

植物分子系统发育与进化研究的基因选择、引物设计与应用策略

王玉国

(复旦大学 生命科学学院 生物多样性科学研究所, 教育部生物多样性与生态工程重点实验室, 上海 200433)

摘要: 引物选择、设计与应用策略是植物分子系统发育与进化研究的关键环节。本文综述了基因选择的原则、引物设计的技巧以及如何有效地利用所涉及的片段获取相应的 PCR 片段的方法。

关键词: 系统发育; 进化; 基因选择; 引物设计; 应用

中图分类号: Q941 文献标识码: A 文章编号: 1000-3142(2010)06-0753-07

Gene selection, primer design and their utility on the studies of plant molecular phylogeny and evolution

WANG Yu-Guo

(Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, Fudan University (IBSFU), Shanghai 200433, China)

Abstract: Primer selection, design and utility are key and absolutely necessary steps on the studies of plant molecular phylogeny and evolution. Here the principles of gene selection, the skills of primer design and the efficient utility of the primers selected to obtain the needed PCR products are summarized.

Key words: phylogeny; evolution; gene selection; primer design; utility

随着分子生物学和生物信息学的快速发展, 核酸序列分析已成为分子系统发育与进化研究中广泛应用的基本方法。而作为 PCR 实验的关键环节, 引物选择与设计策略的好坏却往往可决定研究的成败。

1 植物分子系统发育与进化研究的基因选择原则

以往植物分子系统发育与进化研究, 已经利用了来自三个基因组的基因或 DNA 片段构建了不同类群的系统发育。每个基因组的不同基因可能被应用在不同分类等级的类群。由于不同分类学家对科、属、种等阶元的概念和范围理解不同, 而相同阶元内不同类群的进化速率可能不一致, 因此也会出

现即使同一基因在同一阶元的不同类群中应用效果有所不同的现象, 如 cpDNA *trnL-F* 区可以很好地解释一些属的系统发育, 但在另一些属内则可能很少有分化。尽管如此, 在分子系统发育与进化研究中基因选择还是有基本原则可以遵循的。(1) 通用引物优先原则: 在亲缘关系较远的不同类群中都适用的引物, 往往是通用引物, 其适用范围也广, 如 White 等(1990)设计的扩增核糖体基因(nrDNA) ITS 区域的 ITS5、ITS4、ITS3 和 ITS2 四个引物、Taberlet 等(1991)设计的扩增叶绿体基因(cpDNA) *trnT-trnL-trnF* 区段的 a-f 6 个引物、Demesure 等(1995)设计的扩增线粒体基因(mtDNA) *nad1* 和 *nad4* 内含子的 6 个引物, 这些引物通用性在很多类群都得到很好的验证;(2) 近缘类群优先原则: 通常情况下, 在一个类群中已经成功应用的基因, 往往在

收稿日期: 2010-10-16 修回日期: 2010-11-12

基金项目: 国家自然科学基金(30970199)[Supported by the National Natural Science Foundation of China(30970199)]

作者简介: 王玉国(1970-), 男, 黑龙江甘南人, 博士, 长期从事被子植物分子系统发育与进化研究。(E-mail) wangyg@fudan.edu.cn.

其近缘类群也可以使用。因此,如果在一个类群的近缘类群上已经做过某些基因的系统学分析,这些基因就可以作为优先的选择。以木兰科为例,cpDNA *matK* 基因序列可用来分析木兰科含笑属 (*Michelia*) 的系统发育(金虹等,1999),那么,该基因通常也会适合其近缘类群木兰属 (*Magnolia*) 的系统发育(Azuma 等,1999)。(3)富含信息位点原则:具有关键信息位点或信息位点丰富的基因才有

助于疑难系统发育与进化关系问题的解决。对于较低分类等级的类群,选择长的内含子或基因间隔区,相对短的基因片段而言信息位点可能会多些。因此,在基因选择和引物设计时应作重点考虑。

了解不同基因组的特点、不同基因或 DNA 片段的一般应用分类等级的上下限(图 1),以及具体类群的具体情况,将有助于合理而有效地应用它们去获得系统发育信息(Soltis & Soltis,1998)。

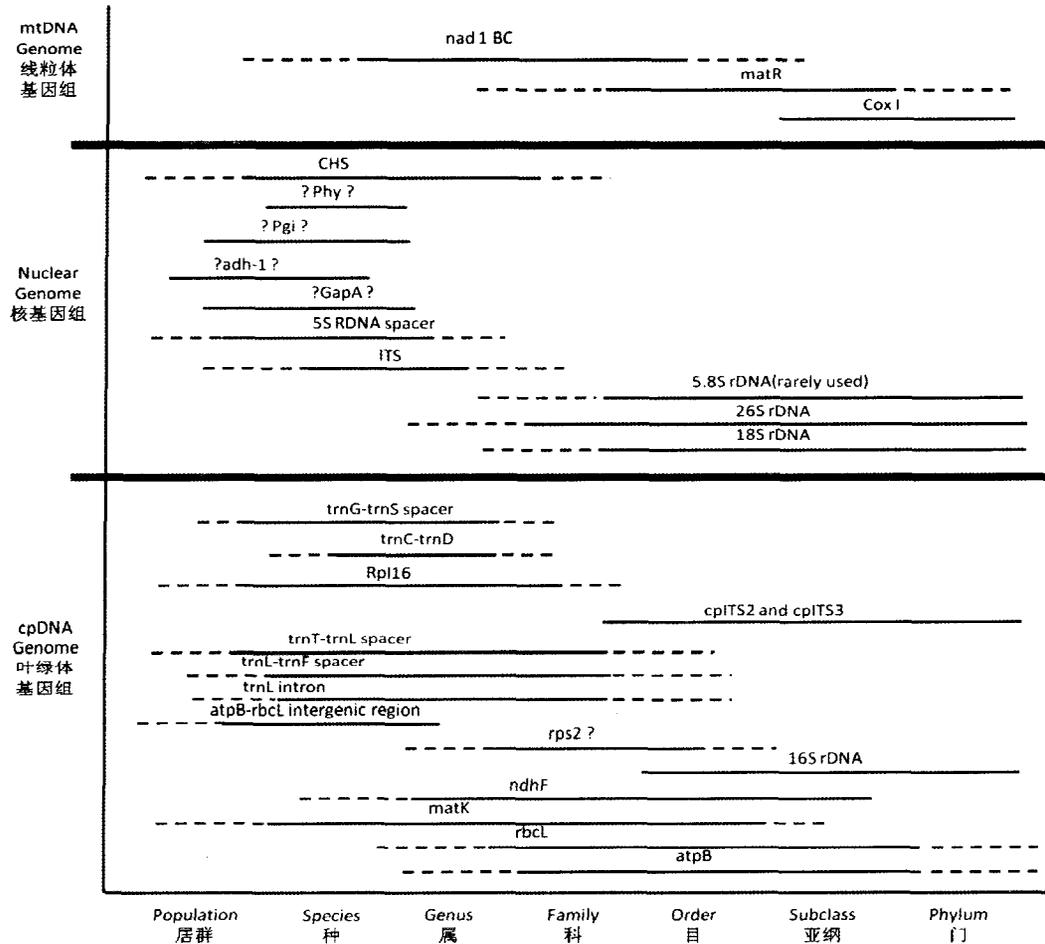


图 1 引物实用范围上下限

Fig. 1 Taxonomic level of utility of primers

根据 Soltis DE & Soltis PS(1998)修改而成。‘——’代表该基因可能的上限或下限,‘?’代表该基因很少被用到
 Figure is modified from Soltis DE & Soltis PS(1998). ‘——’designates the approximate upper or lower limits of applicability. ‘?’Refers to genes that have been rarely used

1.1 叶绿体基因组

主要由单拷贝序列组成,在被子植物中常常单亲遗传,一般母系遗传(猕猴桃属 *Actinidia* 是个例外,为父系遗传),与核基因相比,不受网状进化(reticulate evolution)影响,其进化历史相对容易分析,构建系统树比较容易。叶绿体基因组还具有以下几

个方面的优点受到研究者青睐:(1)cpDNA 在植物总 DNA 中含量较丰富,易于提取和分析;(2)对叶绿体基因遗传背景了解较多,目前已经有超过 70 余种的叶绿体全基因组序列公布,更多物种的叶绿体全基因组序列正在测序之中;(3)不同的叶绿体基因进化速率不同,具有系统学价值的基因很多(Hoot

等,1999),可以选择进化速率快的基因探讨低分类等级的系统发育;也可以选择核酸替换相对保守的基因,去解决较高等级分类群的系统发育问题。例如 *rbcL* 是分子系统学研究常用的基因,通常适合于科以上分类等级的系统发育,有时也用于科内属间;而 *matK* 基因、*trnL-F* 区和 *rpl16* 内含子等则非常适合用于属内、种间的系统发育关系研究。

matK 基因长约 1 500 bp,位于叶绿体赖氨酸 tRNA 基因(*trnK*)的内含子中,编码一种成熟酶(maturase),可能参与 RNA 转录体中 II 型内含子的剪切。*matK* 是叶绿体基因组蛋白质编码基因中进化速率最快的基因之一(Johnson & Soltis,1994; Liang & Hilu,1996)。*matK* 的序列变异中转换和颠换的频率、密码子 3 个位置没有严重偏离,这与 *rbcL* 有较大差别。由于 *matK* 序列的变异均为均一,在构建系统树时不必对不同类型的变异进行加权,极大地增强了分子系统树的可靠性;同时 *matK* 比 *rbcL* 长,核酸替换率是 *rbcL* 的 2~3 倍以上,用于构建分子系统树分辨率大大提高(Johnson & Soltis,1994)。*matK* 一般用于科内、属间、一些类群种间关系的研究(如 Erdogan & Mehlenbacher,2000;Levin 等,2000;Meimber 等,2001)。由于 *matK* 进化速率相对保守,同样也适用于较高等级的分类群研究(Manos & Steele,1997)。

trnL-F 区包括 *trnL*[UAA]内含子和 *trnL*[UAA]-*trnF*[GAA]之间的间隔区(约 0.8 kb)。针对该基因片段所设计的引物比较通用,相对容易扩增而被广泛地应用于低分类等级的系统发育研究之中(如 Levin 等,2000;Sweeney & Price,2000;Wilkström & Kenrick,2000;Yen & Olmstead,2000;Fritsch 等,2001;Manen 等,2002)。

在有详细化石记录可校对系统发育分支的类群,*rbcL*、*matK*、*trnL-F* 等基因可以很好地进行分子钟推导,进一步推测进化历史事件,在物种起源与进化研究中被广泛采用。

rpl16 基因是编码核糖体 L16 的基因,其内含子约 1.0 kb,在绝大多数被子植物中都存在,并成功地应用一些属的系统发育研究,可以很好地解释属内种间的系统发育关系(如 Posno 等,1986;Downie 等,1996,2000;Jordon 等,1996;Kelchner & Clark,1997;Baum 等,1998;Kelchner,2002;Wang 等,2004)。比较烟草和水稻 17 个共有的叶绿体基因内含子碱基序列后发现,*rpl16* 内含子是

分化最快的(Downie 等,1996)。此外,可作为分子系统学研究其他叶绿体 DNA 的分子标记还有:*atpB*、*ndhF*、*rps2*、*rps4*、*rps16*、*cpITS2-4*、*5S-cpITS3*、*trnC-trnD*、非编码序列 PY-IGS 片段等等。CBOL Plant Working Group(2009)曾比较 *matK*、*rbcL*、*rpoB*、*rpoC1*、*atpF-atpH*、*psbK-psbI* 和 *trnH-psbA* 等 7 个基因或基因间隔区,来探索区陆生植物物种的“条形码(Barcoding)”。事实上,可用做 Barcoding 的基因还很多,Heinze(2007)就对已经发表的引物、新设计的引物进行汇总建立一个叶绿体基因引物数据库。文章见 <http://www.plantmethods.com/content/3/1/4>。全部引物可从文中附表获得。研究者可以根据不同目的进行选择。

1.2 线粒体基因组

以往认为线粒体基因重排事件频繁,作为信息分子,同源性受到影响,很难用于分子系统学研究,特别是应用于高等级分类群的研究。但是,一些线粒体基因如 *cox I*、*matR* 和 *mtSSU* 已经在被子植物基部类群的研究中应用,并取得很好的进展(Parkinson 等,1999;Qiu 等,1999)。*nad1* exon BC intron II 也在较低分类等级类群的系统发育和进化研究(Sanjur 等,2002;Schaefer & Renner,2010),甚至在大花草及其宿主的基因水平转移研究中都有报道(Davis & Wurdack,2004)。

1.3 核基因组

尽管核基因组很大,所包含的基因的数量和变异相当大,但是大多数用于系统发育研究的核基因序列却主要限于核糖体 DNA。rDNA 用于植物系统分类和进化研究有其突出的优点:(1)结构功能十分保守,但又具有一定的变异;(2)在细胞中含量大,易获得;(3)与叶绿体基因相比,避免了由于单亲遗传带来的对分支关系的错误解释。因此,rRNA 是进行分子系统分类和进化研究较为理想的指标。最常用基因或片段:18S rRNA 基因、nrDNA ITS 区等。高度保守的拷贝区(18S、26S rDNA)最早被用于科和科以上水平的系统发育研究。而进化速度快的区域,如 ITS,则常用于种间和近缘属间的关系。其他几个核基因,如 *Pgi*、*gapA*、*adh*、*gapC* 显示出一定的应用潜力,甚至可用于居群水平的研究(Soltis & Soltis,1998)。

nrDNA ITS 区(internal transcribed spacer of nuclear ribosome,核糖体 DNA 转录间隔区)这是高等植物中应用最为普遍的 DNA 片段。它位于小

亚基和大亚基 rRNA 基因之间,被 5.8S rRNA 基因分为两段,即 ITS1 和 ITS2。在被子植物中,虽然不同被子植物的 rDNA ITS 区序列差异较大,但 ITS 区的长度差别较小,总长度约为 600~700 bp, ITS-1 的长度为 187~298 bp, ITS2 为 187~252 bp,非常容易进行 PCR 扩增(Baldwin 等,1995)。作为 18S-26S rDNA 的一个组成部分,ITS 在核基因组中是高度重复的,而且通过不等交换和基因转换,这些重复单位间已发生了位点内或位点间的同步进化(concerted evolution)(Elder & Turner, 1995),即不同 ITS 拷贝间的序列趋于相近或完全一致。对于不涉及网状进化(即杂交起源)的类群,通常情况下其 PCR 扩增产物都可以直接测序;对于不易扩增的 DNA 样品,组合使用 ITS5、ITS4、ITS3、ITS2 等通用引物,采用分段 PCR 或巢式 PCR 的方法可以很好地实现 ITS 区 PCR 扩增、测序。

利用 ITS 作分子标记来研究系统发育关系已得到了广泛应用,它在科、亚科、族内的系统发育与分类问题研究方面具有重要作用(Campbell 等,1995;Downie & Katz-Downie,1996;Francisco-Ortega 等,1997)。关系密切的物种之间 ITS 长度接近,而序列有一定程度的变异,因此,该片段特别适合于属、组级等较低等级分类群的系统发育和分类研究(Baldwin,1993;Baldwin 等,1995;Yuan 等,1996;Ainouche & Bayer,1997;Kollipara 等,1997)。

单拷贝与低拷贝核基因 对于杂交或多倍化起源的类群,单拷贝基因和低拷贝核基因是探讨其系统发育与进化的最合适的选择。已有的文献显示,*rpb2-d* 或 *-I copy*、*Adh1*、*Adh2*、*GBSSI(waxy)*、以及花基因 *PISTILLATA(PI)*、*Leafy* 等基因在一些类群中都可以作为单拷贝基因使用(Sang 等,1997;Mason-Gamer 等,1998;Lee 等,2002;Oxelman 等,2004;Kim 等,2008)。由于古多倍化的存在,对于染色体基数较大的类群,寻找完全的单拷贝基因往往是非常困难的。可以将同一基因家族的不同拷贝作为备选基因考虑,但如果研究涉及的取样量较大时,其克隆测序的工作量将明显加大。

适合谱系地理学(Phylogeography)的核基因 谱系地理学的范围一般涉及同一个种的复合体或同种的不同居群。要求的合适基因是在这两个水平上有较高的信息位点变异。通常 nrDNA ITS、cpDNA *trnT-F*、*atpB-rbcL* 等是植物谱系地理学主要的候

选基因。核基因在这方面应有很大潜力,如查尔酮合成酶基因(chalcone synthase)的内含子、*Leafy* 内含子 2、果实液泡转化酶(fruit vacuolar invertase)基因内含子 2 在不同类群中都有很好应用(Caicedo & Schaal,2004;Bartish 等,2006;Liao 等,2010)。优先考虑具有长的内含子或基因间隔区是这类基因筛选的关键。

多基因综合分析是目前被子植物系统发育与进化研究主要的常规手段。不同基因对于揭示植物系统发育过程均具有重要意义。一般情况下,核基因树与叶绿体基因树是一致的,但有时也会发生系统树冲突。不同于单亲遗传的叶绿体基因和线粒体基因。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。对于进化速率不同的基因或 DNA 片段,综合分析可望对所研究类群的系统发育提供更为全面的理解。

2 引物设计软件的使用

并不是在任何时候都有可以借鉴的通用基因,用于植物的系统发育与进化研究。由于研究者所要解决的科学问题不同,近缘类群没有相关研究,往往就需要根据不同研究目的进行引物设计。比如研究植物的杂交与多倍化,就需要设计单拷贝核基因的引物;研究多基因家族进化,就需要设计可以扩增这一基因家族不同基因的引物;做谱系地理学研究,就需要针对进化速率相对较快的基因设计引物;做背景不清的类群的居群遗传结构和遗传多样性研究,有时就需要设计微卫星引物。

无论哪种情况,都要涉及引物设计软件。设计软件种类很多,但其引物设计主要遵循 3 条基本原则:首先,引物与模板的序列要紧密互补;其次,引物和引物之间避免形成二聚体(primer dimer);再次,引物不能在模板的非目的位点引发 DNA 聚合反应(即错位)。在进行引物设计时,还要注意以下的要点(张新宇等,2004):1)引物长度:一般为 15~30 bp。其有效长度可以用公式 $L_n = 2(G+C) + (A+T)$ 来计算,一般不应大于 38 bp,否则产物的特异性将因 PCR 最适延伸温度超过 Taq 酶的最佳作用温度(74 °C)而降低;2)G+C 含量应在 40%~60%之间。PCR 复性温度一般是引物的 T_m 值减去 5~10 °C。当引物长度小于 20 bp 时, $T_m = 4(G+C) + 2$

(A+T);3)碱基应随机分布,应避免连续出现几个同样的碱基。尤其不应在 3'端出现超过 3 个连续的 G 或 C,否则会导致引物在 G+C 富集区错误引发;4)引物 3'端不应发生错配,否则会在很大程度上影响 Taq 酶的延伸效应。引物 3'端最好选用 A、G 或 C,而尽可能避免使用 T(也有文献说明应避免使用 A),尤其是连续 2 个以上的 T;5)引物自身不能含有互补序列,否则会形成发夹结构而导致引物无效;6)两个引物之间不应有多于 4 个互补或同源的碱基,否则会形成引物二聚体,尤其应避免 3'端的互补重叠;7)引物应与扩增目的序列之间存在特异性。与非扩增序列的同源性小于 70%或少于连续 8 个互补碱基。

对于已知序列,可通过特定的引物设计软件如 Primer Premier 5.0 和 Oligo 6.44 来完成引物设计和评估。关于这两个软件的使用,张新宇等(2004)已有较为详尽的说明,在此不再赘述。需要说明的是,即便是软件对给出所设计的引物各项指标得分都很高,也不一定能确保引物一定能扩增出所需的条带。比如,用于克隆目的的 PCR,由于产物序列相对固定,引物设计的选择自由度较低。要想获得效果很好的引物,在自动搜索的基础上还要辅以人工分析,将这两套引物设计软件合并使用可能效果更好。根据以往设计引物的经验,(1)对于特殊的酶退火温度要求较高,如 LA Taq 酶或其他高保真酶,设计时可通过控制引物长度来决定;(2)正向引物和反向引物的 T_m 值、GC 含量值不要相差悬殊;(3)扩增引物不宜在所扩增区段内部有连续的配对区。(4)引物长度可通过 File > Rreference > Default Primer Length 选定,通过手工来调节 3'端碱基组成和两个引物 GC 含量差值。

2.1 利用 mRNA 序列设计引物扩增总 DNA 特殊片段的方法

在实验设计中,常常需要根据已知的 mRNA 序列设计扩增总 DNA 特殊片段的引物。一般原则是从外显子保守区域入手设计。通常的办法:将该 mRNA 序列通过 GenBank 与有基因组全长的物种如拟南芥、水稻等进行 Blast 比对,确定内含子与外显子边界,推测可能的内含子长度;然后在内含子区域用多个(如 20 或以上)n 来代替,将该序列导入 Primer Premier 5.0 之中,再避开 n 区进行引物设计,按引物设计的一般原则,尽可能设计多对扩增引物,以便后续的分段 PCR 或巢式 PCR 扩增目的片段。

2.2 微卫星引物的开发与设计方法

微卫星 DNA(Microsatellite DNA)即简单重复序列(simple repeats sequence, SSR),是一类由 1~6 个碱基组成的基序串联重复而成的 DNA 序列。不同等位基因间的重复数存在丰富差异,且重复序列两侧多是相对保守的单拷贝序列,可通过设计特异引物对基因组总 DNA 进行 PCR 扩增。由于微卫星标记扩增片段的多态性丰富,重复性好,标记一般为共显性,在基因组中分散分布,已成为系统与进化植物学研究的重要分子标记,被广泛应用于植物居群遗传结构、遗传多样性与谱系进化等方面的研究中。

微卫星引物的开发方法多样(Zane 等,2002;张增翠,侯喜林,2004;陈怀琼等,2009)。经典的方法是基于基因组文库构建结合 SSR 探针杂交的筛选法,后来陆续发展了基于 RAPD(Randomly Amplified Polymorphic DNA)的开发策略(Cifarelli 等,1995;Lunt 等,1999)、基于 ISSR-PCR 的 SSR 分离策略(Fisher 等,1996;Lench 等,1996)、基于引物延伸的分离策略(Ostrander 等,1992)、利用选择杂交进行 SSR 分离(Karagyozov 等,1993;Kandpal 等,1994;Zane 等,2002)、建立序列标签文库获得 SSR 引物序列等方法。其中,利用微卫星富集基因组文库构建法与 ISSR-PCR 方法筛选微卫星在实验室已非常广泛(Armour 等,1994;Fleischer & Loew, 1995; Fisher 等, 1996; Lench 等,1996;Matthew 等,1999)。微卫星富集基因组文库构建法(SNX linker 方法)主要是通过基因组酶切、接头(Adaptor)制备与连接、连接产物的扩增、产物与含重复序列的生物素探针杂交、磁珠富集及产物吸收预扩、克隆测序、引物设计来完成;而利用 ISSR 片段筛选 SSR 标记主要依据为 ISSR 是介于 SSR 之间的片段,其中长的 ISSR 片段本身就可能包含有 SSR 片段。通过对这样的片段筛选、克隆测序并进一步设计 SSR 侧翼区引物。

3 多引物组合使用——巢式 PCR 与分段 PCR

对于一些难于扩增的材料,比如长模板或模板 DNA 浓度过低的样品,常规 PCR 往往不能达到扩增目的片段的目的,可以采用巢式 PCR 和分段 PCR 来解决。前者是对 PCR 产物的再扩增,引物至少应有一个为内部引物,否则 PCR 扩增成功率会受影响。需要注意的是以 PCR 产物为模板,模板浓

度是关键,通常情况下,出现上拖带和下拖带分别暗示模板浓度过高或过低,需要调试,否则会达不到扩增的目的;后者使用外部引物和内部引物,内部引物和内部引物的分段扩增,然后再将各段序列拼接起来,如 *trnT-F* 用 *a* 和 *d*, *c* 和 *f* 分别扩增,然后再用 *a*、*d*、*c*、*f* 测序,其效果与用 *a* 和 *f* 扩增,然后再用这 4 个引物测序是相同的。这里的关键是需要有先设计足够的内部引物,各段序列之间应有重叠区域。对于 PCR 扩增产物条带较弱的,除调试模板之外,可以采用多管浓缩与克隆测序相结合的方法先获得序列,然后再设计合适的引物优化实验体系。

基因选择、引物设计与应用策略的好坏,是能否获得高质量 PCR 扩增的前提,只有解决这些问题,才能为后续的序列或遗传多态性分析提供详实可靠的科学数据和关键证据。

致谢 感谢李先琨主编盛情约稿,仅以此文纪念《广西植物》创刊三十周年。本文在撰写过程中,朱永青博士提供宝贵建议,王昊、廖辉提供部分参考资料,张宏彬协助绘图,特此致谢。

参考文献:

- Ainouche ML, Bayer R. 1997. On the origins of the tetraploid *Bromus* species (section *Bromus*, Poaceae): insights from internal transcribed spacer sequences of nuclear ribosomal DNA[J]. *Genome*, **40**: 730–743
- Armour JAL, Neumann R, Gobert S, et al. 1994. Isolation of human simple repeat loci by hybridization selection[J]. *Hum Mol Genet*, **3**: 599–605
- Azuma H, Thien LB, Kawano S. 1999. Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA sequences and evolutionary divergence of the floral scents[J]. *J Plant Res*, **112**: 291–306
- Baldwin BG. 1993. Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA; chromosomal and morphological evolution reexamined[J]. *Am J Bot*, **80**: 222–238
- Baldwin BG, Campbell S, Porter JM, et al. 1995. Utility of nuclear ribosomal DNA internal transcribed spacer sequences in phylogenetic analyses of angiosperms[J]. *Ann Mo Bot Gard*, **82**: 247–277
- Bartish IV, Kaderit JW, Comcs HP. 2006. Late Quaternary history of *Hippophaë rhamnoides* (Elaeagnaceae) inferred from chalcone synthase intron (Chsi) sequences and chloroplast DNA variation[J]. *Mol Ecol*, **15**: 4 065–4 083
- Baum DA, Small RL, Wendel JF. 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets[J]. *Syst Biol*, **47**: 181–207
- CBOL Plant Working Group. 2009. A DNA barcode for land plants[J]. *Proc Natl Acad Sci USA*, **106**: 12 794–12 797
- Caicedo AL, Schaal BA. 2004. Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene[J]. *Mol Ecol*, **13**: 1 871–1 882
- Campbell CS, Baldwin BG, Donoghue MJ, et al. 1995. A phylogeny of the genera of *Maloidae* (Rosaceae): evidence from sequences of the internal transcribed spacers of nuclear ribosomal DNA sequences and congruence with morphology[J]. *Am J Bot*, **82**: 903–918
- Chen HQ (陈怀琼), Sui C (隋春), Wei JH (魏建和), 2009. Summary of strategies for developing SSR primer (植物 SSR 引物开发策略简述)[J]. *Mol Plant Breed* (分子植物育种), **7**: 845–851
- Cifarelli RA, Gallitelli M, Cellini F. 1995. Random amplified hybridization microsatellites (RAHM): isolation of a new class of microsatellite-containing DNA clones[J]. *Nucleic Acids Res*, **23**: 3 802–3 803
- Davis CC, Wurdack KJ. 2004. Host-to-parasite gene transfer in flowering plants; phylogenetic evidence from Malpighiales[J]. *Science*, **305**: 676–678
- Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants[J]. *Mol Ecol*, **4**: 129–131
- Downie SR, Katz-Downie DS. 1996. A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences[J]. *Am J Bot*, **83**: 234–251
- Downie SR, Katz-Downie DS, Cho K-J. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpoCl* intron[J]. *Mol Phylogenet Evol*, **6**: 1–18
- Downie SR, Katz-Downie DS, Watson MF. 2000. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoCl* intron sequences; Towards a suprageneric classification of subfamily Apioideae[J]. *Am J Bot*, **87**: 273–292
- Elder JR, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes[J]. *Q Rev Biol*, **70**: 297–319
- Erdogan V, Mehlenbacher SA. 2000. Phylogenetic relationships of *Corylus* species (Betulaceae) based on nuclear ribosomal DNA ITS region and chloroplast *matK* sequences[J]. *Syst Bot*, **25**: 727–737
- Fisher PJ, Gardner RC, Richardson TE. 1996. Single locus microsatellites isolated using 5'-anchored PCR[J]. *Nucleic Acids Res*, **24**: 4 369–4 371
- Fleischer RC, Loew S. 1995. Construction and screening of microsatellite-enriched genomic libraries[M]//Ferraris J, Palumbi S (eds). *Molecular Zoology: Advances, Strategies and Protocols*. Wiley-Liss, New York: 459–468
- Francisco-Ortega J, Santos-Guerra A, Hines A, et al. 1997. Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae)[J]. *Am J Bot*, **84**: 1 595–1 613
- Fritsch PW, Morton CM, Chen T, et al. 2001. Phylogeny and biogeography of the Styracaceae[J]. *International J Plant Sci*, **162** (6 Suppl.): S95–S116
- Hamilton MB, Pincus EL, Di Fiore A, et al. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites[J]. *Bio Tech*, **27**: 500–507
- Heinze B. 2007. A database of PCR primers for the chloroplast genomes of higher plants[J]. *Plant Methods*, **3**: 4
- Hoot SB, Magallon S, Crane PR. 1999. Phylogeny of basal eudicots based on three molecular data sets: *atpB*, *rbcL*, and 18S nuclear ribosomal DNA sequences[J]. *Ann Mo Bot Gard*, **86**: 1–32
- Jin H, Shi SH, Pan HX, et al. 1999. Phylogenetic relationships between *Michelia* (Magnoliaceae) and its related genera based on the *matK* gene sequence[J]. *Acta Univ Sunyatseni: Sci Nat Edi*, **38**: 93–97
- Johnson LA, Soltis DE. 1994. *matK* DNA sequences and phyloge-

- netic reconstruction in Saxifragaceae[J]. *Syst Bot*, **19**:143—156
- Jordon WC, Courtney MW, Neigel JE. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae) [J]. *Am J Bot*, **83**:430—439
- Kandpal RP, Kandpal G, Weissman SM. 1994. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers[J]. *Proc Natl Acad Sci USA*, **91**:88—92
- Karagyozov L, Kalcheva ID, Chapman VM. 1993. Construction of random small-genomic libraries highly enriched for simple sequence repeats[J]. *Nucleic Acid Research* **21**:3 911—3 912
- Kelchner SA, Clark LG. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae) [J]. *Molecular Phylogenetics Evolution*, **8**:385—397
- Kelchner SA. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints[J]. *American J Bot*, **89**:1 651—1 669
- Kim S-T, Sultan SE, Donoghue MJ. 2008. Allopolyploid speciation in *Persicaria* (Polygonaceae): Insights from a low-copy nuclear region[J]. *Proceedings National Academy of Sci United States of America*, **105**:12 370—12 375
- Kollipara KP, Singh RJ, Hymowitz T. 1997. Phylogenetic and genomic relationships in the genus *Glycine* Willd. based on sequences from the ITS region of nuclear rDNA[J]. *Genome*, **40**:57—68
- Lee J-Y, Mummenhoff K, Bowman JL. 2002. Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* (Brassicaceae) [J]. *Proc Natl Acad Sci USA*, **99**:16 835—16 840
- Lench NJ, Norris A, Bailey A, et al. 1996. Vectorette PCR isolation of microsatellites repeat sequences using anchored dinucleotide repeats primers[J]. *Nucleic Acids Res*, **24**:2 190—2 191
- Levin MM, Thulin M, Labat J-N, et al. 2000. Africa, the odd man out; Molecular biogeography of *Dalbergioid legumes* (Fabaceae) suggests otherwise[J]. *Syst Botany*, **25**:449—467
- Liao P-C, Kuo D-C, Lin C-C, et al. 2010. Historical spatial range expansion and a very recent bottleneck of *Cinnamomum kanehirae* (Lauraceae) in Taiwan inferred from nuclear genes [J]. *BMC Evol Biol*, **10**:124
- Liang H, Hilu KW. 1996. Application of the *matK* gene sequences to grass systematics[J]. *Can J Bot*, **74**:125—134
- Lunt DH, Hutchinson WF, Carvalho GR. 1999. An efficient method for PCR-based identification of microsatellite arrays (PI-MA) [J]. *Mol Ecol*, **8**:893—894
- Manen J-F, Boulter MC, Naciri-Graven Y. 2002. The complex history of the genus *Ilex* (Aquifoliaceae); evidence from the comparison of plastid and nuclear DNA sequences and from fossil data [J]. *Plant Syst Evol*, **235**:79—98
- Manos PS, Steele KP. 1997. Phylogenetic analyses of “higher” Hamamelidae based on plastid sequence data[J]. *Am J Bot*, **84**:1 407—1 419
- Mason-Gamer RJ, Weil CF, Kellogg EA. 1998. Granule-bound starch synthase; structure, function, and phylogenetic utility[J]. *Mol Biol Evol*, **15**:1 658—1 673
- Meimberg H, Ditttrich P, Heubl G. 2001. Molecular phylogeny of Nepenthaceae based on cladistic analysis of plastid *trnK* intron sequence data[J]. *Plant Biol*, **3**:164—175
- Ostrander EA, Jong PM, Rine J, et al. 1992. Construction of small insert genomic DNA Libraries highly enriched for microsatellite repeat sequences[J]. *Proc Natl Acad Sci USA*, **89**:3 419—3 423
- Oxelman B, Yoshikawa N, McConaughy BL, et al. 2004. RPB2 gene phylogeny in flowering plants, with particular emphasis on asteroids[J]. *Mol Phylogenet Evol*, **32**:462—479
- Parkinson CL, Adams KL, Palmer JD. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants[J]. *Curr Biol*, **9**:1 485—1 488
- Posno M, van Vliet A, Grott GSP. 1986. The gene for *Spirodela oligorhiza* chloroplast ribosomal protein homologous to E. coli ribosomal protein L16 is split by a large intron near its 5' end; Structure and expression[J]. *Nucleic Acids Res*, **14**:3 181—3 195
- Qiu YL, Lee J, Bernasconi-Quadroni F, et al. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes[J]. *Nature*, **402**:404—407
- Sanjurjo OI, Andres TC, Wessel-Beaver L. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene; Implications for crop plant evolution and areas of origin[J]. *Proc Natl Acad Sci USA*, **99**:535—540
- Sang T, Donoghue MJ, Zhang D. 1997. Evolution of alcohol dehydrogenase genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species[J]. *Mol Biol Evol*, **14**:994—1 007
- Schaefer H, Renner SS. 2010. A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia [J]. *Mol Phylogenet Evol*, **54**:553—60
- Soltis DE, Soltis PS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis [M]//Soltis DE (eds). *Molecular Systematics of Plants II. DNA Sequencing*. Kluwer Academic Publishers, 1—42
- Sweeney PW, Price RA. 2000. Polyphyly of the genus *Dentaria* (Brassicaceae): Evidence from *trnL* intron and *ndhF* sequence data[J]. *Syst Bot*, **25**:468—478
- Taberlet P, Gelly L, Pautou G, et al. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA[J]. *Plant Mol Biol*, **17**:1 105—1 109
- Wang Y, Fritsch PW, Shi S, et al. 2004. Phylogeny and infrageneric classification of *Symplocos* (Symplocaceae) inferred from DNA sequence data[J]. *Am J Bot*, **91**:1 901—1 914
- White TJ, Bruns T, Lee S, et al. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics [M]//Innis MA (eds). *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego, 315—322
- Wilckström N, Kenrick P. 2000. Relationships of *Lycopodiella* based on combined plastid *rbcL* gene and *trnL* intron sequence data[J]. *Syst Bot*, **25**:495—510
- Yen AC, Olmstead RG. 2000. Molecular systematics of Cyperaceae tribe Cariceae based on two chloroplast DNA regions: *ndhF* and *trnL* intron-intergenic spacer[J]. *Syst Bot*, **25**:479—494
- Yuan Y-M, Kupfer P, Doyle JJ. 1996. Infrageneric phylogeny of the genus *Gentiana* inferred from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA[J]. *Am J Bot*, **83**:641—652
- Zane L, Bargelloni L, Patarnello T. 2002. Strategies for microsatellite isolation; a review[J]. *Mol Ecol*, **11**:1—16
- Zhang X (张新宇), Gao Y (高燕宇). 2004. To design PCR primers with Oligo 6 and primer 5 (PCR 引物设计及软件使用技巧) [J]. *China J Bioinformatics (生物信息学)*, **2(4)**:15—18, 46
- Zhang ZC (张增翠), Hou XL (侯喜林). 2004. Strategies for development of SSR molecular markers (SSR 分子标记开发策略及评价) [J]. *Hereditas (遗传)*, **26**:763—768