

Application of deoxyribonucleic acid barcoding in Lauraceae plants

Zhen Liu^{1,2}, Shi-Lin Chen³, Jing-Yuan Song³, Shou-Jun Zhang⁴, Ke-Li Chen²

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

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.

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

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

Access this article online

Website:

www.phcog.com

DOI:

10.4103/0973-1296.93301

Quick Response Code:



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 MgCl₂, 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 | <i>psbA-trnH</i> | <i>matK</i> | <i>rbcL</i> | 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 | <i>psbA-trnH</i> | <i>matK</i> | <i>rbcL</i> |
|---------------------------------|------------------|-----------------|-----------------|
| 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 |
|------------------|-------------|---|---------------------------------|
| <i>psbA-trnH</i> | <i>matK</i> | W+ = 5565.0, W- = 0.0, n = 105, P < 5.7967E-19 | <i>psbA-trnH</i> >> <i>matK</i> |
| <i>psbA-trnH</i> | <i>rbcL</i> | W+ = 5550.0, W- = 15.0, n = 105, P < 8.9199E-19 | <i>psbA-trnH</i> >> <i>rbcL</i> |
| <i>rbcL</i> | <i>matK</i> | W+ = 1727.0, W- = 1048, n = 74, P < 0.0671 | <i>rbcL</i> = <i>matK</i> |

Table 4: Wilcoxon signed rank test for intraspecific variations

| W + | W - | Intrarerelative Ranks, n, P value | Result |
|------------------|-------------|---|--------------------------------|
| <i>psbA-trnH</i> | <i>matK</i> | W+ = 120.0, W- = 0.0, n = 15, P < 5.5225E-4 | <i>psbA-trnH</i> > <i>matK</i> |
| <i>psbA-trnH</i> | <i>rbcL</i> | W+ = 125.0, W- = 28.0, n = 17, P < 0.0205 | <i>psbA-trnH</i> > <i>rbcL</i> |
| <i>rbcL</i> | <i>matK</i> | W+ = 3.0, W- = 0.0, n = 2, P < 0.1797 | <i>rbcL</i> = <i>matK</i> |

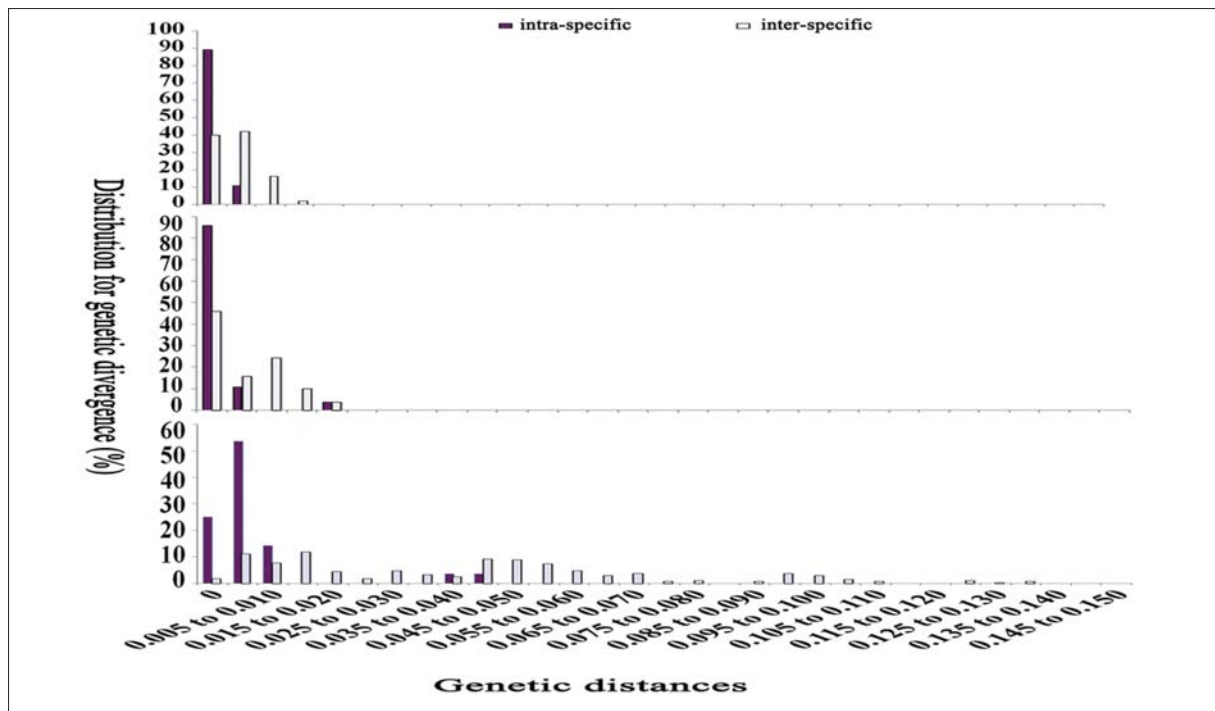


Figure 1: Schematic representation of the deoxyribonucleic acid barcoding gap between interspecific and intraspecific divergences for three candidate DNA barcodes. (a) *matK*; (b) *rbcL*; and (c) *psbA-trnH*

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.

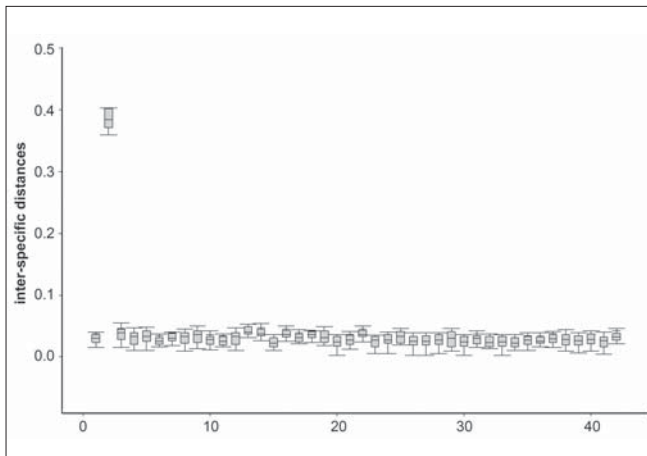


Figure 2: The interspecific divergence of the *psbA-trnH* region in Lauraceae

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 % |
| <i>matK</i> | BLAST1 | 68 | 30.9 | 28.6 | 0 | 0 | 69.1 | 71.4 |
| | Distance | 68 | 27.9 | 31.0 | 0 | 0 | 72.1 | 69.0 |
| <i>rbcL</i> | BLAST1 | 68 | 25.0 | 40.5 | 0 | 0 | 75.0 | 59.5 |
| | Distance | 68 | 25.0 | 42.9 | 0 | 0 | 75.0 | 57.1 |
| <i>psbA-trnH</i> | BLAST1 | 68 | 82.4 | 88.1 | 0 | 0 | 17.6 | 11.9 |
| | Distance | 68 | 64.7 | 81.0 | 0 | 0 | 35.3 | 19.0 |
| <i>rbcL</i> + <i>matK</i> | BLAST1 | 68 | 38.2 | 50.0 | 0 | 0 | 61.8 | 50.0 |
| | Distance | 68 | 36.8 | 50.0 | 0 | 0 | 63.2 | 50.0 |
| <i>matK</i> + <i>psbA-trnH</i> | BLAST1 | 68 | 82.4 | 88.1 | 0 | 0 | 17.6 | 11.9 |
| | Distance | 68 | 66.2 | 81.0 | 0 | 0 | 33.8 | 19.0 |
| <i>rbcL</i> + <i>psbA-trnH</i> | 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 | | |
|----------------------------------|------------------|----------------|-------------------|-------------|-------------|
| | | | <i>psbA-trnH</i> | <i>matK</i> | <i>rbcL</i> |
| <i>Actinodaphne omeiensis</i> | HuBei | PS5014MT01 | HM019377 | HM019307 | HM019447 |
| <i>Actinodaphne omeiensis</i> | HuBei | PS5014MT02 | HM019378 | HM019308 | HM019448 |
| <i>Actinodaphne omeiensis</i> | HuBei | PS5014MT03 | HM019379 | HM019309 | HM019449 |
| <i>Cassytha filiformis</i> | GuangDong | PS5015MT01 | HM019380 | HM019310 | HM019450 |
| <i>Cassytha filiformis</i> | GuangDong | PS5015MT02 | HM019381 | HM019311 | HM019451 |
| <i>Cinnamomum bodinieri</i> | HuBei | PS5016MT01 | HM019382 | HM019312 | HM019452 |
| <i>Cinnamomum bodinieri</i> | HuBei | PS5016MT02 | HM019383 | HM019313 | HM019453 |
| <i>Cinnamomum burmannii</i> | GuangDong | PS5017MT01 | HM019384 | HM019314 | HM019454 |
| <i>Cinnamomum burmannii</i> | GuangXi | PS5018MT01 | HM019385 | HM019315 | HM019455 |
| <i>Cinnamomum camphora</i> | HuBei | PS5019MT01 | HM019386 | HM019316 | HM019456 |
| <i>Cinnamomum camphora</i> | HuBei | PS5020MT01 | HM019387 | HM019317 | HM019457 |
| <i>Cinnamomum cassia</i> | GuangDong | PS5021MT01 | HM019388 | HM019318 | HM019458 |
| <i>Cinnamomum cassia</i> | GuangDong | PS5021MT02 | HM019389 | HM019319 | HM019459 |
| <i>Cinnamomum japonicum</i> | HuBei | PS5022MT01 | HM019390 | HM019320 | HM019460 |
| <i>Cinnamomum jensenianum</i> | HuBei | PS5023MT01 | HM019391 | HM019321 | HM019461 |
| <i>Cinnamomum jensenianum</i> | HuBei | PS5023MT02 | HM019392 | HM019322 | HM019462 |
| <i>Cinnamomum pauciflorum</i> | HuBei | PS5024MT01 | HM019393 | HM019323 | HM019463 |
| <i>Cinnamomum pauciflorum</i> | HuBei | PS5024MT02 | HM019394 | HM019324 | HM019464 |
| <i>Cinnamomum platyphyllum</i> | HuBei | PS5025MT01 | HM019395 | HM019325 | HM019465 |
| <i>Cinnamomum platyphyllum</i> | HuBei | PS5025MT02 | HM019396 | HM019326 | HM019466 |
| <i>Cinnamomum tonkinense</i> | HuBei | PS5027MT01 | HM019397 | HM019327 | HM019467 |
| <i>Cinnamomum wilsonii</i> | HuBei | PS5028MT01 | HM019398 | HM019328 | HM019468 |
| <i>Laurus nobilis</i> | HuBei | PS5029MT01 | HM019399 | HM019329 | HM019469 |
| <i>Laurus nobilis</i> | HuBei | PS5029MT02 | HM019400 | HM019330 | HM019470 |
| <i>Lindera aggregata</i> | JiangXi | PS5030MT01 | HM019401 | HM019331 | HM019471 |
| <i>Lindera aggregata</i> | JiangXi | PS5031MT01 | HM019402 | HM019332 | HM019472 |
| <i>Lindera aggregata</i> | HuBei | PS5031MT02 | HM019403 | HM019333 | HM019473 |
| <i>Lindera chunii</i> | HuBei | PS5032MT01 | HM019404 | HM019334 | HM019474 |
| <i>Lindera fragrans</i> | HuBei | PS5033MT01 | HM019405 | HM019335 | HM019475 |
| <i>Lindera fragrans</i> | HuBei | PS5033MT02 | HM019406 | HM019336 | HM019476 |
| <i>Lindera glauca</i> | GuangDong | PS5034MT01 | HM019407 | HM019337 | HM019477 |
| <i>Lindera glauca</i> | GuangDong | PS5034MT02 | HM019408 | HM019338 | HM019478 |
| <i>Litsea cubeba</i> | GuangDong | PS5036MT01 | HM019411 | HM019341 | HM019481 |
| <i>Litsea glutinosa</i> | GuangDong | PS5037MT01 | HM019412 | HM019342 | HM019482 |
| <i>Litsea honghoensis</i> | HuBei | PS5038MT01 | HM019413 | HM019343 | HM019483 |
| <i>Litsea ichangensis</i> | HuBei | PS5039MT01 | HM019414 | HM019344 | HM019484 |
| <i>Litsea ichangensis</i> | HuBei | PS5039MT02 | HM019415 | HM019345 | HM019485 |
| <i>Litsea monopetala</i> | HuBei | PS5040MT01 | HM019416 | HM019346 | HM019486 |
| <i>Litsea pungens</i> | HuBei | PS5041MT01 | HM019417 | HM019347 | HM019487 |
| <i>Machilus ichangensis</i> | HuBei | PS5042MT01 | HM019418 | HM019348 | HM019488 |
| <i>Machilus leptophylla</i> | HuBei | PS5043MT01 | HM019419 | HM019349 | HM019489 |
| <i>Machilus leptophylla</i> | GuangXi | PS5044MT01 | HM019420 | HM019350 | HM019490 |
| <i>Machilus lichuanensis</i> | HuBei | PS5045MT01 | HM019421 | HM019351 | HM019491 |
| <i>Machilus microcarpa</i> | HuBei | PS5046MT01 | HM019422 | HM019352 | HM019492 |
| <i>Machilus microcarpa</i> | HuBei | PS5046MT02 | HM019423 | HM019353 | HM019493 |
| <i>Machilus oreophila</i> | GuangXi | PS5047MT01 | HM019424 | HM019354 | HM019494 |
| <i>Machilus oreophila</i> | GuangXi | PS5047MT02 | HM019425 | HM019355 | HM019495 |
| <i>Machilus pauhoi</i> | HuBei | PS5048MT01 | HM019426 | HM019356 | HM019496 |
| <i>Machilus rufipes</i> | HuBei | PS5049MT01 | HM019427 | HM019357 | HM019497 |
| <i>Neolitsea aurata</i> | HuBei | PS5050MT01 | HM019428 | HM019358 | HM019498 |
| <i>Neolitsea confertifolia</i> | HuBei | PS5051MT01 | HM019429 | HM019359 | HM019499 |
| <i>Neolitsea confertifolia</i> | HuBei | PS5051MT02 | HM019430 | HM019360 | HM019500 |
| <i>Neolitsea hsiangkweiensis</i> | HuBei | PS5052MT01 | HM019431 | HM019361 | HM019501 |
| <i>Neolitsea hsiangkweiensis</i> | HuBei | PS5052MT02 | HM019432 | HM019362 | HM019502 |
| <i>Neolitsea levinei</i> | HuBei | PS5053MT01 | HM019433 | HM019363 | HM019503 |
| <i>Neolitsea levinei</i> | HuBei | PS5053MT02 | HM019434 | HM019364 | HM019504 |

Table S1: Contd...

| Species | Collection sites | Voucher number | GenBank accession | | |
|---|------------------|----------------|-------------------|-------------|-------------|
| | | | <i>psbA-trnH</i> | <i>matK</i> | <i>rbcL</i> |
| <i>Neolitsea sericea</i> | HuBei | PS5054MT01 | HM019435 | HM019365 | HM019505 |
| <i>Neolitsea wushanica</i> var. <i>pubens</i> | HuBei | PS5055MT01 | HM019436 | HM019366 | HM019506 |
| <i>Persea americana</i> | GuangDong | PS5056MT01 | HM019437 | HM019367 | HM019507 |
| <i>Phoebe bournei</i> | HuBei | PS5057MT01 | HM019438 | HM019368 | HM019508 |
| <i>Phoebe bournei</i> | HuBei | PS5057MT02 | HM019439 | HM019369 | HM019509 |
| <i>Phoebe chekiangensis</i> | HuBei | PS5058MT01 | HM019440 | HM019370 | HM019510 |
| <i>Phoebe chekiangensis</i> | HuBei | PS5058MT02 | HM019441 | HM019371 | HM019511 |
| <i>Phoebe sheareri</i> | HuBei | PS5060MT01 | HM019442 | HM019372 | HM019512 |
| <i>Phoebe sheareri</i> | HuBei | PS5060MT02 | HM019443 | HM019373 | HM019513 |
| <i>Phoebe zhennan</i> | GuangDong | PS5061MT01 | HM019444 | HM019374 | HM019514 |
| <i>Phoebe zhennan</i> | HuBei | PS5061MT02 | HM019445 | HM019375 | HM019515 |
| <i>Sassafras tzumu</i> | 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 |
|----------------------|---------------------------------------|-------------------|
| <i>Actinodaphne</i> | <i>Actinodaphne sesquipedalis</i> | AF268787 |
| <i>Aiouea</i> | <i>Aiouea dubia</i> | EU153942 |
| <i>Aiouea</i> | <i>Aiouea guianensis</i> | AF268780 |
| <i>Alseodaphne</i> | <i>Alseodaphne semecarpifolia</i> | AF268799 |
| <i>Anaueria</i> | <i>Anaueria brasiliensis</i> | AF268800 |
| <i>Aniba</i> | <i>Aniba cinnamomiflora</i> | AF268770 |
| <i>Aniba</i> | <i>Aniba hypoglaucua</i> | AF268771 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia brenesii</i> | AF268809 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia madagascariensis</i> | AF268810 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia ovalis</i> | AF268811 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia pendula</i> | EU153943 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia pendula</i> | EU153944 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia pendula</i> | EU153945 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia sary</i> | AF268812 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia tawa</i> | EU153946 |
| <i>Beilschmiedia</i> | <i>Beilschmiedia velutina</i> | AF268813 |
| <i>Chlorocardium</i> | <i>Chlorocardium rodiei</i> | AF268802 |
| <i>Chlorocardium</i> | <i>Chlorocardium venosum</i> | AF268801 |
| <i>Cinnamomum</i> | <i>Cinnamomum bejolghota</i> | EU153949 |
| <i>Cinnamomum</i> | <i>Cinnamomum camphora</i> | AB331294 |
| <i>Cinnamomum</i> | <i>Cinnamomum camphora</i> | EU153948 |
| <i>Cinnamomum</i> | <i>Cinnamomum japonicum</i> | AF268782 |
| <i>Cinnamomum</i> | <i>Cinnamomum quadrangulum</i> | AF268781 |
| <i>Cinnamomum</i> | <i>Cinnamomum triplinerve</i> | EU153950 |
| <i>Cinnamomum</i> | <i>Cinnamomum triplinerve</i> | EU153951 |
| <i>Cinnamomum</i> | <i>Cinnamomum triplinerve</i> | EU153952 |
| <i>Cinnamomum</i> | <i>Cinnamomum verum</i> | AF268784 |

Table S2: Contd...

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

Table S2: Contd...

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