于获得而受限,但采用 PCR 扩增 cpDNA 来进行比较限制性位点或测序研究可包括那些只能得很少的或保存状况很差的组织的物种,甚至是已绝灭的种(Golenberg et al. 1990)。5) 评价和处理限制性位点数据,在很大程度上仍是手工进行的和繁琐的。目前尚无将限制性位点图谱作为原始数据存贮入计算机的可行方法,以便能使计算机查阅这些图谱并确认同源位点表现出来的有义变异,来进行系统发育分析。在已有的限制性位点数据矩阵中加入新类群需要重新对所有的图谱进行比较,如果没有原始的放射自显影记录,为了重新将原始数据矩阵中的无义位点变异包括在分析中需从所有数据矩阵上比较。与 DNA 测序研究相比,这是限制性位点图谱的一个很大缺点,测序得到的新序列可以很容易地加入到已有的数据矩阵,并且立刻加在一起进行系统发育分析。向已有的图谱研究中加入新数据的困难性,使一些学者只用以前确认的有信息的限制性位点变异分析新类群(Doyle, Doyle & Brown 1990, Rieseberg et al. 1991)。但是这样做就排除了一种可能性,即确认原先在无义限制性位点差异基础上所不能发现的关系。

目前的研究正在扩展限制性位点图谱在被子植物系统学研究中的应用的上限和下限。增加叶绿体基因组内限制性位点的取样密度,或者结合其他不同来源的限制位点数据,包括核 rDNA 重复片段,可克服许多种级水平上的 cpDNA 研究的不完全分辨率。使用其他常用的 6-碱基识别酶(6-内切酶)或 5-碱基、4-碱基识别酶(4-内切酶),通常可在更多位点上切开 DNA,以达到提高限制性位点取样密度的目的。Sytsma 和 Gottlieb(1986)用 29 个 6-碱基内切酶代替 cpDNA 限制性位点研究中常用的 10—20 个酶,得到 *Clarkia* 属 *Peripetasma* 组一个分辨彻底的树系图。大多数 6-碱基内切酶可识别叶绿体基因组内 10—100 个特异位点(只计算反向重复的一个拷贝)。近期在蝶须属 *Antenneria*(菊科)的一个近缘种复合体的研究中(H. Michaels 个人通讯),使用 4-碱基内切酶可识别典型叶绿体基因组上多达 700 个特异位点,尽管其中有许多位点无法分析(因为得到的片段太小,无法检测)以及大多数位点的特异谱带顺序无法确定。在此项研究中,琼脂糖凝胶的浓度上升至 4%,以便分离得到的许多小 cpDNA 片段。通过使用较高频率切割 cpDNA 的酶来

提高取样密度只能在近缘种间才适用。一般而言,小的杂交探针比之于大杂交探针,能允许产生更为精细的图谱,因而使用高频切割的内切酶对趋异分类单元分析,就能得到更为精确的比较图谱(Olmstead & Palmer 1992)。如此小的、邻近的探针只能从烟草叶绿体基因组中获得,但由于烟草基因组与许多其他被子植物 cpDNA 的共线性(colinearity),以及 cpDNA 间核苷酸序列的高度保守性,使得烟草的精细图谱的探针也可用于许多其他被子植物类群。

现今,大多 cpDNA 限制性位点研究使用细胞总 DNA,而不是经提纯的 cpDNA。这也同时允许使用结合有同套 DNA 的杂交滤膜核探针。核 rDNA 重复序列提供了另一个 7—14kb 的序列,其在总 DNA 提取物中的浓度和 cpDNA 大致相当,图谱分析也和 cpDNA 非常类似。例如,Rieseberg, Soltis 和 Palmer(1988)使用 36 个酶分析 cpDNA 与核 rDNA 的限制性位点变异,解决了向日葵属 *Helianthus* 内 3 个近缘种的种间关系及种内居群间关系问题。除了使分析中限制性位点数据得到增加这一优点外,核 rDNA 是一双亲遗传基因组,其进化与叶绿体基因组无关,因而可对用母系遗传的 cpDNA 得到的关系进行检测。这些不同已被成功地用来解释杂交事件(Doyle, Soltis & Soltis 1985, Arnold, Bennett & Zimmer 1990),并提示是否发生了细胞质渗入(Doebley 1989, Rieseberg et al. 1991, Rieseberg & Brunsfeld 1992)。除非在理想条件下(足够量的 DNA 以及使用同源探针,即来自同一个种的探针),单拷贝的核 DNA 序列很难通过杂交进行检测。

cpDNA 限制性位点图谱仍旧是植物系统发育研究中的一个重要手段,而某些动物 mtDNA 研究者们已给予限制性位点分析一个“回顾式的致意”(Wilson et al. 1989),尽管这似乎有些为时过早(Avise 1991, 1994)。一项研究搜集了 cpDNA 限制性位点与序列数据,并比较了二者的应用价值。

在茄科已测定了 18 个种的叶绿体基因 *rbcL* 和 *ndhF* 的序列(Olmstead & Sweere 待发表),其限制性位点数据是从以往的研究中得到的(Olmstead & Palmer 1992)。限制性位点图谱、*rbcL* 和 *ndhF* 基因序列分别提供了 209、63 和 100 个有信息的性状,其一致性指数 CI 值分别为 0.43, 0.50 和 0.59 (Olmstead & Sweere 待发表)。这些结果表明在所分析的 cpDNA 变异水平上,限制性位点分析比 *rbcL* 或 *ndhF* 序列分析所得到的数据都多,但同塑性也较高。然而,在有关菊科的类似研究中,*rbcL* 基因的同塑性高于限制性位点(Kim et al. 1992)。尽管 DNA 测序技术发展迅速,但研究同样数量的 DNA 序列,限制性位点图谱仍不失为一种更为便利的方法。在我们对茄科的研究中(Olmstead & Palmer 1992, 数据未发表),用 10 个酶对 125 种植物(每种间接地取样 7000bp)进行了限制性位点图谱分析。这一项目的实验工作耗时一年,仅由一人完成。若要用测序来获取相同数目的核苷酸数据的话,将需要对相当于 600 个以上的 *rbcL* 基因进行测序!将限制性位点原始数据转换为用于系统发育分析的性状数据,要比转换序列数据费事得多,但叶绿体基因组内普遍存在的性状的随机分布可对限制性位点数据提供一种补偿。

方法 以下论述应被看作是对 Palmer (1986), Palmer 等(1988) 和 Dowling, Moritz 和 Palmer(1990)所讨论的方法的一个补充。早期 cpDNA 研究需要相当大量的,甚至多至 100g 的新鲜叶片材料,程序是首先用蔗糖梯度法分离得到完整的叶绿体,接着从解离的叶绿体中温和地提纯 cpDNA(Palmer 1986)。大多数现行研究采用 Saghai-Maroof (1984)CTAB 法或是其改进的方法(Doyle & Doyle 1987)提取总 DNA,从而对大多

数物种来说,使提取足够多的 cpDNA 所需的组织量降低至 1—3g。这一方法在大多数植物中均适用,但用于某些植物也有过出现问题的报道,这是由于其他物质与 DNA 一起在异丙醇中沉淀出来,从而影响 DNA 的重新悬浮,或沉淀的 DNA 结合进一粘胶层,在离心过程中不能形成絮片(Martin & Dowd 1991, R. Wallace 个人通讯)。如果有上述问题出现,首先建议提高 CTAB 提取缓冲液与叶片材料的比例,同时重复氯仿/异戊醇抽提或增加用酚来进行的其他有机抽提(Chase & Hills 1991, R. Wallace 个人通讯)。CTAB 提取程序的其他改进是在特殊类群的研究者解决特殊问题的过程中得到发展的(Wagner et al. 1987, Palmer et al. 1988, Chase & Palmer 1989, Wendel 1989, Bult, Kallersjo & Suh 1992)。以 PCR 方法为基础的限制性酶切分析(例如 Rieseberg, Hanson & Philbrick, 1992)所采用的 DNA 提取程序中,叶片材料的需要量仅为毫克级(例如 Edwards, Johnstone & Thompson 1991)。

直到最近,细胞总 DNA 仍需要从新鲜组织或经液氮速冻并冰冻保存的组织中提取,这就使得 cpDNA 研究限制在那些容易得到的植物类群上。然而,对用不同方法保存的组织检测表明:快速干燥的组织,如果贮存在干燥环境中,能得到高产率的相对不降解的 cpDNA (Doyle & Dickson 1987, Pyle & Adams 1989, Liston et al. 1990, Chase & Hills 1991)。目前选用硅胶作干燥剂的方法,极大简化了植物组织的野外采集工作。使用标本馆馆藏材料进行以 PCR 为基础的 DNA 研究也越来越普遍(Loockerman & Jansen 待发表),因而在许多研究中不必进行野外采集。

一个理想的研究项目应选取足够的限制性位点来得到足够多的系统发育信息,但同时保证所需的时间和花费也趋于最小。这通常意味着选取最少量必需的限制性内切酶,从而在从限制性内切酶消化经凝胶电泳、滤膜制备、杂交、放射自显影到图谱建成的每一步骤中花费最少的时间和财力。在选择用于比较 cpDNA 图谱的限制性内切酶时应该考虑到:1)所比较 cpDNA 的变异程度;2)酶切割 DNA 的预期次数;3)与预期得到的限制性片段大小和变异相关的杂交探针的大小;4)识别序列的独立性;5)费用。用小杂交探针能得到更为精细的图谱,因此可使用较少的能进行频繁切割的酶。对近缘分类单元的研究要求选取更多的限制性位点,以期获得的分辨率能与研究远缘分类单元的工作进行比较(见上文)。有许多例子说明:一个内切酶的所识别序列可能是另一个较不专一酶的亚集(subset)(如, XhoI, CTCGAG, 和 AvaI, CPyCGPuG),因而单个碱基的置换能相关地影响到两个酶的切割。在选择图谱研究所用的酶之前,应参考识别序列的交叉指数(cross-index)(可以从大多限制性内切酶生产厂家得到)。任何酶在 cpDNA 上限制性位点的相对数目可由它的核苷酸组成粗略地估算出来,例如富含 A/T 的识别位点在 A/T 含量丰富的叶绿体基因组中更经常地出现。为得到更精确的不同内切酶所期望识别位点数目的指数,位点的确切数目可用计算机搜索任何一个全测序的 cpDNAs 进行确定。许多常用酶的清单和它们的位点在烟草 cpDNAs 中的出现频率(Olmstead & Palmer 1992),包括说明及计算机绘制的 11 个酶的图谱,可从 R. G. Olmstead 处得到。

在不同分类单元之间及不同组织之间,cpDNA 与核 DNA 在总 DNA 提取物中的比例是不同的。因而,光学估计(例如用 Et-Br 染色或紫外分光光度法)的 DNA 浓度并非 cpDNA 浓度的可靠指标。一个较为可靠的测量 cpDNA 在总 DNA 制备液中浓度的方法是用一种酶对每一个 DNA 样品进行酶切,电泳并印迹转移,用一个 cpDNA 探针作测试

杂交,然后通过调整将用于酶解的 DNA 量来得到适宜的 cpDNA 杂交信号的强度。这可能需要在用其他酶酶切并跑胶前完成两次或更多次检测凝胶电泳,但应避免 X 射线曝光的时间长短不一,目的是为了更容易对结果进行解释。

在很多限制性位点图谱研究中对所有的酶使用同一浓度的琼脂糖凝胶,而不考虑所期望片段的数目及大小。而对每一个酶预期得到片段数的估计可用来确定胶的浓度,以便最好地分辨所期望大小的片段——高频切割酶使用高浓度,低频切割酶使用低浓度 (Sytsma & Gottlieb 1986, Olmstead & Palmer 1992)。当每块胶中的指示剂跑至同样距离时,改变凝胶浓度,就可以更好地分辨不同大小的片段。用此方法,小至 175bp 的片段在 2.0% 的琼脂糖凝胶上迁移至 12.5cm 处就可以分辨出来。双向印迹转移方法能产生两套同样的滤膜而不必另外跑胶。一个高效率使用放射自显影设备的常用策略是跑一块 20cm 宽的胶,每块上有 35 个孔道(33 个样品与 2 个大小标准),让溴酚蓝标记跑到距起点 10cm 处。接着切开这块胶,印迹转移从起点直到 12.5cm 处,以保留比标记迁移得更远的小片段(Palmer et al. 1988)。两个这样的滤膜可完全覆盖一个标准的 8"×10" X 射线胶片。当期望无变异时,在胶上迁移的距离会更短,于是可将三个滤膜拍在一个 X 射线胶片上(E. Knox 个人通讯)。把杂交探针来源种的 DNA 包括在内(如烟草)将提供另外一个对照和大小标记;如果 DNA 探针来自被研究类群的一个近缘种,这一方法就更为有效,因为二者的许多位点均是被期望为保守的。

在 cpDNA 限制性位点图谱中使用克隆的异源探针,很大程度上使这一方法在植物系统学中得到广泛应用(Palmer 1986, Palmer et al. 1988)。一个最有用的克隆库来自烟草(描述见 Olmstead & Palmer 1992),其叶绿体基因组已全部被测序,并具有维管植物典型的基因次序。此库几乎完全覆盖了整个基因组,并被设计为用一组 40 个一致的小探针(平均为 3.2kb)来研究高度趋异的基因组,并且/或者使用高频切割酶。在某些要求更低、可以更快地覆盖基因组的研究中,相邻的探针可以进行归并,最终只用 5—20 个而非 40 个探针,就能把整个基因组覆盖。其他两个非常有用的克隆库一个来自兰科植物 *Oncidium excavatum* (Chase & Palmer 1989),它是从非禾本科单子叶植物中得到的最好探针组;另一个来自于 *Lactuca sativa* (Jansen & Pamer 1987b),它包含大多数菊科植物具有的 22kb 的倒位。禾草植物美国狼尾草 *Pennisetum americanum* (Thomas et al. 1984)的克隆库已在广泛使用;然而,随着水稻叶绿体基因组全序列的确定(Hiratsuka et al. 1989),使用克隆的水稻 cpDNA 作为杂交探针将会有更多的优点。

Dowling, Moritz 和 Palmer(1990)讨论了叶绿体 DNA 限制性位点图谱研究的一些逻辑与实践。简而言之,构建图谱的过程可分为三步:首先,为用于其中一个要研究的分类单元(参考基因组)的所有酶建立完整 cpDNA 图谱。理想情况下,这一基因组可用来设计杂交探针;否则,用所有酶进行 DNA 双酶解以有利于参考基因组图谱的构建。第二,将 X 射线胶片按酶和探针次序排列,并画出图谱(或者对于保守基因组只标记有变异的位点和片段)。每一分类群或近缘分类群组对每一特定酶在参考图谱的上方(或下方)各自的线上作图。这一步应辨别出大多限制性位点突变以及靠近片段末端的突变是长度突变还是点突变。最后,根据探针片段重新排列 X 射线胶片并根据参考基因组的综合图谱(一个显示所有限制性内切酶排列的图谱)分析它们。这将最终对含糊的长度/位点突变作出明确判断。

另一个限制性位点分析方法不需大量的 DNA 或杂交探针,所用的是 cpDNA 的 PCR 扩增片段(Arnold, Buckner & Robinson 1991)。在相对高置换率区域两翼的保守编码区的引物(*rbcL/orf106*-Rieseberg, Hanson & Philbrick 1992, *rpoC1/rpoC2*-Liston 1992)可用来扩增 2—4kb 的片段,然后用一组限制性内切酶消化(包括 4bp 内切酶)。酶切后的 PCR 产物在高浓度的琼脂糖凝胶(1.5%—2.0%)上分离,分辨出小的片段,并在溴化乙锭染色的胶上直接观察比较,或用 Southern 转移和杂交的方法比较。用两个或更多酶进行双酶解,并和所有其他结果结合在一起能得到扩增片段中限制性位点的线性图谱。

数据分析 许多早期尝试用限制性位点图谱分析进行的系统发育重建,都是根据限制性片段式样差异对遗传距离进行估计(Nei & Li 1979, Nei & Tajima 1981)。那些不用限制性位点(restriction sites),而用限制性片段(restriction fragments)作为距离或简约性分析性状的方法,忽略了长度突变(即插入和缺失)和限制性位点置换之间的差异。不考虑长度突变而分析限制性片段,会导致受同一长度突变影响的不同酶所产生的性状间出现相关性(例如,具同样长度变异的每一分类单元对每一个酶都将会不同),这等于无意中加权了这些性状。当限制性位点位于长度变异之中时,这一问题也会影响片段式样(例如,由于缺失造成不同酶的识别位点丢失)。插入与缺失应通过若干酶尽可能准确地用图谱表现出来,那些在长度变异区域内所没有的位点应记录为丢失数据。通过片段分析(若在遇到相对小的变异时)(Palmer & Zamir 1982),或者用限制性位点图谱(Sytsma & Gottlieb 1986, Coates & Cullis 1987, Dowling, Moritz & Palmer 1990)对限制性位点差异的详细记录(存在或缺失)导致了目前对简约性分析的偏爱(Palmer et al. 1988)。准确的图谱为限制性位点的同源性估计提供了基础(Olmstead et al. 1990)。当长度突变显著时,若记录仅限于片段或限制性位点,说的好一点是没有利用所有的数据,而说的糟一点是产生完全错误的结论(Bremer 1991)。

目前的研究通常完全根据限制性位点的存在与否作为系统发育性状。然而,研究限制性位点图谱也能识别出常常以插入和缺失形式出现的结构重排(Palmer et al. 1988)。这些结构变异很少被应用于系统发育分析中(例如 Sytsma & Gottlieb 1986, Soltis, Soltis & Bothel 1990, Olmstead & Palmer 1992, 但在 Doebley, Renfron & Blanton 1987 中可见到一个例外)。它们未被包括的原因通常认为是由于它们的同源性很难确定。这可以用豌豆属 *Pisum* 的研究为例进行说明(Palmer, Jorgensen & Thompson 1985),该工作发现有 15 个长度变异集中在基因组的两个区域上,并且需要解释在四种情况下的趋同(在同一个研究中 11 个限制性位点变异被确认没有任何同塑性而需要去解释它们的分布)。在其他分类群中也已注意到“热点”(hotspots)的存在,在这些区域长度变异体非常集中(Tassopulu & Kung 1984, Chase & Palmer 1989),而且长度变异几乎是连续的。然而,限制性位点变异的同源性也无法确切得知,许多限制性位点分析的一致性指数小于 0.5,这意味着一个限制性位点的存在与否并不始终是同源性的正确指标。有些研究(例如 Sytsma & Gottlieb 1986, Soltis, Soltis & Bothel 1990, Olmstead & Palmer 1992)发现长度变异体与由限制性位点数据所得的系统树是一致的。在最近对茄科的研究中(Olmstead & Palmer 1992),从不包括自近裔性状在内的限制性位点数据中得到的最短系统树的一致性指数(CI)为 0.36,但五个有信息的长度变异体(即那些为两个或更多分