

植物叶绿体 *ycf15* 基因应用于药用植物鉴定的前景展望

高娜娜, 赵志礼*, 倪梁红

上海中医药大学中药学院 生药学教研室, 上海 201203

摘要: 叶绿体普遍存在于绿色植物中, *ycf15* 基因位于叶绿体基因组 IR 区, 长度一般为 250~550 bp, 高度保守, 目前对该基因研究较少。就有限的研究成果来看, 其独特的序列结构及其显现的丰富多样的遗传信息不仅具有一定的物种鉴定意义, 而且在植物系统进化分析研究中有较大的潜力。同时, *ycf15* 基因具有保守性高、片段长度较小、无需克隆测序等特点, 在进一步构建中药材鉴定 DNA 条形码中具有独特优势。从 *ycf15* 基因的结构特点、物种鉴定与系统学意义几方面进行梳理并展望其在药用植物鉴定中的应用前景。

关键词: *ycf15* 基因; 叶绿体基因组; DNA 分子鉴定; 系统学意义; DNA 条形码

中图分类号: R282.1 **文献标志码:** A **文章编号:** 0253-2670(2017)15-3210-08

DOI: 10.7501/j.issn.0253-2670.2017.15.029

Prospect: Identification of medicinal plant based on chloroplast gene *ycf15*

GAO Na-na, ZHAO Zhi-li, NI Liang-hong

School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

Abstract: The circular double-stranded cpDNA sequence consists of two inverted repeat (IR) regions, a large single-copy region (LSC) and a small single-copy region (SSC). The function of some genes within the genome is clear. However, it is unknown for such chloroplast genes as hypothetical chloroplast reading frames (*ycf*s) and so on. As one of *ycf*s, *ycf15* gene is located within IR regions with length of 250—550 bp, conserved among plants and also, its function is often unclear. The gene has attracted an intense attention during the past decade. This review presents an overview of *ycf15* gene including its basic type, structural features, and the application to phylogenetic analyses. As a potential DNA marker for species identification, future implications of the gene are also introduced.

Key words: *ycf15*; chloroplast genome; DNA-based identification; phylogenetic significance; DNA barcoding

叶绿体在植物细胞内拥有自身完整的一套基因组, 其功能与光合作用密不可分。因叶绿体基因组大小、结构和基因种类一般较为保守, 现已成为研究药用植物系统进化及进行物种鉴定的有力工具^[1]。目前, 一些叶绿体基因的功能已被认识, 但相当一部分尚不清楚, 如一大类称之为开放阅读框 (open reading frames, ORFs) 的基因。其 DNA 序列具编码氨基酸的特性, 但具体编码的蛋白质名称及其功能还在探究之中。起初, 人们发现一段基因片段, 就分别给予命名, 如藜科菠菜 *Spinacia oleracea* L. 的 ORF33 基因及禾本科小麦 *Triticum aestivum* L. 的 ORF38 基因等。进一步研究之后, 注意到 ORF 命名方法有诸多不妥; 因此提出一种新的 “*ycf*”

(hypothetical chloroplast open reading frame) 命名法。就上述菠菜的 ORF33 与小麦的 ORF38, 经分析比对, 发现它们氨基酸序列相似度极高, 故归并为 *ycf8* 基因^[2]。

目前就叶绿体 *ycf* 基因家族研究进展来看, 该类基因在植物系统发育研究领域具有很大的应用潜力及多样的功能。例如, 一般认为 *ycf* 基因具有高度保守的特性, 但在兰科系统发育研究中, *ycf1* 基因显示了丰富的系统发育信息, 并具物种鉴定意义。*ycf1* 各部分高度变异, 在简约信息总数及百分变异方面 *ycf1* 比 *matK* 变异性更高。尽管 nrITS 区比 *ycf1* 变异度更大, 但在对位分析等方面后者更具优势。基于上述研究, *ycf1* 在兰科系统发育研究中的价值

收稿日期: 2017-04-19

基金项目: 国家自然科学基金资助项目 (81173654, 81503354)

作者简介: 高娜娜 (1992—), 女, 硕士研究生, 研究方向为中药 DNA 分子鉴定。E-mail: gnntcm@126.com

*通信作者 赵志礼, 教授, 博士生导师, 主要从事中药资源与品种鉴定工作。Tel: (021)51322202 E-mail: zhilzhao@sohu.com

得到肯定,并有望在其他植物类群内得到应用^[3]。Peter 等^[4]报道了与叶绿素生物合成相关的 2 个 *ycf* 基因 *ycf53* 和 *ycf59* 的功能。*ycf59* 编码 Mg-原卟啉单甲酯环化酶 (Mg protoporphyrin monomethylester cyclase) 的基本酶结构部分,而 *ycf53* 编码叶绿素生物合成中翻译后的调节剂。

作为 *ycf* 家族的另一种常见基因, *ycf15* 位于叶绿体基因组 IR 区,长度为 250~550 bp,高度保守,在烟草^[5]叶绿体基因组中首次被发现,当时注释为“ORF87 基因”。在拟南芥^[6]叶绿体基因组中曾被注释为“ORF77”。然而 *ycf15* 是否具有编码蛋白特性,一直饱受争议^[7-10]。Nguyen 等^[11]研究发现该基因作为新的分子标记可以有效鉴定百合科的秋水仙 *Colchicum autumnale* L. 和嘉兰 *Gloriosa superba* L.; 该基因还具有不可忽视的系统学意义^[9,12]。由此可见, *ycf15* 基因有望作为新的 DNA 条形码 (DNA barcoding) 候选序列发挥其物种鉴定作用; 同时,其功能的探究亦有广阔的研究空间。本文就 *ycf15* 基因的基本结构、物种鉴定与系统学意义进行综述,并展望其在药用植物鉴定中的应用前景。

1 *ycf15* 基因来源与功能

据推测, *ycf15* 基因起源于真核生物^[13],或者起源于基因水平转移^[14-15]。*ycf* 家族普遍被认为是编码蛋白的基因,作为该家族的一员, *ycf15* 也被假定具编码蛋白的功能,这是由于一些菊类植物 *ycf15* 基因序列的确具有完整的编码结构。然而,众多 *ycf15* 假基因存在,一些学者对此提出质疑并开展了相应的研究。

Schmitz-Linneweber 等^[10]研究藜科菠菜 *ycf15* 基因在转录过程中插入片段 (~250 bp) 是否被剪切掉,以此验证该基因是否为编码蛋白的基因。如果是,插入片段即为内含子,转录后会被剪切掉,5'部分和 3'部分则会保留下来;反之,插入片段内部的多个终止密码子会导致整个基因被割裂成多个片段,基因功能丧失,则说明该基因不具编码蛋白的功能。结果显示转录后插入片段没有被剪切掉,由此认为 *ycf15* 高度保守,的确有着重要功能,但并不编码蛋白。Raubeson 等^[9]研究 14 个不同属植物的 *ycf2: trnL-CAA* 基因间区序列 (包括 *ycf15* 基因序列) 以及叶绿体其余部位 (IR 区、LSC 区) 的基因间区序列的相似度,结果发现 *ycf15* 基因序列在不同属之间相似度很高,高度保守; IR 区基因间区序列普遍比 LSC 区更保守^[16],这可能不是因

为 *ycf15* 具编码蛋白功能,而仅仅是因为 *ycf15* 位于叶绿体 IR 区。

2 *ycf15* 基因基本结构

叶绿体基因组结构包括 1 个大单拷贝区 (LSC)、1 个小单拷贝区 (SSC) 和 2 个反向重复区 (IR); *ycf15* 基因位于 IR 区 *ycf2* 与 *trnL-CAA* 之间 (图 1), 多为 250~550 bp。为使分析更具代表性,本文参考维管植物各大类群 *ycf15* 基因序列 (GenBank 数据库), 涉及蕨类植物门 4 科 4 种, 裸子植物门 3 科 5 种, 被子植物门 (APG III 分类系统) 32 目 68 科 155 种共 164 个物种。据此将 *ycf15* 基因归纳为 5 种基本结构类型 (图 2)。其中, 26 条完整 *ycf15* 基因序列 (Ia 型); 29 个物种中该基因丢失; 其余 109 条均为假基因 (pseudogenes)。

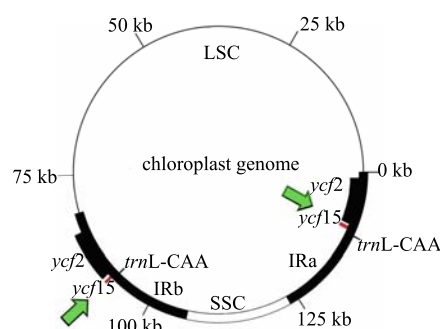


图 1 *ycf15* 基因位置

Fig. 1 Location of *ycf15* gene within chloroplast genome

2.1 完整的 *ycf15* 基因 (intact gene)

Shi 等^[12]研究发现,完整的 *ycf15* 基因主要存在于菊类植物基部类群和蔷薇类植物的一些类群中。本文的分析结果基本一致,即这 26 条完整 *ycf15* 基因,主要存在于菊类的杜鹃花目 (Ericales)、龙胆目 (Gentianales)、唇形目 (Lamiales)、茄目 (Solanales) 和伞形目 (Apiales) 等植物的叶绿体基因组中 (表 1), 占总统计数的 15.9%。其平均大小为 257 bp, 如茄科烟草 *Nicotiana tabacum* L. 总长 264 bp, 以“GTG”为起始密码子,“TAG”为终止密码子,共翻译出 88 个氨基酸。进而以烟草为参照,将此 26 条基因的氨基酸序列进行比对 (图 3), 发现氨基酸序列较为保守,序列间相似度高。但基于 *ycf15* 基因序列的分析显示: 科与科之间差异明显,插入/缺失序列较多,每个科都有其独特的差异位点。根据序列不同区域的变异频率可将 *ycf15* 分为 4 个区段,包括 2 个易变区 (II、IV) 和 2 个保守区 (I、III)。在保守区 I 和 III 中,科与科之间只

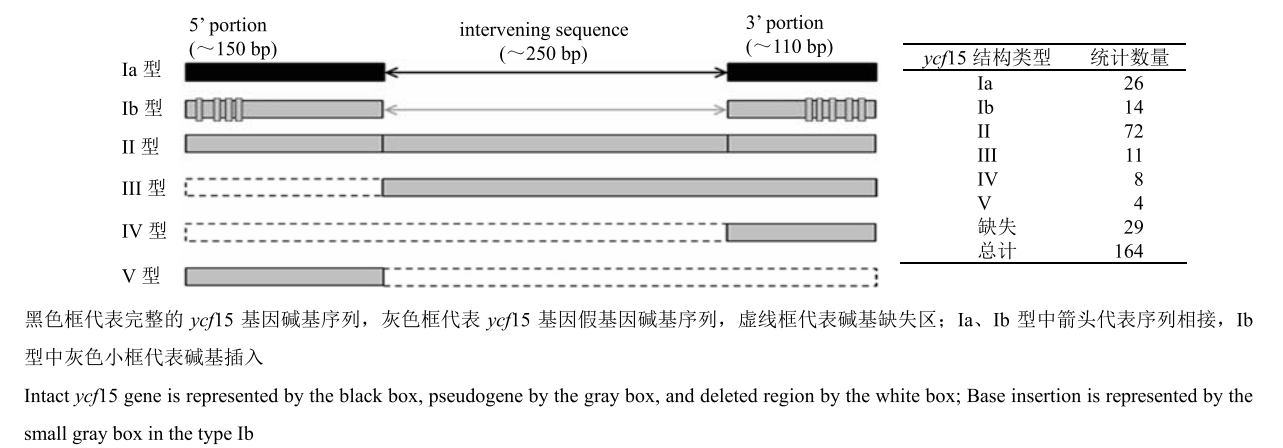


图 2 *ycf15* 基因基本结构类型

Fig. 2 Basic type of *ycf15* gene

表 1 完整 *ycf15* 基因物种信息 (按照 APG III 系统排列)

Table 1 Species information of intact *ycf15* gene (according to APG III)

目	科	种	GenBank 登录号	大小/bp	
杜鹃花目	猕猴桃科（Actinidiaceae）	葛枣猕猴桃 <i>Actinidia polygama</i>	KX345297	249	
		四萼猕猴桃 <i>Actinidia tetramera</i>	KX345298	249	
	报春花科（Primulaceae）	海仙花 <i>Primula poissonii</i>	KF753634	258	
		黄花九轮草 <i>Primula veris</i>	KX639823	249	
	柿科（Ebenaceae）	君迁子 <i>Diospyros lotus</i>	KM522849	249	
		柿 <i>Diospyros kaki</i>	KT223565	249	
	山茶科（Theaceae）	滇山茶 <i>Camellia reticulata</i>	KJ806278	249	
		膜叶茶 <i>Camellia leptophylla</i>	KJ806275	249	
龙胆目	龙胆科（Gentianaceae）	西藏秦艽 <i>Gentiana tibetica</i>	KY213965	231	
		粗茎秦艽 <i>Gentiana crassicaulis</i>	KY214077	237	
		头花龙胆 <i>Gentiana cephalantha</i>	KY214122	237	
唇形目	唇形科（Lamiaceae）	丹参 <i>Salvia miltiorrhiza</i>	JX312195	249	
		黄芩 <i>Scutellaria baicalensis</i>	KR233163	249	
		紫苏 <i>Perilla frutescens</i>	KT220692	249	
		薰衣草 <i>Lavandula angustifolia</i>	KT948988	249	
	木犀科（Oleaceae）	木犀榄 <i>Olea europaea</i>	GU931818	255	
	列当科（Orobanchaceae）	肉苁蓉 <i>Cistanche deserticola</i>	KC128846	249	
	玄参科（Scrophulariaceae）	天目地黄 <i>Rehmannia chingii</i>	KY397677	240	
	茄目	茄科（Solanaceae）	烟草 <i>Nicotiana tabacum</i>	Z00044	264
			茄 <i>Solanum melongena</i>	KU682719	264
			龙葵 <i>Solanum nigrum</i>	KM489055	264
灯笼果 <i>Physalis peruviana</i>			KP295964	264	
颠茄 <i>Atropa belladonna</i>			AJ316582	264	
伞形目	五加科（Araliaceae）	人参 <i>Panax ginseng</i>	KF431956	303	
		三七 <i>Panax notoginseng</i>	KJ566590	303	
		西洋参 <i>Panax quinquefolius</i>	KT028714	303	

发生少量的位点变异，其余位点均高度一致。易变区 II 易发生碱基的插入/缺失，导致基因的长度发生变异，而变异最频繁的区域则是 IV 区，即基因的尾部区。此区域具有科级水平的较高辨识度，即不同科之间差异较大而科内相对保守。根据 *ycf15* 基因序列特征翻译的氨基酸序列分区情况见图 3。

2.2 *ycf15* 假基因 (pseudogene)

本文中的 109 条 *ycf15* 假基因，涉及 5 种结构类型 (图 2)。其中 Ib 型长度在 250 bp 左右，形成原因可能为插入或缺失非 3 倍数的碱基而导致内部含有多个终止密码子 (termination codon)，基因功能丧失 (图 4)。主要发生在双子叶植物杜鹃花目、龙胆目^[17]、唇形目、茄目和伞形目中的

一些进化速率较快的物种中，数量较少。

ycf15 假基因常见类型为 II 型，占统计序列假基因总数的 66.1%。它是在 Ia 型基因 150 bp 左右处插入一段约 250 bp 的片段，即 II 型包括 3 部分：5' 部分 (~150 bp)、插入部分 (~250 bp) 和 3' 部分 (~110 bp)，使总长达 600 bp 左右。单子叶植物^[18-21]以及绝大部分被子植物^[22-23]均存在此类型假基因 (表 2)。

III、IV、V 型 *ycf15* 假基因属于基因退化型，只有部分序列残留，最小的仅剩 50 bp 左右^[24-27]，在植物界中占一小部分。少见禾本科的禾本科，毛茛目的毛茛科，石竹目的石竹科，蔷薇目的蔷薇科、桑科，菊目的菊科^[24-27]和桃金娘目的千屈菜科、野牡丹科等植物中。

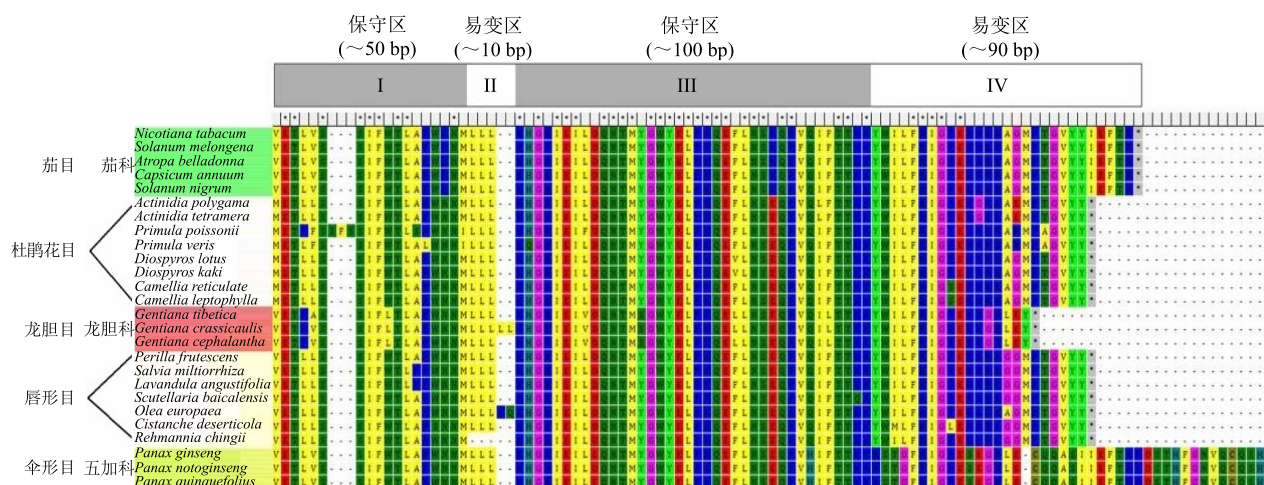
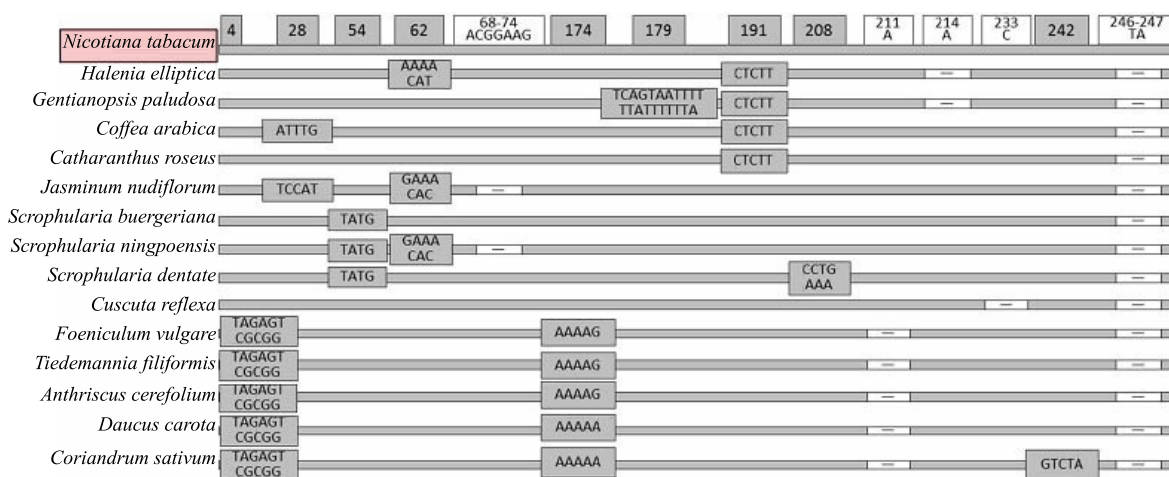


图 3 26 条完整 *ycf15* 基因氨基酸序列比对图

Fig. 3 Amino acid sequence alignment of 26 intact *ycf15* genes



烟草 *Nicotiana tabacum* *ycf15* 基因为参照，灰色块表示碱基插入，白色块表示碱基缺失

Base insertion is represented by the gray box and base deletion by the white box; *ycf15* region of *Nicotiana tabacum* is used as a reference

图 4 14 条 Ib 型 *ycf15* 假基因 indels 变异简图

Fig. 4 Analyses of type Ib *ycf15* pseudogene within 14 species

表 2 II 型 *ycf15* 假基因物种信息 (依 APG III 系统排列)
Table 2 Species information of type II *ycf15* pseudogene (according to APG III)

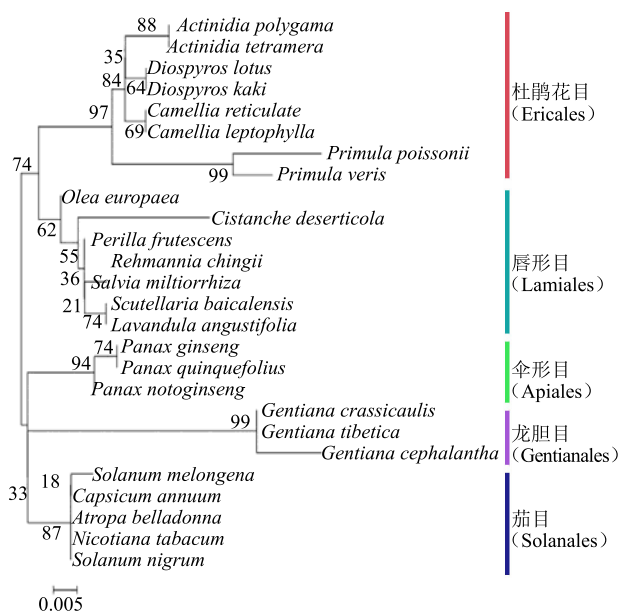
类	目	科名 (物种数)	GenBank 登录号
基本被子植物 (basal angiosperms)	无油樟目 (Amborellales)	无油樟科 (Amborellaceae) (1)	AJ506156
	睡莲目 (Nymphaeales)	睡莲科 (Nymphaeaceae) (2)	DQ354691, AJ627251
木兰类植物 (magnoliids)	樟目 (Laurales)	蜡梅科 (Calycanthaceae) (1)	AJ428413
	白桂皮目 (Canellales)	林仙科 (Winteraceae) (1)	DQ887676
	木兰目 (Magnoliales)	木兰科 (Magnoliaceae) (1)	DQ899947
	胡椒目 (Piperales)	胡椒科 (Piperaceae) (1)	DQ887677
单子叶植物 (monocots)	泽泻目 (Alismatales)	天南星科 (Araceae) (2)	KR270823, KR872391
	薯蓣目 (Dioscoreales)	薯蓣科 (Dioscoreaceae) (2)	KP899622, EF380353
	百合目 (Liliales)	百合科 (Liliaceae) (6)	KU230438, KT722981, KT695605, KM088015, KJ433485, KP125338
	天门冬目 (Asparagales)	鸢尾科 (Iridaceae) (1)	KT626943
		兰科 (Orchidaceae) (6)	KM244734, KT722982, KR919606, KT588924, KX507360, AY916449
	禾本目 (Poales)	凤梨科 (Bromeliaceae) (1)	AP014632
		禾本科 (Poaceae) (3)	X86563, EF115542, AP006714
	姜目 (Zingiberales)	姜科 (Zingiberaceae) (1)	KR967361
真双子叶植物 (eudicots)	毛茛目 (Ranunculales)	小檗科 (Berberidaceae) (2)	DQ923117, KR779994
		罂粟科 (Papaveraceae) (1)	KU204905
		毛茛科 (Ranunculaceae) (3)	KT820667, KM206568, KX347251
	虎耳草目 (Saxifragales)	景天科 (Crassulaceae) (1)	JX427551
		芍药科 (Paeoniaceae) (1)	KJ206533
	石竹目 (Caryophyllales)	石竹科 (Caryophyllaceae) (1)	KF527884
		蓼科 (Polygonaceae) (2)	KR816224, KX085498
	葡萄目 (Vitales)	葡萄科 (Vitaceae) (1)	DQ424856
	金虎尾目 (Malpighiales)	大戟科 (Euphorbiaceae) (1)	EU117376
		杨柳科 (Salicaceae) (1)	KJ624919
	葫芦目 (Cucurbitales)	葫芦科 (Cucurbitaceae) (2)	DQ119058, KY014105
	豆目 (Fabales)	豆科 (Fabaceae) (4)	KR136271, LN555649, AP002983, AC093544
	蔷薇目 (Rosales)	胡颓子科 (Elaeagnaceae) (1)	KP211788
		鼠李科 (Rhamnaceae) (1)	KU351660
		蔷薇科 (Rosaceae) (2)	KP760070, KX768420
	桃金娘目 (Myrtales)	桃金娘科 (Myrtaceae) (4)	HM347959, GQ870669, KC180805, AY780259
		柳叶菜科 (Onagraceae) (4)	AJ271079, EU262890, KX827312, EU262889
	十字花目 (Brassicales)	十字花科 (Brassicaceae) (3)	AP000423, KJ716483, KT844940
	锦葵目 (Malvales)	锦葵科 (Malvaceae) (2)	DQ345959, JF317353
		瑞香科 (Thymelaeaceae) (1)	KT148967
	无患子目 (Sapindales)	楝科 (Meliaceae) (1)	KF986530
		芸香科 (Rutaceae) (2)	DQ864733, KT153018
		无患子科 (Sapindaceae) (1)	KM454982
		黄杨科 (Buxaceae) (1)	EF380351

2.3 ycf15 基因缺失

统计的 29 个物种缺失 *ycf15* 基因^[28-35], 包括蕨类植物门^[28]和裸子植物门中的一些类群^[29-32], 尚有部分被子植物如禾本科 (Poaceae)^[33,34]、豆科 (Fabaceae)、旋花科 (Convolvulaceae)^[35]和桔梗科 (Campanulaceae) 中的一些类群均无 *ycf15* 基因, 占总统计数的 17.7%。

3 ycf15 基因物种鉴定价值与系统学意义

ycf15 基因序列既高度保守又含有丰富的变异位点, 为初步评价 *ycf15* 完整基因的系统学意义, 本文进行了相关的序列分析。参考 APG III (2009) 被子植物分类系统, 对杜鹃花目、龙胆目、唇形目、伞形目和茄目 5 目 11 科共 26 个具有完整 *ycf15* 基因的物种构建系统进化树, 结果显示了清晰的系统学关系, 即每个目各自聚为一大支, 每个科单独聚为一支 (图 5)。



1 000 bootstrap replications, Kimura 2-parameter model

图 5 基于 26 个物种完整 *ycf15* 基因序列构建的 ML 进化树
Fig. 5 ML phylogenetic tree based on intact *ycf15* gene sequences within 26 species

进而基于 45 个物种的 *ycf15* 序列 (既有完整基因又有假基因) 构建系统进化树, 其各大支的拓扑结构与 APG III (2009) 被子植物分类系统亦基本吻合。尽管假基因丧失编码功能, 但仍遗留有一定的系统发育信息。*ycf15* 基因及假基因蕴含的系统学信息为从 DNA 分子水平进一步鉴定物种打下坚实的基础。

同时, 如图 3 所示, 易变区 II 和 IV 均可发现多处变异位点, 包括插入缺失引起的长度多态性及碱基突变引起的位点多态性, 这些有望作为新的候选分子标记。在构建藏药“解吉”基原植物 DNA 条形码的前期工作中, 注意到 *ycf15* 基因具有物种鉴定意义。遗憾的是在一些关于叶绿体基因组系统进化分析的研究中^[36-37], 不包括 *ycf15* 基因序列, 大大忽视了该基因的分析潜力。

4 结语与展望

ycf15 作为新的分子标记具有巨大的研究空间和潜力。其优势在于: (1) 该基因为单亲遗传, 与核基因组片段相比, 无需克隆测序, 利于实验分析。(2) 作为编码基因序列, 保守性高, 根据具体鉴定对象, 可直接在序列中的高度保守区段筛选并设计个性化鉴定引物, 进而获得满意的鉴定结果。(3) 由于加工与贮存时间及条件等原因, 中药材样品中的 DNA 模板往往降解严重, 很难获得适于扩增的目的片段。*ycf15* 具有片段长度较小的特点 (一般为 250~550 bp); 同时, 若直接在序列中的不同位置筛选并设计个性化鉴定引物, 可将有效鉴定序列长度进一步缩小。因此, 相较于其他较长分子标记序列, 用该片段进行物种鉴定, 易于扩增, 获得目的鉴定片段的可能性大大提高。

大量序列资料显示 *ycf15* 基因位于 IR 区 *ycf2* 与 *trnL-CAA* 之间, 但亦有一些工作^[24-27,38-39]将该基因的位置注释在 *rps7* 与 *trnV-GAC* 之间, 这是值得商榷的。

当然该基因仍存在一定的局限性, 如豆科和禾本科等植物内 *ycf15* 基因结构变异较大, 均有 3 种基因结构, 可见其作为该科的鉴定意义有限; 还有, 茄科内各属植物之间 *ycf15* 基因变异很少, 几乎一致, 其作为该科内种间的鉴定意义也不大。故作为新的分子标记, *ycf15* 基因应依据具体类群及物种情况具体分析, 不可一概而论。

参考文献

- [1] 倪梁红, 赵志礼, 米 玛. 药用植物叶绿体基因组研究进展 [J]. 中药材, 2015, 38(9): 1991-1995.
- [2] Hallick R B, Bairoch A. Proposals for the naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes [J]. *Plant Mol Biol Rep*, 1994, 12(2): S29-S30.
- [3] Neubig K M, Whitten W M, Carlswald B S, et al. Phylogenetic utility of *ycf1* in orchids: A plastid gene

- more variable than *matK* [J]. *Plant Syst Evol*, 2009, 277(1/2): 75-84.
- [4] Peter E, Wallner T, Wilde A, *et al.* Comparative functional analysis of two hypothetical chloroplast open reading frames (*yef*) involved in chlorophyll biosynthesis from *Synechocystis* sp. PCC6803 and plants [J]. *J Plant Physiol*, 2011, 168(12): 1380-1386.
- [5] Shinozaki K, Ohme M, Tanaka M, *et al.* The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression [J]. *Eur Mol Biol Org J*, 1986, 5(9): 2043-2049.
- [6] Sato S, Nakamura Y, Kaneko T, *et al.* Complete structure of the chloroplast genome of *Arabidopsis thaliana* [J]. *DNA Res*, 1999, 6(5): 283-290.
- [7] Goremykin V V, Hirsch-Ernst K I, Wolf S, *et al.* Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm [J]. *Mol Biol Evol*, 2003, 20(9): 1499-1505.
- [8] Steane D A. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae) [J]. *DNA Res*, 2005, 12(3): 215-220.
- [9] Raubeson L A, Peery R, Chumley T W, *et al.* Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus* [J]. *BioMed Central Genom*, 2007, 8: 174.
- [10] Schmitz-Linneweber C, Maier R M, Alcaraz J P, *et al.* The plastid chromosome of spinach (*Spinacia oleracea*): Complete nucleotide sequence and gene organization [J]. *Plant Mol Biol*, 2001, 45(3): 307-315.
- [11] Nguyen P A T, Kim J S, Kim J H. The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus [J]. *Planta*, 2015, 242(1): 223-237.
- [12] Shi C, Liu Y, Huang H, *et al.* Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: An exemplary study of *yef15* function and evolution in angiosperms [J]. *PLoS One*, 2013, 8(3): e59620.
- [13] Martin W, Stoebe B, Goremykin V, *et al.* Gene transfer to the nucleus and the evolution of chloroplasts [J]. *Nature*, 1998, 393(14): 162-165.
- [14] Stegemann S, Keuthe M, Greiner S, *et al.* Horizontal transfer of chloroplast genomes between plant species [J]. *Proceed Nat Acad Sci USA*, 2012, 109(7): 2434-2438.
- [15] Bergthorsson U, Adams K L, Thomason B, *et al.* Widespread horizontal transfer of mitochondrial genes in flowering plants [J]. *Nature*, 2003, 424(10): 197-201.
- [16] Perry A S, Wolfe K H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat [J]. *J Mol Evol*, 2002, 55(5): 501-508.
- [17] Ku C, Chung W C, Chen L L, *et al.* The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: Plastid genome evolution, molecular marker identification, and phylogenetic implications in Asterids [J]. *PLoS One*, 2013, 8(6): e68518.
- [18] Goremykin V V, Holland B, Hirsch-Ernst K I, *et al.* Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications [J]. *Mol Biol Evol*, 2005, 22(9): 1813-1822.
- [19] Chang C C, Lin H C, Lin I P, *et al.* The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications [J]. *Mol Biol Evol*, 2006, 23(2): 279-291.
- [20] Asano T, Tsudzuki T, Takahashi S, *et al.* Complete nucleotide sequence of the sugarcane (*Saccharum Officinarum*) chloroplast genome: A comparative analysis of four monocot chloroplast genomes [J]. *DNA Res*, 2004, 11(2): 93-99.
- [21] Saski C, Lee S B, Fjellheim S, *et al.* Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes [J]. *Theoret Appl Gene*, 2007, 115(4): 571-590.
- [22] Daniell H, Wurdack K J, Kanagaraj A, *et al.* The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in malpighiales: RNA editing and multiple losses of a group II intron [J]. *Theoret Appl Gene*, 2008, 116(5): 723-737.
- [23] Cai Z Q, Penaflor C, Kuehl J V, *et al.* Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids [J]. *BioMed Central Evol Biol*, 2006, 6: 77.
- [24] Liu Y, Huo N X, Dong L L, *et al.* Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants [J]. *PLoS One*, 2013, 8(2): e57533.
- [25] Curci P L, Paola D D, Danzi D, *et al.* Complete chloroplast genome of the multifunctional crop globe artichoke and comparison with other Asteraceae [J]. *PLoS One*, 2015, 10(3): e0120589.

- [26] Nie X J, Lv S Z, Zhang Y X, *et al.* Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*) [J]. *PLoS One*, 2012, 7(5): e36869.
- [27] Choi K S, Park S J. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae [J]. *Gene*, 2015, 572(2): 214-221.
- [28] Kim H T, Chung M G, Kim K J. Chloroplast genome evolution in early diverged leptosporangiate ferns [J]. *Mol Cells*, 2014, 37(5): 372-382.
- [29] Wakasugi T, Tsudzuki J, Ito S, *et al.* Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii* [J]. *Proceed National Acad Sci USA*, 1994, 91(21): 9794-9798.
- [30] Wu C S, Lai Y T, Lin C P, *et al.* Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy [J]. *Mol Phylogene Evol*, 2009, 52(1): 115-124.
- [31] Tsudzuki J, Ito S, Tsudzuki T, *et al.* A new gene encoding tRNA (GGG) is present in the chloroplast genome of black pine: a compilation of 32 tRNA genes from black pine chloroplasts [J]. *Curr Gene*, 1994, 26(2): 153-158.
- [32] Li Z H, Qian Z H, Liu Z L, *et al.* The complete chloroplast genome of Armand pine *Pinus armandii*, an endemic conifer tree species to China [J]. *Mitochondrial DNA Part A: DNA Mapping Sequencing Anal*, 2016, 27(4): 2635-2636.
- [33] Morton B R, Clegg M T. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae) [J]. *Curr Gene*, 1993, 24(4): 357-365.
- [34] Ogiwara Y, Isono K, Kojima T, *et al.* Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA [J]. *Mol Gene Genom*, 2002, 266(5): 740-746.
- [35] McNeal J R, Kuehl J V, Boore J L, *et al.* Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta* [J]. *BMC Plant Biol*, 2007, 7: 57.
- [36] Jansen R K, Cai Z Q, Raubeson L A, *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns [J]. *Proceed Nat Acad Sci USA*, 2007, 104(49): 19369-19374.
- [37] Ni L H, Zhao Z L, Xu H X, *et al.* The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion [J]. *Gene*, 2016, 577(2): 281-288.
- [38] Saski C, Lee S B, Daniell H, *et al.* Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes [J]. *Plant Mol Biol*, 2005, 59(2): 309-322.
- [39] Williams A V, Boykin L M, Howell K A, *et al.* Correction: the complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene [J]. *PLoS One*, 2015, 10(9): e0138367.