

亚热带森林乔木树种 DNA 条形码研究 ——以哀牢山自然保护区为例*

卢孟孟^{1,2}, 慈秀芹^{1,2}, 杨国平¹, 李捷^{1**}

(1 中国科学院西双版纳热带植物园 植物系统发育与保护生物学实验室, 云南 昆明 650223;

2 中国科学院大学, 北京 100049)

摘要: 作为新一代植物志 iFlora 的重要组成部分, DNA 条形码已经成为物种鉴定中重要且有效的方法。本研究以亚热带森林的乔木树种为研究对象, 开展了 DNA 条形码的尝试性工作。为评估 DNA 条形码对鉴定亚热带森林树种的有效性, 收集并研究了来自哀牢山自然保护区内 51 科 111 属中 204 个树种的 525 个乔木个体。结果显示, 所选 4 个 DNA 片段 (*rbcL*, *matK*, *trnH-psbA* 和 ITS) 的 PCR 扩增成功率都超过 90%; 测序成功率 *rbcL* 和 *matK* 最高, 分别为 90.7% 和 90.5%, *trnH-psbA* 次之 (83.6%), ITS 最低 (73.5%), 表明 4 个片段在亚热带森林乔木中都具有较好的通用性。应用 BLAST 与 NJ Tree 两种方法, 对物种和属水平的鉴别成功率进行统计, 发现单片段中 ITS 最高, 分别为 68.4% ~ 81.3% 和 99.0% ~ 100%, 核心条码 *rbcL* 和 *matK* 组合的成功率是 52.8% ~ 60.2% 和 86.7% ~ 90.5%, 再与补充条码 *trnH-psbA* 和 ITS 联合, 可以成功鉴别 74.7% ~ 79.6% 哀牢山自然保护区亚热带森林中的乔木物种。由于 ITS 片段在亚热带森林部分重要树种类群 (樟科和壳斗科等) 中的测序成功率较差, 所以对这类植物类群采用 *trnH-psbA* 作为 DNA 条形码是一个更好的选择。

关键词: DNA 条形码; 亚热带森林; 哀牢山自然保护区; 物种鉴别; ITS; 乔木

中图分类号: Q 949, Q 781

文献标识码: A

文章编号: 2095-0845(2013)06-733-09

DNA Barcoding of Subtropical Forest Trees

——A Study from Ailao Mountains Nature Reserve, Yunnan, China

LU Meng-Meng^{1,2}, CI Xiu-Qin^{1,2}, YANG Guo-Ping¹, LI Jie^{1**}

(1 Laboratory of Plant Phylogenetics and Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Science, Kunming 650223, China; 2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: To evaluate how effective DNA barcoding is for the identification of subtropical forest trees, we sampled 525 individuals representing 204 species in 111 genera of 51 plant families that occur in the Ailao Mountains Nature Reserve, and tested the ability of *rbcL*, *matK*, *trnH-psbA* and ITS sequences to discriminate species. PCR success was over 90% for each of these four sequences, while sequencing success rate was highest for *rbcL* and *matK* (90.7% and 90.5%, respectively), followed by *trnH-psbA* (83.6%), and lowest for ITS (73.7%). Thus, all four sequences showed a relatively high level of applicability for subtropical forest trees that occur in the Ailaoshan Mountains Nature Reserve. Using two different “species identification” methods – BLAST and Neighbor Joining (NJ)—the highest rate of success for identification at species (68.4%–81.3%) and genus (99.0%–100%) levels was obtained using ITS when only a single region was used. When two molecular regions were used in combination, *rbcL* and *matK* correctly identified 52.8%–60.2% of species and 86.7%–90.5% of genera, while using all

* 基金项目: 中国科学技术基础性工作专项 (2012FY110400, 2011FY120200), 中国科学院热带森林生态学重点实验室的支持

** 通讯作者: Author for correspondence; E-mail: jieli@xtbg.ac.cn

收稿日期: 2013-09-04, 2013-10-14 接受发表

作者简介: 卢孟孟 (1984-) 男, 在读硕士研究生, 研究方向: 植物系统发育与保护生物学。E-mail: llm505945286@126.com

four regions in combination correctly discriminated 74.7%–79.6% of species. The relatively low sequencing success rate of ITS was mainly due to failure in certain groups (such as Lauraceae and Fagaceae), which play an important role in subtropical forest, suggesting that the ITS region may not be appropriate for DNA barcoding these particular plant groups.

Key words: DNA barcode; Subtropical forest trees; Ailao Mountains Nature Reserve; Species identification; ITS

森林生态系统类型的自然保护区建设, 不仅能有效保护大量的森林资源, 更重要的是保护具有典型性和代表性的森林生态系统, 进而在科学研究和改善生态环境方面发挥重要的作用 (薛达元和蒋明康, 1995)。对于这类保护区的建设, 植物多样性尤其是对构成森林生态系统主体的乔木树种的调查是非常重要的, 它是开展生态学及其他保护区监测工作的基础和必要步骤。然而, 野外工作是一项耗时费力的活动, 而针对高大乔木时难度更大, 物种的鉴定必须依赖有丰富野外经验的分类学家及熟悉该区域的野外工作人员, 但目前这类人才越来越匮乏 (Chase 和 Fay, 2009)。随着现代社会的发展、全球气候变化以及人类活动的加剧, 生物多样性正以空前的速率丧失 (Sala, 2000), 生物多样性保护的迫切性凸显出分类学力量的匮乏和传统分类学的局限。目前, 依赖 DNA 序列差异来识别物种的 DNA 条形码技术 (Hebert 等, 2003), 以其数据分析显示出了可靠性和客观性, 已成为运用分子生物学技术鉴定物种最经济和最有效的工具之一, 被广泛用于生物多样性的调查和评估 (Costion 等, 2011; Dincă 等, 2011; De Vere 等, 2012; Armenise 等, 2012; De Mattia 等, 2012; Dolman 等, 2012; Jin 等, 2013; Ji 等, 2013)。然而, 应用 DNA 条形码来鉴定和识别亚热带森林乔木树种的研究目前还鲜有报道。

国际生命 DNA 条形码联盟植物工作组 (CBOL Plant Working Group) 在 2009 年提出植物 DNA 条形码以叶绿体基因 *rbcL* 和 *matK* 作为核心条码, 同时以叶绿体基因间隔区 *trnH-psbA* 和核基因 ITS 作为补充条形码 (CBOL Plant Working Group, 2009), 这 4 个 DNA 片段都有各自的优缺点。核心条码 *rbcL* 具有很好的通用性, 但对物种的识别能力不强 (Gonzalez 等, 2009; CBOL Plant Working Group, 2009); 核心条码 *matK* 在 Lahaye 等 (2008) 和 De Vere 等 (2012) 的研究中显示

具有很好的物种鉴别作用, 但在鉴别小檗属 (*Berberis*) 和女贞属 (*Ligustrum*) 时却效果不好 (Roy 等, 2010; Gu 等, 2011); 而补充条码 *trnH-psbA* 也广受争议, Chase 等 (2007) 以及 Roy 等 (2010) 都认为 *trnH-psbA* 是一个很好的 DNA 条形码, 而 Sass 等 (2007) 的研究则反对把它列为 DNA 条形码; 补充条码 ITS 作为唯一的一个核基因片段, 受到质疑主要是因为其存在多拷贝、易受真菌污染以及测序困难等问题, 然而, 在最近的很多研究中 ITS 被认为是最有效的 DNA 条形码。Li 等 (2011) 通过对中国种子植物 75 科 141 属 1 757 种共约 6 286 个样本的研究, 发现 ITS 具有最高的物种分辨率; Tripathi 等 (2013) 的研究也表明 ITS 和 *trnH-psbA* 对印度热带树种来说是最适合的 DNA 条形码。尽管诸多学者的研究结果显示不同区域的物种鉴定成功率有差异, 但所有研究都认为 DNA 条形码对评估植物多样性是有效的。

位于滇中的哀牢山自然保护区分布着我国目前面积最大 (34 483 hm²)、且以云南特有植物物种为优势的亚热带常绿阔叶林, 保存着完整而稳定的亚热带山地森林生态系统。其中森林植被类型多样、生物多样性高、生物资源丰富, 蕴藏着许多珍稀濒危动植物资源, 如滇金丝猴、长臂猿、银杏、水青冈等, 具有重要的保护意义和研究价值 (哀牢山自然保护区综合考察团, 1988)。根据 2009 年出版的《云南哀牢山种子植物》记录, 该地区有种子植物 199 科、956 属和 2 242 种及 206 变种 (亚种) (朱华和闫丽春, 2009), 经统计发现其中乔木树种约为 600 种。为评估 DNA 条形码在亚热带森林乔木物种鉴定中的实际应用能力, 我们对该地区开展了常见乔木树种的 DNA 条形码尝试性研究, 其目标主要有两个: 一是检测 DNA 条形码在亚热带森林乔木树种中的通用性; 二是评估 DNA 条形码在哀牢山常见树种中的鉴定成功率。通过开展该项研究, 确立

DNA 条形码能否用于哀牢山地区常见乔木树种的鉴定；同时，以 DNA 条形码的大批量标准数据为核心，结合现代植物学知识和成果以及生物信息学，用 iFlora 的理念，形成具有图片信息、地理信息及 DNA 分子序列等为核心的可利用数据库（李德铎等，2012），以期为亚热带森林生物多样性调查和生态监测平台建设提供基础数据。

1 材料和方法

1.1 研究地点与样品采集

本研究以哀牢山自然保护区西坡徐家坝和东坡千家寨为主要研究区域，具体的采样地点信息见表 1。以哀牢山自然保护区亚热带森林中常见的乔木树种为研究对象，采样标准借鉴 Li 等（2011）采取的方法，尽可能每个物种选择 2 个或 2 个以上的个体。实验材料以硅胶干燥的新鲜叶片为主，少数无法采集到叶片的高大乔木或者采集季节落叶的树种选取树皮韧皮部作为 DNA 提取的材料，共计获得乔木 525 个个体。经专家鉴定，确定其隶属于 204 种（其中有 2 个个体未能鉴定到种）、111 属和 51 科（按 APG III 系统），本研究采集物种数目约占哀牢山地区乔木树种的三分之一（204/600）。由于各种条件的限制，其中有 66 个物种仅采到 1 个个体。

1.2 DNA 提取、扩增及测序

本研究中样品 DNA 的提取使用天根公司的植物总 DNA 提取试剂盒。选择叶绿体 *rbcL*、*matK* 以及基因间隔区 *trnH-psbA* 与核基因 ITS 4 个分子序列作为研究的扩增片段。参照国际生命 DNA 条形码联盟植物工作组推荐的 PCR 反应体系（CBOL Plant Working Group, 2009），并作出了优化和调整，研究中用到的引物信息见表 2。

所有扩增产物均送北京六合华大基因科技有限公司（BGI）完成测序工作。所得序列使用 Sequencher 4.14（Bromberg 等，1995）进行拼接和校对，*rbcL*、*matK* 和 ITS 这 3 个 DNA 片段在 Geneious 6.1.5（Biomatters Ltd.）中选择 ClustalX（Larkin 等，2007）排序，而 *trnH-psbA* 由于序列长度变化大，选用了软件 SATe（Liu 等，2009）进行排序。

1.3 数据分析

对 PCR 扩增成功率和测序成功率的数据统计，主要参考了 Kress 等（2009）的计算方法。PCR 扩增成功率是 PCR 扩增成功的个体数与总个体数的百分比值，测序成功率是获得的高质量序列数与总个体数的百分比值。对 DNA 条形码在哀牢山自然保护区亚热带森林乔木树种中的鉴定能力评估，研究中采用了两种常用方法，即 BLAST 和邻接树分析（Neighbor-Joining Tree, NJ）。对于 BLAST，首先在 Geneious 6.1.5（Biomatters Ltd.）中将 4 个 DNA 片段分别建立一个本地数据库（Burgess 等，2011），再把每条序列与数据库内的所有序列进行 BLAST，以相同位点的百分比（Identical Sites）作为量化标准。如果相同物种间的所有个体经过 BLAST，发现 Identical Sites 值都大于与其他所有物种个体间的 Identical Sites 值，那么我们认为这个物种的序列得到了准确鉴定。对于邻接树分析（NJ Tree）方法，通过构建 Tamura-Nei 距离下的 NJ 树，并完成 1 000 次运算以获得拓扑结构上的节点支持率，在评估一个物种是否被成功鉴定时，只有同一个物种的所有个体在构建的 NJ 系统树上形成单系分支，并且 Bootstrap value $\geq 70\%$ ，我们才认为该物种被成功鉴定。这两种方法的物种鉴定成功率是通过成功鉴定物种数与各自获得序列的总物种数的百分比值得到的。

表 1 哀牢山自然保护区采样信息表

Table 1 The information of samples collected in Ailao Mountains Nature Reserve

采集区 Collection areas	植被类型 Vegetation type	采集地 Locality	海拔 Attitude/m	纬度 Latitude	经度 Longitude
徐家坝(西坡)	季风常绿阔叶林	响水河	1393	24°27'9.5"	100°54'55"
		道班、三棵桩	1460	24°26'13"	100°54'09"
		大黑丫口	2015	24°30'12"	100°53'01"
	中山湿性常绿阔叶林	生态站入口旁	2395	24°31'07"	101°00'53"
		杜鹃湖	2500	24°32'12"	101°01'36"
		小新厂低	2522	24°32'57"	101°01'37"
	苔藓矮林	三棵树	2666	24°32'10"	101°01'52"
		小新厂顶	2720	24°33'48"	101°01'26"
千家寨(东坡)	半湿性常绿阔叶林	千家寨	2000 ~ 2400	24°14'09"	101°17'08"
	中山湿性常绿阔叶林	千家寨	2400 ~ 2600	24°14'45"	101°17'49"

表 2 DNA 条形码片段扩增使用的引物信息表

Table 2 The information of primers amplified DNA barcodes

DNA 片段 Region	引物对 Primers	序列 Sequences (5'-3')	来源 References
<i>rbcL</i>	<i>rbcL1F</i>	ATGTCACCACAAACAGAGACTAAAGC	Ivanova 等, 2008
	<i>rbcL724R</i>	TCGCATGTACCTGCAGTAGC	Ivanova 等, 2008
<i>matK</i>	<i>matK3F</i>	CGTACAGTACTTTTGTGTTTACGA	Kim (unpublished)
	<i>matK1R</i>	ACCCAGTCCATCTGGAAATCTTGG	Kim (unpublished)
	<i>matK472F</i>	CCCRTYCATCTGGAAATCTTGGTTC	Yu 等, 2011
	<i>matK1248R</i>	GCTRTRATAATGAGAAAGATTCTGC	Yu 等, 2011
<i>trnH-psbA</i>	<i>trnH2</i>	CGCATGGTGGATTACAATCC	Fazekas 等, 2010
	<i>psbA</i>	GTTATGCATGAACGTAATGCTC	Sang 等, 1997
ITS	ITS5P	GGAAGGAGAAGCTGTAACAAGG	Möller and Cronk, 1997
	ITS8P	CACGCTTCTCCAGACTACA	Möller and Cronk, 1997
	ITS4	TCCTCCGCTTATTGATATGC	White 等, 1990

2 结果

2.1 序列通用性

对哀牢山自然保护区亚热带森林中采集到 204 个物种的 525 个个体进行序列统计, 发现 4 个 DNA 片段共获得可用序列 1 776 条, 有 98.5% 的个体得到至少一条序列, 仅有 8 个个体没能得到任何一条序列。在使用的 4 个 DNA 片段中, PCR 扩增成功率是 ITS 最高, 为 97.0%, 其次是 *matK* (93.1%) 和 *rbcL* (93.0%), *trnH-psbA* (90.5%) 最低。但测序成功率却是 *rbcL* 和 *matK* 最高, 分别为 90.7% 和 90.5%, *trnH-psbA* (83.6%) 次之, 最低的是核基因 ITS, 仅有 73.5% (表 3)。317 个个体 (60.4%) 得到了 4 个基因片段的序列, 代表了在哀牢山自然保护区亚热带森林中收集到的 76.5% 的物种和 84.7% 的属。

2.2 物种鉴定成功率

本研究应用 BLAST 和邻接树 (NJ Tree) 两种分析方法, 比较了每个片段及其相互组合对哀牢山自然保护区亚热带森林树种的物种和属的鉴定成功率。BLAST 分析结果显示, 在至少得到一条序列的 517 个个体中, 单个 DNA 片段中 ITS 对物种和属水平的鉴别成功率最高为 81.3% 和 99.0%, 其次依次为 *trnH-psbA* (64.6%, 93.3%) 和 *matK* (51.5%, 91.7%), *rbcL* 最低, 只有 49.5% 和 82.6%。NJ 树的统计结果是 ITS 的物种鉴别率最高 (69.1%), 其次是 *rbcL* (47.7%) 和 *trnH-psbA* (43.6%), 而 *matK* (42.3%) 最低 (图 1); 对属的鉴别率依然是 ITS (100%) 最高, *matK* (81.1%) 次之, 而 *rbcL* (76.3%)

和 *trnH-psbA* (75.0%) 最低 (图 1)。与此同时, 研究结果显示, 在数据库完全一致的情况下, 单个片段的物种鉴定成功率, BLAST 的结果普遍高于 NJ 树的结果。

对多片段的不同组合进行统计分析, 发现核心条码组合, 即 *rbcL+matK*, 可以成功鉴定 52.8% ~ 60.2% 的物种和 86.7% ~ 90.5% 的属 (图 1); 补充条码 *trnH-psbA* 与核心条码的组合, 使物种水平的鉴别能力上升到了 60.0% ~ 63.0%, 提高了 2.8% ~ 7.2%; ITS 与核心条码的组合, 物种的鉴别能力为 72.7% ~ 78.0%。四个片段的联合使用, 可以对 95% 以上的属完成准确鉴定, 物种水平的鉴别能力也提升到了 74.7% ~ 79.6% (图 1)。这些数据表明 DNA 条形码对哀牢山自然保护区亚热带森林乔木树种的鉴定具有很好的实用性。

在研究中注意到并不是所有的物种都可以被 DNA 条形码有效地区分, 即使是 4 个 DNA 片段的组合。研究发现诸如樟科 (Lauraceae)、壳斗科 (Fagaceae)、冬青科 (Aquifoliaceae)、山矾科 (Symplocaceae) 等植物类群在单片段、多片段的组合条件下都不能被有效地区分开 (表 4)。我们尤其注意到, 在樟科和壳斗科等植物类群中 ITS 的测序成功率较低。樟科和壳斗科的样本量占总样品数的 24.2% (127/525), 但 ITS 测序成功率却仅有 30.7% (39/127) (表 5)。在这些无法获得高质量 ITS 序列的特定植物类群中, *trnH-psbA* 片段比核心条码表现出了较高的种间分辨率, 在一定程度上可以提高对这些类群的物种识别能力 (表 4)。

表 3 哀牢山自然保护区亚热带森林树种的扩增和测序成功率

Table 3 PCR and sequencing results for subtropical forest trees in Ailao Mountains Nature Reserve

	<i>rbcL</i>		<i>matK</i>		<i>trnH-psbA</i>		ITS		<i>rbcL+matK</i>		<i>rbcL+matK</i> <i>+trnH-psbA</i>		<i>rbcL+matK</i> <i>+ITS</i>		All four
	PCR	SEQ	PCR	SEQ	PCR	SEQ	PCR	SEQ	Either SEQ	Both SEQ	Either SEQ	Three SEQ	Either SEQ	Three SEQ	All SEQ
No. of samples (525)/%	488	476	489	475	475	439	504	386	498	449	501	417	516	335	317
	0.93	0.907	0.931	0.905	0.905	0.836	0.97	0.735	0.949	0.855	0.954	0.794	0.983	63.8	0.604
No. of species (204)/%	201	199	201	197	196	189	201	172	203	191	203	184	204	162	153
	0.986	0.975	0.986	0.966	0.961	0.926	0.986	0.843	0.995	0.937	0.995	0.902	1.0	0.794	0.75
genus(111)	—	109	—	109	—	104	—	102	111	105	111	102	111	99	95

表 4 DNA 条形码在樟科、壳斗科、山矾科和冬青科的物种鉴别能力

Table 4 The ability of species identification for DNA barcodes in Lauraceae, Fagaceae, Symplocaceae and Aquifoliaceae

	樟科 Lauraceae	壳斗科 Fagaceae	山矾科 Symplocaceae	冬青科 Aquifoliaceae
哀牢山地区物种数 No. of species in Ailao Mountains	51	43	14	17
采集到的物种数 No. of species collected	25	21	8	6
BLAST 成功 鉴定物种数 No. of species identified success	<i>rbcL</i>	6	2	0
	<i>matK</i>	8	2	0
	<i>trnH-psbA</i>	14	6	2
	ITS	12	14	2
	<i>rbcL+matK</i>	10	4	0
	<i>rbcL+matK+trnH-psbA</i>	13	5	2
	All	10	12	3

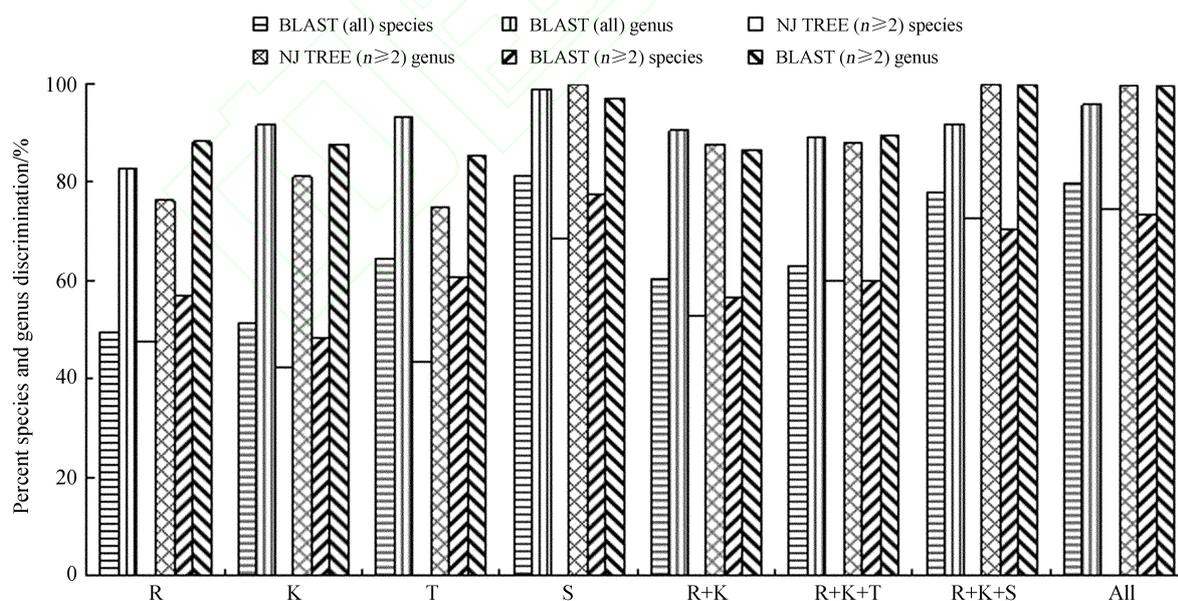


图 1 单片段和多片段组合在哀牢山自然保护区物种和属水平的鉴别成功率

R、K、T、S 分别代表 *rbcL*、*matK*、*trnH-psbA*、ITS；all 代表由所有物种组成的数据库； $n \geq 2$ 代表由含有 2 个或 2 个以上个体的所有物种组成的数据库

Fig. 1 Species and Genus discrimination rate of all single- and multi-DNA regions in Ailao Mountains Nature Reserve

R, *rbcL*; K, *matK*; T, *trnH-psbA*; S, ITS; NJ tree, neighbor-joining tree; all, the database consistedof all species; $n \geq 2$, the database consisted of species including two or more samples

表5 研究中涉及到的物种数超过5的科的ITS通用性分析

Table 5 The universality analysis of ITS for families including five or more species in this study

	樟科 Lauraceae	壳斗科 Fagaceae	杜鹃花科 Ericaceae	蔷薇科 Rosaceae	山茶科 Theaceae	唇形科 Lamiaceae	五列木科 Pentaphylacaceae	山矾科 Symplocaceae	冬青科 Aquifoliaceae
哀牢山地区物种数 No. of species in Ailao Mountains	51	43	56	50	19	34	11	14	17
采集物种数 No. of species collection	25	21	15	14	9	8	8	8	6
采集个体数 No. of samples collection	64	63	46	34	28	15	26	32	19
ITS 测序成功物种数 No. of species of ITS sequencing success	11	13	15	9	7	8	8	8	6
ITS 测序成功个体数 No. of samples of ITS sequencing success	19	20	43	25	17	14	26	29	19

3 讨论

本研究是针对哀牢山自然保护区亚热带森林乔木物种开展区域植物 DNA 条形码研究的一次尝试,其目的在于评估 DNA 条形码对一个地区植物类群在引物通用性、序列质量和物种鉴定能力等方面的总体表现。

3.1 DNA 条形码通用性

引物的通用性是评价一个 DNA 条形码理想与否的重要指标 (Cowan 等, 2006; CBOL Plant Working Group, 2009)。在我们的研究中发现 *matK* 片段的扩增和测序成功率高达 93.1% 和 90.5%, 与 Hollingsworth (2008)、Chase 等 (2007) 以及 Kress 等 (2009) 的研究相比, 表现出更高的通用性和成功率, 即使与 Fazekas 等 (2008) 用 10 对引物扩增得到的序列结果 (88%) 相比, 也要高出不少; 该结果接近 Li 等 (2011) 对分布于中国的 1757 种种子植物 DNA 条形码研究所报道的 *matK* 的扩增和测序成功率 (91%, 95.3%)。这在一定程度上应该归功于具有较高通用性的不同引物的使用。我们所使用的两对引物是近年来在 DNA 条形码研究中筛选和设计出来的, 具有更短的序列长度, 避免了扩增片段中单核苷酸重复序列的出现, 从而提高了成功率。在这两对引物中, 我们优先使用 *matK3F* 和 *matK1R*, 可以对哀牢山自然保护区亚热带森林中 60% 左右的乔木物种进行成功扩增与测序, 具有普遍的适用性, 应作为 *matK* 片段扩增的首选引物; 而增加另一对引物—*matK472F* 和

matK1248R (Yu 等, 2011), 可以把该片段对哀牢山亚热带森林乔木物种的引物通用性提升到 90% 左右, 因此 *matK* 所存在的通用性不好这一缺点可以通过采用尝试不同引物的方法来解决。

在本研究中, 核心条码 (*rbcL* 和 *matK*) 在亚热带森林乔木树种中表现出了最好的通用性, 与之相比, 补充条码的测序成功率不太理想。补充条码 *trnH-psbA* 测序不成功的主要原因是部分物种中存在单核苷酸重复序列, 通常表现为几个甚至是十几个碱基 A 或 T 的连续重复。目前所使用的 Sanger 测序方法, 容易造成重复序列后的峰图混乱, 导致多核苷酸重复序列数目的不确定及该结构之后的序列不可用 (Devey 等, 2009)。ITS 测序成功率在所有 DNA 片段中最低 (73.5%), 但从已发表的数据来看, Gonzalez 等 (2009) 对亚马逊热带森林 285 个树种的研究结果显示, ITS 扩增和测序成功率为 41%; Tripathi 等 (2013) 的研究表明, ITS 序列在印度热带森林树种的扩增与测序成功率是 74.0% 和 62.0%, 与之相比, 我们的结果要理想很多。哀牢山自然保护区内的亚热带常绿阔叶林主要是由壳斗科、樟科、山茶科和木兰科等组成, 在我们的研究中 ITS 片段测序成功率低, 主要是受到壳斗科和樟科的影响。这两个科的个体数约占总样本的 1/4, 却仅有 30.7% 的测序成功率。如果去除樟科和壳斗科的物种, ITS 的测序成功率会出现显著升高, 从现在的 73.5% (386/525) 上升到 91.0% (362/398)。其他学者 Simeone 等 (2013) 和 Denk 和

Grimm (2010) 对壳斗科栎属 (*Quercus*) 的研究以及 Abeysinghe 等 (2009) 对樟科樟属 (*Cinnamomum*) 的研究, 都表明 ITS 是一个多拷贝片段, 存在不完全的一致性进化现象 (incomplete concerted evolution)。因此, 在这些特定类群中, 对 ITS 扩增产物直接测序往往得不到好的结果, 可能需要克隆才能得到高质量序列。

3.2 物种鉴定能力

研究发现, 不同的分析方法对评估 DNA 条形码鉴别物种的能力上是存在差异的。BLAST 方法主要是通过比较两两序列间的位点变化来区分物种, 对只有一个个体的物种也具有适用性, 而邻接树方法适用于含有多个个体的物种。在数据库一致的情况下, 对单个 DNA 片段物种的鉴定成功率比较发现, BLAST 得到的物种鉴定率均高于 NJ 树的结果, 这与两种方法本身的工作原理有关。前者仅考虑位点之间的差异, 后者还要考虑到遗传距离模型以及是否构成单系类群。事实上, 目前很多类群的分类系统无法保证形态学上的种都是单系发生的类群, 此外还存在隐存种 (cryptic species) 现象 (Liu 等, 2011), 这无疑也会使得利用 NJ 树的方法容易获得较低的物种鉴定成功率。多片段的组合分析在运用 BLAST 方法时, 有时反而会降低物种的鉴定成功率, 如 *rbcL+matK* 在属水平的鉴定成功率 (90.5%) 比 *rbcL+matK+trnH-psbA* (89.2%) 的组合要高, 主要原因是 *trnH-psbA* 存在较多的插入和缺失位点, 降低了同种个体间 Identical Sites 值, 从而影响了物种鉴定成功率。

3.2.1 核心条码具有较低的物种鉴定成功率

虽然有学者曾报道核心条码具有较高的物种鉴定能力, 如 Burgess 等 (2011) 在加拿大温带植物群的研究中, 发现核心条码可以成功鉴定 93% 的物种, 以及 De Vere 等 (2012) 发现 *rbcL+matK* 可以鉴别威尔士境内 69.4% ~ 74.1% 的有花植物, 但是我们的研究结果没能达到那么高的物种鉴别水平。在属水平上, 核心条码的组合对哀牢山自然保护区的亚热带森林乔木具有较好的识别作用 (86.7% ~ 90.5%), 但在物种水平的鉴别能力偏低, 仅有 52.8% ~ 60.2%, 接近 Li 等 (2011) 对分布于中国的 1 757 种种子植物的研究结果 (BLAST: 60.8%)。对此存在以下几种可能

的解释, 一是该地区核心条码分子序列的遗传分化率低, 缺乏足够的信息位点, 因而不能将物种有效地区分开 (裴男才等, 2011), 这与哀牢山地区的一些类群近缘种较多, 种间分化时间较短有关, 核心条码较好的通用性也从侧面支持这种解释, 上述可能是造成核心条码对哀牢山地区乔木物种鉴定成功率低的主要原因; 二是对有些样品的形态学鉴定可能不太准确, 主要是采集的部分标本只采到了营养枝条, 缺乏花和果等重要的形态分类特征, 使得对其鉴定到物种存在很大的难度; 三是形态学上物种的界定和认识与 DNA 序列本身存在差异, 而在哀牢山地区对相关类群的分类与系统学研究工作开展得相对较少, 形态学与系统发育物种之间的关系没有得到有效的纠正。

3.2.2 补充条码具有较高的物种鉴定成功率

补充条码基因间隔区 *trnH-psbA* 片段是备受学者争议的片段 (Hollingsworth 等, 2011)。尽管 *trnH-psbA* 在不同植物类群间碱基数目差异很大, 存在大量的插入和缺失造成排序困难以及在某些类群中存在单核苷酸重复序列等特殊结构造成测序困难, 但我们的研究发现该片段在哀牢山自然保护区亚热带森林乔木树种中具有较核心条码 *rbcL* 和 *matK* 更高的种间分辨率。前人已开展的植物 DNA 条形码研究, 也发现 *trnH-psbA* 对乔木类群的鉴别是最具有希望与潜力的条形码之一 (Gonzalez 等, 2009; Tripathi 等, 2013)。*trnH-psbA* 片段与核心条码的组合, 可以实现优势互补, 相比于核心条码 *rbcL+matK* 的联合, *trnH-psbA* 的加入对哀牢山自然保护区亚热带乔木物种水平的解决能力提高到了 60.0% ~ 63.0%, 应用 BLAST 方法对物种鉴定成功率提高了 7.2% (图 1), 与 Kress 等 (2009) 在巴拿马样地的研究 (提高 6%) 相近, 高于 Burgess 等 (2011) 的研究 (提高 2.2%)。此外, 在一些特殊的植物类群中, 如樟科、壳斗科等, 很难获得高质量的 ITS 序列, *trnH-psbA* 要比核心条码 *rbcL* 或 *matK* 更能发挥积极作用 (表 4), 因此, *trnH-psbA* 应该被选为哀牢山亚热带森林的乔木 DNA 条形码之一。

补充条码核基因 ITS 尽管在本研究中有较低的通用性, 但我们的研究显示其在物种水平的鉴别上有着最好的可用性 (图 1)。ITS 在所有单片段中具有最高的物种识别能力 (68.4% ~

81.3%)，与核心条码联合可以提高约 17.8% ~ 19.3% 的物种鉴定成功率。ITS 高鉴定能力主要是因为其进化速率快可以鉴别和区分近缘种，与 Kress 等 (2005)、Sass 等 (2007) 的研究认为 ITS 是最可能被定为植物 DNA 条形码的结论是一致的。现在 ITS 被否决作为一个植物条形码的主要因素有以下几点：不完全的协同进化、真菌污染以及低扩增与测序成功率等。但我们的研究发现，上述问题主要是集中在樟科、壳斗科等特定类群，在绝大多数其他类群中并没有出现，所以，本研究支持 Li 等 (2011) 将补充条码 ITS 并入到核心条码的结论。此外，ITS 作为植物 DNA 条形码唯一的一个核基因片段，遗传信息来自于双亲，较之叶绿体基因可以提供更多的信息。因此，应该把 ITS 片段列为亚热带森林乔木树种的 DNA 条形码之一，但在特定植物类群 (如壳斗科、樟科等) 中应谨慎使用。

4 小结

通过对哀牢山自然保护区乔木树种的 DNA 条形码研究发现，核心条码 *rbcL+matK* 物种鉴定成功率较低，辅以补充条形码，*rbcL+matK+trnH-psbA+ITS* 的组合可以鉴别哀牢山自然保护区亚热带森林中 74.7% ~ 79.6% 的乔木物种和 95% 以上的属，具有很好的实用价值。在哀牢山自然保护区开展区域 DNA 条形码研究，是该地区 iFlora 工作的重要组成部分，不仅为国家级哀牢山自然保护区的资源普查提供本底分子数据资料，同时也促进对亚热带森林的生物多样性、群落生态学和保护生物学等领域的研究。该项研究再一次验证了 DNA 条形码作为一种标准、快速识别物种的方法是可行的。区域性 DNA 条形码的研究工作能够促进 iFlora 的快速完成，在地区生物多样性的保护及生态学研究发挥极大的作用。在实际应用中，DNA 条形码通过结合地理学或者形态学数据，即使是在研究背景相对不充分的森林群落中，也能帮助我们提高物种识别率 (Gonzalez 等, 2009)，对目前分类学人才匮乏的现状来说是一种经济且有效的手段，而 DNA 条形码遇到的各种难题，随着生物信息学和新一代测序技术的不断发展，也会得到逐步地解决和完善。

致谢 野外考察得到了云南省林业厅、哀牢山国家级自然保护区景东管理局和镇沅管理局、中国科学院西双版纳热带植物园哀牢山生态站的大力支持；物种鉴定工作由中国科学院昆明植物研究所李锡文研究员完成；采样工作得到了植物系统发育与保护生物学实验室李朗博士、李恒、黄小翠、刘志芳等人的协助，在此一并表示感谢。

[参 考 文 献]

- 哀牢山自然保护区综合考察团, 1988. 哀牢山自然保护区综合考察报告集 [M]. 昆明: 云南民族出版社
- 朱华, 闫丽春, 2009. 云南哀牢山种子植物 [M]. 云南: 云南科技出版社
- Abeyasinghe PD, Wijesinghe KGG, Tachida H *et al.*, 2009. Molecular characterization of Cinnamon (*Cinnamomum verum* Presl) accessions and evaluation of genetic relatedness of Cinnamon species in Sri Lanka based on *trnL* intron region, intergenic spacers between *trnT-trnL*, *trnL-trnF*, *trnH-psbA* and nuclear ITS [J]. *Research Journal of Agriculture and Biological Sciences*, **5** (6): 1079—1088
- Armenise L, Simeone MC, Piredda R *et al.*, 2012. Validation of DNA barcoding as an efficient tool for taxon identification and detection of species diversity in Italian conifers [J]. *European Journal of Forest Research*, **131** (5): 1337—1353
- Bromberg C, Cash H, Curtis P *et al.*, 1995. Sequencher. Gene Codes Corporation. Ann Arbor, Michigan
- Burgess KS, Fazekas AJ, Kesanakurti PR *et al.*, 2011. Discriminating plant species in a local temperate flora using the *rbcL* plus *matK* DNA barcode [J]. *Methods in Ecology and Evolution*, **2**: 333—340
- CBOL Plant Working Group, 2009. A DNA barcode for land plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (31): 12794—12797
- Chase MW, Cowan RS, Hollingsworth PM *et al.*, 2007. A proposal for a standardised protocol to barcode all land plants [J]. *Taxon*, **56** (2): 295—299
- Chase MW, Fay MF, 2009. Barcoding of plants and fungi [J]. *Science*, **325** (5941): 682—683
- Costion C, Ford A, Cross H *et al.*, 2011. Plant DNA barcodes can accurately estimate species richness in poorly known floras [J]. *PLoS ONE*, **6** (11): e26841
- Cowan RS, Chase MW, Kress WJ *et al.*, 2006. 300 000 species to identify: problems, progress, and prospects in DNA barcoding of land plants [J]. *Taxon*, **55** (3): 611—616
- De Mattia F, Gentili R, Bruni I *et al.*, 2012. A multi-marker DNA barcoding approach to save time and resources in vegetation surveys [J]. *Botanical Journal of the Linnean Society*, **169** (3): 518—529
- De Vere N, Rich TCG, Ford CR *et al.*, 2012. DNA barcoding the native flowering plants and conifers of Wales [J]. *PLoS ONE*, **7** (6): e37945
- Denk T, Grimm GW, 2010. The oaks of Western Eurasia: traditional classifications and evidence from two nuclear marks [J]. *Taxon*, **59** (2): 351—366

- Devey DS, Chase MW, Clarkson JJ, 2009. A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions [J]. *Taxon*, **58** (1): 7—15
- Dincú V, Zakharov EV, Hebert PDN *et al.*, 2011. Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe [J]. *Proceedings of the Royal Society B: Biological Sciences*, **278** (1704): 347—355
- Dolman PM, Panter CJ, Mossman HL, 2012. The biodiversity audit approach challenges regional priorities and identifies a mismatch in conservation [J]. *Journal of Applied Ecology*, **49** (5): 986—997
- Fazekas AJ, Burgess KS, Kesanakurti PR *et al.*, 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well [J]. *PLoS ONE*, **3** (7): e2802
- Fazekas AJ, Steeves R, Newmaster SG, 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats [J]. *Biotechniques*, **48** (4): 277—285
- García-Robledo C, Erickson DL, Staines CL *et al.*, 2013. Tropical plant-herbivore networks: reconstructing species interactions using DNA barcodes [J]. *PLoS ONE*, **8** (1): e52967
- Gonzalez MA, Baraloto C, Engel J *et al.*, 2009. Identification of Amazonian trees with DNA barcodes [J]. *PLoS ONE*, **4** (10): e7483
- Gu J, Su JX, Lin RZ *et al.*, 2011. Testing four proposed barcoding markers for the identification of species within *Ligustrum* L. (Oleaceae) [J]. *Journal of Systematics and Evolution*, **49** (3): 213—224
- Hebert PDN, Cywinska A, Ball SL *et al.*, 2003. Biological identifications through DNA barcodes [J]. *Proceedings of the Royal Society of London Series B: Biological Sciences*, **270** (1512): 313—321
- Hollingsworth PM, 2008. DNA barcoding plants in biodiversity hot spots: progress and outstanding questions [J]. *Heredity*, **101** (1): 1—2
- Hollingsworth PM, Graham SW, Little DP, 2011. Choosing and using a plant DNA barcode [J]. *PLoS ONE*, **6** (5): e19254
- Ivanova NV, Fazekas AJ, Hebert PDN, 2008. Semi-automated, membrane-based protocol for DNA isolation from plants [J]. *Plant Molecular Biology Reporter*, **26** (3): 186—198
- Ji Y, Ashton L, Pedley SM *et al.*, 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding [J]. *Ecology Letters*, **16** (10): 1245—1257
- Jin Q, Han H, Hu X *et al.*, 2013. Quantifying species diversity with a DNA barcoding-based method: Tibetan Moth species (Noctuidae) on the Qinghai-Tibetan plateau [J]. *PLoS ONE*, **8** (5): e64428
- Kress WJ, Wurdack KJ, Zimmer EA *et al.*, 2005. Use of DNA barcodes to identify flowering plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (23): 8369—8374
- Kress JW, Erickson DL, Jones FA *et al.*, 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (44): 18621—18626
- Larkin MA, Blackshields G, Brown NP *et al.*, 2007. Clustal W and Clustal X version 2.0 [J]. *Bioinformatics*, **23** (21): 2947—2948
- Lahaye R, van der Bank M, Bogarin D *et al.*, 2008. DNA barcoding the floras of biodiversity hotspots [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (8): 2923—2928
- Li DZ, Gao LM, Li HT *et al.*, 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (49): 19641—19646
- Li DZ (李德铎), Wang YH (王雨华), Yi TS (伊廷双) *et al.*, 2012. The next generation Flora: iFlora [J]. *Plant Diversity and Resources* (植物分类与资源学报), **34** (6): 525—531
- Liu K, Raghavan S, Nelesen S *et al.*, 2009. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees [J]. *Science*, **324** (5934): 1561—1564
- Möller M, Cronk Q, 1997. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences [J]. *American Journal of Botany*, **84** (7): 956—956
- Pei NC (裴男才), Zhang JL (张金龙), Mi XC (米湘成) *et al.*, 2011. Plant DNA barcodes promote the development of phylogenetic community ecology [J]. *Biodiversity Science* (生物多样性), **19** (3): 284—294
- Roy S, Tyagi A, Shukla V *et al.*, 2010. Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species [J]. *PLoS ONE*, **5** (10): e13674
- Sala OE, 2000. Global biodiversity scenarios for the year 2100 [J]. *Science*, **287** (5459): 1770—1774
- Sang T, Crawford D, Stuessy T, 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paonia* (Paeoniaceae) [J]. *American Journal of Botany*, **84** (8): 1120
- Sass C, Little DP, Stevenson DW *et al.*, 2007. DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of cycads [J]. *PLoS ONE*, **2** (11): e1154
- Simeone MC, Piredda R, Papini A *et al.*, 2013. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, Fagaceae): problems, prospects and phylogenetic implications [J]. *Botanical Journal of the Linnean Society*, **172**: 478—499
- Tripathi AM, Tyagi A, Kumar A *et al.*, 2013. The internal transcribed spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India [J]. *PLoS ONE*, **8** (2): e57934
- White TJ, Bruns T, Lee S *et al.*, 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics [J]. *PCR Protocols: A Guide to Methods and Applications*, **18**: 315—322
- Xue DU (薛达元), Jiang MK (蒋明康), 1995. Contributions of nature reserves in China to biodiversity conservation [J]. *Journal of Natural Resources* (自然资源学报), **10** (3): 286—292
- Yu J, Xue JH, Zhou SL, 2011. New universal *matK* primers for DNA barcoding angiosperms [J]. *Journal of Systematics and Evolution*, **49** (3): 176—181