

· 研究报告 ·

樟属植物ITS序列多态性分析

黄建峰^{1, 2}, 李朗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, 昆明 650223; ²中国科学院大学, 北京 100049

摘要 对樟科樟属(*Cinnamomum* Schaeffer) 17个代表样本的核糖体DNA内转录间隔区(nrDNA ITS)进行克隆测序。对获得的87条不同ITS序列的长度变异、GC含量、5.8S区二级结构的稳定性、遗传距离、进化模式以及系统发育关系进行了相关分析。研究结果显示, ITS序列在樟属植物内存在明显的多态性, 87条序列中的22条序列被鉴定为假基因序列, 其余65条序列为功能基因序列; 假基因序列采用中性进化模式, 变异明显大于功能序列。ITS序列在樟属植物中出现一致性进化不完全和假基因现象也可能发生在樟科其它类群中, 这可能是导致樟科植物ITS序列直接测序方式成功率低的重要原因。

关键词 樟属, ITS, 多态性, 假基因

黄建峰, 李朗, 李捷 (2016). 樟属植物ITS序列多态性分析. 植物学报 51, 609–619.

真核生物的核糖体DNA (nuclear ribosomal DNA, nrDNA)是由一些高度重复序列组成的多基因家族, 在基因组内往往有成千上万的拷贝, 这些拷贝在基因组内串联重复, 分布于1对或多对染色体上(Eickbush and Eickbush, 2007)。每个重复单元由编码区18S rDNA、5.8S rDNA、26S rDNA和位于编码区之间的转录间隔区(internal transcribed spacer, ITS1和ITS2)共同组成。nrDNA (包括ITS)通常采用一致性进化(concerted evolution)模式。在这种进化模式下, 不同的ITS拷贝序列会趋向基本一致或完全一致, 便于对ITS的PCR扩增产物进行直接测序, 同时ITS序列适合于较低分类群的系统关系重建, 使其成为现今被子植物系统学研究中最常用的分子标记之一(Bailey et al., 2003; Eickbush and Eickbush, 2007); 此外, ITS序列在DNA条形码研究中虽被列为补充条码(CBOL Plant Working Group, 2009), 但在最近的很多研究中它被认为是最有效的DNA条形码(Li et al., 2011a; Tripathi et al., 2013; 卢孟孟等, 2013)。然而越来越多的研究发现, 某些物种甚至个体内的ITS序列存在一致性进化不完全现象, 拷贝之间存在明显的差异(Muir et al., 2001; Zheng et al., 2008; Xiao et al., 2010; Chen et al., 2015)。此外, 一些ITS拷贝可能会由于功能退化而变成假基因(pse-

udogene)。假基因在序列结构上与功能基因非常相似, 但已经丧失了正常的蛋白质编码功能(Vanin, 1985)。由于假基因与功能基因的进化速率不同, 不能反映真实的基因进化速率, 所以用假基因或包含假基因的序列为依据构建的系统树并不可靠(Mayol and Rossello, 2001; 马长乐和周浙昆, 2006)。总结以往的研究, 发现ITS假基因一般有以下特征: (1) GC含量降低; (2) 在5.8S编码区长度有变化, 碱基替代率高, 即使在高保守区也是如此; (3) RNA二级结构稳定性较低; (4) 进化速率加快(Buckler et al., 1996; Mayol and Rossello, 2001; Álvarez and Wendel, 2003; 周阿涛等, 2013)。已有研究证实, 在某些植物类群中存在ITS假基因的现象, 包括苏铁属(*Cycas*) (Xiao et al., 2010)、梨属(*Pyrus*) (Zheng et al., 2008)、栎属(*Quercus*) (Muir et al., 2001)、落叶松属(*Larix*) (Wei and Wang, 2004)和山茶属(*Camellia*) (徐颖等, 2011; 周阿涛等, 2013; 尤欢等, 2014)等。

素有“四大金刚”之称的樟科、壳斗科、木兰科和山茶科是我国南方常绿阔叶林的主要建群成分, 为我国常绿阔叶林的重要特征(Li et al., 2011a)。而DNA条形码的研究显示, 樟科、壳斗科和山茶科等植物类群在ITS序列上的测序成功率不仅明显低于另外3个条码(*rbcL*、*matK*和*trnH-psbA*), 同时也明显低于其它

收稿日期: 2015-08-11; 接受日期: 2015-11-13

基金项目: 国家自然科学基金(No.31370245)

* 通讯作者。E-mail: jieli@xtbg.ac.cn

植物类群。例如, 卢孟孟等(2013)在对以常绿阔叶林为主体的云南省哀牢山自然保护区乔木树种DNA条形码研究中, 发现在4个DNA条形码中ITS的扩增成功率最高, 但测序成功率却最低。他们统计分析了样本中物种数目超过5的科中ITS的通用性, ITS测序成功个体数与采集个体数比值最低的前3名依次为樟科、壳斗科和山茶科。特别是樟科和壳斗科类群的ITS测序成功率极低, 仅有30.7%, 严重降低了ITS在该地区乔木树种中DNA条形码的应用。如果排除樟科和壳斗科这两个类群, ITS的测序成功率显著提高。对于这些植物类群, 通过直接测序方式得到的ITS序列往往出现严重的杂合现象, 达不到使用标准, 这是导致其测序成功率低的主要原因, 而这种序列的杂合现象最有可能来自ITS序列多态性。虽然采用克隆测序的方式能够解决序列杂合的问题, 但会大大增加科研成本。研究已证实壳斗科和山茶科两个类群的ITS序列存在一致性进化不完全和假基因现象(Muir et al., 2001; 徐颖等, 2011; 周阿涛等, 2013; 尤欢等, 2014), 这可能是导致两个类群测序成功率低的重要原因。对于木兰科, 由于存在高的种间杂交亲和性, 至少在人工培育条件下, 木兰科内类群之间很容易发

生杂交(Qiu et al., 1995; 龚洵等, 2001), 这限制了作为双亲遗传的ITS在木兰科中的运用。而被誉为我国常绿阔叶林“四大金刚”之一的樟科的ITS序列测序成功率低的原因, 至今还没有合理的解释。

在樟科植物的分子系统学研究中, ITS序列被认为是解决能力较强的分子标记(Chanderbali et al., 2001; Ho and Hung, 2011; Li et al., 2011b)。我们对樟属(*Cinnamomum*)的分子系统学研究发现, 其中17个样本的ITS序列无法通过直接测序的方式获得, 测序胶图显示出严重的杂合现象(Huang et al., 2016)。在本研究中, 我们对17个樟属样本采用克隆测序的方式, 通过对克隆测序结果进行分析, 探讨ITS序列的多态性, 并采用不同的分析方法检验ITS假基因在樟属植物中是否存在。

1 材料与方法

1.1 材料

本研究选取樟属(*Cinnamomum* Schaeffer Bot. Exped.) 14个代表物种(其中有1个未知种)共17个样本(表1), 并选取樟属的近缘属植物月桂(*Laurus nobilis*

表1 材料信息

Table 1 Details of materials

Code	Speices	Voucher	Locality
H1C	<i>Cinnamomum appelianum</i>	Huang J. F. & Ding X., 14043005 (HITBC)	Guizhou, China
D16	<i>C. baileyianum</i>	Crayn D., CSIRO38 (CNS)	Queensland, Australia
H5A	<i>C. bodinieri</i>	Li J. et al., 2007212 (HITBC)	Zhejiang, China
H53	<i>C. crenulicupulam</i>	Huang J. F., QSBG10 (HITBC)	Chiengmai, Thailand
H18	<i>C. heyneanum</i>	Huang J. F., H-BN012 (HITBC)	Yunnan, China
H21A	<i>C. chekiangense</i>	Huang J. F., H-KZ30 (HITBC)	Yunnan, China
H21B	<i>C. chekiangense</i>	Ci X. Q. et al., CXQ279 (HITBC)	Zhejiang, China
H26	<i>C. liangii</i>	Ci X. Q. et al., 20100043 (HITBC)	Zhejiang, China
H35B	<i>C. pauciflorum</i>	Huang J. F., H-BN005 (HITBC)	Hunan, China
B17	<i>C. rhynchophyllum</i>	Huang J. F. et al., 2013082923 (HITBC)	Bogor, Indonesia
B20	<i>C. sintoc</i>	Huang J. F. et al., 2013082920 (HITBC)	Bogor, Indonesia
B24	<i>C. sp.</i>	Huang J. F. et al., 2013082910 (HITBC)	Bogor, Indonesia
H43	<i>C. subavenium</i>	Ci X. Q. et al., CXQ303 (HITBC)	Zhejiang, China
H43A	<i>C. subavenium</i>	Huang J. F. & Li L., 2013024 (HITBC)	Hainan, China
H43B	<i>C. subavenium</i>	Ci X. Q. et al., CXQ257 (HITBC)	Hunan, China
H50	<i>C. tsangii</i>	Huang J. F. & Li L., H-WZS07 (HITBC)	Hainan, China
H47A	<i>C. tsoi</i>	Huang J. F. & Li L., H-JFL21 (HITBC)	Hainan, China
LN	<i>Laurus nobilis</i>	Huang J. F. & Ding X., LN (HITBC)	Yunnan, China

CNS: 澳大利亚热带标本馆; HITBC: 中国科学院西双版纳热带植物园标本馆。

CNS: Australian Tropical Herbarium; HITBC: Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences.

L.)作为外类群。野外采集经硅胶干燥后的叶片材料用于总DNA的提取。

1.2 总DNA提取、扩增、纯化、载体连接和测序

使用天根生化科技(北京)有限公司的Pant Genomic DNA Kit (Cat No.DP305-03)提取植物总DNA。采用ITSF (5'-GCTACGTTCTTCATCGATGC-3') (Chanderbali et al., 2001) 和 ITS4 (5'-TCCTCCGCT-TATTGATATGC-3') (White et al., 1990)对樟属植物ITS序列进行扩增。PCR扩增体系为25 μL : 2.5 μL 10 $\times$ PCR buffer, 2.5 μL MgCl_2 (25 $\text{mmol}\cdot\text{L}^{-1}$), 上下游引物各1 μL (10 $\mu\text{mol}\cdot\text{L}^{-1}$), 2.0 μL dNTP (2.5 $\text{mmol}\cdot\text{L}^{-1}$), 2.0 μL DMSO (二甲基亚砜), 0.3 μL rTaq 聚合酶, 13.7 μL DNA模板+ddH₂O。PCR扩增产物在1%琼脂糖凝胶上进行电泳检测, 对扩增成功的产物用Omega Bio-Tek公司的Cycle-Pure Kit (Cat No. D6493-01)进行DNA纯化。使用北京全式金生物技术有限公司的pEASY[®]-T3 Cloning Kit (Cat No. CT301-02)将纯化产物与载体连接、转化, 然后进行阳性克隆筛选, 每个样本挑取5–6个阳性克隆送华大基因(BGI)测序。

1.3 数据分析

1.3.1 ITS数据矩阵构建、处理和分析

使用Seqman (DNASar, Madison, WI)软件, 通过将测序序列与相应胶图对比完成单向序列的校正, 然后将正向和反向互补序列进行对比拼接, 以保证所得序列的准确性。将比对好的数据导入Clustal_X 1.8.1 (Thompson et al., 1997), 使用默认参数对序列矩阵进行排序, 之后使用BioEdit v7.2.5 (Hall, 1999)软件对排序结果进行人工校对, 以形成最终的ITS序列矩阵。

使用MEGA 6软件(Tamura et al., 2013)计算ITS序列长度和GC含量, 同时采用K2P模型(Kimura 2-parameter model)计算樟属植物的种内和种间遗传距离(p-distance), 并以月桂作为外类群, 统计17个樟属样本中的每个单克隆序列相对于月桂的遗传距离。利用在线软件Mfold (<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>)预测5.8S区域的二级结构和最小自由能(ΔG at 37 $^{\circ}\text{C}$)。利用DnaSP软件(Rozas and Rozas, 1999)进行核苷酸差异平均值

(average number of nucleotide differences, K)、核苷酸多样性(nucleotide diversity, π)和中性检验(Tajima's D)测定。采用SPSS软件对数据进行显著性检验。

1.3.2 系统发育分析

采用PAUP* 4.0b10 (Swofford, 2003)软件对ITS矩阵构建最大简约树(maximum parsimony tree), 使用启发式搜索与100次重复的随机加入, 二分树再接法进行分支交换寻找最简约树, 利用100次重复的自展分析检验各分支的置信度(bootstrap support, BS) (Felsenstein, 1981)。利用MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003)构建贝叶斯树(Bayesian tree), 运用Modeltest 3.7 (Posada and Crandall, 1998; Posada and Buckley, 2004)筛选合适的进化模型, 使用马尔科夫链的蒙特卡洛方法, 同时起始4条马尔科夫链、3条热链和1条冷链, 以随机树为起始树, 每隔500代保存1棵树, 弃去前25%预热树, 剩余的树用于计算一致性树和各分支的后验概率(posterior possibility, PP)。

2 结果与讨论

2.1 ITS序列长度、GC含量、最小自由能分析和假基因的鉴别

本研究从17个樟属样本中共得到87条不同的ITS克隆序列。通过对每条序列的碱基组成、长度变异、5.8S区域的二级结构和最小自由能进行统计分析(表2), 发现其中22条序列符合假基因的特征, 因此鉴别为假基因序列, 其余65条为功能序列。在17个样本中, 包括*C. chekiangense* H21A、*C. crenulicupulam* H53、*C. pauciflorum* H35B、*C. sintoc* B20、*C. sp.* B24、*C. subavenium* H43、*C. subavenium* H43A和*C. tsangii* H50的8个样本既得到了ITS功能序列, 也得到了假基因序列。而*C. tsoi* H47A得到的4个序列全部为假基因序列, 在其余的8个样本中得到的全部为功能序列。其中在*C. crenulicupulam* H53和*C. tsoi* H47A两个样本中分别得到了2个完全一致的假基因序列, 在统计分析时计为1条。65条功能序列的长度变异范围为639–689 bp, 22条假基因序列的长度变异范围为586–687 bp, 后者的长度变异明显大于前

表2 樟属17个样本ITS序列特征信息

Table 2 Characteristic information of ITS regions of 17 *Cinnamomum* samples

Species+Code	Type	No.	Length (bp) (SE)		GC content (%) (SE)		5.8S ΔG (SE) (kcal·mol ⁻¹)		π	K
			ITS1	ITS2	ITS1	ITS2	5.8S	5.8S		
<i>C. appelianum</i> H1C	F	5	214.2(1.7)	292.4(2.1)	156.0(0.0)	72.7(0.7)	72.5(0.4)	57.8(0.4)	0.06268	41.500
<i>C. baileyianum</i> D16	F	6	230.0(0.0)	290.0(0.0)	156.0(0.0)	73.4(0.3)	70.9(0.1)	57.6(0.2)	0.00799	5.400
<i>C. bodinieri</i> H5A	F	4	197.0(0.0)	286.0(0.0)	156.0(0.0)	76.0(0.2)	74.2(0.2)	57.7(0.0)	0.00626	4.000
<i>C. chekiangense</i> H21A	F	1	234.0	299.0	156.0	75.2	71.6	58.3	—	—
	P	4	223.8(9.6)	291.0(5.0)	156.0(0.0)	70.4(2.7)	67.3(2.9)	54.0(0.6)	0.13148	85.333
<i>C. chekiangense</i> H21B	F	5	234.4(3.6)	290.0(7.3)	156.0(0.0)	74.1(0.4)	72.1(0.4)	58.0(0.6)	0.01605	9.000
<i>C. orenalicupulam</i> H53	F	3	236.0(0.0)	293.0(0.0)	156.0(0.0)	74.3(0.1)	71.7(0.0)	57.7(0.0)	0.00681	4.667
	P	1	236.0	295.0	142.0	75.4	66.1	54.2	—	—
<i>C. heyneanum</i> H18	F	6	237.2(0.4)	293.5(0.2)	156.0(0.0)	75.0(0.2)	72.2(0.1)	57.9(0.2)	0.01217	8.333
<i>C. liangjii</i> H26	F	5	234.4(1.4)	293.0(0.2)	155.8(0.2)	74.7(0.2)	72.1(0.1)	58.0(0.1)	0.01187	8.000
<i>C. pauciflorum</i> H35B	F	3	237.7(0.7)	293.7(0.3)	156.0(0.0)	74.5(0.2)	72.3(0.2)	57.9(0.2)	0.00583	4.000
	P	2	208.5(7.5)	285.5(0.5)	146.0(10.0)	52.3(0.9)	67.8(0.8)	52.3(0.9)	0.11576	72.000
<i>C. rhynchophyllum</i> B17	F	5	231.0(0.0)	292.6(1.1)	156.0(0.0)	74.8(0.3)	71.0(0.3)	57.6(0.1)	0.00737	5.000
<i>C. sintoc</i> B20	F	5	210.2(2.9)	292.8(0.2)	156.0(0.0)	72.6(0.6)	72.1(0.2)	57.7(0.2)	0.01272	8.000
	P	1	218.0	289.0	140.0	67.0	63.0	50.0	—	—
<i>C. sp.</i> B24	F	1	235.0	293.0	156.0	75.3	72.4	57.7	—	—
	P	5	223.2(3.3)	288.6(0.8)	150.0(3.7)	67.9(1.0)	64.9(0.7)	51.8(0.8)	0.10513	67.600
<i>C. subavenium</i> H43	F	4	235.3(1.3)	293.8(4.8)	155.8(0.3)	76.0(0.2)	72.3(1.0)	57.3(0.4)	0.02848	19.167
	P	1	196.0	287.0	156.0	61.7	65.5	55.8	—	—
<i>C. subavenium</i> H43A	F	4	213.5(8.9)	295.3(0.63)	156.0(0.0)	73.0(0.8)	71.8(0.3)	57.9(0.4)	0.03225	19.833
	P	1	217.0	296.0	156.0	71.4	71.6	55.8	—	—
<i>C. subavenium</i> H43B	F	5	230.60(4.4)	297.0(0.0)	156.0(0.0)	75.6(0.4)	72.8(0.3)	58.0(0.2)	0.00961	6.400
<i>C. tsangjii</i> H50	F	3	236.0(0.0)	293.0(0.0)	156.0(0.0)	74.3(0.1)	71.7(0.0)	57.7(0.0)	0.01611	10.667
	P	3	210.7(0.0)	272.3(20.2)	156.0(0.0)	63.1(1.2)	66.2(3.3)	55.3(0.6)	0.14202	80.667
<i>C. tsoi</i> H47A	P	4	219.0(6.3)	271.0(11.9)	138.0(9.3)	72.5(0.5)	69.7(1.0)	52.8(1.4)	0.12436	64.667
All	F	65	226.7(1.6)	292.7(0.4)	155.97(0.0)	74.4(0.2)	72.1(0.1)	57.8(0.1)	0.04095	22.031
	P	22	219.1(2.5)	284.4(3.5)	148.8(2.3)	69.3(0.9)	66.9(0.6)	53.4(0.4)	0.13518	55.423
T test			P<0.05	P<0.01	P<0.01	P<0.01	P<0.01	P<0.01	P<0.01	P<0.01

F: ITS功能序列; P: ITS假基因序列; SE: 标准误; K: 核苷酸多样性; π : 核苷酸多样性; ΔG : 5.8S区最小自由能; — 表示缺失数据; T检验: 检验功能序列和假基因序列两组数据间是否存在显著性差异。
F: Functional ITS sequences; P: ITS pseudogene sequences; SE: Standard error; K: Average number of nucleotide differences; π : Nucleotide diversity; ΔG indicates the minimal folding free energy of 5.8S region; — indicates the missing data; T test: To tests if the dataset of functional sequences is significantly different from the dataset of pseudogene sequences.

者。T检验显示,假基因序列的ITS1、ITS2和5.8S三个区域长度和GC含量,都显著低于功能序列,而其5.8S区域的最低自由能显著高于功能序列(表2),表明这些假基因序列的二级结构稳定性降低。

2.2 核苷酸多样性及中性检测

核苷酸多样性分析显示,对于每个同时得到功能序列和假基因序列的样本,假基因序列的核苷酸多样性(π)和平均核苷酸差异数(K)都明显高于功能序列(表2)。22条假基因序列的 π 和 K 值大小分别是65条功能序列的3.3倍和2.5倍。

Tajima's D 检验结果显示,功能序列($D=-1.849\ 21$)和假基因序列($D=-1.329\ 11$)的 D 值均为负值。功能序列的 D 值的绝对值较大,且差异显著($P<0.05$),表明ITS功能序列偏离中性进化;而假基因序列的 D 值绝对值较小,且未呈现显著差异($P>0.10$),表明假基因序列倾向于中性进化(Fu, 1996, 1997)。

2.3 遗传距离

利用MEGA 6软件,分析了87条克隆序列相对于月桂

ITS序列的遗传距离(图1)。T检验结果显示,65条功能序列相对于月桂的遗传距离和22条假基因序列相对于月桂遗传距离2组数据间具有极显著差异($P<0.001$);功能序列和假基因序列相对于月桂的遗传距离平均值分别为0.062 5和0.135 0,表明与功能序列相比,假基因序列的进化速率明显加快。从图1也可看出,22个假基因序列(X轴标记为P的序列)中,有18个与月桂的遗传距离(Y轴)明显大于功能序列与月桂的遗传距离;而只有4个假基因序列(加 ψ 标记)与月桂的遗传距离,相对于功能序列与月桂的遗传距离差异不明显。

2.4 系统发育关系

以月桂为外类群,将采用Modeltest 3.7获得的最适核苷酸替代模型GTR+G和相关参数用于MrBayes 3.1.2软件,对87条不同的ITS序列构建贝叶斯树,将后验概率低于50%的分支作为多歧分支处理(图2)。最大简约法得到的一致性树与贝叶斯树拓扑结构基本一致,本文未显示简约树,只将其靴带支持率显示在贝叶斯树上(图2,假基因序列用“ ψ ”标记)。从系统

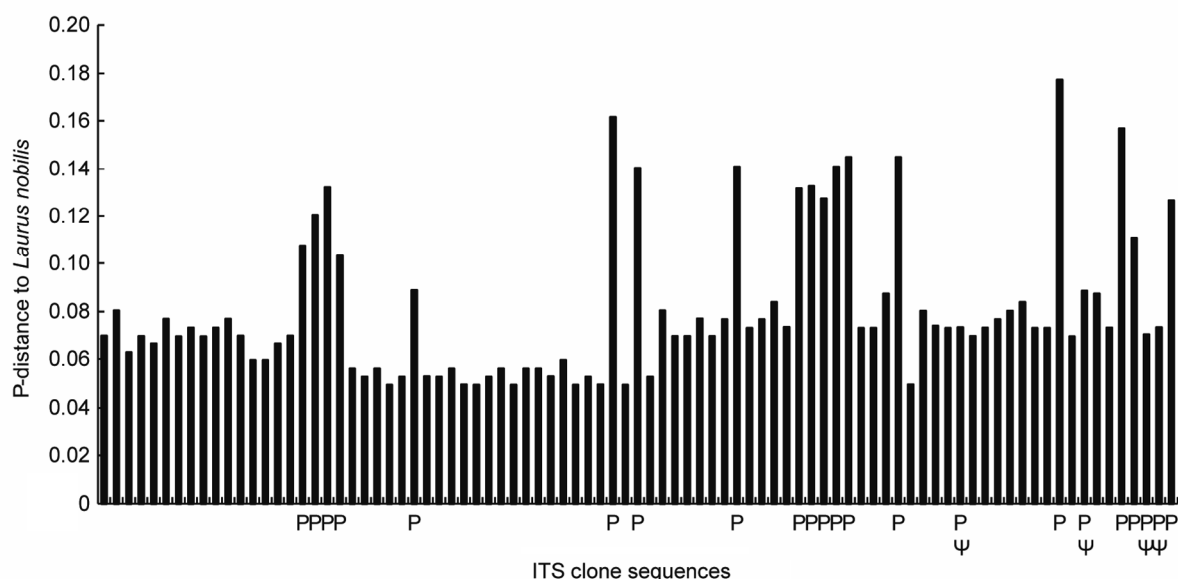


图1 ITS克隆序列相对于月桂的遗传距离

P和 ψ 表示ITS假基因序列。

Figure 1 Distribution of p-distance from ITS clone sequences to *Laurus nobilis*
P and ψ indicate ITS pseudogene sequences.

树中可以看出, 包括*C. baileyianum* D16、*C. bodinieri* H5A、*C. crenulicupulam* H53和*C. rhynchophyllum* B17在内的4个样本各自的功能序列能很好地聚为单系, 并得到很高的支持率(BS: 88%–100%; PP: 100%)。而包括*C. chekiangense* H21B、*C. heyneanum* H18、*C. liangii* H26、*C. pauciflorum* H35B、*C. sintoc* B20、*C. subavenium* H43、*C. subavenium* H43B和*C. tsangii* H50在内的8个样本各自的功能序列虽然不能很好地聚为单系, 但在树图上显示出非常近的亲缘关系(图2)。同时该8个样本内不同功能序列间的遗传距离也很近, 例如, *C. chekiangense* H21B的5条功能序列间的平均遗传距离为0.014 0, *C. liangii* H26的6条功能序列之间的平均遗传距离为0.012 4, 均远小于本文中14个樟属物种的种间平均遗传距离(0.050 4)。因此这8个样本各自的功能序列不能形成单系, 可能仅仅是由于系统发育信息位点不足所致。而*C. appelianum* H1C和*C. subavenium* H43A两个样本各自的功能序列不能聚为单系, 同时, 在系统树上也显示出非常远的系统关系。二者个体内不同功能序列间的平均遗传距离分别为0.074 1和0.033 9, 与樟属内种间平均遗传距离(0.050 4)相比差别不大, 特别是*C. appelianum* H1C的不同拷贝间遗传距离大于樟属内种间平均遗传距离, 显然这2个物种在系统树上不能形成单系的现象不能用信息位点的不足来解释。*C. chekiangense* H21A和*C. sp.* B24各自得到1个功能序列, 前者的功能序列*C. chekiangense* H21A-21与其假基因序列*C. chekiangense* H21A-9I形成姐妹群, 而后者*C. sp.* B24的唯一功能序列*C. sp.* B24-82I和*C. sintoc* B20的4个功能序列形成姐妹群。目前尚未发现在同个样本中所有功能序列和假基因序列聚为单系的情况, 而来自不同样本的假基因序列显示出非常近的亲缘关系。

2.5 讨论

2.5.1 ITS序列多态性

本研究在所选取的17个樟属样本中, 均发现存在明显的ITS序列多态性。一方面, ITS功能序列存在广泛变异, 其中包括*C. appelianum* H1C、*C. chekiangense* H21B、*C. heyneanum* H18、*C. liangii* H26、*C. pauciflorum* H35B、*C. rhynchophyllum* B17、*C. sintoc* B20、*C. subavenium* H43、*C. subavenium*

H43A和*C. subavenium* H43B在内的10个样本的功能序列存在长度多态性; 而在*C. baileyianum* D16、*C. crenulicupulam* H53、*C. bodinieri* H5A和*C. tsangii* H50四个样本的功能序列中未发现长度多态性, 但存在碱基多态性; 其余3个样本, 可能由于实验选取的阳性克隆数目太少, 只得到1个功能序列(*C. chekiangense* H21A和*C. sp.* B24), 或者未得到功能序列(*C. tsoi* H47A), 因此无法判断其ITS功能序列是否存在多态性。另一方面, 在17个樟属样本的9个样本中发现有假基因序列的存在, 而且假基因序列间的差异更大(大于功能序列之间的差异)(表2)。显然, 樟属植物的nrDNA ITS序列存在不完全一致性进化。

以往的研究表明, 多倍化、异源多倍体、杂交以及基因家族在染色体上的结构等因素都有可能导致一致性进化不完全(Muir et al., 2001; Xiao et al., 2010; 周阿涛等, 2013)。而在樟属植物的研究中至今未见有关杂交的报道, 也未发现多倍化和异源多倍体现象(Okada and Tanaka, 1975)。ITS序列多态性还有可能来源于同一细胞内不同ITS拷贝基因座位的不同。ITS在同一基因组内有成千上万的拷贝, 而驱动ITS一致性进化的分子机制有不等交换(unequal crossover)、基因转换(gene conversion)和基因扩增(gene amplification)等(Hillis et al., 1991; Gonzalez and Sylvester, 2001; Kovari et al., 2004)。许多研究表明, 位于同一条染色体上的序列比位于不同染色体位点上的序列更容易发生一致性进化(Copenhaver and Pikaard, 1996; Parkin and Butlin, 2004); 不等交换在姐妹染色单体之间比非姐妹染色单体之间更加频繁, 所以位点之间的基因重组频率要比位点内低很多(Copenhaver and Pikaard, 1996; Eickbush and Eickbush, 2007)。对山茶属的研究表明, 在该类群的rDNA上存在大量位点, 且大都不在同一染色体上, 可能会阻碍不等交换的发生, 这可能是导致其ITS序列多态性的原因之一(Gu and Xiao, 2003; 徐颖等, 2011; 周阿涛等, 2013)。然而在樟属植物中由于缺乏相关研究, 我们难以对其ITS序列多态性产生原因进行更深入的探讨。

此外, 在本研究的17个样本中, *C. appelianum* H1C和*C. subavenium* H43A两个样本各自的功能序列在系统发育关系重建中不能形成单系, 同时也不能

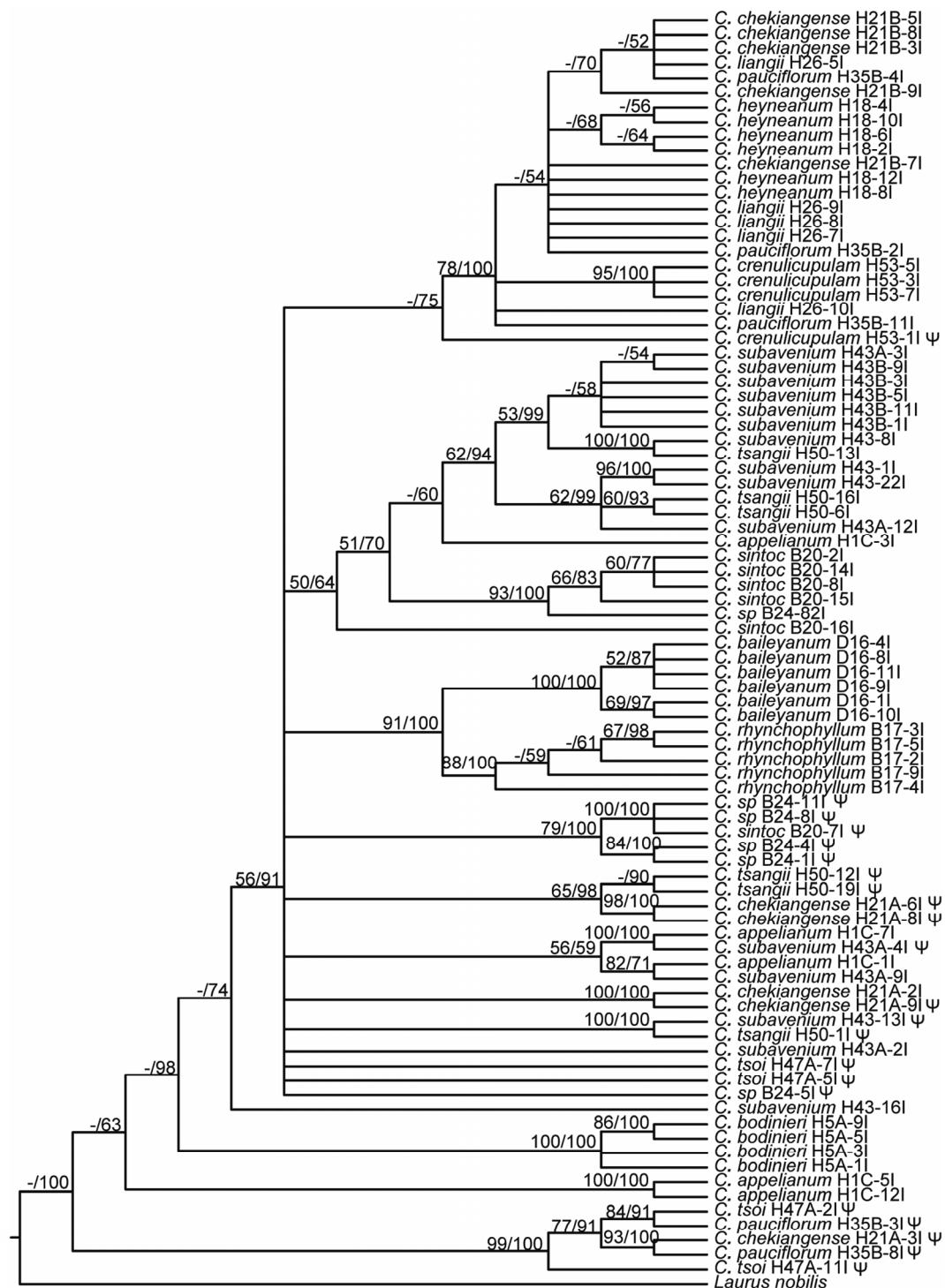


图2 基于Bayesian算法对ITS克隆测序样本构建的50%一致性系统进化树

大于50%的靴带支持率/大于50%的后验率标注在进化支上(bootstrap support/posterior possibility). C: *Cinnamomum*; ψ: ITS假基因序列

Figure 2 Fifty percent majority rule consensus tree from Bayesian analysis of ITS clone sequence dataset

Bootstrap support values ($\geq 50\%$)/Bayesian posterior possibilities ($\geq 50\%$) are shown above the branches. C: *Cinnamomum*; ψ: ITS pseudogene sequences

以信息位点不足来解释,表明ITS序列在这2个样本中存在高度的分化;而在*C. subavenium*的另外2个样本*C. subavenium* H43和*C. subavenium* H43A中未发现类似情况,可能是实验中选取的阳性克隆数目不足所致。

2.5.2 ITS假基因

nrDNA ITS基因家族有成千上万个拷贝,在进化过程中,一些拷贝可能会由于功能退化而演变成假基因。本研究在樟科类群中发现了ITS假基因的存在,与功能序列相比,樟属植物的ITS假基因序列GC含量明显降低;长度变异明显增大,特别是在高度保守的5.8S编码区的长度变化大,碱基替代率高;5.8S编码区的最小自由能增大,二级结构稳定性降低。由于假基因不受选择压力的约束,以中性进化模式进化,因而比功能基因的进化速率要快。本文在遗传距离的分析中也显示,樟属植物的假基因序列相对于外类群的遗传距离显著大于功能序列相对于外类群的遗传距离。因此用于重建系统发育关系时,假基因可能会因为长支吸引而聚在一起。我们的系统发育分析(图2)也显示,除了少数假基因序列与功能序列聚在一起(例如,*C. chekiangense* H21A-9I、*C. crenulicupulam* H53-1I和*C. subavenium* H43A-4I)之外,绝大部分假基因序列相互聚在一起,与功能序列分离开来。因此,若将假基因序列与功能序列一起用于系统发育关系的重建会对系统树拓扑结构的稳定性造成影响,扰乱我们对类群间系统发育关系的准确理解,因此应将它们排除在外。

此外,在进行PCR扩增实验时,模板的高GC含量区域在变性后的复性过程中容易形成二级结构,DNA聚合酶跨越这些结构,得到较短的非特异扩增产物,这在我们对樟科植物ITS序列的PCR扩增实验中经常出现。对樟属植物ITS序列GC含量的统计分析(表2)显示其GC含量非常高,特别是ITS1 (74.4%)和ITS2区域(72.1%)。而由于假基因序列的GC含量和二级结构的稳定性都明显下降,因而在普通PCR条件下会因首先变性而被优先扩增。而在反应体系中加入DMSO可以破坏DNA中的氢键,使GC含量高的模板易于完全变性,有利于减少DNA的二级结构,这样聚合酶就可以无障碍地读取DNA模板上的信息,从而得到理想的特异扩增产物,同时能够有效防止假基因

序列的扩增,提高功能序列的扩增效率(Buckler and Holtsford, 1996; Zheng et al., 2008; Xiao et al., 2010; 尤欢等, 2014)。我们在对樟属129个样本的ITS序列进行扩增时(Huang et al., 2016),在25 μ L的PCR扩增体系中添加了2 μ L的DMSO,最后只在其中9个样本中发现了假基因序列,这在一定程度上说明DMSO能够有效减少假基因的扩增成功率,但并不能完全杜绝。

另外,由于ITS1区域的变异相对于ITS2更大,特别是高频率地出现缺失/插入,会增加排序的困难,尤其是当类群之间的亲缘关系较远时(如科内属间关系),而5.8S保守区的系统发育信息非常有限。相对于ITS1和5.8S区,ITS2区域不仅能够提供较多的信息量,同时也能让排序变得简单易行,然而只有ITS2区的数据明显增加了ITS假基因判断的困难。因此对于只有ITS2片段的序列,怀疑有可能存在ITS假基因的类群,建议增加ITS1和5.8S区的数据,特别是5.8S保守区的序列对假基因的判断至关重要。

本研究首次对樟科植物中ITS不完全一致性进化和假基因现象进行了报道,虽然目前只是在樟属植物中发现了上述现象,但在樟科其它类群中可能也存在类似情况,应该引起学者的注意,特别是在叶绿体基因组分子标记对樟科类群系统发育关系解决能力极其有限的情况下(Rohwer, 2000; Chanderbali et al., 2001; Rohwer and Rudolph, 2005),ITS序列会首先成为学者的选择。本研究结合了不同的分析方法和策略,能够很好地鉴定ITS假基因,具有一定的借鉴价值。由于ITS序列的不完全一致性进化和假基因的存在,导致通过直接测序方式获得的序列出现严重的杂合现象而无法使用,这也许是在樟科植物DNA条形码研究中ITS序列测序成功率低的重要原因之一。

致谢 本研究承蒙中国科学院昆明植物研究所李锡文研究员在樟属植物的分类鉴定中提供的重要帮助;澳大利亚热带标本馆(Australian Tropical Herbarium, ATH)馆长Darren Crayn教授提供的澳大利亚樟属样本;慈秀芹和丁鑫在样品采集中提供支持;徐武美在数据统计分析方面提供了帮助。在此一并致谢。

参考文献

龚洵, 潘跃芝, 杨志云 (2001). 木兰科植物的杂交亲和性. 云南植物研究 23, 339–344.

- 卢孟孟, 慈秀芹, 杨国平, 李捷 (2013). 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. *植物分类与资源学报* **35**, 733–741.
- 马长乐, 周浙昆 (2006). ITS 假基因对柝属系统学研究的影响及其对分子系统学研究的启示. *云南植物研究* **28**, 127–132.
- 徐颖, 徐晶, 高继银, 张文驹 (2011). 山茶属植物 ITS 的多态性——一个广泛逃离一致性进化的实例. *植物学报* **46**, 162–169.
- 尤欢, 周阿涛, 岳亮亮, 寸东义, 丁元明 (2014). 山茶属植物 ITS 的扩增及其序列特征分析. *植物研究* **34**, 403–408.
- 周阿涛, 岳亮亮, 李旻, 刘迪秋, 丁元明 (2013). 云南山茶 (*Camellia reticulata*) nrDNA ITS 序列多态性分析. *植物科学学报* **31**, 1–10.
- Álvarez I, Wendel JF (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* **29**, 417–434.
- Bailey CD, Carr TG, Harris SA, Hughes CE (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Mol Phylogenet Evol* **29**, 435–455.
- Buckler ES, Holtsford TP (1996). Zea systematics: ribosomal ITS evidence. *Mol Biol Evol* **13**, 612–622.
- Buckler ES, Ippolito A, Holtsford TP (1996). The evolution of ribosomal DNA: divergent paralogous and phylogenetic implications. *Genetics* **145**, 821–832.
- CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc Natl Acad Sci USA* **106**, 12794–12797.
- Chanderbali AS, van der Werff H, Renner SS (2001). Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. *Ann Missouri Bot Gard* **88**, 104–134.
- Chen ZY, Xiong ZJ, Pan XY, Shen SQ, Geng YP, Xu CY, Chen JK, Zhang WJ (2015). Variation of genome size and the ribosomal DNA ITS region of *Alternanthera philoxeroides* (Amaranthaceae) in Argentina, the USA, and China. *J Syst Evol* **53**, 82–87.
- Copenhaver GP, Pikaard CS (1996). Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J* **9**, 273–282.
- Eickbush TH, Eickbush DG (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**, 477–485.
- Felsenstein J (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368–376.
- Fu YX (1996). New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**, 557–570.
- Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.
- Gonzalez IL, Sylvester JE (2001). Human rDNA: evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. *Genomics* **73**, 255–263.
- Gu ZJ, Xiao H (2003). Physical mapping of the 18S-26S rDNA by fluorescent *in situ* hybridization (FISH) in *Camellia reticulata* polyploidy complex (Theaceae). *Plant Sci* **164**, 279–285.
- Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser (Oxf)* **41**, 95–98.
- Hillis DM, Moritz C, Porter CA, Baker RJ (1991). Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**, 308–310.
- Ho KY, Hung TY (2011). Cladistic relationships within the genus *Cinnamomum* (Lauraceae) in Taiwan based on analysis of leaf morphology and inter-simple sequence repeat (ISSR) and internal transcribed spacer (ITS) molecular markers. *Afr J Biotechnol* **10**, 4802–4815.
- Huang JF, Li L, van der Werff H, Li HW, Rohwer JG, Crayn DM, Meng HH, van der Merwe M, Conran JG, Li J (2016). Origins and evolution of cinnamon and camphor: a phylogenetic and historical biogeographical analysis of the *Cinnamomum* group (Lauraceae). *Mol Phylogenet Evol* **96**, 33–44.
- Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.
- Kovari A, Matyasek R, Lim KY, Skalická K, Koukalova B, Knapp S, Chase M, Leitch AR (2004). Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. *Biol J Linn Soc* **82**, 615–625.
- Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW (2011a). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci USA* **108**, 19641–19646.
- Li L, Li J, Rohwer JG, van der Werff H, Wang ZH, Li HW

- (2011b). Molecular phylogenetic analysis of the *Persea* group (Lauraceae) and its biogeographic implications on the evolution of tropical and subtropical amphi-Pacific disjunctions. *Am J Bot* **98**, 1520–1536.
- Mayol M, Rossello JA** (2001). Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Mol Phylogenet Evol* **19**, 167–176.
- Muir G, Fleming CC, Schlotterer C** (2001). Tree divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Mol Biol Evol* **18**, 112–119.
- Okada H, Tanaka R** (1975). Karyological studies in some species of Lauraceae. *Taxon* **24**, 271–280.
- Parkin EJ, Butlin RK** (2004). Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Mol Biol Evol* **21**, 1595–1601.
- Posada D, Buckley TR** (2004). Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* **53**, 793–808.
- Posada D, Crandall KA** (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Qiu YL, Chase MW, Parks CR** (1995). A chloroplast DNA phylogenetic study of the eastern Asia-eastern North America disjunct section *Rytidospermum* of *Magnolia* (Magnoliaceae). *Am J Bot* **82**, 1582–1588.
- Rohwer JG** (2000). Toward a phylogenetic classification of the Lauraceae: evidence from *matK* sequences. *Syst Bot* **25**, 60–71.
- Rohwer JG, Rudolph B** (2005). Jumping genera: the phylogenetic positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) based on different analyses of *trnK* intron sequences. *Ann Missouri Bot Gard* **92**, 153–178.
- Ronquist F, Huelsenbeck JP** (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
- Rozas J, Rozas R** (1999). DnaSP version 3, an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.
- Swofford DL** (2003). PAUP*: Phylogenetic Analysis Using Parsimony (and other methods), Version 4.0b10. Sunderland (Massachusetts), Sinauer.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S** (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG** (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Tripathi AM, Tyagi A, Kumar A, Singh A, Singh S, Chaudhary LB, Roy S** (2013). The internal transcribed spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India. *PLoS One* **8**, e57934.
- Vanin EF** (1985). Processed pseudogenes: characteristics and evolution. *Ann Rev Genet* **19**, 253–272.
- Wei XX, Wang XQ** (2004). Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. *Mol Ecol* **13**, 3115–3123.
- White TJ, Bruns TD, Lee SB, Taylor JW** (1990). Amplification and Direct Sequencing of Ribosomal RNA Genes and the Internal Transcribed Spacer in Fungi. PCR Protocols and Applications—a Laboratory Manual. Orlando: Academic Press. pp. 315–322.
- Xiao LQ, Möller M, Zhu H** (2010). High nrDNA ITS polymorphism in the ancient extant seed plant *Cycas*: incomplete concerted evolution and the origin of pseudogenes. *Mol Phylogenet Evol* **55**, 168–177.
- Zheng X, Cai D, Yao L, Teng Y** (2008). Non-concerted ITS evolution, early origin and phylogenetic utility of ITS pseudogenes in *Pyrus*. *Mol Phylogenet Evol* **48**, 892–903.

Polymorphism of the Internal Transcribed Spacer of nrDNA in *Cinnamomum* Schaeffer (Lauraceae)

Jianfeng Huang^{1,2}, Lang Li¹, Jie Li^{1*}

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

Abstract In this study, nuclear ribosomal DNA (nrDNA) ITS fragments of 17 *Cinnamomum* samples were cloned and sequenced. A total of 87 different ITS sequences were obtained. Their length polymorphism, GC content, secondary structure stability of the 5.8S region, p-distance, evolutionary pattern and phylogenetic relationships were analyzed. The ITS sequences had a high degree of polymorphism. Among the 87 ITS sequences, 22 were pseudogenes and the other 65 were functional genes. The pseudogenes showed a neutral evolutionary pattern and had a higher degree of polymorphism than functional genes. The incomplete concerted evolution of the ITS region and pseudogenes within the genus *Cinnamomum* may have also occurred in other Lauraceae plants, which could be responsible for the low sequencing success rate of the ITS region in Lauraceae plants.

Key words *Cinnamomum*, ITS, polymorphism, pseudogene

Huang JF, Li L, Li J (2016). Polymorphism of the internal transcribed spacer of nrDNA in *Cinnamomum* Schaeffer (Lauraceae). *Chin Bull Bot* **51**, 609–619.

* Author for correspondence. E-mail: jieli@xtbg.ac.cn

(责任编辑: 朱亚娜)