

植物DNA条形码研究现状及应用前景

唐建阳,周先治

(福建省农科院农业生物资源研究所,福州 350003)

摘要:生物DNA条形码已成为近年来生物学研究的热点,该技术在动物研究中已得到广泛应用,在植物中研究进展缓慢,植物DNA条形码筛选区域主要集中在叶绿体和核糖体内转录间隔区(ITS),目前还未获得广泛认同的植物DNA条形码。主要综述了不同单片段条形码及组合条形码的研究及应用现状,并展望了植物DNA条形码应用前景。

关键词:植物DNA条形码;研究现状;应用前景

中图分类号:S321

文献标识码:A

论文编号:2009-1956

Progress and Outstanding of DNA Barcoding in Plant

Tang Jianyang, Zhou Xianzhi

(Institute of Agrobiological Resources, Fujian Academy of Agricultural Sciences, Fuzhou 350003)

Abstract: DNA barcode of life has become one of hotspots of biology research in the recent years. DNA barcoding is now well established for animals. However, DNA barcoding in plants is going on slowly. And the search for an analogous region to plant has focused on chloroplast DNA and the internal transcribed spacers of nuclear ribosomal. The quest for a universal DNA barcode in plants is still disputed. In this paper, the advance of these barcoding regions and their combinations are reviewed, the advantages, criterion, drawbacks, and existed dispute, and the prospect of application of DNA barcoding in plant are summarized.

Key words: DNA barcoding in plants, research status, prospect of application

0 引言

2003年加拿大动物学家Paul Hebert首次将“DNA条形码”引入生物界^[1]。2003年3月和9月在美国冷泉港召开的两次国际会议拟定了国际生物条形码计划(International Barcode of Life Project)的发展蓝图^[1-2]。2004年6月成立生物条形码联盟(Consortium for the Barcode of Life, CBOL),该联盟计划在接下来20多年中建立真核生物条形码数据库,至2012完成涵盖鱼类和鸟类所有物种的条形码^[3]。截至2009年12月已有50多个国家的170多个研究单位参与其中^[4]。2007年5月10日,世界上第一个DNA barcoding鉴定中心在加拿大University of Guelph成立。生物条形码网络数据系统www.barcodinglife.org、鱼类条形码数据库www.fishbol.org和鳞翅目条形码数据库www.lepbarcoding.org^[5]相继建立。

org^[5]相继建立。

许多的研究表明线粒体细胞色素c氧化酶亚基I(Cytochrome c oxidase I, CO I)基因可以作为生物条形码^[6-8]。由于植物线粒体基因组进化速率远慢于叶绿体基因组和核糖体基因组^[9],因此CO I基因不适合作为大多数植物条形码基因。近年来在植物DNA条形码方面进行了诸多研究^[10-12],遗憾的是至今尚未获得广泛认同条形码标准片段^[13-14]。笔者主要综述了植物DNA条形码概念、应用范围和意义及标准,植物DNA条形码研究的现状和存在的问题,并展望了植物DNA条形码研究的前景,以期为国内相关工作者提供些许资料。

1 植物DNA条形码概念、应用意义及标准

植物DNA条形码就是利用一个或少数几个标准

基金项目:福建省科技计划项目“福建中药材种质资源标准化整合研究与示范”(2009R10037-2);福建省产业技术开发项目“名贵南药福建春砂仁产业化生产关键技术研究开发”。

第一作者简介:唐建阳,男,1961年出生,副研究员,从事作物良种资源研究与综合利用方面的工作。通信地址:350003福建省福州市五四北路福建省福州市五四北路247号福建省农科院农业生物资源研究所。E-mail: tjy836@163.com。

收稿日期:2009-09-22,修回日期:2009-12-03。

的DNA片段快速、准确识别和鉴定植物种类^[1]。应用植物DNA条形码可以准确、快速大规模的鉴别植物,不受植物个体形态、生长阶段限制。植物DNA条形码数据库的建立,可以加快已知物种的识别速度和新物种的发现;如果DNA条形码扫描仪能实现,将大大减少对植物分类的人力和物力的投入,加快植物分类进展,植物DNA条形码利用DNA序列进行物种鉴定,与限制片段长度多态性(Restriction Fragment Length Polymorphism, RFLP)、随机扩增多态性DNA(Random Amplified Polymorphic DNA, RAPD)、随机引物聚合酶链反应(Arbitrarily Primed Polymerase Chain Reaction, AP-PCR)以及基因芯片技术相比,具有极高的可重复性,并且所需要的引物数量及投入的人力、物力、时间显著减少。

理想的植物DNA条形码应当符合以下四个标准:(1)种间差异显著,种内差异小;(2)条形码要标准化,必须能用相同的DNA片段区分不同的分类类群;(3)片段大小300~800 bp^[1],高度可靠易于基因分离和PCR扩增,尤其是对于存在DNA降解的材料;(4)序列两侧翼保守便于引物设计^[15]。

2 植物DNA条形码候选片段及组合

叶绿体单亲遗传,避免重组,同时包含许多变异区域,核糖体内转录间隔区(Internal Transcribed Spacers, ITS)是最常用的植物分子系统调查序列^[11],植物叶绿体和核糖体的进化速率快于线粒体^[9],因此植物学家推荐的潜在条形码区域主要集中在核糖体间隔区和叶绿体基因组的编码和非编码区域^[10-12,16]。生物条形码联盟(Consortium for the Barcode of Life, CBOL)植物工作组最初建议的植物条形码片段均为叶绿体片段:成熟酶K蛋白(maturase K protein, *matK*),核糖核酸聚合酶C1亚基(Ribonucleic acid polymerase C1' Subunit, *rpoC1*),核糖核酸聚合酶B亚基(Ribonucleic acid polymerase B' Subunit, *rpoB*),乙酰辅酶A水解酶(beta subunit of acetyl-coenzyme A carboxylase, *accD*),还原型烟酰胺腺嘌呤二核苷酸脱氢酶J亚基(Nicotinamide adenine dinucleotide dehydrogenase subunit J, *ndhJ*)和预测叶绿体阅读框5(Hypothetical chloroplast reading frame 5, YCF5),推荐这些基因作为条形码的原因在于这些基因存在用通用引物扩增和种间具有足够差异性的潜能,可以单独或组合作为条形码来识别物种,后来研究发现*accD*在禾本科植物中缺失,*ndhJ*在松属植物中缺失,在部分兰花中变短或功能丧失,*YCF5*在苔藓类植物中缺失,因此在第二阶段更新中被排除^[17]。Kress等^[11]的研究显示两个非编码区域(核糖

体内转录间隔区ITS和质体[18]以及质体[19]作为条形码。由于单一基因或片段未能彻底解决植物DNA条形码问题。2007年Kress等推荐1,5-二磷酸核酮糖羧化酶(*ribulose-1,5-bisphosphate carboxylase, rbcL*)基因和质体[20]。同年Chase等^[21]提出两套组合方案:*rpoC1+rpoB+matK*或*rpoC1+matK+trnH-psbA*。Lahaye等^[22]提出采用*matK*基因单独或者与matK+atpF-atpH+psbK-psbI和*matK+atpF-atpH+trnH-psbA*两个组合条形码^[23]。根据植物DNA条形码的标准,对单片段和多片段组合候选植物DNA条形码的研究现状及存在问题等分别进行介绍,单片段扩增常用的部分引物序列见表1。

3 植物DNA条形码的研究现状

3.1 单片段

3.1.1 *matK* 与叶绿体其他编码基因相比*matK*基因的进化速率快,*matK*片段大小合适,种间差异度高和碱基转换/颠换数低^[21-22],Labaye等^[24]推荐该基因作为植物DNA条形码,Labaye等^[24]采用Cuénoud等^[25]设计的一对针对*matK*基因的引物扩增1667个植物材料的成功率达到100%,直到现在其他条形码的引物的仍未达到这一水平。但Kress等^[26]对此提出异议,原因是他们扩增材料96%都是兰科植物,其他科植物是否有这么高的扩增成功率还有待验证。该基因不同分支类群很难进行扩增和测序,引物的通用性很差^[21,27],Fazekas等^[28]采用10对引物对32属92种植物的251个个体的扩增成功率达到87.6%,*matK*单片段对物种的正确识别率仅达到56%,其中25.5%的物种测序的双向覆盖度小于80%。Sass等^[29]采用生物条形码联盟植物工作组(Plant Working Group of the Consortium for the Barcode of Life, PWG-CBOL)建议的*matK*引物扩增苏铁目植物的成功率仅达到24%,Kress等^[20]对48属96种植物的成功扩增率仅为39.3%,正确识别率为14.6%。这些年,在开发该片段通用引物方面进行了大量工作,但至今未取得理想的结果^[29]。因此*matK*基因通用性引物的设计和效果验证将是下一步工作重点。*matK*基因已有的部分引物见表1。

3.1.2 *rpoC1* Sass等^[29]采用生物条形码联盟植物工作组(Plant Working Group of the Consortium for the Bar-

表1 植物DNA条形码片段的引物序列

| 基因或间隔区 | 引物 | 方向 | 序列5'-3' |
|-----------------------|--|-----------------------------|-----------------------------|
| <i>matK</i> | 2.1 ^[17] | f | CCTATCCATCTGGAAATCTTAG |
| | 2.1a ^[17] | f | ATCCATCTGGAAATCTTAGTTC |
| | 5 ^[17] | r | GTTCTAGCACAAAGAAAGTCG |
| | 3.2 ^[17] | r | CTTCCTCTGTAAAGAATTC |
| | Agiosperms_KIM* | f | ATCCATCTGGAAATCTTAGTTC |
| | | r | GTTCTAGCACAAAGAAAGTCG |
| | Plants_KIM* | f | CRATCWATTCATTCAATATT |
| | | r | CGTACAGTACTTTTGTGTTT |
| | <i>matK</i> -KEW ^[17] | f | AATATCCAAATACCAAATCC |
| | | r | ACCCAGTCCATCTGGAAATCTTGGTTC |
| | 390F ^[26] | f | CGATCTATTCATTCAATATTC |
| | 1326R ^[26] | r | TCTAGCACACGAAAGTCGAAGT |
| | X ^[20] | f | TAATTTACGATCAATTCATTC |
| | F (<i>Equisetum</i>) ^[17] | f | ATACCCCAITTTATTCATCC |
| | R (<i>Equisetum</i>) ^[17] | r | GTACTTTTATGTTTACGAGC |
| | F (<i>Adiantum</i>) ^[17] | f | GATGTTGCAGTCTATTCATTC |
| | pkF1 ^[28] | f | TTTCTGATGAACAARTGGAA |
| | pkF3 ^[28] | f | CTACGATACTGGGTNAARGA |
| | pkF4 ^[28] | f | CCCTATTCTATTCAAYCCNGA |
| | pkF5 ^[28] | f | CCTCATTTTATTCAAYCCNGA |
| | pkF7 ^[28] | f | CTAATACCCTACCCNATHCA |
| | pkR1 ^[28] | r | CGTATCGTGCTTTTRIGYTT |
| 3F_KIM* | | CGTACAGTACTTTTGTGTTTACGAG | |
| 1R_KIM* | | ACCCAGTCCATCTGGAAATCTTGGTTC | |
| <i>rpoC1</i> | 1 ^[17] | f | GTGGATACACTTCTTGATAATGG |
| | 2 ^[17] | f | GGCAAAGAGGGAAGATTTCG |
| | 3 ^[17] | r | TGAGAAAACATAAGTAAACGGGC |
| | 4 ^[17] | r | CCATAAGCATATCTTGAGTTGG |
| | LP1 ^[17] | f | TATGAAACCAGAATGGATGG |
| | LP5 ^[17] | r | CAAGAAGCATATCTTGASTYGG |
| | ajfMossF ^[38] | f | GGCAAAGAAGGACGTTTTTCG |
| | ajfMossR ^[38] | r | CCAGAAGCATATCTTGACTTGG |
| <i>rpoB</i> | 1 ^[17] | f | AAGTGCATTGTTGGAACCTGG |
| | 2 ^[17] | f | ATGCAACGTCAAGCAGTTCC |
| | 3 ^[17] | r | CCGTATGTGAAAAGAAGTATA |
| | 4 ^[17] | r | GATCCCAGCATCACAAATCC |
| | LP1.1 ^[17] | f | TCTAATATGCARCGTCAAGG |
| | LP3 ^[17] | r | TTTACCCAAYRAAACATCHCC |
| | LP4.3 ^[17] | r | ATAATACCTTTATTWCCATG |
| | LP5.2 ^[17] | r | AAATAAGGCATATCTTGCTCT |
| | ajfF1 ^[29] | f | TCTAATATGCAICGTCAAGC |
| ajfR1 ^[29] | r | GAGGIGTTAITTTACCTAC | |

(续表 1)

| 基因或间隔区 | 引物 | 方向 | 序列 5'-3' |
|--|--|----|-----------------------------|
| <i>rbcL-a</i> | <i>rbcL-a</i> F ^[11] | f | ATGTCACCACAAACAGAGACTAAAGC |
| | <i>rbcL</i> .ajf634R ^[29] | r | GAAACGGTCTCTCCAACGCAT |
| | <i>rbcL-a</i> R ^[11] | r | CTTCTGCTACAAATAAGAATCGATCTC |
| <i>cox1</i> | 42F ^[28] | f | GGATCTTCTCCACTAACCACAA |
| | ajf699R ^[39] | r | CCGAAAGAGATGCTGGTATA |
| 23S rDNA | p23SrV F1 ^[18] | f | GGACAGAAAGACCCTATGAAGCTT |
| | p23SrV R ^[18] | r | TCAGCCTGTTATCCCTAGAGTAAC |
| <i>trnL</i> | c ^[40] | f | CGAAATCGGTAGACGCTACG |
| | d ^[40] | r | GGGGATAGAGGGACTTGAAC |
| | g ^[19] | f | GGGCAATCCTGAGCCAA |
| | h ^[19] | r | CCATTGAGTCTCTGCACCTATC |
| <i>trnK-rps16</i> | <i>trnK5'r</i> ^[41] | | TACTCTACRRTTGAGTTAGCAAC |
| | <i>rps16-4547mod</i> ^[41] | | AAAGGKGCTCAACCTACARGAAC |
| <i>trnH-psbA</i> | <i>trnH</i> ^[36] | f | CGCGCATGGTGGATTCAAAATCC |
| | <i>psbA3</i> ^[42] | r | GTTATGCATGAACGTAATGCTC |
| <i>trnH-psbA</i> (including protein coding region) | <i> fwd</i> ^[11] | f | CGAGCCTGTTTCTGGTTCTC |
| | <i> rev</i> (short-fragment) ^[11] | r | GGGGTGTGGGTAGAGCAGT |
| | <i> rev</i> (long-fragment) ^[11] | r | CCGACGACGAACTAACATTTG |
| <i>atpF-atpH</i> | <i> atpF</i> * | f | ACTCGCACACTCCCTTTCC |
| | <i> atpH</i> * | r | GCTTTTATGGAAGCTTAAACAAT |
| <i>rpl36-infA-rps8</i> | <i> rpl36f</i> ^[11] | f | CACAAATTTTACGAACGAAG |
| | <i> rps8r</i> ^[11] | r | TAATGACAGAYCGAGARGCTCGAC |
| <i>trnV-atpE</i> | <i> trnV5f</i> ^[11] | f | GTGTAACGAGTTGCTCTACCA |
| | <i> S1022</i> ^[43] | r | CGACATTTGCACATTTAGATGCTAC |
| <i>trnC-ycf6</i> | <i> trnC</i> ^[44] | f | CCAGTTCAAATCTGGGTGTC |
| | <i> petN1r</i> ^[45] | r | CCCAAGCAAGACTTACTATATCC |
| <i>ycf6-psbM</i> | <i> petN1</i> ^[45] | f | GGATATAGTAAGTCTTGCTTGGG |
| | <i> psbM2r</i> ^[45] | r | TTCTTGCATTTATTGCTACTGC |
| <i>psbM-trnD</i> | <i> psbM1</i> ^[45] | f | GCGGTAGGAACTAGAATAAATAG |
| | <i> trnD</i> ^[41] | r | GGGATTGTAGTTCAATTGGT |
| <i>atpB-rbcL</i> | <i> S2r</i> ^[41] | f | AGAAGTAGTAGGATTGATTCTCATA |
| | <i> RBCL1</i> ^[41] | r | GAATCCAACACTTGCTTTAGTCTCT |
| <i>trnL-F</i> | e ^[40] | f | GGTTCAAGTCCCTCTATCCC |
| | f ^[40] | r | ATTTGAACTGGTGACACGAG |
| <i>rbcL</i> | 1f ^[46] | f | ATGTCACCACAAACAGAAAC |
| | 724r ^[46] | r | TCGCATGTACCTGCAGTAGC |
| <i>psbK-psbI</i> | <i> psbK</i> * | f | TTAGCCTTTGTTTGCCAAG |
| | <i> psbI</i> * | r | AGAGTTTGAGAGTAAGCAT |
| <i>nrITS</i> | <i> ITS5a</i> ^[47] | f | CCTTATCATTTAGAGGAAGGAG |
| | <i> ITS4</i> ^[48] | r | TCCTCCGCTTATTGATATGC |

注: *Ki-Joong Kim 未公开发表引物。

code of Life, PWG—CBOL) 建议的 *rpoC1* 片段特异引物(表 1) 扩增苏铁目植物的成功率达到 100.0%, 没有非

特异性扩增, 该片段适于做条形码。而 Chase 等^[21]和 Lahaye 等^[24]则认为该片段不适宜做植物 DNA 条形码,

因为该片段在被子植物中非常保守。Fazekas等^[28]采用3对引物(表1)对32属92种植物的251份样品材料的测序成功率达到94.8%,但*rpoC1*单片段对所测物种的正确识别率仅为29%。*rpoC1*片段种间差异度偏小是其作为条形码的最大障碍,由于该片段易于扩增,引物通用性好,因此可考虑和其他片段组合作为条形码。

3.1.3 *rpoB* Sass等^[29]的研究发现*rpoB*不适于作条形码,采用PWB-CBOL推荐引物(表1)扩增苏铁目植物的成功率为33%,非特异性扩增达到67%。而Chase等^[21]和Lahaye等^[24]研究同样发现该片段不适于做植物DNA条形码,在被子植物中非常保守。

3.1.4 *rbcL* *rbcL*被提议作为植物DNA条形码是因为该片段在GeneBank中有近50 000条序列,该片段易扩增和比对。Lahaye等^[24]比较了不同的植物DNA条形码片段,发现该片段对物种的区分度比*matK*和*trnH-psbA*小。*rbcL*全长1300 bp左右,需使用4对引物双向测序才能获得全长,而作为条形码片段长度在300~800 bp^[11]。Kress等^[23]选取其中的一段*rbcL-a*进行扩增,设计引物(表1)通用性好,扩增成功率达90%以上,单片段在属水平上的区别率达到69.8%。Fazekas等^[28]用2对引物(表1)对32属92种植物的251份样品材料的扩增成功率达到100.0%,该片段对物种的正确识别率为48%,其他许多研究也发现*rbcL*在种水平的差异不够大,不适于单独作条形码^[11-12,30]。

3.1.5 *trnH-psbA* *psbA*在叶绿体的反向重复序列中位置变动或重复导致*trnH-psbA*间隔区长度或拷贝变异较大,这也给非同属植物的比对带来一定困难^[31-33]。Kress等^[11]分析的53科80属99种植物中*trnH-psbA*的扩增长度为247~1221 bp,间隔区为119~1094 bp,扩增成功率达到100.0%,在所选取的10个片段中序列属水平差异度最高,92%的扩增片段在340~660 bp之间,99种植物均具有独特的间隔序列,符合理想条形码标准。Kress等^[20]以及Fazekas等^[28]的研究结果也表明*trnH-psbA*在备选片段中的扩增成功率和物种正确识别率都是最高的。在近缘种的区分上*trnH-psbA*也是表现最好的,*trnH-psbA*对兰科植物的识别率达90%以上^[24],对内毛茛属植物的识别达到70%^[34]。但Sass等^[29]采用Kress等^[11]推荐的引物(表1)在苏铁目植物中除苏铁*Cycas*外均扩增出两条带,*trnH-psbA*间隔区不适于作为苏铁目植物的条形码。在伞形科独活属和禾本科甜茅属植物中,*trnH-psbA*片段变异度不够不足以识别物种^[35-36]。

3.1.6 nrITS 核糖体内转录间隔区ITS(18S-5.8S-26S)是植物分子鉴定中最常用的片段^[35]。该片段在光合真

核生物植物(除蕨类植物)和真菌中广泛存在,被视为植物条形码的备选片段^[11,36]。ITS是最常用作种水平区分的片段,截至2004年10月在GenBank中被被子植物的ITS序列超过36 000条^[11]。ITS可以扩增出ITS1和ITS2两个连接5.8S的短小片段,这有利于对降解材料的识别。在种水平上ITS1比ITS2的识别效果好^[10]。ITS区段进化速率快,同源性比对复杂,5.8S片段非常保守,ITS近缘性比对不易区分或没有匹配的情况下,可以比对5.8S片段,限定比对范围^[11]。Sass等^[29]研究发现ITS在属水平上对苏铁目植物的80.9%,适于作为苏铁目植物的条形码,不过由于序列存在长的Poly-G、Poly-C和Poly-A,导致测序成功率低。由于核基因组具有多拷贝特性以及二级结构问题导致该片段扩增、测序、比对困难^[35,37]。Kress等^[20]扩增48属96种植物发现ITS1片段的扩增成功率为60.4%(扩增引物见表1),对所有研究材料的正确识别率为45.8%,种内变异大,进一步降低了该片段作为条形码的可能性。

3.1.7 其他片段 Kress等^[11]研究发现*rp136-rps8*、*trnK-rps16*、*trnV-ycf6*、*psbM-trnD*、*atpB-rbcL*、*trnL-F*这8个片段与ITS以及*trnH-psbA*相比在种间的差异度偏小,不适于作条形码。Taberlet等^[19]分析了叶绿体内含子*trnL*及其P6环的部分片段,发现这两个片段种间差异度偏小,对物种识别率低,不适于作条形码^[18-20]。Fazekas等^[28]研究发现23S rDNA片段(UPA)、*cox1*对32属92种251份植物的正确识别率分别为7%和10%,不合作条形码,*psbK-psbI*片段的正确识别率为44%,可以作为组合条形码使用。

3.2 多片段组合

许多研究表明单片段识别率低,不符合条形码的标准^[11-12,20-21,24,28-29],在筛选条形码时不应只关注单片段,还可考虑多片段组合。Newmaster等^[12]在分析了GenBank中10 000条*rbcL*序列在属水平上区分度接近85%,因此可以把该片段作为条形码的核心区段(core region),然后根据不同的类群区分需要选择二级区段进一步分析。Kress等^[20]分析发现*trnH-psbA*、*rbcL-a*、*rpoC1*、*ITS1*、*rpoB2*、*accD*、*ycf5*、*ndhJ*、*matK*等9个片段对48属96种植物的正确识别率都低于80%,组合条形码*trnH-psbA+rbcL*、*trnH-psbA+rpoB2*、*trnH-psbA+rpoC1*的正确识别率提高到88%,采用Blastn比对GenBank中的序列数据发现,*rbcL-a*单位点种水平的正确识别率达到76.3%,*trnH-psbA*的正确识别率为83%,组合条形码的*rbcL-a+trnH-psbA*的正确识别率为95.0%,*rpoB*和*rpoC*单位点识别率较低,因此推荐*trnH-psbA+rbcL*为最佳组合条形码。Fazekas等^[28]也证实两片段组合

中这对组合的扩增和识别效果最好,同时还提出条形码的识别能力与组合数目相关。Chase 等^[21]在分析单片段条形码优缺点的基础上针对陆生植物提出了 *rpoC1+rpoB+matK* 或 *rpoC1+matK+psbA-trnH* 的条形码组合。Kim 等^[23]在第二届国际生物条形码大会上提出了 *matK+atpF-atpH+psbK-psbI* 或 *matK+atpF-atpH+trnH-psbA* 两个条形码组合。Fazekas 等^[28]采用 32 属 92 种 251 份植物材料对对这 5 个组合条形码分析显示在 5 个组合的正确识别率在 61%~69% 之间,都不是很高,没有特别理想的组合。Seberg 等^[49]研究了不同组合条形码对番红花属 *Crocus* 86 种植物的鉴定情况,发现组合条形码 *ndhF+matK+trnH-psbA+rps8-rpl36* 可以区分出其中的 79 种,正确识别率达到 92%,区分效果最好。Fazekas 等^[28]建议多片段联合时,编码片段在 *rpoC, rpoB, matK* 中选择,而非编码片段则在 *trnH-psbA* 和 *atpF-atpH* 中选择。

4 植物 DNA 条形码在植物学研究中的应用

4.1 植物 DNA 条形码在植物分类学研究中的应用

植物 DNA 条形码技术可以用来快速鉴定植物样品和新植物物种的发现。DNA 分子的三个特点决定了它可作为分子鉴定有力工具。第一, DNA 极其稳定可长期保存,即使是在干植物组织中也可以提取到 DNA,许多有效、简便的提取和保存 DNA 技术得到了广泛应用^[50-52];第二, DNA 存在于植物组织或果实的核细胞中,也存在于非核细胞的质体或线粒体中,几乎所有类型的植物材料都可以用来分析,例如植物种子;第三,由于蛋白比 DNA 易衰变,并且核酸序列具有大量的非编码区,核酸序列比蛋白能提供更多的信息。植物 DNA 条形码鉴定可以克服传统形态鉴定的缺陷,第一,由于研究特征的表型易变或者是隐秘类群的存在而造成的错误;第二,形态鉴定只适用于植物生长的特定生长阶段。Nitta^[53]利用 *rbcL* 和 *trnH-psbA* 组合条形码成功鉴定出膜蕨新种 *Polyphlebium borbonicum*, Kress 等^[11]研究发现采用 ITS 和 *trnH-psbA* 组合条形码可以区分显花植物。

4.2 植物 DNA 条形码在植物多样性评估研究中的应用

植物 DNA 条形码技术将来可能在植物群落多样性评估中扮演重要角色。在植物多样性评估中,植物 DNA 条形码技术比传统形态分类更快捷、更廉价,并且可以克服传统形态分类的缺陷。利用植物 DNA 条形码技术不仅可以评估传统的植物群落,例如藏北高寒草原,还可以用来评估植物多样性异常丰富的热带雨林,这些植物群落利用传统的形态分类方法鉴定在

有限的时间内几乎是不可能完成的。植物 DNA 条形码技术可以用来大规模的植物多样性评估,因为条形码可以同时对于一个给定的生境小区内大多数物种进行鉴定。目前,植物 DNA 条形码技术在植物多样性评估中还未曾应用。

4.3 植物 DNA 条形码在古植物生物学研究中的应用

由于古植物化石保存数量较低,保存完整性较低,因此通过植物化石数据来重现古植物群落是非常困难的事情。大多数情况下,通过形态鉴定到种水平是非常困难,甚至是不可能的,在这种情况下,植物 DNA 条形码技术可以帮助科学家重现古生物群落和重建过去的生境。

只有在干燥或者是寒冷的环境中分子保存较好的情况下,适于利用植物 DNA 条形码技术研究古植物生物学。从收集的更新世和全新世时期的西伯利亚冰川冻土来分析古生物群落,发现在这两个时期的植物组成发生了变化^[54]。距今 450 000 年的格陵兰冰盖底部的淤泥冰抽提样品分析发现,南格陵兰在那时候被森林覆盖,森林群落主要由 *Picea*、*Pinus* 和 *Alnus* 三个属的植物构成,这三个属的植物在现在的南斯堪的纳维亚森林中有发现^[55]。

4.4 植物 DNA 条形码在植物系统发育研究中的应用

植物 DNA 条形码技术获得的标准条形码序列不仅可以用于植物的分类鉴定,还可以用于植物的系统发育分析。基于多标记的组合条形码的植物系统发育分析比单位点更可靠。李洪涛等^[56]基于 ITS、*rpl16* 和 *trnS-trnR* DNA 序列探讨了锥形果属的系统位置,发现锥形果属是一个自然属,雪胆属 *Hemsleya* 的短柄雪胆 *H. delavayi* 和圆锥果雪胆 *H. macrocarpa* 曾经被作为锥形果属的种,分子证据表明它们确实隶属于雪胆属,锥形果属单独构成雪胆属的姊妹群,而非是与绞股蓝属 *Gynostemma* 共同构成。张金梅等^[57]基于叶绿体基因组 *ndhF*、*rps16-trnQ*、*trnL-F* 和 *trnS-G* 四个基因以及核基因组 GPAT 基因等 DNA 条形码研究了芍药属牡丹组物种系统发育关系,结果发现四川牡丹、矮牡丹、卵叶牡丹和紫斑牡丹在内的进化线与根据形态所建立的物种限定不吻合。出现这种情况的原因是因为这些居群系统之间存在一定程度的地理隔离,但不存在生殖隔离,出现这种两个基因组数据背离的现象可能是由于不同居群系统在进化历史中捕获了其他居群系统的叶绿体基因组。

5 植物 DNA 条形码研究存在的问题及对策

植物中很难找到一个像动物中 COI 序列一样在物种鉴定和新物种发现中广泛通用的 DNA 条形码的

有两方面的原因:其一植物线粒体的进化速率远慢于叶绿体或核糖体,变异度不够;其二植物DNA条形码所选择的两个区域主要是叶绿体和核糖体区域,叶绿体基因的进化速率大约只有核基因的四分之一,比哺乳动物的线粒体基因的进化速率低10~20倍^[9,58]。DNA条形码选择区域过窄,也是造成植物DNA条形码研究进展缓慢的重要原因,可以通过大规模的测序比对分析,选择其他新的片段作为条形码,筛选其他的组合条形码。

由于叶绿体具有单亲遗传特性,不存在多拷贝的问题,因此植物DNA条形码推荐序列主要集中在叶绿体上,叶绿体基因虽然测序简便,但在某些情况下(例如木材和枯死植物)得不到叶绿体DNA,这也限制了其应用范围。不过这可以通过拓宽植物DNA条形码筛选区域,来解决该问题。

在下列两种情况下,单序列DNA条形码有可能造成错误识别,一种情况是杂交和网状进化有可能造成不同物种带有相同叶绿体或线粒体基因组,还有一种情况是基因组之间存在遗传物质交流,这导致单序列不能准确识别这些物种,这可以通过筛选组合条形码序列加以解决。

Fazekas等^[28]认为15%的单系分析失败不是因为没有足够差异造成的,而是因为在不同的“并系”分析进化树中没有足够的比对信息造成的。这需要加大、加快植物样品测序规模,尤其是同属内不同近缘种的序列测序,在大规模数据分析的基础上,筛选出更合适的条形码片段或组合条形码,同时植物DNA条形码也可以借鉴生物形态分类学方法,在门、纲、目、科、属及种不同的分类单元中筛选出不同级别的条形码。

6 植物DNA条形码应用前景

近年来,国内许多学者也开始关注该项技术,陈士林等^[59]、宁淑萍等^[60]分别就植物DNA条形码的相关研究及应用前景进行了综述。石林春等^[61]通过筛选分析,发现ITS2片段适合作为杜鹃属植物的DNA条形码。福建省农科院农业生物资源研究所收集了大量的植物种质资源,本研究所正在开展植物DNA条形码筛选和应用研究,发现在所有条形码中,ITS的区分效果最好,笔者利用ITS序列研究了南瓜属、西瓜属和葫芦属作物的系统发育,发现中国南瓜与美洲南瓜的亲缘关系较近,中国南瓜与黑籽南瓜的亲缘关系较远,中国南瓜种下分化与地理来源有一定关系,西瓜具有地理分化现象,非洲来源的西瓜和来自中国的西瓜栽培种分属不同分支,葫芦属作物同样也存在地理分化现象(待发表)。笔者所在研究所准备针对药用植物,筛选

适合的DNA条形码,将该技术应用于药用植物的快速鉴定和系统分化研究。

尽管植物DNA条形码的研究进展缓慢,不及动物条形码那样应用广泛,存在诸多争议,但随着信息技术的发展和测序技术的改进,植物DNA条形码的研究将日臻完善,为植物的分类鉴定及新物种的发现提供快捷、准确的信息,加快植物分类学的研究进展,成为植物分类学学家有益的辅助工具;在植物多样性评估中,植物DNA条形码可以作为一种新的植物多样性评估标记,作为传统的生物多样性标记如物种丰富度、Simpson's指数和Shannon's指数的一种互补标记,尤其在古植物生物学研究中将会得到广泛应用,可以通过分析化石样本或其他样本材料的总DNA中某种类型的DNA序列的相对数量来替代传统的生物多样性评估标记分析;植物DNA条形码技术还可以应用于外来入侵植物的风险评估以及外来物种的识别等诸多方面。植物DNA条形码应用前景广阔。

参考文献

- [1] Hebert P D N, Cywinska A, Ball SL, et al. Biological identification through DNA barcodes [J]. Proceedings of the Royal Society of London[J]. Series B,Biological Sciences,2003,270:313-321.
- [2] Hebert P D N, Ratnasingham S, de Waard J R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species [J]. Proceedings of the Royal Society of London. Series B, Biological Sciences,2003,270(S):S96-S99.
- [3] Marshall E. Will DNA bar codes breathe life into classification?[J]. Science,2005,307:1037.
- [4] Consortium for the barcode life [EB/OL].[2009-12-03].http://www.barcoding.si.edu/.
- [5] Ratnasingham S, Herbert P D N. BOLD: The barcode of life data system (www.barcoding.org) [J].Molecular Ecology Notes,2007,7: 355-364.
- [6] Halibabaei M, Janzen D H, Burns J M, et al. DNA barcodes distinguish species of tropical Lepidoptera [J]. Proceedings of the National Academy of Sciences, USA, 2006,103:968-971.
- [7] Yancy H F, Zemlak T S, Mason J A, et al. Potential use of DNA barcodes in regulatory science: Applications of the regulatory fish encyclopedia [J].Journal of Food Protection,2008,71:210-217.
- [8] Yoo H S, Eah J Y, Kim J S, et al. DNA barcoding Korean birds [J]. Molecules and Cells,2006,22:323-327.
- [9] Wolfe K H, Li W H, Sharp M H. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs [J].Proceedings of the National Academy of Sciences, USA.1987,84:9054-9058.
- [10] Chase M W, Salamin N, Wilkinson M, et al. Land plants and DNA barcodes: short-term and long-term goals [J].Philosophical Transactions of the Royal Society,2005,360:1889-1895.

- [11] Kress W J, Wurdack K J, Zimmer E A, et al. Use of DNA barcodes to identify flowering plants [J]. *Proceedings of the National Academy of Sciences, USA*, 2005, 102: 8369-8374.
- [12] Newmaster S G, Fazekas A J, Ragupathy S. DNA barcoding in the land plants: evaluation of *rbcL* in a multigene tiered approach [J]. *Canadian Journal of Botany*, 2006, 84: 335-341.
- [13] Pennisi E. Wanted: A barcode for plants [J]. *Science*, 2007, 318: 190-191.
- [14] Ledford H. Botanical identities: DNA barcoding for plants comes a step closer [J]. *Nature*, 2008, 451: 616.
- [15] Stoeckle M. Taxonomy, DNA, and the Bar Code of Life [J]. *BioScience*, 2003, 53: 796-797.
- [16] Cowan R S, Chase M W, Kress W J, et al. 300 000 species to identify: problems, progress, and prospects in DNA barcoding of land plants [J]. *Taxon*, 2006, 55: 611-616.
- [17] Royal Botanic Gardens, Kew, UK (on-wards) DNA Barcoding Phase 2 update [EB/OL]. [2007-03-06]. <http://www.kew.org/barcoding/update.html>.
- [18] Presting G G. Identification of conserved regions in the plastid genome: implications for DNA barcoding and biological function [J]. *Canadian Journal of Botany*, 2006, 84: 1434-1443.
- [19] Taberlet P, Coissac E, Pompanon F, et al. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding [J]. *Nucleic Acids Research*, 2006, 35: e1-e8.
- [20] Kress W J, Erickson D L. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region [J]. *PLoS ONE*, 2007, 2(6): e508.
- [21] Chase M W, Cowan R S, Hollingsworth P M, et al. A proposal for a standardized protocol to barcode all land plants [J]. *Taxon*, 2007, 56: 295-299.
- [22] Selvaraj D, Sarma R K, Sathishkumar R. Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding [J]. *Bioinformatics*, 2008, 3(1): 24-27.
- [23] Lee H-L, Yi D-K, Kim J-S, et al. Development of plant DNA barcoding markers from the variable noncoding regions of chloroplast genome. Abstract presented at the Second International Barcode of Life Conference. Academia Sinica, Taipei, Taiwan, 2007: 18-20.
- [24] Lahaye R, van der Bank M, Bogarin D, et al. DNA barcoding the floras of biodiversity hotspots [J]. *Proceedings of the National Academy of Sciences, USA*, 2008, 105: 2923-2928.
- [25] Cuénoud P, Savolainen V, Chatrou L W, et al. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences [J]. *American Journal of Botany*, 2002, 89: 132-144.
- [26] Kress W J, Erickson D L. DNA barcodes: Genes, genomics and bioinformatics [J]. *Proceedings of the National Academy of Sciences, USA*, 2008, 105: 2761-2762.
- [27] Hollingsworth P M. DNA barcoding plants in biodiversity hotspots: Progress and outstanding questions [J]. *Heredity*, 2008, 101: 1-2.
- [28] Fazekas A J, Burgess K S, Kesanakurti P R, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well [J]. *PLoS One*, 2008, 3: e2802.
- [29] Sass C, Little D P, Stevenson D W, et al. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads [J]. *PLoS One*, 2007, 2: e1154.
- [30] Wolf P G, Rowe C A, Sinclair R B, et al. Complete nucleotide sequence of the chloroplast genome from a Leptosporangiate Fern, *Adiantum capillus-veneris* L. [J]. *DNA Research*, 2003, 10: 59-65.
- [31] Lidholm J, Szmidt A, Gustafsson P. Duplication of the *psbA* gene in the chloroplast genome of two *Pinus* species [J]. *Molecular and General Genetics*, 1991, 226: 345-352.
- [32] Newmaster S G, Fazekas A J, Steeves R A D, et al. Testing candidate plant barcode regions in the Myristicaceae [J]. *Molecular Ecology Resources*, 2008, 8: 480-490.
- [33] Álvarez I, Wendel J F. Ribosomal ITS sequences and plant phylogenetic inference [J]. *Molecular Phylogenetics and Evolution*, 2003, 29: 417-434.
- [34] Stoeckle M. Taxonomy DNA and the barcode of life [J]. *BioScience*, 2003, 53: 2-3.
- [35] Baldwin B G, Sanderson M J, Porter J, et al. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on Angiosperm phylogeny [J]. *Annals of the Missouri Botanical Garden*, 1995, 82: 247-277.
- [36] Tate J A, Simpson B B. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species [J]. *Systematic Botany*, 2003, 28: 723-737.
- [37] Sang T, Crawford D J, Stuessy T F. Chloroplast DNA phylogeny, reticulate evolution, and Biogeography of *Paeonia* (Paeoniaceae) [J]. *American Journal of Botany*, 1997, 84: 1120-1136.
- [38] Cho Y, Qiu Y-L, Kuhlman P, et al. Explosive invasion of plant mitochondria by a group I intron [J]. *Proceedings of the National Academy of Sciences, USA*. 1998, 95: 14244-14249.
- [39] Ivanova N V, Fazekas A J, Hebert P D N. Semi-automated, membrane based protocol for DNA isolation from plants [J]. *Plant Molecular Biology Reporter*, 2008, 26: 186-198.
- [40] Taberlet P, Gielly L, Pautou G, et al. Universal primers for amplification of three noncoding regions of chloroplast DNA [J]. *Plant Molecular Biology*, 1991, 17: 1105-1109.
- [41] Johnson L A, Soltis D E. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequence [J]. *Annals of the Missouri Botanical Garden*, 1995, 82: 149-175.
- [42] Sang T, Crawford D J, Stuessy T F. Chloroplast DNA phylogeny, reticulate evolution, and Biogeography of *Paeonia* (Paeoniaceae) [J]. *American Journal of Botany*, 1997, 84: 1120-1136.
- [43] Hoot S B, Culham A, Crane P R. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae [J]. *Annals of the Missouri Botanical Garden*, 1995, 82: 194-207.
- [44] Demesure B, Sodji N, Petit R J. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants [J]. *Molecular Ecology*, 1995, 4: 129-131.
- [45] Lee C, Wen J. Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants [J]. *Molecular Phylogenetics and Evolution*, 2004, 31:

- 894-903.
- [46] Lledo M D, Crespo M B, Cameron K M, et al. Systematics of Plum-baginaceae based upon cladistic analysis of *rbcL* sequence data [J]. Systematic Botany, 1998, 23: 21-29.
- [47] Stanford A M, Harden R, Parks, C R. Phylogeny and Biogeography of *Juglans* (Juglandaceae) based on *matK* and ITS sequence [J]. American Journal of Botany, 2000, 87: 872-882.
- [48] White T, White T J, Burns T, Lee S, Taylor J: Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics// PCR protocols, a Guide to Methods and Applications[M]. Innis M, Gelfand D, Sninsky J, et al. California, Academic Press, 1990: 315-322.
- [49] Seberg O, Petersen G. How many loci does it take to DNA barcode a *Crocus*? [J]. PLOS one, 2009, 4: e4598.
- [50] Nsubuga A M, Robbins M M, Roeder A D, et al. Factors affecting the amount of genomic DNA extracted from ape feces and the identification of an improved sample storage method [J]. Molecular Ecology, 2004, 13: 2089-2094.
- [51] Smith L M, Burgoyne L A. Collecting, archiving and processing DNA from wildlife samples using FTA(R) databasing paper [J]. BMC Ecology, 2004, 4: 4.
- [52] Rogers N L, Cole S A, Lan H C, et al. Newsaliva DNA collection method compared to buccal cell collection techniques for epidemiological studies [J]. American Journal Human Biology, 2007, 19: 319-326.
- [53] Nitta J H. Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea [J]. TAXON, 2008, 57 (3): 725-736.
- [54] Willerslev E, Hansen A J, Binladen J, et al. Diverse plant and animal genetic records from Holocene and Pleistocene sediments [J]. Science, 2003, 300: 791-795.
- [55] Willerslev E, Cappellini E, Boomsma W, et al. Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland [J]. Science, 2007, 317: 111-114.
- [56] 李洪涛, 李德铎. 基于 ITS、*rpl16* 和 *trnS-trnR* DNA 序列讨论锥形果属的系统位置 [J]. 植物分类学报, 2008, 46(4): 595-599.
- [57] 张金梅, 王建秀, 夏涛, 等. 基于系统发育分析的 DNA 条形码技术在澄清芍药属牡丹组物种问题中的应用 [J]. 中国科学 C 辑: 生命科学, 2008, 38(12): 1166 - 1176.
- [58] Palmer J D, Herbon LA. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence [J]. Journal of molecular Evolution, 1988, 28: 87-97.
- [59] 陈士林, 姚辉, 宋经远, 等. 基于 DNA barcoding(条形码)技术的中药材鉴定 [J]. 世界科学技术-中医药现代化, 2007, 9(3): 7-12.
- [60] 宁淑萍, 颜海飞, 郝刚, 等. 植物 DNA 条形码研究进展 [J]. 生物多样性, 2008, 16(5): 417-425.
- [61] 石林春, 梁宗锁, 韩建萍, 等. 基于杜鹃属植物的 DNA 条形码序列筛选 [J]. 世界科学技术-中医药现代化, 2009, 11(1): 54-57.