

植物 DNA 条形码技术的发展及应用

刘宇婧¹, 刘越¹, 黄耀江¹, 龙春林^{1,2,①}

(1. 中央民族大学生命与环境科学学院, 北京 100081; 2. 中国科学院昆明植物研究所, 云南 昆明 650204)

摘要: 在对 DNA 条形码技术的发展过程进行归纳分析的基础上, 对植物 DNA 条形码技术的研究进展、工作流程及分析方法、影响其鉴定准确性的因素及其在植物分类学研究中的应用现状及存在的争议进行了综合分析和阐述, 并展望了植物 DNA 条形码技术的发展趋势及应用前景。通过具体实例说明将植物 DNA 条形码技术与传统植物学知识相结合可作为民族植物学的研究手段之一。认为: 目前常用的植物 DNA 条形码主要有单一片段和多片段组合 2 种方式, 这 2 种方式各有优缺点; 常用的 DNA 序列有 *matK*、*trnH-psbA*、*rbcL* 和 *ITS* 等, 但均有一定的局限性; 针对不同的使用目的, 应选择不同的植物 DNA 条形码标准; 影响植物 DNA 条形码鉴定准确性的因素包括物种的类型和数量、系统树构建方法、杂交/基因渗入、物种起源时间的差异、分子进化速率差异等; 当前植物 DNA 条形码研究工作的重点是选择合适的 DNA 片段并对其进行评价。

关键词: DNA 条形码; 发展进程; 分类学; 植物物种鉴定; 民族植物学

中图分类号: Q789; Q949 文献标志码: A 文章编号: 1674-7895(2011)01-0074-09

Progress and application of DNA barcoding technique in plants LIU Yu-jing¹, LIU Yue¹, HUANG Yao-jiang¹, LONG Chun-lin^{1,2,①} (1. College of Life and Environmental Sciences, Minzu University of China, Beijing 100081, China; 2. Kunming Institute of Botany, the Chinese Academy of Sciences, Kunming 650204, China), *J. Plant Resour. & Environ.* 2011, 20(1): 74-82, 93

Abstract: Based on summarization and analysis of development process of DNA barcoding technique, research progress of DNA barcoding technique in plants, its working process and analysis method, influencing factors on its identification accuracy, its application status and dispute existing in plant taxonomic study were comprehensively analyzed and described, and further development trend and application prospect of DNA barcoding technique in plants were proposed. By means of some research examples, it was indicated that combining method of DNA barcoding technique in plants and traditional botanical knowledge could be taken as one of studying ways for ethnobotany. And also, it was suggested that common DNA barcoding in plants mainly were two modes of single fragment and multiple fragments combining, and both of them had their respective advantages and disadvantages. Common DNA sequences included *matK*, *trnH-psbA*, *rbcL* and *ITS*, etc, but they all had a certain limitation. Different standards of DNA barcoding in plants should be selected in order to different application aims. Influencing factors on its identification accuracy included type and number of species, construction method of phylogenetic tree, hybridization and gene introgression, variance of species origin time and variance of molecular evolution rate. Current focus on DNA barcoding in plants is how to select suitable DNA fragments and to evaluate their values.

Key words: DNA barcoding; development progress; taxonomy; plant species identification; ethnobotany

收稿日期: 2010-08-16

基金项目: 国家教育部和国家外专局高等学校创新引智工程项目(B08044); 中央民族大学“985”工程项目(MUC-985); 中国科学院大科学装置开放研究项目(O92441112F)

作者简介: 刘宇婧(1985—), 女, 满族, 山西大同人, 硕士研究生, 主要从事植物 DNA 条形码技术和民族植物学研究。

①通信作者 E-mail: chunlinlong@hotmail.com

人类社会的发展至今,已经通过分类学方法鉴定和描述了170万种生物,然而这个数字仅仅是分类学家预计的全球物种总数的15%,因此,对地球上众多的物种进行分类和鉴定仍然是全人类的一项长期而艰巨的任务。由于物种的形态会受到生物性别和发育阶段的影响,因此,对物种进行分类客观上存在一定的困难;同时,传统分类学方法无法鉴定出许多群体中普遍存在的隐种(即外观相似但遗传特征截然不同的物种),可能会导致物种划分有误,需要对这些物种进行订正;另外,物种的分类需要长期的经验和知识积累,往往要很多年才能培养出一位擅长某一类群分类的专家。鉴于以上因素并基于对时间和成本的综合考虑,开发一种更灵活的物种标准化鉴定系统已成为当前生物分类学研究领域的当务之急^[1-3]。

DNA条形码(DNA barcoding)技术是以染色体组为基础,用1段标准DNA序列作为标记实现快速、准确和自动化的物种鉴定方法,类似于超市用条形码扫描识别成千上万种商品,该方法是在近几年来迅速发展起来的一项全新的物种鉴定技术,为解决传统分类学面临的难题提供了新方法。除物种鉴定功能外,DNA条形码技术还有助于发现新种和隐种,也可应用于其他学科(如民族植物学)的研究。该技术在动物研究中已得到广泛的应用,采用的标准片段是线粒体*COI*基因中1段约650 bp的DNA片段,然而,*COI*基因在植物中的进化速率远慢于在动物中的进化速率,因此*COI*基因只适用于低等植物中某些藻类的鉴定^[4-5]。迄今为止,在陆生植物中还没有找到通用性DNA条形码^[6]。尽管如此,对植物DNA条形码的研究依然在迅速有效地进行。构建DNA条形码标准数据库、利用该技术鉴别已知物种和发现新种及隐种、重建物种和高级阶元的演化关系,是目前分子生物学和分类学发展的最新研究方向,对植物保护生物学和生物多样性研究具有重大意义。

DNA条形码具有诸多优点:①以信息稳定的DNA序列为检测对象,每一物种具有各自特定的DNA序列信息,同种生物不同生长时期具有相同的DNA序列信息,即使经过加工使其形态发生变化,但其DNA序列信息不会改变,而传统形态学鉴别特征会因趋同性和变异导致鉴定错误;与传统分类方法相比,DNA条形码技术扩大了检测样本的范围,即使样本部分受损也不会影响鉴别结果。②能够抛开形态相似的假象,从基因水平上对传统分类方法难以区分

的类群进行物种鉴定。③建立DNA条形码数据库,不但可以一次性鉴定大量物种,而且可以提供各物种的明确信息;不仅能够弥补传统分类学对物种形态描述的不足,而且还可以加快对已知物种的识别速度;同时还便于发现新物种,对分类学科的快速深入发展有一定的推动作用。④在DNA条形码技术的基础上研制简便和高效的条形码扫描仪,可以加快物种鉴定和进化研究的步伐,有益于推进分类学研究基础薄弱的国家,尤其是发展中国家物种鉴定及进化研究的步伐^[2,7]。

1 植物DNA条形码技术的研究进展

1.1 DNA条形码的发展轨迹

从2003年至2009年,陆续召开了多次有关DNA条形码的国际会议。2003年,20多位分类学专家、分子生物学家和生物信息学家在冷泉港召开了2次研讨会,会议题目分别为“Taxonomy and DNA”和“Taxonomy, DNA and the Barcode of Life”,以期利用某个特定基因实现物种鉴定的目标。2004年,由Sloan基金会赞助,在华盛顿史密森尼研究院(Smithsonian Institution)成立了生命条形码联盟(CBOL, the Consortium for the Barcode of Life),大部分国家的自然历史博物馆、标本馆、相关研究机构和私人机构等都加入了该组织,致力于鉴定生物物种全球标准的研制。2005年,由CBOL和英国自然历史博物馆主办,在伦敦召开了生命条形码协会第一届国际DNA条形码会议,决定为1000万个物种建立DNA条形码编码库,并将其设置在GenBank中,作为一个向公众开放的DNA序列数据库。2007年,在台北召开的第二届国际DNA条形码会议对理想DNA条形码的选择展开了深入的讨论,并给予了高度的重视。2009年,在墨西哥城召开的第三届国际DNA条形码会议上,与会代表对植物DNA条形码达成共识,决定将叶绿体基因片段*rbcl*和*matK*作为植物DNA标准条形码的核心码,同时建议将叶绿体基因片段*trnH-psbA*和核基因片段*ITS*作为植物DNA条形码的补充码;并指出*ITS2*片段虽然在DNA条形码方面具有较大的潜力,但要警惕操作过程中的真菌污染问题。此外,要继续开展通用单拷贝核基因片段的筛选和测试,这对推进全球DNA条形码计划有重要作用^[8-10]。可见,DNA条形码技术正成为国际上生物

分类学研究的一个重要方向和主要研究手段。

加拿大圭尔夫大学(University of Guelph)教授、加拿大皇家学会会员 Hebert 率先将条形码技术引入生物界并提出了“DNA 条形码”的概念。Hebert 等采用线粒体细胞色素 C 氧化酶亚基 I(COI)上的 1 段基因序列,先后对脊椎动物和无脊椎动物 11 个门共 13 320 个物种进行比较分析,发现除腔肠动物(Cnidaria)外,98% 的物种种内遗传距离差异为 0%~2%,种间遗传距离差异平均值为 11.3%,说明这段 COI 基因序列具有良好的分类鉴别能力,并据此提出可以建立以 1 段长度为 650 bp 的 COI 基因序列为基础的 DNA 条形码识别方法^[11-12]。

近些年来,国内外的许多学者都对植物中适合作为 DNA 条形码的基因序列进行了积极的探索和研究^[6,11-15]。线粒体 COI 基因在植物中的进化速率远远慢于动物,因此,线粒体基因不适合作为大多数植物的 DNA 条形码;核基因组通常具有多拷贝的特性,种内变异较大,引物通用性较差,并且扩增时对模板 DNA 的质量要求较高,不适宜于存在 DNA 降解的材料;叶绿体基因组相对保守,但包含许多变异区域,而且还具有一些自身的优势,如单亲遗传避免了基因重组、植物个体中均有大量的叶绿体,即使模板 DNA 高度降解也容易扩增成功。所以,从叶绿体基因组中选择植物 DNA 条形码是可行的。

目前,研究者主要通过单一片段或多片段组合 2 种植物 DNA 条形码对相关植物物种进行分类鉴定研究,并尝试建立植物 DNA 条形码标准。

1.2 单一片段植物 DNA 条形码研究进展及其优缺点分析

1.2.1 单一片段植物 DNA 条形码的研究进展

Newmaster 等^[16-17]通过对部分植物叶绿体基因组全面详细的分析总结出以下几点:①UPA 片段是适合于藻类的 DNA 条形码,同时也可以作为陆地植物 DNA 条形码的组合条形码之一;②[6,18-19]。在对陆地植物进行大尺度取样研究后,Kress 和 Erickson 提出:虽然

rbcL 片段的变异率低但通用性较好,可以作为 DNA 条形码的核心片段,对不能识别的物种可以采用高变异率的 trnH-psbA 片段进行进一步的细分^[20]。但是,在一些特殊的植物类群中,如伞形科(Umbelliferae)的独活属(*Heracleum* L.)和禾本科(Poaceae)的甜茅属(*Glyceria* R. Br.)中, trnH-psbA 片段的变异率较低^[21-22],因而 trnH-psbA 片段的使用应根据物种而定。Lahaye 等^[23]用 8 个叶绿体 DNA 片段对兰科(Orchidaceae)植物进行了研究,通过比较最终认为 matK 片段可以作为有花植物通用的 DNA 条形码。Fazekas 等^[6]采用 matK、trnH-psbA、rbcL、ITS1、UPA、rpoB、rpoC1、atpF-atpH 和 psbK-psbI 等片段对 32 属 92 种共 251 株植物进行了研究,结果表明:在所有片段中 trnH-psbA 片段的扩增成功率和物种识别率均最高。Newmaster 等^[24]指出:采用 matK、trnH-psbA 和 rbcL 片段可以对金合欢属(*Acacia* Mill.)类群进行有效的分类鉴定。Song 等^[25]采用 rbcL、trnH-psbA、ndhJ、rpoB、rpoC1、accD、ycf5 和 nrITS 等片段对蓼科(Polygonaceae)的 38 个类群进行分析,指出 trnH-psbA 片段最适合作为蓼科植物鉴定的 DNA 条形码。

在 Hebert 提出“DNA 条形码”概念后不久,中国学者也开始关注 DNA 条形码技术。陈士林领导的研究小组针对 753 属 4 800 种植物 6 600 个样品的研究结果显示:ITS2 序列具有良好的通用性,在种级水平上的分辨成功率可以达到 92.7%,为此,该研究小组建议将 ITS2 序列作为植物 DNA 条形码之一^[26]。作者所在的课题组利用目前植物 DNA 条形码研究中的热点候选序列 ITS2、trnH-psbA、rbcL、rpoC1 和 ycf5 对芸香科(Rutaceae)72 属 192 个类群的 300 个样本进行了研究和比较,结果显示:ITS2 序列具有良好的扩增效率和测序成功率,在所考察的候选序列中具有最大的种间变异和较小的种内变异,且种间变异明显大于种内变异,同时,ITS2 序列的物种鉴定成功率最高,其各个评价指标均优于其他候选序列^[27];此外,ITS2 序列在黄芪属(*Astragalus* L.)类群中也具有很好的通用性,扩增效率为 100%,测序成功率为 85.7%,而且种间变异远大于种内变异^[28]。还有研究者对全世界桦木科(Betulaceae)赤杨属(*Alnus* Mill.)所有类群(共 26 种)的 131 个单株进行了分析,研究结果显示:rbcL、matK、trnH-psbA 和 ITS 片段在种级水平上的分辨能力差异较大,分辨率分别为 10.0%、31.2%、63.6% 和 79.6%^[29]。

1.2.2 常用单一片段植物DNA条形码的优缺点
随研究的不断深入,研究人员发现一些单一片段DNA条形码存在一定的局限性,尤其是 *matK*、*trnH-psbA*、*rbcL* 和 *ITS* 等常用植物DNA条形码的缺点在一定程度上制约了DNA条形码技术的发展。

与其他片段相比,*matK* 片段的进化速率较快,但该片段的引物通用性较差,鉴定不同类群时往往需要设计不同的引物^[2,30]。Lahaye 等^[23]采用 *matK* 片段的 390F/1326R 引物对 1 667 个兰科植物进行扩增,扩增成功率达到 100%;但 Fazekas 等^[6]采用相同引物获得的扩增成功率却低于 50%,因此,*matK* 片段的引物通用性还有待更多的实验检验。

trnH-psbA 片段是叶绿体基因进化速率最快的片段之一^[31],但该片段具有较高的突变率和简单的序列重复和重排,在进行序列分析时需要人工校正^[2],从而限制了该片段在系统发育研究中的应用。尽管如此,也有研究者认为:虽然 *trnH-psbA* 片段具有一定的缺陷,但仍然可以作为潜在的DNA条形码^[20]。

在大部分植物中,*rbcL* 片段都比较容易扩增和测序,因而该片段被许多学者作为研究对象并开展了相关研究^[2,16,32-33],但植物中 *rbcL* 片段的进化率远远低于动物,且其变异主要存在于种级以上水平,种级水平的变异很小,因此,有研究者提出可以将该片段作为组合DNA条形码之一^[20]。

ITS 序列在陆地植物中的引物通用性较差^[20],但是可将该片段作为一个潜在的DNA条码用于相关研究^[26,34],已有学者开始关注 *ITS2* 片段。*ITS2* 是 *ITS* 序列中的 1 个片段,位于 5.8S 和 26S rRNA 之间;*ITS2* 片段较短,易于扩增和测序,尤其对发生DNA降解的材料更为有利^[26-27,35];该片段在物种水平上变异较大,已被广泛用于物种分类及系统进化研究。尽管 *ITS* 序列比 *ITS2* 片段更早受到学者们的关注,但是在某些植物类群中 *ITS* 序列的扩增效率比较低,不符合DNA条形码通用性的要求^[26,28,36]。

叶绿体 *trnL*(UAA)内含子全序列及其 p6 环种间变异太小,限制了其在系统进化研究中的应用,不适合作为陆地植物种级水平鉴别的DNA条形码。虽有这些不足,但该片段仍有许多优势,例如:高度保守、扩增体系稳定、p6 环在DNA高度降解的样本中仍可以成功扩增,因此,该片段在植物DNA条形码中的应用需要进一步研究^[37]。

此外,还有一些非重点研究的片段,例如:在非被

子植物中,*rpoB2* 片段具有较低的扩增率;*rpoC1* 片段虽然具有较高的扩增率,但突变率较低。这些片段均不适合用于种级以上水平的植物物种鉴定研究^[2,38]。

1.3 多片段组合植物DNA条形码研究进展

大量的研究表明:单独使用某一个片段对所有的植物物种进行准确鉴定是不太可能的,为此,研究者们又相继提出了不同的植物DNA条形码的多序列组合方案。Kress 等早在 2005 年就提出了多序列组合观点,他们预测将 *trnH-psbA* 和 *ITS* 序列组合用于被子植物种类鉴定将有广泛的应用前景,但并未做具体的实验研究^[34]。同年,Chase 等提出条形码分析的“交通灯法”:首先使用叶绿体条形码进行初步分析,用绿灯表示能被准确鉴定的物种;用黄灯表示存在问题的物种,再由分析者根据实际情况判定是否需要进一步的精确鉴定;用红灯表示识别非常不准确的物种^[39]。Newmaster 等^[17]则提出将等级分类的方法应用于植物物种鉴定的观点:首先选择 1 个核心片段作为第 1 级条形码,然后再根据不同类群选择不同片段作为第 2 级条形码。Chase 等也提出了使用条形码组合的建议,并指出:在 *rpoC1* 和 *rpoB* 与 *matK* 的组合中,*rpoC1* 和 *rpoB* 序列的引物通用性较好,扩增成功率也较高,虽然进化速率较慢,但也能区分出许多物种;*matK* 序列变异较大,能够提供更多的识别信息,但是 *matK* 序列的引物通用性还有待进一步研究;在 *rpoC1* 和 *matK* 与 *trnH-psbA* 的组合中,保守的 *rpoC1* 序列和长度一定的 *matK* 序列可以保证大部分物种的种间鉴定,再加上高变异率的 *trnH-psbA* 序列就可以识别出更多的物种^[30]。2005 年, Kress 和 Erickson^[20]通过对 *rbcL*、*rpoC1*、*rpoB2*、*trnH-psbA*、*matK* 和 *ITS1* 序列的两两组合分析,最终确定 *rbcL* 和 *trnH-psbA* 组合适用于对陆地植物的识别和鉴定;在这些叶绿体基因片段中,*rpoB2*、*rpoC1*、*rbcL*、*matK* 属于编码区片段,*trnH-psbA* 属于非编码区片段。Fazekas 等^[6]认为:DNA 条形码的识别能力与组合片段的数目有一定关系;在两两片段组合中,*rbcL* 和 *trnH-psbA* 组合的扩增率和物种识别率最高。通过对 7 个候选序列进行单个分析、两两组合分析和三个组合分析后,CBOL 植物工作组指出:三个组合与两两组合的鉴别力一样,而且增加了鉴别工作量和投入成本,建议将 *rbcL* 和 *matK* 组合作为鉴别陆地植物的通用DNA条形码^[2]。任保青等^[29]认为:在整个植物界,若想用 1 个基因或 1 个短的DNA片段来区别所有物种几乎

是不可能的,但是可以通过1个条形码组合在基因组范围内实现阶层条形码,进而逐级缩小鉴定范围,最终达到物种自动鉴定的目的;在陆地植物中比较可行的DNA条形码组合是*rbcL*、*matK*、*ITS*与*trnH-psbA*的组合,分别适于科级、属级、种级类群的鉴别。

多序列组合方案为植物DNA条形码的研究指出了一个新的方向,但根据现有的研究结果筛选出的片段并不能完全满足DNA条形码的标准,仍需要大量的实验去验证它们的物种识别和鉴定能力。

1.4 植物DNA条形码标准的筛选

迄今为止,已有大量的DNA片段被作为DNA条形码用于物种鉴定,有的是单一片段方式,有的则是序列组合方式,但目前对于植物DNA条形码的选择还没有一个明确的标准。根据当前的发展状况,Kress等^[34]和Taberlet等^[37]提出了理想的DNA条形码标准:①种间应有明显的变异以鉴别所有物种,同时种内变异应足够小;②DNA条形码应尽可能标准化,即采用同一DNA片段来鉴定物种;③DNA片段应足够短,长度约700 bp,便于单向测序,也有利于PCR扩增,尤其是对DNA易降解的材料更有利;④存在高度保守区域,便于设计通用引物,并且该DNA片段易扩增和测序;⑤目标DNA条形码应包括足够的种系进化信息,以便对物种进行系统分类。

Erickson等^[40]提出了可进行量化和优化的陆生植物DNA条形码标准:①PCR扩增成功率至少90%;②选择1个在系统进化研究中已经被研究多年的DNA片段作为条形码;③DNA条形码之间应具有互补性且不相关联,将准确鉴定的可能性最大化;④使用的DNA条形码在分类学应用中应具有普遍意义;⑤具有可行的生物信息学分析方法。

事实上,完美的植物DNA条形码并不存在,对于不同的使用目的,以上标准并非都适用。例如,对于分类学家而言,DNA条形码具有足够种系进化信息和高度变异非常重要;而对于法医鉴定或者加工食品的鉴定,DNA条形码的标准化更为重要^[39]。

2 植物DNA条形码技术的应用范畴

2.1 物种鉴定以及新种和隐种的发现

除了可以进行物种鉴定外,利用DNA条形码技术还可以发现许多新种和隐种。例如,Lahaye等^[23]单独使用*matK*片段对分布在中美洲的1000多种兰

科植物进行分析,显示单独使用*matK*片段能够揭示隐种并且证明了DNA条形码的可行性;2010年,Newmaster等^[41]通过DNA条形码技术发现了感应草属(*Biophytum* DC.)中的3个隐种,并发现了草沙蚕属(*Tripogon* Roem. et Schult.)的1个新种。协助传统分类方法发现那些形态相似但存在遗传分化的隐种是DNA条形码技术对分类学研究的重要贡献,可显著提高实地生态学考察研究的准确性和效率。

2.2 用于解决一些生物学问题

Jurado-Rivera等^[38]认为:DNA条形码还可用于解决一些复杂的生物学问题,如寄生关系等。通过扩增草食性动物牛的*trnL*基因,发现*Peltoschema*与4种寄生植物关系密切。因此,利用DNA条形码可以正确识别寄生植物并确定营养关系。

2.3 在民族植物学研究中的应用

值得一提的是,DNA条形码技术在民族植物学研究中也得到了一定的应用。民族植物学以传统社会管理和利用的植物为主要研究对象,在研究过程中常常通过借助民间分类学家(parataxonomist)的知识来获得当地人对植物进行分门别类的信息。这些民间分类学家的知识有些是传承自长辈、有些则结合了自己的观察和实践,他们不必像分类学家那样一般需要获得植物的繁殖器官才能有效鉴定物种,而是基于他们长期积累的知识、通过植物某个生长阶段的某个(些)器官(通常是营养器官),就能准确说出植物的当地名称、生长习性、资源和分布状况、用途和用法等。在开展民族植物学研究的过程中,由植物分类学家和民间分类学家分别获得的植物种数往往存在一定的差异,或前者数量多于后者,或后者数量多于前者。到底哪一个物种数是准确的?在没有足够的时间等待植物开花和结果的情况下,采集植物叶片通过DNA条形码技术进行分类鉴定,是一种理想的快速鉴别方法。基于这样的设想,加拿大圭尔夫大学的Newmaster博士及其合作者,将DNA条形码技术应用于民族植物学的研究。印度南部西高止山脉的2个山地部落Iruilas和Malasars把在当地被称为“Sunai pul”的1种草沙蚕属禾草用于祭祀和其他经济用途,这种植物对当地社区居民十分重要,因而当地的民间分类学家(或信息提供者 informant)能轻易识别这一物种。有趣的是,这是1个植物分类学上的隐种,植物分类学家难以通过形态特征来识别。Newmaster等采用DNA条形码技术,不仅证实了民间分类学家鉴

定该物种的正确性,而且发现 *Sunai pul* 是植物学上的1个新种,将其命名为 *Tripogon cope* Newmaster。因此,他们认为DNA条形码技术可以应用于一些民间分类群(ethnotaxon)的鉴定和识别^[42]。Newmaster等还将DNA条形码与基因组学的相关技术相结合,应用于印度一些部落的民间分类群,包括民族药用植物。在此基础上,他们提出了“民族植物基因组学”(ethnobotany genomics)的概念,认为通过民族植物基因组学的研究,可以识别一些隐种,还能确定一些珍稀植物的分布及其生态学特性^[41]。

清风藤属(*Sabia* Colebr.)中的小花清风藤(*Sabia parviflora* Wall. ex Roxb.)是1种非常重要的民间药用植物,被贵州和云南的布依族民众广泛用于治疗黄疸型肝炎、止血和消炎,布依族民众也使用同属植物作为代用品,但是一些形态相似的混淆品已经进入市场。龙春林领导的研究组用 *trnH-psbA*、*rbcL* 和 *matK* 片段对6种清风藤属植物和7种市场混淆品进行鉴定^[43],结果表明:这3个DNA片段的扩增成功率和物种识别率都很高,且小花清风藤与同属替代品之间的序列差异率较低(仅为0.00%~0.86%),但与混淆品之间的序列差异率却较高(达24.50%),说明布依族民众识别清风藤属植物的传统知识是可靠的。研究结果还表明,DNA条形码技术不仅可以有效鉴别市场上小花清风藤的混淆品,而且对民族传统医药文化的保护有重要作用。

2.4 构建AIT系统

2009年,Newmaster等^[44]提出开发和研制AIT(automated identification technology)系统,即:将1片未知的植物叶片插入AIT机,通过无线网络,就可以获得该叶片所代表植物的名称、分布、化学成分、食用价值等,既省时又经济,可代替昂贵又费时的传统分类方法,同时也减少了通过形态特征识别物种的一些限制因素。预计AIT系统可能会给生物界带来革命性的改变,也会对人类产生重要影响。

3 植物DNA条形码技术的工作流程及分析方法

3.1 工作流程

植物DNA条形码技术的工作流程主要包括采集样品、提取DNA、选择DNA条形码、设计引物或者使

用通用引物、扩增序列、编辑与人工校正序列、分析结果、将最终结果提交至相应数据库等八大步骤。其中,采样是全部工作的基础,需要制定标准化流程以便能进一步复核研究结果。序列分析是DNA条形码研究中最重要的一步,经过对序列进行比对和人工校正后,采用MEGA或PAUP法计算物种的种内和种间K2P距离(kimura-2-parameter distance),以确定遗传距离的阈值,比较种、属和科3个水平上的序列差异。然后,根据上述分析结果建立系统进化树,最后依据DNA条形码分析的聚类结果就能对未知样本进行分类和鉴定,也可以发现隐种^[38,45-50]。

3.2 分析方法

由于植物的种间杂交现象比较普遍,因此植物DNA条形码的分析方法也处于不断的摸索过程中,目前报道的分析方法主要包括以下几种:

3.2.1 DNA条形码校对机制 在分析过程中,可能会由于多种原因导致扩增的序列不正确,如采样时存在污染、PCR污染、测序过程错误等,可以通过DNA条形码序列分析发现这些错误,并重新进行PCR扩增、测序和人工校正以纠正错误序列,或找多位专家对凭证标本进行重新鉴定。DNA条形码序列具有三大校对防错机制,即序列校对防错、系统树分析防错、barcoding gap检验防错,以确保DNA条形码序列和物种的正确对应关系^[14]。

3.2.2 DNA条形码序列分析 向GenBank提交获得的序列后,用Clustal X软件对序列进行排序并用DNAMAN软件对序列进行比对后,使用Bioedit软件完成格式转换,再用MEGA4.0软件进行统计分析并构建系统树,最后,用bootstrap分析法检验(1000次重复)严格一致性树中各分支的可信度^[2,14]。

3.2.3 DNA条形码序列的鉴定效率评价 一般采用4种方法评价DNA条形码序列的物种鉴定能力。

一是BLAST搜索法(相似性搜索算法)。该方法首先需要通过实验建立物种的序列数据库,并确定一个阈值,将样本的DNA序列与全体序列进行比对,如果可以在数据库中搜索到仅包含query序列及同物种的序列,则认为该query序列可以准确鉴定到种。该方法具有运行速度快、精确度高的特点。

二是序列之间遗传距离的比较法。通过比较样本DNA序列与全体序列的K2P距离,确定种内和种间变异的阈值,即barcoding gap,进而评估其物种鉴定效果。常用柱形图来呈现种内及种间遗传距离的

分布频度。该方法比较耗时,并且对于缺失位点和变异位点的等价处理会造成一些数据误差^[51]。

三是系统进化树的建立。建树的目的是为了检验每个物种的单系性,即同一物种的不同个体能否紧密聚集在一起,当不同物种聚在一起时则认为该DNA条形码不能够正确鉴定物种^[14,23,52]。

四是在筛选DNA条形码的研究过程中,需要对各个片段的效果进行评估,因此需要对不同片段在种内和种间的变异进行比较,目前采用Wilcoxon signed rank tests法进行比较。

4 影响植物DNA条形码技术鉴定准确性的因素

4.1 物种的类型和数量

评价种内距离通常采用3个参数,即K2P距离、平均 θ 值和平均溯祖度,这3个参数均会随样本数的增多而增大,因此应尽可能增加物种内的取样个体数。然而,出于对成本的考虑,通常认为每个物种内的个体采集数量应尽量不超过10个,并且最好包括5个居群的个体,这就要求研究者必须对拟作为DNA条形码使用的基因片段有充分的了解^[53]。此外,还可以应用种间距离平均值(average interspecific distance)、种间平均变异 θ 值(average theta prime)和种间最小距离(smallest interspecific distance)^[26,51]这3个参数来对种间差异进行评估。

4.2 系统树构建方法

系统树又叫进化树,构建进化树的方法主要分为独立元素法(discrete character method)和距离依靠法(distance method)2种。其中,独立元素法包括最大简约性法(maximum parsimony method)和最大可能性法(maximum likelihood method);距离依靠法包括除权配对法(UPGMA)和邻位相连法(neighbor-joining)。Lahaye等^[23]对进化树进行了分析,认为MP和UPGMA的物种正确识别率相对较高,但结果相差不大时选用NJ树则能达到快速简便的效果。不同建树方法得到的结果有一定差异,只有采用适当方法构建的进化树才会接近真实的“进化树”。

4.3 杂交/基因渗入

通常,物种间的边界会由于杂交或基因渗入等原因而显得模糊不清,这种情况下需要通过DNA条形

码的不同组合方式来增加标记数,以达到准确区分物种的目的。

4.4 物种起源时间的差异

如果物种之间产生生殖隔离的时间非常短,那么采用DNA条形码技术来鉴定这些物种就比较困难。然而,化石记录显示大多数物种分化时间都超过了100万年,因此,对于大部分物种而言,采用DNA条形码技术进行鉴定工作比较简单^[54-55]。

4.5 分子进化速率差异

有研究表明,如果某个类群中的各个谱系分子进化速率不一致且差异达到100倍时,由于进化快的谱系存在2次突变,在这个类群中鉴定出进化速率慢的物种将变得非常困难^[56]。因此,分子进化速率差异对DNA条形码技术鉴定物种的准确率有较大影响。

5 植物DNA条形码技术研究存在的问题与展望

自Hebert提出“DNA条形码”概念以来,该技术得到了许多生物学家的支持并有较大进展。由于COI基因在动物鉴定中几乎百分之百的成功,因此许多学者认为DNA条形码技术在一定条件下是可行的^[11,57-58];然而,也有部分学者持怀疑和反对态度,认为采用DNA条形码技术将是一种倒退^[59],会将分类学退回到类型学^[60]。虽然有人认为传统分类学在未来将被DNA条形码取而代之,但多数学者认为DNA条形码可弥补传统分类学方法的不足,是对日渐萎缩的传统形态分类学强有力的补充。此外,还有一些学者认为如此短的DNA片段不能提供物种水平的可靠信息,完全依靠DNA条形码会导致鉴定错误。

同一DNA条形码片段在不同类群中的应用效果不一样,因而目前尚未获得一致的植物DNA条形码标准片段,因此,当前的研究热点仍然是对可能的DNA条形码片段进行选择 and 评价,并进行更大规模的分析 and 整体评价。将DNA条形码技术与其他分类学方法结合使用,有助于区别个体及加快发现新物种的速度,然而,DNA条形码只能作为一种初步筛选技术达到快速识别的目的,若要最终确定一个新物种,还需要经过很多缜密的检测手段。目前,植物DNA条形码的分析方法还不够成熟,需要生物信息学领域的专家开发出适合多片段和针对某些特殊片段的分

析方法。另外,对于植物DNA条形码技术的价值还存在一些争议,关于该技术能否适用于动植物近缘和近期分化物种的鉴定也存在较大的争议,这也是DNA条形码技术研究的难点^[41,61-65]。

虽然目前在DNA条形码研究中投入的资金比较大,但是一旦成功构建出一个标准、统一的DNA条形码序列数据库,设计出AIT系统,不仅能方便不同人群和民众准确地采集和鉴别植物,而且还能满足不同背景的专业人员快速鉴定植物的需求,并能降低鉴定物种所需的人力和资金成本。随着生物技术的发展,测序过程将会更快速、更便宜,从这个意义上讲,DNA条形码技术是一个快捷、经济又实用的技术^[24]。

序列标记、DNA库和组织库(分类凭证标本)是实现条形码标准化的3个关键环节。随着条形码序列数据的不断积累及研究范围的不断扩大,同时伴随大规模测序所产生的大量数据,条形码研究会逐渐实现与世界范围内其他分类学研究计划的协调发展,作为一种物种鉴定工具必将广泛应用于植物系统分类学的研究,并在其他领域(如生药学、民族植物学等)得到应用。同时,包括质体、核糖体、较短的低拷贝核基因等多个基因位点在内的组合DNA条形码的出现,将有效克服在被子植物物种分类和鉴定过程中经常出现的由于杂交和多倍体现象所带来的问题。

参考文献:

- [1] 莫帮辉,屈莉,韩松,等. DNA条形码识别I. DNA条形码研究进展及应用前景[J]. 四川动物, 2008, 27(2): 303-306.
- [2] Hollingsworth P M, Forrest L L, Spouge J L, et al. A DNA barcode for land plants [J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(31): 12794-12797.
- [3] Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes[J]. BMC Plant Biology, 2009, 7: 1-17.
- [4] Blaxter M, Mann J, Chapman T, et al. Defining operational taxonomic units using DNA barcode data[J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2005, 360: 1935-1943.
- [5] Rach J, DeSalle R, Sarkar I N, et al. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata [J]. Proceedings of the Royal Society B: Biological Sciences, 2008, 275: 237-247.
- [6] 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(7): e2802.
- [7] 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.
- [8] 陈士林,姚辉,宋经元,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.
- [9] Armstrong K F, Ball S L. DNA barcodes for biosecurity: invasive species identification [J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2005, 360: 1813-1823.
- [10] Chu K H, Xu M, Li C P. Rapid DNA barcoding analysis of large datasets using the composition vector method [J]. BMC Bioinformatics, 2009, 10(Suppl. 14): S8.
- [11] Hebert P D N, Cywinska A, Ball S L, et al. Biological identifications through DNA barcodes [J]. Proceedings of the Royal Society B: Biological Sciences, 2003, 270: 313-321.
- [12] Hebert P D N, Penton E H, Burns J M, et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(41): 14812-14817.
- [13] 陈念,赵树进. 入侵种的DNA条形码鉴定[J]. 生物技术通讯, 2009, 20(1): 135-137.
- [14] 陈士林,宋经元,姚辉,等. 药用植物DNA条形码鉴定策略及关键技术分析[J]. 中国天然药物, 2009, 7(5): 322-327.
- [15] 陈念,付晓燕,赵树进,等. DNA条形码: 物种分类和鉴定技术[J]. 生物技术通讯, 2008, 19(4): 629-631.
- [16] 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.
- [17] 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(3): 480-490.
- [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] 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(11): e1154.
- [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] Logacheva M D, Valiejo-Roman C M, Pimenov M G. *ITS* phylogeny of West Asian *Heracleum* species and related taxa of Umbelliferae-Tordylieae W. D. J. Koch, with notes on evolution of their *psbA-trnH* sequences [J]. Plant Systematics and Evolution, 2008, 270: 139-157.
- [22] Whipple I G, Barkworth M E, Bushman B S. Molecular insights into the taxonomy of *Glyceria* (Poaceae: Meliceae) in North America [J]. American Journal of Botany, 2007, 94(4): 551-557.

- [23] 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 of the United States of America, 2008, 105(8): 2923–2928.
- [24] Newmaster S G, Ragupathy S, Janovec J. A botanical renaissance: state-of-the-art DNA barcoding facilitates an Automated Identification Technology system for plants [J]. International Journal of Computer Applications in Technology, 2009, 35(1): 50–60.
- [25] Song J Y, Yao H, Li Y, et al. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique [J]. Journal of Ethnopharmacology, 2009, 124(3): 434–439.
- [26] Chen S L, Yao H, Han J P, et al. Validation of the *ITS2* region as a novel DNA barcode for identifying medicinal plant species [J]. PLoS One, 2010, 5(1): e8613.
- [27] 罗焜. 基于芸香科、天南星科的植物 DNA 条形码通用序列研究 [D]. 武汉: 湖北中医药大学, 2010.
- [28] 高婷. 利用 DNA 条形码技术鉴定药用双子叶植物: 以菊科、豆科为例 [D]. 北京: 中国医学科学院北京协和医院药用植物研究所, 2010.
- [29] 任保青, 陈之端. 植物 DNA 条形码技术 [J]. 植物学报, 2010, 45(1): 1–12.
- [30] Chase M W, Cowan R S, Hollingsworth P M, et al. A proposal for a standardised protocol to barcode all land plants [J]. Taxon, 2007, 56: 295–299.
- [31] Štorchová H, Olson M S. The architecture of the chloroplast *psbA-trnH* non-coding region in angiosperms [J]. Plant Systematics and Evolution, 2007, 268: 235–256.
- [32] 李耀利, 朱华, 杨俊波. 从 *rbcL* 序列探讨单室茛菪属的系统位置 [J]. 云南植物研究, 2002, 24(3): 352–358.
- [33] 王彦涵, 张寿州, 高建平, 等. 从叶绿体 DNA *rbcL* 序列分析探讨五味子科的系统发育 [J]. 复旦学报: 自然科学版, 2003, 42(4): 550–554.
- [34] 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 of the United States of America, 2005, 102(23): 8369–8374.
- [35] Chiou S J, Yen J H, Fang C L, et al. Authentication of medicinal herbs using PCR-amplified *ITS2* with specific primers [J]. Planta Medica, 2007, 73(13): 1421–1426.
- [36] Gonzalez M A, Baraloto C, Engel J, et al. Identification of Amazonian trees with DNA barcodes [J]. PLoS One, 2009, 4(10): e7483.
- [37] 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, 2007, 35(3): 1–8.
- [38] Jurado-Rivera J A, Vogler A P, Reid C A M, et al. DNA barcoding insect-host plant associations [J]. Proceedings of the Royal Society B: Biological Sciences, 2009, 276: 639–648.
- [39] 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 B: Biological Sciences, 2005, 360: 1889–1895.
- [40] Erickson D L, Spouge J, Resch A, et al. DNA barcoding in land plants: developing standards to quantify and maximize success [J]. Taxon, 2008, 57: 1304–1316.
- [41] Newmaster S G, Ragupathy S. Ethnobotany genomics: discovery and innovation in a new era of exploratory research [J]. Journal of Ethnobiology and Ethnomedicine, 2010, 6: 2.
- [42] Ragupathy S, Newmaster S G, Murugesan M, et al. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India [J]. Molecular Ecology Resources, 2009, 9(Suppl. 1): 164–171.
- [43] Sui X Y, Huang Y, Tan Y, et al. Molecular authentication of the ethnomedicinal plant *Sabia parviflora* and its adulterants by DNA barcoding technique [J]. Planta Medica, 2010, 76: 1–5.
- [44] Newmaster S G, Ragupathy S. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae) [J]. Molecular Ecology Resources, 2009, 9(Suppl. 1): 172–180.
- [45] Kumar S, Hahn F M, McMahan C M, et al. Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines [J]. BMC Plant Biology, 2009, 9: 1–12.
- [46] 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.
- [47] 杨绪勤, 邓传良, 刘丽盈, 等. 基于 *matK* 序列的木犀属植物系统发育初步研究 [J]. 北京林业大学学报, 2009, 31(6): 9–14.
- [48] 成玉, 陈成彬, 薛梅, 等. 应用 AFLP 技术探讨半夏属五个种的亲缘关系 [J]. 云南植物研究, 2006, 28(6): 559–564.
- [49] 葛家文, 王好转. DNA 条形码与生物分类学研究 [J]. 畜牧与饲料科学, 2009, 30(5): 52–53.
- [50] 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.
- [51] Meier R, Zhang G Y, Ali F. The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification [J]. Systems Biology, 2008, 57(5): 809–813.
- [52] Steinke D, Vences M, Salzburger W, et al. Tax I: a software tool for DNA barcoding using distance methods [J]. Philosophical Transactions of the Royal Society B: Biological Sciences, 2005, 360: 1975–1980.