

The Applications of Chloroplast Genome Analysis in Plant System Development

Hui Zhou¹, Shengli Jin², Gang Li¹, Lei Zhang¹, Rui Qin¹, Hong Liu^{1*}

¹Engineering Research Center of Protection and Utilization for Biological Resources in Minority Regions, South-Central University for Nationalities, Wuhan

²State Key Laboratory of Hybrid Rice, College of Life Science, Wuhan University, Wuhan
Email: liuhong@mail.scuec.edu.cn

Received: Oct. 8th, 2013; revised: Nov. 12th, 2013; accepted: Nov. 25th, 2013

Copyright © 2014 Hui Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for Hans and the owner of the intellectual property Hui Zhou et al. All Copyright © 2014 are guarded by law and by Hans as a guardian.

Abstract: Chloroplast is a specific organelle structure of green plants and algae microorganisms, which is widely used in comparative genomics and plant system development. In this paper, according to the characteristics of chloroplast genome, we analyzed the applications of chloroplast genes including encoding genes: *rbcL*, *makt*, *rps4* and *Ndhhf*; non-coding region interval sequences: *trnF-L*, *rps4-trns* and *psbA-trnH*; non-coding introns: *rpl16* and *rps16*. We also analyzed the principle, method, application, advantages and disadvantages of chloroplast DNA technologies in plant system development.

Keywords: Chloroplast Genome; Plant System Development; DNA Analysis Technology

叶绿体基因组分析在植物系统发育中的应用

周 会¹, 荆胜利², 李 刚¹, 张 磊¹, 覃 瑞¹, 刘 虹^{1*}

¹中南民族大学, 南方少数民族地区生物资源保护及综合利用工程中心, 武汉

²武汉大学生命科学院, 杂交水稻国家重点实验室, 武汉

Email: liuhong@mail.scuec.edu.cn

收稿日期: 2013 年 10 月 8 日; 修回日期: 2013 年 11 月 12 日; 录用日期: 2013 年 11 月 25 日

摘 要: 叶绿体是所有绿色植物和藻类微生物特有的细胞器结构, 其基因组信息被广泛运用于比较基因组学和植物系统发育研究。本文对叶绿体基因组的结构特征, 包括编码区基因如 *rbcL*、*makt*、*rps4* 和 *Ndhhf* 等叶绿体基因, 非编码区中的间隔序列 *trnF-L*、*rps4-trns* 和 *psbA-trnH* 等以及内含子 *rpl16*、*rps16* 等在内的叶绿体基因组应用于植物系统发育方面研究概况与进展进行了概述, 探讨了叶绿体各种 DNA 分析技术的原理、方法、应用及其优缺点, 并展望了叶绿体基因组分析在植物系统发育中的应用前景。

关键词: 叶绿体基因组; 植物系统发育; DNA 分析技术

1. 引言

当前, 在对植物系统发育研究中, 主要从传粉方

*通讯作者。

式、繁育、化石以及分子水平上对包括叶绿体, 线粒体与核基因组的分析比较来进行。由于叶绿体基因组(chloroplastDNA, cpDNA)是仅次于核基因组的第二大

基因组且叶绿体的核酸置换率适中,故植物叶绿体基因组在种及种以上较高阶元的系统发育研究有着显著的优势,加上叶绿体基因组较小,均为闭环的双链DNA,总量约占植物总DNA量的10%~20%,长度在120~210 kb之间^[1],在分子水平上差异明显,多为单拷贝序列,在序列与结构上都是相当保守的,分子性状的同源关系更容易确定。此外cpDNA的编码区和非编码区的分子进化速度差异显著,具有独立的进化路线,可适用于不同层次的系统学研究,不依赖其他数据就可以用于构建系统进化树,从而保证了类群间的可比性。

2. 在植物系统发育方面常用的叶绿体DNA分析技术

2.1. cpDNA的限制性内切酶酶切图谱分析法

大多数叶绿体(高等植物叶绿体)基因组长度变异主要由2个反向重复序列(inverted repeats, IR)引起的。这2个反向重复序列长约22~25 kb,并将整个cpDNA分为一个长单拷贝区(large single copy, LSC)和一个短单拷贝区(small single copy, SSC)^[2],由于不同物种的cpDNA识别特定限制性内切酶的位点位置不同,所以得到的片段长度不同,通过比较,可用来研究物种间的亲缘关系。植物cpDNA的所有酶切位点位置的简图的研究,通过将植物的叶绿体中提取和纯化的cpDNA分别用不同的限制性内切酶消化,并用凝胶电泳分离消解DNA片段,然后比较不同个体、群体或种间的差异,再依据这些数据构建系统进化树。对于那些不易提纯cpDNA的植物的研究,采取从叶片中提取总DNA的方法,同样地用限制性内切酶消化并凝胶电泳,将分离的DNA片段转移到滤膜上,与含有叶绿体基因的片段(即探针)杂交,最后比较杂交带谱的差异进而构建系统进化树。因为叶绿体基因组相对较小,易于提纯,异质性弱,并且在进化上比较保守,所以cpDNA限制性酶切图谱分析法在不同的植物演化水平上都得到了广泛的应用,为探讨植物间的亲缘关系提供了极珍贵的信息。

Kato等^[3]用7种限制性内切酶构建了红小豆(*Vigna umbellata*)的限制性内切酶图谱,发现它的杂草型类群在大单拷贝区由于重复序列分子内重组导致了一个96bp的缺失,这在遗传学上研究植物系统进

化有着重要意义。彭隽敏等^[4]对云南小麦、西藏半野生小麦及普通小麦的酶切图谱作了分析,结果表明它们之间虽属于不同的亚种,形态差异明显,但是它们的cpDNA的进化仍然处于同一水平,这为研究它们之间的亲缘关系提供了有力的依据,同时也说明了cpDNA在进化上比较保守,不易发生变异。另外,此方法还可用于证实天然准性杂种和研究多倍体的起源。Kim等^[5]对*Krigia*属做的酶切图谱表明二倍体*K. biflora*和四倍体*K. montata*可能是六倍体*K. montata*的亲本种,并通过ITS序列证明了此观点,从而弄清了该物种的起源途径。

2.2. DNA杂交技术

不同物种DNA之间退火的不同程度可用来衡量DNA的同源性,从而推导物种间的亲缘关系。DNA杂交可分为滤膜杂交和溶液杂交两种。由于液相杂交比较简便,快捷,在植物系统学研究中应用较多。在溶液中DNA分子的热稳定性或熔解温度(T_m)与两条子链的互补性成正相关关系,且杂交DNA分子的T_m值低于同源DNA分子的,通过测定同源DNA分子和杂交DNA分子的T_m值,就可推测和比较不同类群DNA间相同序列区域所占的比例,进而判断它们之间亲缘关系的远近^[1]。因而cpDNA杂交技术可应用于推测植物科、属、种的系统发育及种的分化形成时期等方面的研究。将待测物种cpDNA提取出来,经过适当的酶切后,通过琼脂糖凝胶电泳,将不同大小的片段分开;然后将凝胶上的DNA片段转移到硝酸纤维素膜上;再用放射性同位素标记的另一物种cpDNA片段作为探针与硝酸纤维素膜上的DNA进行杂交,然后再测定它们之间的T_m值,比较数据。

通过来自一种叶绿体基因组的克隆片段与另一种结合在滤膜上的DNA进行的异源滤膜杂交,发现碧冬茄属和芸苔属的叶绿体基因组内没有重排,而绿豆属的基因组中有一个约50 kb的倒位^[6]。Palmer等利用DNA杂交技术发现,在豆科蝶形花亚科的6个族和12个属中都有一个50 kb的大倒位,禾本科的3个属都有一个20 kb的倒位,菊科的6个属都有一个30 kb的倒位^[7]。但由于该技术实验步骤繁琐,获得的杂交资料难以定性,很难排除错配、插入和丢失等因素的影响,且对DNA样品量的要求大,又因为存在同位素的安全性问题,所以应用受到限制。

2.3. 核酸顺序分析

核酸分子包含着生物遗传信息的同时也记载着生物进化的历史,即核酸结构不仅指导生物的形态建成和个体发育,而且还能够反映生物之间的亲缘关系。这种技术是生物遗传物质最全面,最彻底的比较,可用于任何等级水平(包括科级,属级,种级等)亲缘关系的研究。通过生成互相独立的若干组带放射性标记的寡核苷酸(每组寡核苷酸都有固定的起点,但却随机终止于特定的一种或者多种残基上),由于 DNA 上的每一个碱基出现在可变终止端的机会均等,因此上述每一组产物都是一些寡核苷酸混合物,这些寡核苷酸的长度由某一种特定碱基在原 DNA 全片段上的位置决定。在可区分长度仅差一个核苷酸的不同 DNA 分子条件下,对各组寡核苷酸进行电泳分析,只要把几组寡核苷酸加样于测序凝胶中若干个相邻的泳道上,即可从凝胶的放射自显影片上直接读出 DNA 上的核苷酸顺序,因此 DNA 序列分析是检测遗传物质变异最直接的方法。

俸宇星等^[8]测定了悬铃木科(Platanaceae)和金缕梅科(Hamamelidaceae)5个亚科6个代表种的 *rbcL* 基因序列,对低等金缕梅植物及新近的相关类群进行了分子系统发育分析,获得了四个最简约树,简约树的步长为 893,其 CI 值和 RI 值分别为 0.558 和 0.591,为金缕梅科植物的系统发育研究提供了依据。张文衡^[9]对广义川续断目(Dipsacales)的 21 种植物 cpDNA 的 *trnL-F* 区进行了测序,并建立了系统发育树状图,结果显示,败酱科、川续断科、双参属、刺参属和广义忍冬科的 4 个属(双盾木属 *Dipelta*、虫胃实属 *Kolkwitzia*、六道木属 *Abelia* 和北极花属 *Linnaea*)形成了一个单系群;双参属和川续断科之间有着更近的关系,并建议将其作为一个亚科置于川续断科;广义忍冬科为一多系类群。Zhang 等^[10]利用新一代测序技术对六种木本竹类的叶绿体全基因组进行了测序,并结合已测序的 18 个禾本科叶绿体全基因组进行了系统发育分析,以此评估叶绿体系统发育基因组学在竹亚科植物系统发育重建中的有效性。研究结果在叶绿体基因组水平上证实了禾本科 BEP 分支中竹亚科和早熟禾亚科为姊妹关系。与传统的分子系统学研究结果相比,Zeng 等^[11]利用叶绿体全基因组序列所构建的系统发育树在竹亚科内部具有更高的分辨率和支持率。

由此可见核酸顺序分析对揭示植物系统进化的分子规律有着极其重要的作用。但是这是一项相当艰巨的工作且 DNA 测序工作的难易程度及成本与片段长度有密切关系,在短时间内还很难广泛地应用到植物系统学研究上,但是随着新一代测序技术的发展,测序技术的产业化,成本降低,相信伴随着测序技术的发展,叶绿体系统发育基因组学在低阶元系统分类中的应用价值将得到更多的体现。

2.4. RFLP 分析

RFLP 分析即限制性内切酶酶切片段长度多态性分析指用限制性内切酶处理不同生物个体的 DNA 所产生的大分子片段的大小差异,是目前植物系统发育研究中使用较多的一种 DNA 分析技术。植物在长期进化过程中,种属间甚至品种间同源 DNA 序列上的限制性内切酶识别位点不同或者由于点突变、重组等原因,引起限制性内切酶识别位点上核苷酸的替换、插入或缺失变化从而引起 DNA 线性分子某一特定内切酶的识别位点发生变化,通过处理可得到限制性内切酶酶切片段长度多态性。此技术需将待研究的植物 cpDNA 经适当的限制性内切酶切割成不同的限制性片段,再通过凝胶电泳,由于不同大小的 DNA 片段在凝胶上的泳动速率不同,在凝胶上便形成了连续涂片,然后经 Southern 杂交、放射自显影得到 DNA 的 RFLP。由于 RFLP 来源于基因组 DNA 的变异、不受显隐性关系、环境条件和发育阶段的影响,具有稳定遗传和特异性的特点,可以获得能够反映种属遗传差异的大量多态性,故可用来研究植物类群特别是属间、种间甚至品种间的亲缘关系、系统发育和演化^[11]。

Remiaw^[12]利用 RFLP 分析技术证明:如果利用总的 cpDNA 作探针,与消化的总基因组进行 southern 杂交,可对 cpDNA 的保守区和可变区同时进行分析,从而能获得丰富的多态性,可为各种水平上的系统重建提供信息。Wattier 等^[13]用标记的总 cpDNA 作探针,对 *Ceramium* 属的两种红藻做 RFLP 分析,在种内发现了 cpDNA 存在多态性,并在种的水平上对两种红藻进行了系统发育研究。近年来,已有多项研究运用 cpDNA-RFLP 分析来探索微进化过程,包括被子植物的遗传结构、杂种地带、种系地理学等方面。

2.5. PCR-RFLP 分析

根据保守序列设计出通用引物再通过 PCR 反应扩增出特定基因,然后酶切纯化扩增片段。由于扩增的长度不大,经溴化已锭染色即可观察到限制性片段长度多态性。根据需要可选择少数基因或个别基因,用尽可能多的酶进行酶切,以获得单一基因的多限制性内切酶酶切片段长度多态性信息,根据需要有时可通过切点的拼接读出完整的基因序列。也可选择较多的基因,用少量的酶分析可获得多基因的限制性内切酶片段长度多态性信息。从代表性角度看,多基因分析方法优于单一基因序列分析的方法。

Wang Xiao-ru 等^[14]对松科 7 属、38 种植物的 4 个 cpDNA 基因进行 PCR-RFLP 分析,提示了很多重要类群的系统位置并构建了系统树。Ciprian 等运用 PCR-RFLP 方法对猕猴桃属 20 个种群的 5 个 cpDNA 基因进行了研究,发现至少有 3 个物种起源于种间的自然杂交^[15]。由于 PCR-RFLP 分析不需提纯 cpDNA,且不需要使用放射性元素或其它标记物,不受同位素半衰期的限制,从而减少实验室的污染;需要的植物材料很少,提取小于 50 mg 的植物材料中的总 DNA 即可满足 PCR 的需要量,20~25 个循环,从理论上就可以将靶序列扩增约 104 倍,即可检测出在总 DNA 中仅存在 1/1013 量的单拷贝序列;其次对植物材料要求不高,新鲜的叶片、快速干燥的叶片或保存较好的化石均可以用于总 DNA 的提取;再者扩增产物的特异性高,使得分析结果更为可靠。故此技术是目前叶绿体 DNA 系统学分析中使用最广泛的,也是遗传研究、物种鉴别、种间杂种品系鉴定、物种混杂分析等的有力工具。

2.6. 微卫星序列分析

真核生物基因组中,微卫星(短序列)是一类由 1~6 个碱基为重复单位串联而成的短片段,长度一般小于 100 bp,整个基因组都有分布,其上各位点不同等位基因的重复单位的数目高度变异,而且重复单位的序列可能也不完全相同,因而微卫星序列有高度的多态性。微卫星序列分析一般采用聚合酶链反应扩增相应的 cpDNA 的微卫星 DNA,扩增产物经纯化、测序反应、上机测序等步骤进行序列分析,得到微卫星序列多态性。微卫星序列分析在区别亲缘关系很近的基因

型上优势明显:1) 不需要用限制性酶切分析长度变化可直接可以在胶上显示出来。2) 在同一凝胶泳道上可对几个位点同时分析,而得出叶绿体单元型数据。3) 叶绿体单元型和微卫星等位基因可通过荧光标记的多重 PCR 同时分析及对单亲遗传标记和双亲遗传标记同时进行分析,在对种群间基因流动、渐渗杂交、多倍体起源等方面起着重要作用。但是微卫星序列在叶绿体细胞器基因组中比较罕见,进一步广泛应用受到引物设计的限制,需以叶绿体微卫星序列两侧的共有序列作为引物队,从而在不同的植物之间扩增出叶绿体微卫星序列,当然还有其他的方法。再者由于单核苷酸重复序列引起的长度多态性不一定适合于探索物种之间的系统发育关系,特别是当物种本身在该区域内存在高突变率或为平行演化时。

kurt 等^[16]设计了 10 对共有序列的引物队,其中 8 对引物在双子叶被子植物中普遍适用,并且揭示了叶绿体微卫星序列在烟草(*Nicotiana* spp.)、番茄(*Lycopersicon* spp.)、猕猴桃(*Actinidia* spp.)等属植物种间、种内的多态性,这表明以 cpDNA 微卫星序列两侧的共有序列作为通用引物是可行的。Provan 等^[17]在单子叶植物水稻 cpDNA 完全测序的基础上,设计了 6 对引物,对 6 种野生型水稻类群和 12 种栽培型水稻类群的微卫星序列进行了分析,在种内水平上同样检测到显著的长度多态性。

2.7. DNA 单链构像多态性(SSCP)分析

由于 DNA 链的碱基顺序不同,单链 DNA 在自然状态下会折叠弯曲形成一定的空间结构,不同物种的 cpDNA 的单链形成不同的空间结构,显示出多态性。在 SSCP 分析中,设计特异引物后用 PCR 技术定点扩增基因组上的某一目的序列,再将扩增片段变性处理成为单链 DNA,最后进行非变性的聚丙烯酰胺凝胶电泳。如果被扩增的目的片段任意一条 DNA 链中的碱基序列发生变化,则可能会由于这种变化而影响其构像和电泳迁移率使其带纹出现在电泳图谱不同位置,从而显示不同生物个体的 DNA 多态性。虽然 PCR-SSCP 分析是当前真核基因突变分析的一种十分灵敏的方法,但是该技术比较复杂,对操作人员的技术要求较高,故目前在叶绿体 DNA 系统学研究中还没有得到很广泛应用。

对 *Bostrychia moritziana*、*B. tenuissima*、*Caloglossa leprieurii* 这 3 种热带红藻叶绿体 DNA 中编码 Rubisco 大、小亚基基因之间的间隔区序列进行 PCR-SSLP 分析, 揭示了该序列在它们种内和种群内的多态性, 并且从地理学上构建了 *C. leprieurii* 的单元型多样性^[18]。Isoda 等对松科的 *A. veitchii* 和 *A. homolepis* 的叶绿体基因进行了 PCR-SSCP 分析, 并配合其它的分子标记技术, 得出 *A. veitchii* 和 *A. homolepis* 能够产生自然杂种的结论, 并提出应该重新考虑它们间的系统关系^[19]。

3. 叶绿体基因组系统发育分析常用序列

3.1. 叶绿体编码区基因分析

3.1.1. *rbcL* 基因

高等植物的 *rbcL* 基因在结构上和原核生物基因相似, 由 5'非编码区、编码区和 3'非编码区三部分组成。5'非编码区具有可以和叶绿体 16SrRNA 3'端附近互补的 SD 序列; 3'非编码区具反向重复序列, 能形成典型的茎环结构作为转录终止信号^[20]; 编码 1,5-二磷酸核酮羧化酶/氧化酶的大亚基, 以单拷贝的形式存在, 不发生基因转变, 长达 1.4kb, 能提供较多的分子性状, 且由于核酮糖-1,5-双磷酸羧化酶/加氧酶(Rubisco)在光合作用中所起的关键作用, *rbcL* 基因的进化速率慢^[21], 比较适合远缘间及科级以上分类群的系统关系的研究。Chase 等对 499 种种子植物的 *rbcL* 基因序列进行了分析, 从而对整个种子植物进行系统发育重建, 他们构建的分支图将来利用分子或非分子性状重建系统发育提供了一个有用的框架^[22]。

3.1.2. *matK* 基因

matK 基因编码酪氨酸蛋白激酶, 是位于 *trnK* 基因(叶绿体 tRNA 编码基因)的内含子内的单一拷贝序列, 长约 1500 个 bp, 参与 RNA 转录中 II 型内含子的剪切。到目前为止, 该基因被认为是所有编码蛋白基因中进化速率最快的基因, 且其序列变异较为均一, 在构建系统树时不必对不同类型的变异进行加权, 极大的增强了分子系统树的可靠性, 在被子植物科、属水平的近缘关系研究中已被使用^[23]。此外, *matK* 基因的 3'端较保守, 而 5'端相对多变, 因此, 3'端可用于科间水平分析, 而 5'端可用于科内、属间, 甚至种间的系统发育研究, 整个基因均可以用来做多水平的系统发育分析。吴世安等^[24]对黄精族(Polygonateae)6

属 13 种的 *matK* 基因进行了 PCR-RFLP 分析, 结果表明: *trnK* 基因的 PCR 产物在各类群间几乎不存在长度变异, 均约 2600 bp, 而 *rpl16* 基因则在各属之间及黄精属内表现出长度变异, 变异范围在 1140~1320 bp 之间; 限制性酶切位点同源性分析显示, 黄精属、鹿药属、竹根七属和舞鹤草属组成的狭义黄精族与铃兰族中的铃兰属有较近的亲缘关系, 并支持了将扭柄花属和万寿竹属从广义百合科黄精族中分出来的观点; 在狭义黄精族之内, 鹿药属与舞鹤草属组成一支, 黄精属与竹根七属组成另一支, 为探讨黄精族内属间的系统演化关系提供了分子生物学上的依据。何兴金等^[25]对国产葱属(*Allium*)9 个组中的 18 个代表种的 *matK* 扩增产物进行了酶切分析, 共获得 303 个酶切变异位点, 其中 163 个为信息位点。通过 PAUP (Version 3.1.1) 软件对所得的数据进一步分析表明: 葱属可分为 6 个亚属级的分类阶元, 对宽叶组的划分是合理的; 根茎组、单生组和洋葱组均是不自然的分类群, 得出应对该属下分类系统做相应的调整的结论^[1]。

3.1.3. *rps4* 基因

rps4 基因, 长约 600 bp, 合成叶绿体中核糖核蛋白小亚基(30 s), 与邻近的 *rps4-trnS* 基因间隔区片段进化都较快, 比较适合探讨属下等级的亲缘关系。Nadot 等^[26]选择了 39 种禾本科植物, 28 种其他的单子叶植物、11 种双子叶植物等, 分别用这些植物的 *rps4* 基因进行了序列分析, 并构建了系统树, 结果发现 *rps4* 基因树和 *rbcL* 基因树是一致的, 并证明了 *rps4* 基因在禾本科甚至在单子叶植物目级系统学研究中都是一个很好的工具。

3.1.4. *atpB* 基因

atpB 基因是光合作用中的重要基因, 编码 ATP 合酶的 β 亚基, 进化速率与 *rbcL* 非常相似, 其长度约为 1.5 kb, 对其测序方便且能提供足够的系统发育信息, 适合于较高分类阶元的系统关系研究^[27]。崔杰等^[28]在甜菜中扩增到一段全长为 2293 bp 含有 *atpB* 完整基因的序列, 其中包括 1497 bp 的编码区序列, 推测其编码 498 个氨基酸, 并建立了氨基酸序列的系统进化树^[29]。

3.1.5. *infA* 基因

infA 基因是一个极其活跃的基因, 编码一个约 70

个氨基酸的蛋白质翻译起始因子 IF1 (细胞器中蛋白质翻译起始的重要组分),参与核糖体蛋白 50S 亚基的合成的 *rpl36* 基因,其变异率比 *infA* 基因低。由于其基因收到的选择压力小,变异快,适合研究低阶元的系统发育过程。刘畅等^[30]利用小麦叶绿体基因组中 *infA-rpl36* 区域的序列设计引物,对小麦族的 12 个二倍体和多倍体的物种进行了 *infA* 和 *rpl36* 序列测定、系统树的建立和分析。

3.1.6. *ndhhF* 基因

ndhhF 基因编码叶绿体中电子传递系统组分^[31],通过异源杂交实验发现,在地钱,烟草和水稻上,cpDNA 分子分别有 6 个和 7 个基因与人的线粒体上的 DANH 脱氢酶亚基 ND1, 2, 3, 4, 4l, 5 和 6 的基因同源。通过 Northern 杂交证明 *ndhhF* 基因是活跃转录的,可能编码了一些有功能的 NADH-泛醌氧化还原酶,推测这些酶是属于叶绿体呼吸电子传递链的。由于该基因与光合作用有关,受到的选择压力大,进化速率较慢,因此在属以上的系统学研究中有一定的应用价值^[1]。

此外:

trnK^[31], 转录翻译系统基因,编码一种 tRNA,适用于种以下的低阶元的系统发育研究^[32]。

psbA 光合作用的基因,编码光系统 II 蛋白复合体的中心重要组成部分,广泛使用于较高水平如属以上水平的分子系统发育关系的比较^[33]。

psbB, 光合作用的基因,编码光系统 II 蛋白复合体的组分。

ndhF, 编码电子传递系统组分。

rps12 和 *rps7*, 编码核糖核蛋白体小亚基组分,通过烟草叶绿体相关序列设计引物,从杨树的叶绿体基因组中克隆出 2 个相邻的 DNA 片段,分析发现,这 2 个 DNA 片段含有核糖体蛋白 3'*rps12*、*rps7* 基因和 NADH 脱氢酶第二亚基 *ndhB* 基因片段。用 DNAMAN 等软件,将扩增的杨树叶绿体 DNA 片段与烟草、拟南芥、玉米和黑松的相关序列进行比较,表明所扩增的片段具有较高的保守性,尤其是在这 3 个基因的编码区,同源性都在 90%以上,插入或缺失常发生在基因间隔区,同源性约 80%;对其编码区所推导的氨基酸序列进行了比较,同源性均在 92%以上^[34]。

ndhB^[34], 编码电子传递系统组分。当然其它一些

基因在系统学中也有较好的应用。

3.2. 叶绿体非编码区基因

3.2.1. 间隔区序列

间隔区序列由于不受功能的限制,受到的选择压力小,变异速度快,其进化速率约为编码区的 10 倍左右^[35],因此对其扩增片段采用多重酶切等,则可产生更丰富的多态性,从而确定种间、种内类群的关系。

如 *trnL-F* 间隔区,在被子植物科下水水平的系统学研究中,*trnL-F* 间隔区的使用频率较高,它包含 *trnL* 内含子和 *trnL-trnF* 基因间隔区序列,尽管只有 800~1000 bp 的长度,由于不受功能的限制,受外界选择压力小,进化速率相对较快,可较好地用于属间或属下等亲缘关系和系统位置的研究^[36]。王玉金等^[37]测定了菊科风菊属 5 个亚属 37 种植物和川木香属的 1 种样品的 *trnL-F* 序列,长度在 773~838 bp 之间,排序后长 859 bp,当空位作缺失处理时,变异位点仅为 82 个,其中信息位点 21 个(2.44%),建了一个初步代表整个属下形态的分类系统。侯鑫等^[38]对小叶、中间和柠条锦鸡儿的 *trnL-F* 片段进行直接测序,长度为 1009~1021 bp,对位排列后长度为 1053 bp,包括 87 个变异位点,其中 54 个为信息位点,中间锦鸡儿与小叶的序列完全一致,而与柠条有明显差异,据此进行了它们的种间关系分析。

rps4-trnS 基因间隔区,该片段进化较快,比较适合于研究属下等级的亲缘关系。陆树刚等^[39]对篦齿蕨及其近缘类群的 *rps4-trnS* 区的序列进行 PCR 扩增和序列测定,对位排列后序列长度为 1278 bp,其中 525 个可变位点,312 个信息位点,探讨了篦齿蕨属的系统位置。

psbA-trnH 片段是位于叶绿体 DNA 基因组上 *psbA* 和 *trnH* 基因之间的一段长约 300 bp 的非编码序列,其进化速率大大快于 *matK* 基因,可用于植物属间及种间的系统发育研究^[40],孙华钦等^[41]对不同类群蕨种的 *psbA-trnH* 基因间区进行 PCR 扩增并测序,获得了该区间的完整序列,在种间具有明显的较大差异,并进行了分子系统学分析。

atpB-rbcL 间隔区在系统分析中也能有较好的应用,可用于种属间系统发育的研究^[42],刘明珍等^[43]应用 *atpB-rbcL* 间隔区序列来评价蓝药蓼和大铜钱叶

蓼 2 个物种的亲缘关系和属的归属,发现该片段为 753~902 bp,种间差异很大,存在丰富的系统发育变异位点,其中变异位点为 196 个,非信息位点 145 个,用最大简约法寻找得到 1 棵最简约树。

petG-trnP, 一段位于 *petG* (细胞色素 b/f 复合物) 和 *trnP* (脯氨酸转运 RNA) 基因之间的长约 320 bp 的间隔序列,且 *trnW* (色氨酸转运 RNA, 74bp) 分隔成两部分,已被广泛地应用于植物属间及种间的系统发育研究^[44]。

此外, *petD-rpoA*, *trnD-Y*, *trnS-G*, *trnT-L* 等大量的非编码区在系统学上均有较好的应用,多被运用于分子系统学方面。

3.2.2. 内含子

内含子不直接参与编码,收到的功能限制小,选择压力小,进化速率快,相对于编码区而言可用于较低阶元的研究。

rpl16: 约 1000 kb, 编码叶绿体中的核糖体蛋白,其内含子序列的差异主要体现在长度上,由于 *rpl16* 在 cpDNA 中被认为是进化最快的 DNA 片段之一,通常用于科间和属间的分析^[45,46]。杨俊波等应用 *rpl16* 等 4 个 DNA 片段对山茶属的 21 种植物进行序列测定和分析,建立的分子系统树支持山茶属为一单系这一结论^[47]。

rps16: 编码核糖体蛋白 S16, 位于植物 cpDNA 大单拷贝区, *rps16* 内含子一般适用于较高分类阶元的系统发育研究^[48]。在前人的基础之上,董文攀从不同亚科不同属不同种上对叶绿体整个基因组进行了系统的研究,从单个基因到整个叶绿体基因组,研究出 21 对叶绿体通用引物,并通过分析蜡梅科叶绿体基因组的进化,探讨了被子植物叶绿体基因组插入/缺失、短片段倒位与重复等微结构变异式样,同时在不同分类水平上探讨蜡梅科叶绿体基因组序列分化^[49]。

4. 叶绿体基因组用于植物系统发育发展前景

分子生物学技术发展迅速,新的技术和方法层出不穷,适于叶绿体基因研究植物系统发育的 DNA 分析技术将会更多,实用性会更强,这将会为研究植物的系统发育做出更大贡献。同时也需要研究者依据需

要结合各种技术的应用特点和使用范围,选择合适的方法,以便得到最佳效果。但也存在许多问题。首先,虽相同 cpDNA 序列在不同类群间的进化速率有差异,但序列本身在植物系统学研究中总有一相对稳定的适用范围,这使其涵盖的研究内容和层次有限。其次,虽然基因之间在序列变异方面存在较大差异,但一个或少数几个基因片段提供的系统发育信息有限,构建的系统树可能反映的仅仅是该片段的基因树,因在许多情况下基因树并不等同于物种树,不一定代表真正的物种系统发育关系。分子片段仅是分类群诸多性状的一个来源,它虽然能为分类群的系统重建提供不可忽视的信息,但并不能完全地反映其真实地历史演化。再者,叶绿体基因组是母性遗传的,只能够帮助推断杂交物种形成的母系来源并不能单靠叶绿体基因组来解释种群间的杂交现象,虽然有越来越多的叶绿体被用作分子标记来研究类群间的系统发育关系,但只有将这些分子片段提供的信息与其他分子片段信息、传统的形态及生理特征结合起来,获得更多的信息,才能更接近系统发育的真相^[50]。将具有单亲遗传特性的叶绿体和线粒体基因组与具有双亲遗传特性的核基因组联系起来,有利于构建更加全面的系统进化树。一般情况下,核基因树与叶绿体基因树是一致的,但有时也会出现系统树冲突现象。由于核基因是双亲遗传的,利用其序列变异探讨植物的系统发育过程,特别是解决网状进化问题具有明显的优势,这在一定程度上弥补了单遗传基因的不足,因此对于遗传背景不同、进化速率不同的基因或 DNA 片段进行综合分析,可望对所研究类群的系统发育提供更佳全面的信息。

随着分子系统发育研究方法的深入,生物学家已将生物信息大分子看作重要的演化依据,并在不断寻找新的、检测功能良好的分子标记及检测手段,随着技术的不断进步,来自分子方法的数据可能成为系统学研究最主要的数据来源,并将引起分类、系统、发育和进化研究中的又一次革命性的变化,而且关于种群遗传结构方面的研究必将为保护生物学提供遗传变异证据,在生物多样性的研究和保护中起到重要指导作用^[51]。

叶绿体编码区的变化会带来很大的表型改变,进化速度较慢适用于较高阶元科、目乃至更高的系统发

育学研究；而非编码区的突变给表型带来的影响很小，进化速度较快，适用于较低阶元种、属的系统发育学研究，而其它各种 DNA 序列虽已得到越来越广泛的应用，但由于受许多因素如杂交和渗入、谱系分选、水平转移、重组、位点间互作及致同进化等影响，使得基于单一遗传位点 DNA 片段构建的基因树并不必然与物种的真实进化途径相一致。另外叶绿体是母系遗传，用叶绿体全基因组的结果更可能代表全基因组(物种)的进化关系。基于上述特点，叶绿体基因组分析在植物系统发育研究中成为不可或缺的研究内容。新一代高通量测序技术给大量测定叶绿体基因组带来了机遇，利用新一代测序技术高通量的优点，可在较短时间内实现多个物种叶绿体基因组测序，为比较叶绿体基因组学和系统发育叶绿体基因组学研究提供更多的数据。本实验室在前人做的基础上对国内国外 34 个红花品种的叶绿体基因组进行了 PCR 扩增，发现各个品种之间存在差异，可能预示着前人发现的 21 对引物不仅可用于高级元系统发育分析，也可用于某些种的低级元分析如品种与品种之间。同时，还对猕猴桃的不同品种进行了扩增分析，对前人的实验结论进行了验证。

线粒体和叶绿体都具有内共生起源，所含的基因数量少，无内含子，有自己的遗传物质和遗传体系，母系遗传，为细胞提供能量的同时还参与诸如细胞分化、细胞信息传递和细胞凋亡等过程，并拥有调控细胞生长和细胞周期的能力。与呼吸相关的基因可保持一定的稳定性，虽然目前受技术的限制，随着技术的不断更新，其在植物系统方面的应用范围将不断扩大。本实验室也正致力于此方面的研究。

项目基金

湖北省自然科学基金(2010CDZ011)。

参考文献 (References)

- [1] 李健仔, 李思光, 罗玉萍, 万璐 (2002) 叶绿体 DNA 分析技术及其在植物系统学研究中的应用. *江西科学*, **3**, 183-189.
- [2] 石开明 (2002) DNA 序列在植物系统进化研究中的应用. *湖北民族学院学报(自然科学版)*, **4**, 5-10.
- [3] Kato, S., Yamaguchi, H., Shimaoto, Y. and Mikami, T. (2000) The chloroplast genomes of azuki bean and its close relatives: A deletion mutation found in weed azuki bean. *Heredity*, **132**, 43-48.
- [4] 彭隽敏, 孔清, 徐乃瑜 (1995) 云南小麦、西藏半野生小麦及

- 普通小麦叶绿体 DNA 限制性内切酶图谱的研究. *遗传*, **6**, 4-6.
- [5] Kim, K.J. and Jansen, R.K. (1994) Comparison of phylogenetic hypothesis among different data sets in dwarf dandelions (*Krigia*): Additional information from interal transcribed spacer sequences of nuclear ribosomal DNA. *Plant Systematics and Evolution*, **190**, 157-185.
- [6] Palmer, J.D., Shield, C.R., Cohen, D.B. and Orton, T.J. (1983) Chloroplast DNA evolution and the origin of amphidiploid Brassica species. *Theoretical and Applied Genetics*, **65**, 181-189.
- [7] Poulsen, C.R. (1983) Comments on the structure and function of the large subunit of the enzyme ribulose biphosphate carboxylase-oxygenase. *Carlsberg Research Communications*, **48**, 48-57.
- [8] 俸宇星, 汪小全, 潘开玉 (1998) *RbcL* 基因序列分析对连香树科和交让木科系统位置的重新评价兼论低等金缕梅类的关系. *植物分类学报*, **5**, 411-422.
- [9] 张文衡, 陈之端, 陈虎彪, 汤彦承 (2001) 从叶绿体 DNA *trnL* 序列论双参属的归属问题. *植物分类学报*, **4**, 337-344.
- [10] Zhang, Y.J., Ma, P.F. and Li, D.Z. (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *Plosone*, **6**, Article ID: e20596.
- [11] Zeng, C.X., Zhang, Y.X., Triplett, J.K., Jimmy, K., Yang, J.B. and Li, D.Z. (2010) Large multilocus plastid phylogeny of the tribe Arundinarieae (Poaceae: Bambusoideae) reveals ten major lineages and low rate of molecular divergence. *Molecular Phylogenetics and Evolution*, **56**, 821-839.
- [12] Remiaw, A.W., Alistair, L.D. and Barbara, A.W. (2001) cpDNA-RFLP in *Ceramium* (Rhodophyta): Intraspecific polymorphism and species-level phylogeny. *American Journal of Botany*, **88**, 1209-1213.
- [13] Wattier, R.A., Davidson, A.L., Ward, B.A., Ward, B.A. and Maggs, C.A. (2001) cpDNA-RFLP in *Ceramium* (Rhodophyta): Interspecific polymorphism and species-level phylogeny. *American Journal of Botany*, **88**, 1209-1203.
- [14] Wang, X.-R., Yoshihiko, T., Hiroshi, Y., Kazutoshi, N. and Alfred, E.S. (1999) Phylogenetic relationships of Eurasian pines (*Pinus* Pinaceae) based on chloroplast *rbcl*, *matk*, *rp120-rps18* spacer, and *trnV* intron sequences. *American Journal of Botany*, **86**, 1742-1753.
- [15] Ciprini G, Testolin R, Garder R. (1998) Restriction site variation of PCR-amplified chloroplast DNA regions and its implication for the evolution and taxonomy of Actinidia. *Theoretical and Applied Genetics*, **96**, 389-396.
- [16] Kurt, W. and Richard, C.G. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **40**, 9-19.
- [17] Provan, J., Corbett, G., Waugh, R., McNicol, J.W., Morgante, M. and Powell, W. (1996) DNA fingerprints of rice (*Oryza sativa*) obtained from hypervariable chloroplast simple sequence repeats. *Proceedings of the Royal Society B: Biological Sciences*, **1375**, 1275-1281.
- [18] Zuccarello, G.C., Bartlett, J. and Yeates, P.H. (2000) Differentiation of *Caloglossa leprieurii* (Rhodophyta) populations in northern and eastern Australia using plastid haplotypes [P]. *European Journal of Phycology*, **35**, 357-363.
- [19] Isoda, K., Shiraishi, S., Watanabe, S. and Kitamura, K. (2000) Molecular evidence of natural hybridization between *Abies veitchii* and *A. homolepis* (Pinaceae) revealed by chloroplast, mitochondrial and nuclear DNA markers. *Molecular Ecology*, **9**, 1965-1974.
- [20] Sugiura, M. (1992) The chloroplast genome. *Plant Molecular Biology*, **19**, 149-168.
- [21] 王艇, 苏应娟, 朱建明 (1999) 叶绿体 *rbcl* 基因序列在植物系统学研究中的应用. *武汉植物学研究*, 增刊, 8-14.
- [22] Chase, M.W., Soltis, D.E., Olmstead, R.G. and Morgan, D. (1993) Phylogenetic of seed plants: an analysis of nucleotide se-

- quences from the plastid gene *rbcl*. *Annals of the Missouri Botanical Garden*, **80**, 528-586.
- [23] Johnsom, S.A. and Soltis, D.E. (1995) Phylogenetic inference in Saxi fragaceae sensus and Gilia (polemoniaceae) using *matK* sequence. *Annals of the Missouri Botanical Garden*, **82**, 149-175.
- [24] 吴世安, 吕海亮, 杨继, 饶广远, 尤瑞麟, 葛颂, 钟阳 (2000) 叶绿体 DNA 片段的分析在黄精族系统学研究中的应用. *植物分类学报*, **2**, 97-110.
- [25] 何兴金, 葛颂, 许介眉, 洪德元 (2000) 中国葱属系统发育的 PCR-RFLP 分析. *中国科学(C 辑)*, **2**, 183-191.
- [26] Nadot, S., Bitter, G., Cater, L., Lacroix, R. and lejeune, B. (1995) A phylogenetic analysis of monotyledonsbsed on the chloroplast gene *rps4* using parsimony and a new numerical phenetics method. *Molecular Phylogenetics and Evolution*, **4**, 257-282.
- [27] Sasa, S., Lori, K. and Richard, G. (2002) Monophyly of the convol vulaceae and circumscription of their major lineages based on DNA sequence of multiple chloroplastloci. *American Journal of Botany*, **89**, 1510-1522.
- [28] 崔杰, 徐德昌, 李滨胜, 杨谦, 孙璟晗 (2006) 甜菜 ATP 合酶 β 亚基基因 *atpB* 的克隆、序列分析及进化. *植物研究*, **5**, 583-588.
- [29] 王化坤, 姜晓鸣, 章镇 (2006) 叶绿体微卫星在植物种质资源研究中的应用. *分子植物育种*, **3**, 92-98.
- [30] 刘畅, 杨足君, 李光蓉, 冯娟, 邓科君, 黄健, 任正隆 (2006) 叶绿体基因 *infA-rpl36* 区域在小麦族物种中的序列变异分析. *遗传*, **10**, 1265-1272.
- [31] 黄冲, 吴乃虎 (1990) 叶绿体生物分子生物学研究进展. *生物工程进展*, **226**, 140-150.
- [32] Lehnebach CA, Cano A, Monsalve C, McLenachan P, Horandel E, Lockhart P. (2007) Phylogenetic relationships of the monotypic Peruvian genus *Laccopetalum* (Ranunculaceae). *Plant Systematics and Evolution*, **46**, 109-116.
- [33] Sousa, L.O.F., Wendt, T., Brown, G.K., Tuthill, D.E. and Evans, T.M. (2007) Mmonophyly and phylogenetic ralationships in Lymania(Bromeliaceae: Bromelioideae)based on morphology and chloroplast DNA Sequences. *Systematic Botany*, **32**, 264-270.
- [34] 周奕华, 侯丙凯, 石东乔, 肖宇红, 石锐, 胡赞民, 陈正华 (2001) 杨树叶绿体 3'rps12 基因、*rps7* 基因和 *ndhB* 基因片段的克隆及序列分析. *中国生物化学与分子生物学报*, **5**, 606-616.
- [35] 刘志文, 韩旭, 李莉, 李宪臻, 陈温福 (2008) 叶绿体和线粒体 DNA 在植物系统发育中的应用进展. *河南农业科学*, **7**, 5-9.
- [36] Elspeth, M.H., Gwilym, P.L. and Julie, A.H. (2005) A phylogenetic reappraisal of the peltophorum group (Caesal pinieae: Leguminosae) based on the chloroplast *trnL-F*, *rbcL* and *rps16* sequence data. *American Journal of Botany*, **92**, 1359-1371.
- [37] 王玉金, 刘建全 (2004) 利用叶绿体 DNA *trnL-F* 序列初步探讨菊科风毛菊属的系统发育. *植物分类学报*, **2**, 136-153.
- [38] 侯鑫, 刘俊娥, 赵一之, 赵利清 (2006) 基于 ITS 序列和 *trnL-F* 序列探讨小叶锦鸡儿、中间锦鸡儿和柠条锦鸡儿的种间关系. *植物分类学报*, **2**, 126-134.
- [39] 陆树刚, 李春香 (2006) 用叶绿体 *rbcL* 和 *rps4-trnS* 区序列确定亚洲特有单型属-篦齿蕨属的系统位置. *植物分类学报*, **5**, 494-502.
- [40] Charles, A. and Robert, S. (2004) phylogenetic studies of mam-millaria(cactaceae) insight from chloroplast sequence variation and hypothesis testing using the parametricboot strap. *American Journal of Botany*, **91**, 1086-1098.
- [41] 孙华钦, 罗科, 邹文俊, 邓思杨 (2006) 穿龙薯蓣、黄山药和盾叶薯蓣 *psbA-trnH* 片段序列分析. *应用与环境生物学报*, **6**, 792-797.
- [42] Su, Y., Wang, T., Zheng, B., Jiang, Y., Chen, G. and Gu, H.Y. (2004) Population genetic structure and phylogeographical pattern of a relict tree fern, *Alsophila spinulosa*(Cyatheaceae), inferredfrom cpDNA *atpB-rbcL* intergenic spacers. *Theoretical and Applied Genetics*, **109**, 1459-1467.
- [43] 刘明珍, 周忠泽, 邱英雄, 孙伟, 董翔 (2007) 分子证据支持蓝药蓼和大铜钱叶蓼归入冰岛蓼属. *植物分类学报*, **2**, 227-233.
- [44] Goremykin V, Hirsch KI, Wolff S, Hellwing FH. (2003) The chloroplast genome of the 'basal' angiosperm *Calycanthus fertilis*-structural and phylogenetic analyses. *Plant Systematics and Evolution*, **242**, 119-135.
- [45] Charles, A. and Robert, S. (2004) Phylogenetic studies of mam-millaria(cactaceae) insight from chloroplast sequence variation and hypothesis testing using the parametric bootstrap. *merican Journal of Botany*, **91**, 1086-1098.
- [46] Joey, S. and Randall, L. (2005) Chloroplast DNA phylogeny and phylogeography of the north american plums (*Prunus* subgenus *prunussection prunuocerasus*, rosaceae). *American Journal of Botany*, **92**, 2011-2030.
- [47] 杨俊波, 李洪涛, 杨世雄, 李德铎, 杨莹燕 (2006) 四个 DNA 片段在山茶属分子系统学研究中的应用. *云南植物研究*, **2**, 108-114.
- [48] Elspeth, M.H., Gwilym, P.L. and Julie, A.H. (2005)A phylogenetic reappraisal of the peltophorum group (Caesalpinieae: leguminosae) based on the chloroplast *trnL-F*, *rbcL* and *rps16* sequence data. *American Journal of Botany*, **92**, 1359-1371.
- [49] Dong, W.P., Liu, J., Yu, J., Wang, L. and Zhou, S.L. (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *Plose One*, **7**, Article ID: e35071.
- [50] 田欣, 李德株 (2002) DNA 序列在植物系统学研究中的应用. *云南植物研究*, **2**, 50-53.
- [51] 徐宏发, 王静波 (2001) 分子系统学研究进展. *生态学杂志*, **3**, 41-46.