Application of deoxyribonucleic acid barcoding in Lauraceae plants

Zhen Liu[1,2], Shi-Lin Chen[3], Jing-Yuan Song[3], Shou-Jun Zhang[4], Ke-Li Chen[2]

[1]Department of Pharmacy, The 309th Hospital of Chinese People's Liberation Army, Beijing , [2]Key Laboratory of Traditional Chinese Medicine Resource and Compound Prescription, Ministry of Education, Hubei University of Chinese Medicine, Wuhan, [3]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, [4]Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Republic of China

Submitted: 12-04-2011 Revised: 07-05-2011 Published: 28-02-2012

ABSTRACT

Background: This study aims to determine the candidate markers that can be used as DNA barcode in the Lauraceae family. **Material and Methods:** Polymerase chain reaction amplification, sequencing efficiency, differential intra- and interspecific divergences, DNA barcoding gap, and identification efficiency were used to evaluate the four different DNA sequences of *psbA-trnH, matK, rbcL*, and ITS2. We tested the discrimination ability of *psbA-trnH* in 68 plant samples belonging to 42 species from 11 distinct genera and found that the rate of successful identification with the *psbA-trnH* was 82.4% at the species level. However, the correct identification of *matK* and *rbcL* were only 30.9% and 25.0%, respectively, using BLAST1. The PCR amplification efficiency of the ITS2 region was poor; thus, ITS2 was not included in subsequent experiments. To verify the capacity of the identification of *psbA-trnH* in more samples, 175 samples belonging to 117 species from the experimental data and from the GenBank database of the Lauraceae family were tested. **Results:** Using the BLAST1 method, the identification efficiency were 84.0% and 92.3% at the species and genus level, respectively. **Conclusion:** Therefore, *psbA-trnH* is confirmed as a useful marker for differentiating closely related species within Lauraceae.

Key words: Deoxyribonucleic acid barcoding, ITS2, Lauraceae, *matK*, *psbA-trnH*, *rbcL*

Access this article online

Website: www.phcog.com

DOI: 10.4103/0973-1296.93301

Quick Response Code:

INTRODUCTION

Lauraceae is a large family of woody plants (except the herbaceous parasite, *Cassytha*) with about 50 genera and 2500 to 3000 species distributed throughout tropical to subtropical latitudes. Lauraceae plants have the extremely important economic value. A great number of them are important resource in the construction timber, spice, essential oil, and medicinal plants. Simultaneously, as their crowns are spacious, they have immense ecological value for virescence and environment protection. Boasting of various kinds and widespread distribution, Lauraceae plants are known to have an ancient origin with a fossil record dating back to the mid-Cretaceous period.[1] However, the evolution and developing process of these plants are very slow. Since boundaries of many species in the family are quite unclear, it is difficult to identify them while the traditional morphological methods are used. Thus, it is significant to develop a quick, simple, and effective method to identify the species in the Lauraceae family.

Address for correspondence:
Prof. Ke-Li Chen, Key Laboratory of Traditional Chinese Medicine Resource and Compound Prescription, Ministry of Education, Hubei University of Chinese Medicine, Wuhan 430065, Republic of China. E-mail: kelichen@126.com

Deoxyribonucleic acid (DNA) barcoding is the researching focus on biodiversity in the world in recent years. The core of the research is to choose a universal barcode in order to appraise the species quickly and accurately. In 2003, Herbert analyze the order of the genes of the cytochrome c oxidase subunit 1 (*CO1*) belonging to 11 phyla from 13320 species.[2] Then, as regards animals, most researchers agree that the mitochondrial gene encoding *CO1* is a favorable region for use as the standard DNA barcode in the world. Compared with the excellent study in the animal barcode, the study in the plants barcode is relatively slow.

The plant working group of the Consortium for the barcode of life recommended the two-locus combination of *rbcL* + *matK* for plant barcoding.[3] Chen *et al.*, tested the discrimination ability of ITS2 in more than 6600 plant samples belonging to 4800 species from 753 distinct genera; they found that the ITS2 region possesses many

advantages compared with plastid loci, including *rbcL* and *matK* region. They also recommended for *psbA-trnH* to be a complementary barcode to ITS2 for a broad series of plantae.[4]

Despite some scholars having carried out DNA barcoding research for related species and genera,[5-9] none had referred to multiple samples in the Lauraceae family. In this study, four potential DNA regions (*psbA-trnH*, *matK*, *rbcL*, and ITS2) were tested for their suitability as DNA barcodes for the Lauraceae family (68 samples belonging to 42 species from 11 genera). The true ability of the candidate sequences to identify species of Lauraceae as a universal DNA barcode is assessed in spite of many closely related species in the samples.

MATERIALS AND METHODS

Experimental materials (68 samples belonging to 42 species from 11 diverse genera) were collected from the Chinese provinces of Hubei, Jiangxi, Guangdong, and Guangxi. The materials are authenticated by Prof. Panhong Lin of Hubei College of Traditional Chinese Medicine and Engr. Zhang Shoujun of Wuhan Botanical Garden at the Chinese Academy of Sciences. All specimen and image vouchers were maintained at the herbarium of Hubei College of Traditional Chinese Medicine. To increase further the number of species represented, *psbA-trnH* sequences from the taxonomy database of the National Centre for Biotechnology Information (NCBI) were included in the reference database.

Leaf tissues were firstly dried in silica gel. A total of 10 mg of each of the dried tissues was rubbed for 1 min at a frequency of 30 times/second in a FastPrep bead mill (Retsch MM400, Germany). Total DNA was extracted as instructed by the Plant Genomic DNA Kit (Tiangen Biotech Co., China). The polymerase chain reaction (PCR) reaction mixture consisted of 1 μL (~30 ng) DNA, 2 μL of 25 mM $MgCl_2$, 2.5 μL of 10×PCR buffer, 1.0 U of Taq DNA polymerase, 2 μL of 2.5 mM dNTPs mix (Biocolor BioScience and Technology Co., China), 1.0 μL of 2.5 μM primers (Synthesized by Sangon Co., China); the final volume was 25 μL. Sequences of the universal primers for the tested DNA barcode, including those for *psbA-trnH*, *matK*, *rbcL*, and ITS2, as well as general PCR reaction conditions, were obtained from previous studies.[4] PCR products were purified using the Gel Band Purification Kit (Tiangen Biotech Co., China) and sequenced on an ABI 3730XL sequencer (Applied Biosystems, USA). The sequences were submitted to GenBank.

Sequence editing and contig assembly were conducted by CodonCode Aligner (CodonCode Co., Germany). Sequences were aligned using CLUSTALW and analyzed by the MEGA 4.0 software program. Average interspecific distances, theta prime, and smallest interspecific distances were used to characterize interspecific divergences.[4,10,11] Average intraspecific distances, theta, and coalescent depth were calculated to determine intraspecific variations using Kimura 2-parameter (K2P) distances.[10] Wilcoxon signed rank tests were performed as described previously.[12,13] Barcoding gap was calculated by TAXON DNA.[14] To estimate the reliability of species identification using the DNA barcoding technique, two methods (BLAST1 and the nearest genetic distance) were carried out.[15]

RESULTS

PCR amplification and sequencing efficiency

Results showed that *psbA-trnH*, *matK,* and *rbcL* sequences were successfully amplified and sequenced at 100%. However, in our pilot study, the PCR amplification efficiency of the ITS2 region was poor; thus, ITS2 was not included in subsequent experiments [Table 1].

Analysis of intraspecific variations and interspecific divergences

A favorable barcode should own low intraspecific variations and high interspecific divergence in order to distinguish different species. First, upon comparison of interspecific genetic distances among congeneric species for three candidate barcodes, it was observed that the chloroplast noncoding region of *psbA-trnH* exhibited the highest interspecific divergence for all three metrics, followed by *rbcL*, while *matK* provided the lowest divergence [Table 2]. Moreover, Wilcoxon signed rank tests confirmed that *psbA-trnH* provided the highest interspecific divergence among congeneric species [Table 3].

Second, it was found that *matK* showed the lowest level of intraspecific variation for all three parameters, followed

Table 1: Efficiency of polymerase chain reaction amplification and success rate of sequencing of potential barcodes in total number of samples

Marker	*psbA-trnH*	*matK*	*rbcL*	ITS2
Number of samples	68	68	68	68
The efficiency of PCR amplification (%)	100	100	100	32.35
The success rate of sequencing (%)	100	100	100	27.27
The rate of obtained sequences (%)	100	100	100	8.823

PCR: Polymerase chain reaction

by *rbcL*, while *psbA-trnH* provided the highest variation [Table 2]. Wilcoxon signed rank tests showed that *rbcL* and *matK* have the lowest variation between conspecific individuals, whereas *psbA-trnH* showed the highest [Table 4].

Assessment of the barcoding gap

Ideally, barcoding involves separate distributions and without overlap between intra- and interspecific variations.[10,16] Results of the present study showed that *psbA-trnH* have a faint gap, whereas *matK* and *rbcL* exhibited significant overlap without any gaps [Figures 1 and 2].

Evaluation of identifying ability of barcodes

In the BLAST1 method, results showed that *psbA-trnH*

Table 2: Analysis of interspecific divergence between congeneric species and intraspecific variation of candidate barcodes

Markers	*psbA-trnH*	*matK*	*rbcL*
All inter-specific distance	0.0176 ± 0.0111	0.0024 ± 0.0034	0.0034 ± 0.0041
Theta prime	0.0178 ± 0.0102	0.0032 ± 0.0034	0.0032 ± 0.0021
Minimum inter-specific distance	0.0047 ± 0.0080	0.0011 ± 0.0026	0.0008 ± 0.0013
All intra-specific distance	0.0032 ± 0.0065	0.0001 ± 0.0005	0.0007± 0.0033
Theta	0.0036 ± 0.0070	0.0001 ± 0.0005	0.0008 ± 0.0035
Coalescent depth	0.0036 ± 0.0070	0.0001 ± 0.0005	0.0008 ± 0.0035

Table 3: Wilcoxon signed rank test for interspecific divergences

W +	W -	Interrelative ranks, n, *P* value	Result
psbA-trnH	*matK*	W+ = 5565.0, W- = 0.0, n = 105, $P < 5.7967E^{-19}$	*psbA-trnH* ≫ *matK*
psbA-trnH	*rbcL*	W+ = 5550.0, W- = 15.0, n = 105, $P < 8.9199E^{-19}$	*psbA-trnH* ≫ *rbcL*
rbcL	*matK*	W+ = 1727.0, W- = 1048, n = 74, $P < 0.0671$	*rbcL* = *matK*

Table 4: Wilcoxon signed rank test for intraspecific variations

W +	W -	Intrarelative Ranks, n, *P* value	Result
psbA-trnH	*matK*	W+ = 120.0, W- = 0.0, n = 15, $P < 5.5225E^{-4}$	*psbA-trnH* > *matK*
psbA-trnH	*rbcL*	W+ = 125.0, W- = 28.0, n = 17, $P < 0.0205$	*psbA-trnH* > *rbcL*
rbcL	*matK*	W+ = 3.0, W- = 0.0, n = 2, $P < 0.1797$	*rbcL* = *matK*

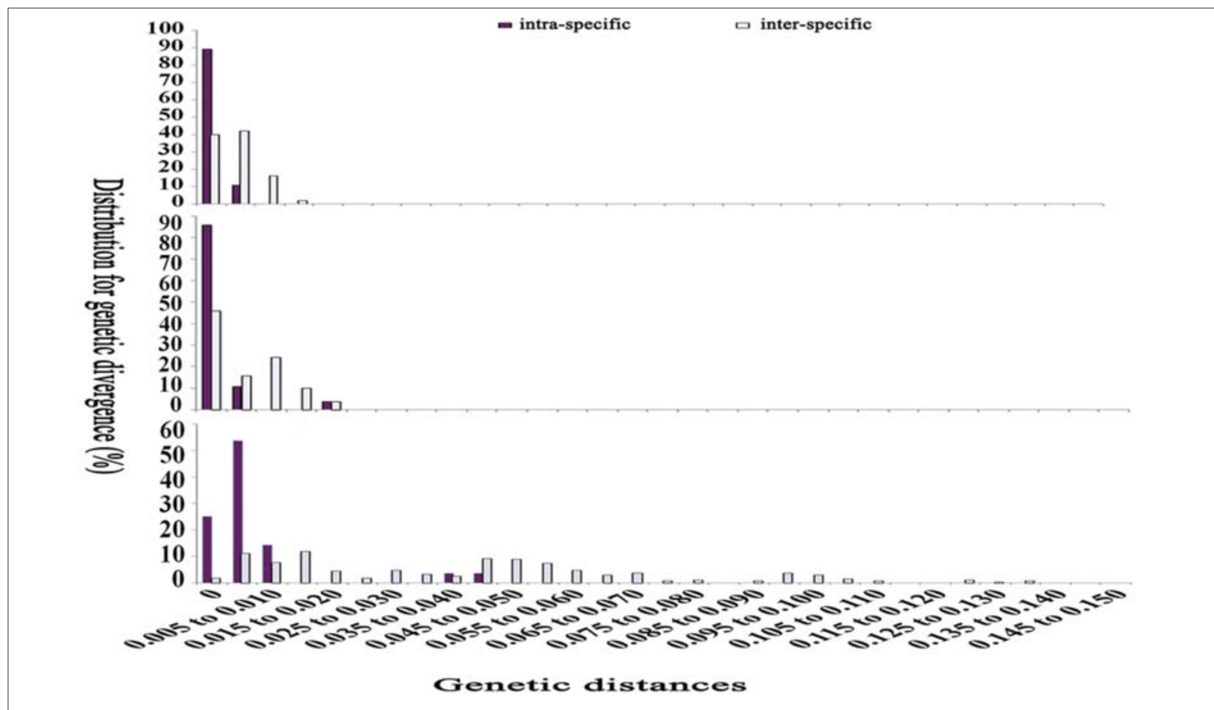

Figure 1: Schematic representation of the deoxyribonucleic acid barcoding gap between interspecific and intraspecific divergences for three candidate DNA barcodes. (a) *matK*; (b) *rbcL*; and (c) *psbA-trnH*

identified correctly 82.4% of the samples at the species level and 88.1% at the genus level. In contrast to *psbA-trnH*, the correct identification for *matK* and *rbcL* were much lower at the species level, as identified by both BLAST1 and nearest genetic distance methods. At the species level, the correct identification of the two-locus combination of *rbcL* + *matK*, *matK* + *psbA-trnH,* and *rbcL* + *psbA-trnH* were 38.2%, 82.4%, and 82.4%, respectively, using BLAST1 [Table 5]. To verify the capacity of the identification of *psbA-trnH* in more samples, 175 samples belonging to 117 species from the experimental data and from the GenBank database of the Lauraceae family were tested [Tables S1 and S2]. Using the BLAST1 method, the identification efficiency were 84.0% and 92.3% at the species and genus level, respectively.

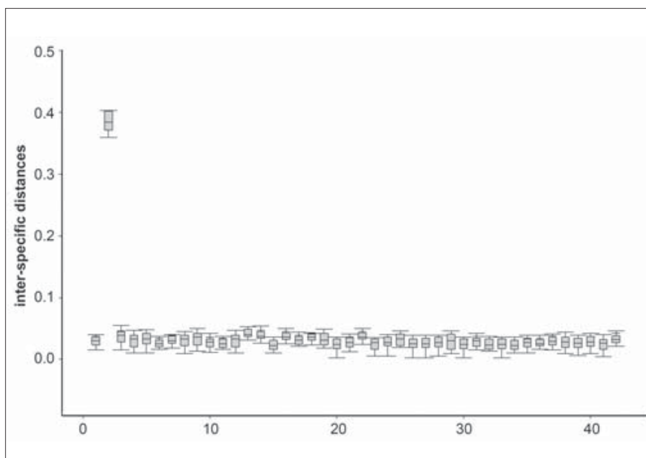

Figure 2: The interspecific divergence of the *psbA-trnH* region in Lauraceae

DISCUSSION

This work, which focused on four popular candidate sequences of *matK*, *rbcL*, *psbA-trnH,* and nrDNA ITS2, has conducted a comparative study of 11 genera 42 species from 68 samples of Lauraceae. In the experiments, it was found that *matK, rbcL, rbcL* + *matK,* and ITS2 were not suitable as a barcode for the Lauraceae family. The *psbA-trnH* region presented itself with short length, easy sequencing, and powerful ability of species identification for Lauraceae plants. By comparing *matK*, *rbcL,* and ITS2, it was found that the *psbA-trnH* region is the best marker for the identification of Lauraceae species.

Selection of the DNA barcode for the Lauraceae family

In the present research, it was found that *psbA-trnH,* as a barcode sequence, showed excellent results. First, the *psbA-trnH* region has a short length in the 195–423 base pairs, which can then be easily amplified and sequenced. The success rate of PCR amplification and sequencing for the *psbA-trnH* of 68 samples from 11 genera of Lauraceae were 100%. Second, the determination of genetic divergences using six metrics and statistical tests confirmed that the *psbA-trnH* region possesses sufficient high interspecific variation. There existed significant differences between interspecific and intraspecific variations. Third, according to BLAST1, the identification efficiency using the *psbA-trnH* region was 84.0% at the species level for the 175 samples from 117 species in 35 genera of Lauraceae. Moreover, the two loci combination of *matK* + *psbA-trnH* and *rbcL* + *psbA-trnH* did not show any improved abilities for identification. The *psbA-trnH* can identify all the species, which were identified by *matK, rbcL,* and the two-locus combination of *rbcL* + *matK*.

Table 5: Comparison of identification efficiency for potential deoxyribonucleic acid barcodes loci using different methods of species identification

Marker	Methods of species identification	Number of samples	Correct identification		Incorrect identification		Ambiguous identification	
			Species level %	Genus level %	Species level %	Genus level %	Species level %	Genus level %
matK	BLAST1	68	30.9	28.6	0	0	69.1	71.4
	Distance	68	27.9	31.0	0	0	72.1	69.0
rbcL	BLAST1	68	25.0	40.5	0	0	75.0	59.5
	Distance	68	25.0	42.9	0	0	75.0	57.1
psbA-trnH	BLAST1	68	82.4	88.1	0	0	17.6	11.9
	Distance	68	64.7	81.0	0	0	35.3	19.0
rbcL + matK	BLAST1	68	38.2	50.0	0	0	61.8	50.0
	Distance	68	36.8	50.0	0	0	63.2	50.0
matK + psbA-trnH	BLAST1	68	82.4	88.1	0	0	17.6	11.9
	Distance	68	66.2	81.0	0	0	33.8	19.0
rbcL + psbA-trnH	BLAST1	68	82.4	88.1	0	0	17.6	11.9
	Distance	68	67.6	83.3	0	0	32.4	16.7

Table S1: Samples for testing potential barcodes and accession numbers in GenBank

Species	Collection sites	Voucher number	GenBank accession		
			psbA-trnH	*matK*	*rbcL*
Actinodaphne omeiensis	HuBei	PS5014MT01	HM019377	HM019307	HM019447
Actinodaphne omeiensis	HuBei	PS5014MT02	HM019378	HM019308	HM019448
Actinodaphne omeiensis	HuBei	PS5014MT03	HM019379	HM019309	HM019449
Cassytha filiformis	GuangDong	PS5015MT01	HM019380	HM019310	HM019450
Cassytha filiformis	GuangDong	PS5015MT02	HM019381	HM019311	HM019451
Cinnamomum bodinieri	HuBei	PS5016MT01	HM019382	HM019312	HM019452
Cinnamomum bodinieri	HuBei	PS5016MT02	HM019383	HM019313	HM019453
Cinnamomum burmannii	GuangDong	PS5017MT01	HM019384	HM019314	HM019454
Cinnamomum burmannii	GuangXi	PS5018MT01	HM019385	HM019315	HM019455
Cinnamomum camphora	HuBei	PS5019MT01	HM019386	HM019316	HM019456
Cinnamomum camphora	HuBei	PS5020MT01	HM019387	HM019317	HM019457
Cinnamomum cassia	GuangDong	PS5021MT01	HM019388	HM019318	HM019458
Cinnamomum cassia	GuangDong	PS5021MT02	HM019389	HM019319	HM019459
Cinnamomum japonicum	HuBei	PS5022MT01	HM019390	HM019320	HM019460
Cinnamomum jensenianum	HuBei	PS5023MT01	HM019391	HM019321	HM019461
Cinnamomum jensenianum	HuBei	PS5023MT02	HM019392	HM019322	HM019462
Cinnamomum pauciflorum	HuBei	PS5024MT01	HM019393	HM019323	HM019463
Cinnamomum pauciflorum	HuBei	PS5024MT02	HM019394	HM019324	HM019464
Cinnamomum platyphyllum	HuBei	PS5025MT01	HM019395	HM019325	HM019465
Cinnamomum platyphyllum	HuBei	PS5025MT02	HM019396	HM019326	HM019466
Cinnamomum tonkinense	HuBei	PS5027MT01	HM019397	HM019327	HM019467
Cinnamomum wilsonii	HuBei	PS5028MT01	HM019398	HM019328	HM019468
Laurus nobilis	HuBei	PS5029MT01	HM019399	HM019329	HM019469
Laurus nobilis	HuBei	PS5029MT02	HM019400	HM019330	HM019470
Lindera aggregata	JiangXi	PS5030MT01	HM019401	HM019331	HM019471
Lindera aggregata	JiangXi	PS5031MT01	HM019402	HM019332	HM019472
Lindera aggregata	HuBei	PS5031MT02	HM019403	HM019333	HM019473
Lindera chunii	HuBei	PS5032MT01	HM019404	HM019334	HM019474
Lindera fragrans	HuBei	PS5033MT01	HM019405	HM019335	HM019475
Lindera fragrans	HuBei	PS5033MT02	HM019406	HM019336	HM019476
Lindera glauca	GuangDong	PS5034MT01	HM019407	HM019337	HM019477
Lindera glauca	GuangDong	PS5034MT02	HM019408	HM019338	HM019478
Litsea cubeba	GuangDong	PS5036MT01	HM019411	HM019341	HM019481
Litsea glutinosa	GuangDong	PS5037MT01	HM019412	HM019342	HM019482
Litsea honghoensis	HuBei	PS5038MT01	HM019413	HM019343	HM019483
Litsea ichangensis	HuBei	PS5039MT01	HM019414	HM019344	HM019484
Litsea ichangensis	HuBei	PS5039MT02	HM019415	HM019345	HM019485
Litsea monopetala	HuBei	PS5040MT01	HM019416	HM019346	HM019486
Litsea pungens	HuBei	PS5041MT01	HM019417	HM019347	HM019487
Machilus ichangensis	HuBei	PS5042MT01	HM019418	HM019348	HM019488
Machilus leptophylla	HuBei	PS5043MT01	HM019419	HM019349	HM019489
Machilus leptophylla	GuangXi	PS5044MT01	HM019420	HM019350	HM019490
Machilus lichuanensis	HuBei	PS5045MT01	HM019421	HM019351	HM019491
Machilus microcarpa	HuBei	PS5046MT01	HM019422	HM019352	HM019492
Machilus microcarpa	HuBei	PS5046MT02	HM019423	HM019353	HM019493
Machilus oreophila	GuangXi	PS5047MT01	HM019424	HM019354	HM019494
Machilus oreophila	GuangXi	PS5047MT02	HM019425	HM019355	HM019495
Machilus pauhoi	HuBei	PS5048MT01	HM019426	HM019356	HM019496
Machilus rufipes	HuBei	PS5049MT01	HM019427	HM019357	HM019497
Neolitsea aurata	HuBei	PS5050MT01	HM019428	HM019358	HM019498
Neolitsea confertifolia	HuBei	PS5051MT01	HM019429	HM019359	HM019499
Neolitsea confertifolia	HuBei	PS5051MT02	HM019430	HM019360	HM019500
Neolitsea hsiangkweiensis	HuBei	PS5052MT01	HM019431	HM019361	HM019501
Neolitsea hsiangkweiensis	HuBei	PS5052MT02	HM019432	HM019362	HM019502
Neolitsea levinei	HuBei	PS5053MT01	HM019433	HM019363	HM019503
Neolitsea levinei	HuBei	PS5053MT02	HM019434	HM019364	HM019504

Table S1: Contd...

Species	Collection sites	Voucher number	GenBank accession		
			psbA-trnH	*matK*	*rbcL*
Neolitsea sericea	HuBei	PS5054MT01	HM019435	HM019365	HM019505
Neolitsea wushanica var. pubens	HuBei	PS5055MT01	HM019436	HM019366	HM019506
Persea americana	GuangDong	PS5056MT01	HM019437	HM019367	HM019507
Phoebe bournei	HuBei	PS5057MT01	HM019438	HM019368	HM019508
Phoebe bournei	HuBei	PS5057MT02	HM019439	HM019369	HM019509
Phoebe chekiangensis	HuBei	PS5058MT01	HM019440	HM019370	HM019510
Phoebe chekiangensis	HuBei	PS5058MT02	HM019441	HM019371	HM019511
Phoebe sheareri	HuBei	PS5060MT01	HM019442	HM019372	HM019512
Phoebe sheareri	HuBei	PS5060MT02	HM019443	HM019373	HM019513
Phoebe zhennan	GuangDong	PS5061MT01	HM019444	HM019374	HM019514
Phoebe zhennan	HuBei	PS5061MT02	HM019445	HM019375	HM019515
Sassafras tzumu	HuBei	PS5062MT01	HM019446	HM019376	HM019516

Table S2: Samples for determining the ability of the *psbA-trnH* barcode to identify species and accession numbers in GenBank

Genus	Species	GenBank accession
Actinodaphne	*Actinodaphne sesquipedalis*	AF268787
Aiouea	*Aiouea dubia*	EU153942
Aiouea	*Aiouea guianensis*	AF268780
Alseodaphne	*Alseodaphne semecarpifolia*	AF268799
Anaueria	*Anaueria brasiliensis*	AF268800
Aniba	*Aniba cinnamomiflora*	AF268770
Aniba	*Aniba hypoglauca*	AF268771
Beilschmiedia	*Beilschmiedia brenesii*	AF268809
Beilschmiedia	*Beilschmiedia madagascariensis*	AF268810
Beilschmiedia	*Beilschmiedia ovalis*	AF268811
Beilschmiedia	*Beilschmiedia pendula*	EU153943
Beilschmiedia	*Beilschmiedia pendula*	EU153944
Beilschmiedia	*Beilschmiedia pendula*	EU153945
Beilschmiedia	*Beilschmiedia sary*	AF268812
Beilschmiedia	*Beilschmiedia tawa*	EU153946
Beilschmiedia	*Beilschmiedia velutina*	AF268813
Chlorocardium	*Chlorocardium rodiei*	AF268802
Chlorocardium	*Chlorocardium venenosum*	AF268801
Cinnamomum	*Cinnamomum bejolghota*	EU153949
Cinnamomum	*Cinnamomum camphora*	AB331294
Cinnamomum	*Cinnamomum camphora*	EU153948
Cinnamomum	*Cinnamomum japonicum*	AF268782
Cinnamomum	*Cinnamomum quadrangulum*	AF268781
Cinnamomum	*Cinnamomum triplinerve*	EU153950
Cinnamomum	*Cinnamomum triplinerve*	EU153951
Cinnamomum	*Cinnamomum triplinerve*	EU153952
Cinnamomum	*Cinnamomum verum*	AF268784

Table S2: Contd...

Genus	Species	GenBank accession
Cryptocarya	*Cryptocarya rhodosperma*	AF268817
Cryptocarya	*Cryptocarya sclerophylla*	AF268818
Cryptocarya	*Cryptocarya thouvenotii*	AF261997
Dicypellium	*Dicypellium manausense*	AF268775
Endiandra	*Endiandra microneura*	AF268814
Endlicheria	*Endlicheria chalisea*	AF268756
Endlicheria	*Endlicheria citriodora*	AF268757
Endlicheria	*Endlicheria reflectens*	AF268758
Eusideroxylon	*Eusideroxylon zwageri*	AF268820
Kubitzkia	*Kubitzkia mezii*	AF268772
Laurus	*Laurus azorica*	EU153958
Laurus	*Laurus nobilis*	AF268785
Laurus	*Laurus nobilis*	EU153959
Laurus	*Laurus nobilis*	FJ493285
Licaria	*Licaria cannella*	AF268773
Licaria	*Licaria triandra*	AF268774
Lindera	*Lindera benzoin*	AF268788
Lindera	*Lindera benzoin*	EF491227
Lindera	*Lindera umbellata*	AF268789
Litsea	*Litsea coreana*	AF268791
Litsea	*Litsea cubeba*	EU153961
Litsea	*Litsea glaucescens*	AF129063
Litsea	*Litsea krukovii*	AB331293
Mezilaurus	*Mezilaurus triunca*	AF268804
Nectandra	*Nectandra purpurea*	EU153972
Nectandra	*Nectandra purpurea*	EU153973
Nectandra	*Nectandra purpurea*	EU153974
Neocinnamomum	*Neocinnamomum mekongense*	AF268806
Neolitsea	*Neolitsea aciculata*	EU153977
Neolitsea	*Neolitsea sericea*	AF268792
Ocotea	*Ocotea botrantha*	AF268776
Ocotea	*Ocotea bullata*	AF268778
Ocotea	*Ocotea calophylla*	EU153978
Ocotea	*Ocotea cernua*	EU153979

Table S2: Contd...

Genus	Species	GenBank accession
Ocotea	*Ocotea cernua*	EU153980
Ocotea	*Ocotea cernua*	EU153981
Ocotea	*Ocotea floribunda*	EU153982
Ocotea	*Ocotea guianensis*	AF268761
Ocotea	*Ocotea guianensis*	EU153983
Ocotea	*Ocotea leucoxylon*	AF268763
Ocotea	*Ocotea malcomberi*	AF268779
Ocotea	*Ocotea oblonga*	EU153984
Ocotea	*Ocotea odorifera*	AF268762
Ocotea	*Ocotea pauciflora*	AF268764
Ocotea	*Ocotea puberula*	EU153985
Ocotea	*Ocotea puberula*	EU153986
Ocotea	*Ocotea quixos*	AF261999
Ocotea	*Ocotea rhynchophylla*	AF268766
Ocotea	*Ocotea tomentella*	AF268765
Parasassafras	*Parasassafras confertiflora*	AF268790
Persea	*Persea americana*	AF268794
Persea	*Persea americana*	EU153989
Persea	*Persea caerulea*	AF268795
Persea	*Persea caerulea*	EU153990
Persea	*Persea lingue*	AF268796
Persea	*Persea meridensis*	AF268797
Persea	*Persea thunbergii*	AF268798
Pleurothyrium	*Pleurothyrium cinereum*	AF268769
Potameia	*Potameia micrantha*	AF268815
Potameia	*Potameia microphylla*	AF268816
Potoxylon	*Potoxylon melagangai*	AF268821
Rhodostemonodaphne	*Rhodostemonodaphne crenaticupula*	AF268759
Rhodostemonodaphne	*Rhodostemonodaphne kunthiana*	EU153991
Rhodostemonodaphne	*Rhodostemonodaphne penduliflora*	EU153992
Rhodostemonodaphne	*Rhodostemonodaphne praeclara*	AF268760
Sassafras	*Sassafras albidum*	AF268793
Sassafras	*Sassafras albidum*	EF491223
Sassafras	*Sassafras albidum*	EF491224
Sassafras	*Sassafras albidum*	EF491225
Sassafras	*Sassafras albidum*	EF491226
Sassafras	*Sassafras albidum*	EU153993
Sassafras	*Sassafras randaiense*	EF491221
Sassafras	*Sassafras randaiense*	EF491222
Sassafras	*Sassafras tzumu*	EF491217
Sassafras	*Sassafras tzumu*	EF491218
Sassafras	*Sassafras tzumu*	EF491219
Sassafras	*Sassafras tzumu*	EF491220
Sextonia	*Sextonia pubescens*	AF262000
Sextonia	*Sextonia rubra*	AF268805
Umbellularia	*Umbellularia californica*	AF268777

The *rbcL* sequence possesses advantages of versatility, easy amplification, and alignment. However, the variation in the *rbcL* region mainly exists for the above-species level, as the variation in the species level is insufficient to discriminate the different species.[12,13,17,18] The evolutionary rate of *matK* segment is faster than the coding regions of others, but Rohwer *et al*.,[19] reported that the *matK* sequence has low-evolutionary rates for Lauraceae (ie, the informative sites are only 9.7%). In this study, the two loci can be easily amplified and sequenced, but it was also found that they were too conservative for Lauraceae plants-their interspecific divergence were very low. Although *matK* and *rbcL* provided good PCR efficiency (both at 100%) and satisfactory sequencing efficiency (both at 100%), the successful identification rate of *matK* and *rbcL* were 30.9% and 25.0%, respectively, according to BLAST1. The success rate was only 38.2% at the species level when the two loci combination was used.

Many researchers have proposed the use of ITS2 as a suitable marker applicable for phylogenetic reconstruction and taxonomic classification.[4,20,21] In our study, the success rate of PCR amplification with ITS2 was poor in Lauraceae; thus, ITS2 was not included in subsequent experiments. We strictly observe the standard operating program of PCR, during the test, and similar experiment was repeated three times. The success rates for ITS2 sequences were 32.35%, 32.35%, and 30.88%, respectively. Then, we compared the success rate of PCR amplification of Lauraceae and Caprifoliaceae, used the same primers of ITS2 and PCR reaction conditions. Results showed that ITS2 sequences are relatively easy to amplify in Caprifoliaceae. In contrast to Caprifoliaceae, the success rate of PCR amplification of Lauraceae were much lower. Furthermore, in our experiments, ITS2 provided not satisfactory PCR efficiency (32.35%) and bad sequencing efficiency (27.27%), because homologous sequences existed. Our much work shows that in the direct PCR amplification and sequencing ITS2 produce a high success rate in some taxonomy group but the low success rate in another taxonomy group. It is found that ITS2 region produced a low success rate in direct PCR amplification and sequencing in Lauraceae species and it is also unsuitable to be DNA barcode of Lauraceae.

Discussion on samples with unsuccessful identification

In our study, the *psbA-trnH* sequence was chosen as a DNA barcode in identifying the species of Lauraceae family. Among the 175 samples tested, 28 samples could not be identified. At present, there is no stated consensus on the taxonomy of Lauraceae, and the relationships among the species of the family are still poorly understood.[22] The present study found that ambiguous identification mainly occurred in five genera (*Persea*, *Ocotea*, *Litsea*, *Machilus,* and *Cinnamomum*) which have always been as source of dispute in taxonomy. It was difficult to distinguish species in the same genus because they show little differences in morphology. The relationship among species of these genera is complex and the boundaries across groups are

vague, which could result in improper classification.[23-27] These species could not be identified by *matK*, *rbcL,* and the two-locus combination of *rbcL* + *matK*, could also not be identified by *psbA-trnH* in this study. A possible method for the species of these genera identification may be whole chloroplast genome sequencing.

The present research made a new exploration in the application of DNA barcode technology, as well as provided new approaches and evidences for the classification and phyletic evolution of Lauraceae plants. However, because of sampling constraints, lack of duplication of some species individuals, and the presence of those highly related species (ie, from sister species) not included in the analysis, some flaws in the research still exist. Hopefully, with the increasing number of materials and the progress of the study, DNA barcode technology can provide more effective information and more reliable method for the identification of Lauraceae plants.

REFERENCES

1. Drinnan AN, Crane PR, Friis EM, Pedersen KR. Lauraceous Flowers from the Potomac Group (Mid-Cretaceous) of Eastern North America. Bot Gaz 1990;151:370-84.
2. Hebert PD, Ratnasingham S, de Waard JR. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci 2003;270 Suppl 1:S96-9.
3. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci U S A 2009;106:12794-7.
4. Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, *et al*. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. Plos One 2010;5:e8613.
5. Gao T, Chen SL. Authentication of the medicinal plants in Fabaceae by DNA barcoding technique. Planta Med 2009;75:417.
6. Pang XH, Chen SL. Using DNA barcodes to identify Rosaceae. Planta Med 2009;75:417.
7. Song JY, Yao H, Li Y, Li XW, Liu C, Han JP, *et al*. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. J Ethnopharmacol 2009;124:434-9.
8. Yao H, Song JY, Ma XY, Liu C, Li Y, Xu HX, *et al*. Identification of Dendrobium species by a candidate DNA barcode sequence: The chloroplast psbA-trnH intergenic region. Planta Med 2009;75:667-9.
9. Zhu YJ, Chen SL, Yao H, Tan R, Song JY, Luo K, *et al*. DNA barcoding for the identification plants of the genus Paris. Yao Xue Xue Bao 2010;45:376-82.
10. Meyer CP, Paulay G. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol 2005;3:2229-38.
11. Meier R, Zhang GY, Ali F. The use of mean instead of smallest interspecific distances exaggerates the size of the “Barcoding Gap” and leads to misidentification. Syst Biol 2008;57:809-13.
12. Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One 2007;2:e508.
13. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, *et al*. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A 2008;105:2923-8.
14. Slabbinck B, Dawyndt P, Martens M, De Vos P, De Baets B. TaxonGap: A visualization tool for intra- and inter-species variation among individual biomarkers. Bioinformatics 2008;24:866-7.
15. Ross HA, Murugan S, Li WL. Testing the reliability of genetic methods of species identification via simulation. Syst Biol 2008;57:216-30.
16. Moritz C, Cicero C. DNA Barcoding: Promise and Pitfalls. PLoS Biol 2004;2:e354.
17. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, *et al*. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 2008;3:e2802.
18. Newmaster SG, Fazekas AJ, Steeves RA, Janovec J. Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour 2008;8:480-90.
19. Rohwer JG. Toward a phylogenetic classification of the Lauraceae: Evidence from matK sequences. Syst Bot 2000;25:60-71.
20. Schultz J, Maisel S, Gerlach D, Müller T, Wolf M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 2005;11:361-4.
21. Miao M, Warren A, Song WB, Wang S, Shang HM, Chen ZG. Analysis of the internal transcribed spacer 2 (ITS2) region of scuticociliates and related taxa (Ciliophora, Oligohymenophorea) to infer their evolution and phylogeny. Protist 2008;159:519-33.
22. Li J, Li XW. Advances in lauraceae systematic research on the world scale. Acta Bot Yunnan 2004;26:1-11.
23. Kojoma M, Kurihara K, Yamada K, Sekita S, Satake M, Iida O. Genetic identification of cinnamon (*Cinnamomum* spp.) based on the trnL-trnF chloroplast DNA. Planta Med 2002;68:94-6.
24. Van der Werff H. A synopsis of Persea (Lauraceae) in Central America. Novon St. Louis Mo. 2002;12:575-86.
25. Van der Werff H. A synopsis of Ocotea (Lauraceae) in Central America and Southern Mexico. Ann Mo Bot Gard 2002;89: 429-51.
26. Li J, Christophel DC, Conran JG, Li HW. Phylogenetic relationships within the ‘core’ Laureae (Litseacomplex, Lauraceae) inferred from sequences of the chloroplast gene matK and nuclear ribosomal DNA ITS regions. Plant Syst Evol 2004;246:19-34.
27. Wei FN, Tang SC. On the circumscription of Machilus and of Persea (Lauraceae). Acta Phytotaxon Sin 2006;44:437-42.

Cite this article as: Liu Z, Chen S, Song J, Zhang S, Chen K. Application of deoxyribonucleic acid barcoding in Lauraceae plants. Phcog Mag 2012;8:4-11.

Source of Support: Nil, **Conflict of Interest:** None declared.