JSE Journal of Systematics and Evolution

Journal of Systematics and Evolution 49 (3): 252–260 (2011) doi: 10.1111/j.1759-6831.2011.00116.x

**Research Article**

# Molecular identification of species in Juglandaceae: A tiered method

[1,2]Xiao-Guo XIANG [1,2]Jing-Bo ZHANG [1]An-Ming LU [1]Rui-Qi LI*

[1](*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences,* Beijing 100093, China)

[2](*Graduate University of Chinese Academy of Sciences,* Beijing 100049, China)

**Abstract** DNA barcoding is a method of species identification and recognition using DNA sequence data. A tiered or multilocus method has been recommended for barcoding plant species. In this study, we sampled 196 individuals representing 9 genera and 54 species of Juglandaceae to investigate the utility of the four potential barcoding loci (*rbcL*, *matK*, *trnH-psbA*, and internal transcribed spacer (ITS)). Our results show that all four DNA regions are easy to amplify and sequence. In the four tested DNA regions, ITS has the most variable information, and *rbcL* has the least. At generic level, seven of nine genera can be efficiently identified by *matK*. At species level, ITS has higher interspecific p-distance than the *trnH-psbA* region. Difficult to align in the whole family, ITS showed heterogeneous variability among different genera. Except for the monotypic genera (*Cyclocarya, Annamocarya, Platycarya*), ITS appeared to have limited power for species identification within the *Carya* and *Engelhardia* complex, and have no power for *Juglans* or *Pterocarya*. Overall, our results confirmed that a multilocus tiered method for plant barcoding was applicable and practicable. With higher priority, *matK* is proposed as the first-tier DNA region for genus discrimination, and the second locus at species level should have enough stable variable characters.

**Key words** DNA barcoding, ITS, Juglandaceae, *matK*, *rbcL*, *trnH-psbA*.

DNA barcoding is a system to aid species recognition and identification through the characteristics of a standard gene region across all organisms (Hebert et al., 2003). For animals, the mitochondrial gene *cox1* as a barcode shows great promise in identifying cryptic species, accelerating biodiversity inventories and identifying species from degraded material (Smith et al., 2005; Hajibabaei et al., 2006; Rubinoff, 2006). However, mitochondrial DNAs evolve too slowly to provide enough informative sites to discriminate species in plants. Identification of a locus or a combination of loci that can serve as an effective DNA barcode for plant species is challenging.

Most of the previous plant barcoding studies were carried out on a large scale, attempting to find universal and consistent markers for all angiosperms or land plants. For example, Kress et al. (2005) compared nine DNA markers and suggested that internal transcribed spacer (ITS) and *trnH-psbA* were potentially usable DNA regions for barcoding flowering plants. Newmaster et al. (2006) analyzed more than 10 000 *rbcL* sequences and suggested that *rbcL* can be used at generic level, and an implementation of a second locus would be more applicable for land plants. Kress & Erickson (2007) suggested that a combination of the non-coding *trnH-psbA* spacer region and the coding *rbcL* gene was a two-locus global barcode for land plants. Chase et al. (2007) outlined two combinations of three regions (*rpoC1*, *rpoB*, and *matK*; *rpoC1, matK*, and *trnH-psbA*) as candidate markers for land plant barcoding. Lahaye et al. (2008a, 2008b) suggested *matK* could be used as a universal and variable DNA barcode for flowering plants. Recently, CBOL Plant Working Group (2009) formally suggested the combination of *rbcL* and *matK* as the core barcode for land plants. Yao et al. (2010) and Chen et al. (2010) proposed ITS2 as a powerful DNA barcoding marker for plants and animals. However, some authors used one or several candidate markers to test their utility through dense sampling within a single family or genus. For instance, Sass et al. (2007) used seven cpDNA regions and ITS to discriminate cycads (Cycadaceae), showing that ITS had the most variability. Newmaster et al. (2008) suggested that *matK* and *trnH-psbA* provided resolution among all the *Compsonuera* species (Myristicaceae). Ren et al. (2010) combined *trnH-psbA* and ITS to discriminate 88.5% species of *Alnus* (Betulaceae).

Received: 9 November 2010 Accepted: 14 November 2010

* Author for correspondence. E-mail: liruiqi@ibcas.ac.cn; Tel.: 86-10-62836447; Fax: 86-10-62590843.

Recently, a tiered or multilocus method has been proposed for plant DNA barcoding (Newmaster et al., 2006). That is, one region is selected for discriminating genera or families, and the other(s) for identifying species. However, there are few empirical studies that have tested this method in a specific taxonomic group.

The family Juglandaceae consists of nine genera and ca. 60 species, and is distributed mainly in the Northern Hemisphere (Manchester, 1987, 1999). *Annamocarya*, *Cyclocarya*, *Engelhardia*, *Platycarya*, and *Pterocarya* are restricted to the Old World. *Alfaroa* and *Oreomunnea* are restricted to the New World. *Carya* and *Juglans* occur in both continents (Lu, 1982). There are seven genera and ca. 27 species distributed in China. Taxonomy of the genera was largely based on flower and fruit characters (Manning, 1978; Manchester, 1987). Although the genera are easy to identify by morphological characters, species identification in certain genera are difficult using morphological characters alone. For instance, species of *Carya* distributed in eastern North America are almost indistinguishable because of their highly similar morphological characters. Juglandaceae is confirmed to be monophyletic based on molecular analyses (Manos & Steele, 1997; Li et al., 2004). Recent phylogenetic studies based on morphology and DNA sequences (ITS, *trnL-F*, *atpB-rbcL*) have partially resolved relationships among genera and some groups within genera of Juglandaceae (Manos & Stone, 2001; Manos et al., 2007).

In this study, we use four DNA regions (*rbcL*, *matK*, *trnH-psbA*, and ITS) that were frequently recommended as potential DNA barcodes in previous researches to differentiate species of Juglandaceae. Our objectives are: (i) to test the universality of the four DNA regions in Juglandaceae; (ii) to evaluate the potential of barcodes to broadly identify genera and species across the family; and (iii) to provide some suggestions on future DNA barcoding regions for plants.

## 1 Material and methods

### 1.1 Taxon sampling

Multiple samples of each species recognized in *Flora of China* (Lu & Stone, 1999) were included in this study, covering both morphological and geographical range of each taxon. Most of the species distributed in the Americas and Europe were collected. In total, we sampled 196 individuals representing nine genera and 54 species. There are 88 sequences of *rbcL, matK, trnH-psbA*, and ITS downloaded from GenBank. The detailed information was included in Table S1.

### 1.2 DNA extraction, amplification, and sequencing

Total DNAs were isolated from silica gel-dried leaves following the protocol of Doyle & Doyle (1987). The amplification of DNA regions was carried out using standard polymerase chain reaction (PCR). Primer sequences for amplification and sequencing are presented in Table S2. The PCR products were sequenced on an ABI 3730 DNA sequencer (Applied Biosystems, Foster City, CA, USA). The primers for ITS are from Baldwin (1992); primers for *trnH-psbA* are from http://www.kew.org/barcoding/protocols.html; primers for *matK* are from Cuenoud et al. (2002); and primers for *rbcL* are from Kress & Erickson (2007).

### 1.3 Data analysis

The sequences were first aligned using ClustalX (Thompson et al., 1997) software then manually adjusted in Bioedit version 7 (Hall, 1999).

In order to obtain an estimate of variation in the four regions examined, pairwise Kimura 2-parameter (K2P) distances were calculated in Mega 4.0 (Kumar et al., 2008). Indels were coded with the simple indel coding method of Simmons & Ochoterena (2000). Neighbor-joining trees of ITS, *matK, rbcL*, and *trnH-psbA* were generated using MEGA version 4 with K2P (Kumar et al., 2008) with 1000 replicates to determine clade support and tree resolution.

To ensure accurate species assignments in our datasets, we used the "Best Match", "Best Close Match", and "All Species Barcodes" methods of the program TaxonDNA (Meier et al., 2006). This program determines the closest match of a sequence from comparisons to all other sequences in an aligned dataset. Comparing with the "Best Match", "Best Close Match" strategy requires a threshold similarity value that defines how similar a sequence match needs to be before it can be identified. This value can be estimated for a given dataset by obtaining a frequency distribution of all intraspecific pairwise distances and determining the threshold distance below which 95% of all intraspecific distances are found. These sequence identification methods were carried out on the *rbcL*, *matK*, *trnH-psbA*, and ITS datasets.

## 2 Results

### 2.1 Characteristics of the four DNA regions

Of the 196 individuals included in this study, PCR amplification was successful for all loci, except ITS for *Engelhardia hainanensis* Chen. Success rates for bidirectional sequencing were highest for *rbcL* (100%) and *trnH-psbA* (100%), followed by ITS (94%) and

**Table 1** Evaluation of four DNA markers

| | *rbcL* | *matK* | *trnH-psbA* | ITS | ITS2 |
|---|---|---|---|---|---|
| Universal ability to primer | Yes | Yes | Yes | Yes | — |
| Percentage PCR success | 100 | 90 | 100 | 94 | — |
| Percentage sequencing success | 100 | 95 | 100 | 100 | — |
| Aligned sequence length (bp) | 742 | 792 | 299 | 749 | 240 |
| Indels length (bp) | — | 6 | 1–29 | 1–157 | 1–25 |
| No. information sites/variable sites | 20/88 | 51/123 | 26/39 | 198/273 | 29/115 |
| Distribution of variable sites | Dispersive | Dispersive | Dispersive | Dispersive | — |
| No. samples (individuals) | 62 | 68 | 196 | 106 | 106 |
| Interspecific distance mean (range) | 0.0080 (0–0.0450) | 0.0106 (0–0.0280) | 0.0228 (1–0.0800) | 0.0608 (0–0.1300) | 0.0889 (0–0.2130) |
| Intraspecific distance mean (range) | 0.0062 (0–0.0750) | 0.0069 (0–0.0260) | 0.0038 (0–0.0100) | 0.0035 (0–0.0150) | 0.0058 (0–0.0320) |
| Ability to discriminate (NJ) (genera) | — | 6/9 | — | — | — |
| Ability to discriminate (sequence analysis) (genera) | 4/9 | 7/9 | — | — | — |
| Ability to discriminate (NJ) (species) | — | 6/25 | — | 6/48 | 5/48 |
| Ability to discriminate (sequence analysis) (species) | — | — | 4/35 | 9/48 | 7/48 |

—, Missing data. ITS, internal transcribed spacer; NJ, neighbor-joining; PCR, polymerase chain reaction.

*matK* (90%). The universality of primer and success of sequencing of the four regions are summarized in Table 1.

We obtained 47 *rbcL* sequences for 10 species, 43 *matK* sequences for 10 species, 189 *trnH-psbA* sequences for 49 species, and 52 ITS sequences for 14 species. A total number of 331 new sequences were generated in this study. The aligned *rbcL* sequences were 742 bp long without indels, and the 20 informative sites and 88 variable sites were dispersed across the entire alignment. For the *matK* matrix, aligned sequence length was 792 bp, and 51 informative sites and 123 variable sites were dispersed across the matrix, with two 6-bp indels. The aligned *trnH-psbA* sequences were 299 bp in length, and the 39 variable sites were dispersed across the matrix, with seven indels 1–29 bp long. For the ITS matrix, aligned sequences were 729 bp long, and the distribution of 198 informative sites and 273 variable sites was dispersive, with 21 indels that were 1–157 bp long (Fig. 1, Table 1). The aligned sequence of ITS2 was 240 bp, and there were 29 informative sites and 115 variable sites across the matrix, with 11 indels that were 1–25 bp long (Table 1).

All DNA regions tested here showed marginal amounts of variation across all samples (Fig. S1). In these regions, ITS2 had the highest interspecific p-distance, followed by ITS, *trnH-psbA*, *matK,* and *rbcL* (Fig. S1). Some different species presented identical sequences, and the mean interspecific p-distance was higher than their intraspecific p-distance (Table 1).

According to the TaxonDNA results, the species identification rates were generally low. Even the highest species identification rates (*rbcL*, ITS) are just over 30%. Unexpectedly, the barcode with the lowest identification rates was *trnH-psbA,* which was a recommended barcode for species delimitation in previous studies (Table 2).

## 2.2 Identification of genera

**2.2.1** ***rbcL*** The distribution of interspecific and intraspecific distance in the *rbcL* dataset is shown in Fig. 2A. The mean value of the interspecific distance was slightly higher than the intraspecific distance (Table 1), with only seven stable variable sites among the matrix. *Engelhardia* has a unique character state (position 656: G) that differentiates it from other genera of Juglandaceae, and *Platycarya* has three. Both *Alfaroa* and *Oreomunnea* have a three-combination diagnostic site (Table S3). However, no monophyletic groups at generic level were detected with strong support.

**2.2.2** ***matK*** The distribution of interspecific and intraspecific distance in the *matK* dataset is shown in Fig. 2B. The mean value of genetic distance between species was much higher than the intraspecific distance (Table 1) with 22 stable variable sites in the matrix. Five genera tested here have unique variable sites. For instance, *Carya* has a unique diagnostic site (position 514: G), *Pterocarya* (position 390: A), *Annamocarya* (position 601–606: CTAAAT), and the genus *Engelhardia* has three unique diagnostic sites (position 376:

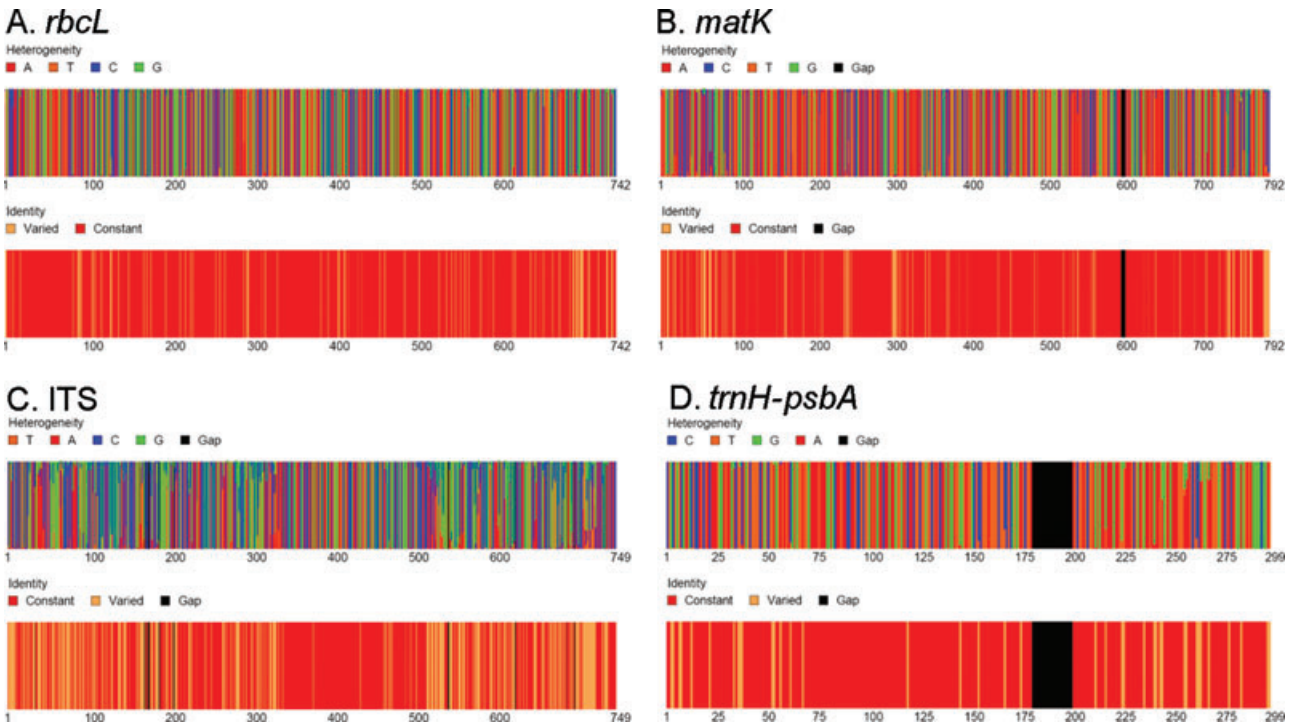


**Fig. 1.** Nucleotide compositions and variability of the four DNA markers. **A,** *rbcL*; **B,** *matK*; **C,** Internal transcribed spacer (ITS); **D,** *trnH-psbA*.

T, position 738: C, position 769: A). *Platycarya* has five unique character states that differentiate it from other genera of Juglandaceae (position 131–136: CTAATT, position 351: T, position 535: C, position 631: A, position 678: A). *Juglans* and *Cyclocarya* can be distinguished from other genera by combined sites (e.g., *Juglans* by the combination of position 126 (G) and position 390 (G) (Table 3)). Both *Alfaroa* and *Oreomunnea* share the same characters. Moreover, based on the tree-building analysis, all genera except *Juglans* formed monophyletic groups with strong or slightly weak supporting values (Fig. S2).

## 2.3 Identification of species

**2.3.1** ***trnH-psbA*** The *trnH-psbA* region was easy to align within the family Juglandaceae, and several indels were found dispersed across the alignment. The distribution of interspecific and intraspecific distance is shown in Fig. 2C. The mean value of genetic p-distance between species was much higher than the intraspecific distance (Table 1). However, no regular variable sites across the family were detected and only four species have unique characters (Table S4). *Cyclocarya paliurus* has a unique site at position 259 (C), *Juglans australis* has a unique site at position 244 (T), *Engelhardia colebrookiana* has a combination state (position 90–117: CCAAT—TGA and a deletion located at position 120–142).

**2.3.2 ITS** The distribution of interspecific and intraspecific distance for the ITS dataset is shown in Fig. 2D. The mean value of genetic distance between species was 40 times more than that of intraspecific distance (Table 1). It was difficult to align ITS among genera. We carried out the alignment both within a single genus and between closely related genera. Within the closely related *Juglans* and *Pterocarya,* there is no unique character state that differentiates them from other species within these two genera. Within the genus *Carya,* five of eight species have diagnostic characters. *Carya cathayensis* has four unique sites (position 240: C or position 463–472: CGT—— or position 533: C or position 567: G), and the other four species with unique

**Table 2** Species identification success rate based on TaxonDNA method

| | *rbcL* | *matK* | *trnH-psbA* | ITS | ITS2 |
|---|---|---|---|---|---|
| Best Match (%) | 19 (30.64) | 13 (27.08) | 39 (20.2) | 25 (30.48) | 25 (23.58) |
| Best Close Match (%) | 19 (30.64) | 13 (27.08) | 39 (20.2) | 25 (30.48) | 25 (23.58) |
| All Species Barcodes (%) | 9 (14.51) | 2 (4.16) | 38 (19.68) | 9 (10.97) | 30 (28.3) |

ITS, internal transcribed spacer.

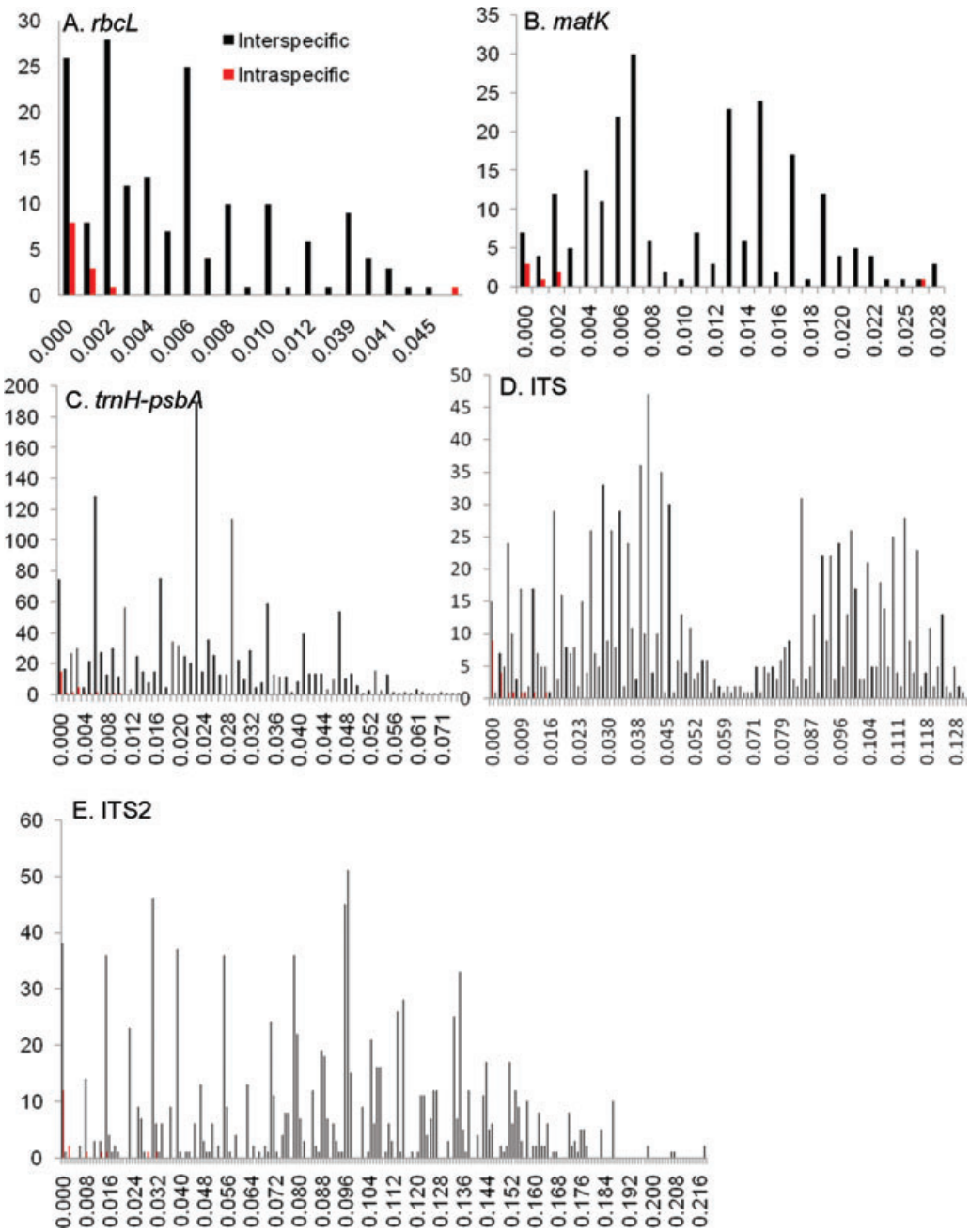


**Fig. 2.** Relative distribution of interspecific and intraspecific distances of the five DNA markers. **A,** *rbcL*; **B,** *matK*; **C,** *trnH-psbA*; **D,** Internal transcribed spacer (ITS); **E,** ITS2.

character states included *C. glabra* (position 214: T), *C. myristiciformis* (position 194: G), *C. illinoinensis* (position 94: A), and *C. tonkinensis* (position 463–472: CGTGTGCGCG). Additionally, three species share the same characters: *C. cordiformis, C. ovata*, and *C. tomentosa* (Table S5). Within *Engelhardia*, five of seven species can be identified, such as *E. acerofolia* (position 513: G), *E. fenzelii* (position 465: T), *E. unijuga* (position 28: C, position 464: G, position 531: G), and *E. roxburghiana* (position 197: A, position 250: G) (Table S6). The monophyletic genera such as *Cyclocarya, Annamocarya,* and *Platycarya,* have their own unique states. Based on the tree-building analysis, only six species across the tested species are monophyletic with strong or weak supporting values (neighbor-joining–bootstrap: 63–98) (Fig. S3).

**Table 3** Character-based DNA database for Juglandaceae species from *matK* region

| Taxon (n) | Position | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 126 | 131–136 | 152 | 172 | 187 | 208 | 351 | 376 | 390 | 492 | 514 | 535 | 544 | 601–606 | 631 | 670 | 678 | 712 | 738 | 741 | 761 | 769 |
| *Alfaroa* | T | —— | G* | A | A | C* | C | C | G | T* | T | T | C* | —— | T | C | G | A* | A | A | T* | G |
| *Annamocarya* | T | —— | T | T | C | T | C | C | G | A | T | T | A | CTAAAT | C | C | G | G | A | G | C | G |
| *Carya* | T | —— | T | T | C | T | C | C | G | A | G | T | A | —— | C | C | G | G | A | G | C | G |
| *Cyclocarya* | T | —— | T | T | A | T | C | C | G | A | T | T | A | —— | C | C | G | G | A | G | C | G |
| *Engelhardia* | T | —— | T | A | A | C* | C | T | G | A | T | T | C* | —— | T | C | G | A* | C | G | T* | A |
| *Juglans* | G | —— | T | T | A | T | C | C | G | A | T | T | A | —— | C | C | G | G | A | G | C | G |
| *Oreomunnea* | T | —— | G* | A | A | C* | C | C | G | T* | T | T | C* | —— | T | C | G | A* | A | A | T* | G |
| *Platycarya* | T | CTAATT | T | T | A | T | T | C | G | A | T | C | A | —— | C | A | A | G | A | G | C | G |
| *Pterocarya* | G | —— | T | T | A | T | C | C | A | A | T | T | A | —— | C | C | G | G | A | G | C | G |

——, Deletions. *Genera share the same characters. Shaded areas indicate that the genus can be identified by these character(s).

The distribution of interspecific and intraspecific distance of ITS2 is shown in Fig. 2E. The mean value of genetic distance between species was more than that of intraspecific distance (Table 1). ITS2 is 162–216 bp long, and it is alignable in the whole family. We did the alignment both within each genus and between closely related genera, and found the variability of ITS2 displayed a heterogeneity among genera, e.g. high variation was examined in the genus *Carya*, and very low in *Engelhardia, Pterocarya*, and *Juglans.*

## 3 Discussion

### 3.1 Evaluation of methods

As published results were based on different criteria for measuring species resolution of DNA barcoding, we first re-evaluated their success in discriminating described animal and plant species using two universally acceptable criteria: (i) the samples of a species form a monophyletic group (e.g. Hajibabaei et al*.,* 2006; Fazekas et al*.,* 2008; Lahaye et al., 2008a, 2008b); and (ii) differences between intraspecific and interspecific genetic distances (the genetic distance gap; see Hebert et al., 2004; Barrett & Hebert, 2005). Our results suggested that all four DNA markers plus ITS2 have higher interspecific divergence than the intraspecific divergence. However, only a few species formed monophyletic groups on tree construction.

TaxonDNA is an effective channel to analyze identification rates for DNA barcodes (Meier et al., 2006). Customized for testing barcoding efficiencies, this program offers more perspicuous results than some other methods. Although with a high correct identification success rate when using “Best Match”and “Best Close Match”, it has a low correct identification rate by “All species barcoding”. For example, *matK* shows species identification rates as 27.08% of “Best Match” and 4.16% of “All species barcoding” (Table 2). Because the matrices of TaxonDNA should include all species within a genus, and each individual species in the dataset should be collected from the full geographical ranges, it is difficult to judge the correct identification rate of DNA markers due to incomplete sampling. Moreover, these methods, by comparing the similarity between query and other sequences, cannot identify species that have only one sample in a dataset.

### 3.2 Identification of genera

The gene *rbcL* was easy to amplify and sequence, as in other land plants (Kress et al., 2005; Newmaster et al., 2008; Ren et al., 2010). It was also evaluated as a potential standard core barcoding region, given its

universality and ease of amplification and alignment (Newmaster et al., 2006; Chase et al., 2007). This gene has been shown to differentiate a large percentage of congeneric taxa in land plants and was suggested as a first-tier marker (Newmaster et al., 2006). However, its discriminating ability in Juglandaceae is limited, as it failed to identify all genera of the family.

As well as simple sequence variability and crucial characteristics, barcoding loci should include primer universality and easy amplification and sequencing (Chase et al., 2007; Fazekas et al., 2008). Lahaye et al. (2008a) analyzed 1084 plant species (nearly 96% of orchid species) and showed that a portion of the plastid *matK* gene could be a universal DNA barcode for flowering plants. Dunning & Savolainen (2010) designed 26 primer sets for *matK* in order to ensure that it can be used by plant DNA barcoding. Our results show that the primers of *matK* (390F and 1326R), suggested by Lahaye et al. (2008a), are universally applicable within Juglandaceae and sequences are easy to align across the entire family. In addition, *matK* has enough stable variable sites to discriminate seven of the nine genera tested in this study. It is proposed as the first-tier DNA barcoding marker for Juglandaceae.

### 3.3 Identification of species

Newmaster et al. (2006) suggested a tiered method to barcode plants, that is, a first-tier coding region shared by all land plants provides resolution at a certain rank (e.g., family or genus) and a more variable (coding or non-coding) region provides resolution at species level. The ITS region is the most variable among the DNA regions in our study, but it is difficult to align. The higher variability of the ITS region has also been observed in other genera (Ren et al., 2010). Unfortunately, there are limited stable diagnostic characters to identify the close species among Juglandaceae. ITS2 was proposed as a DNA barcode for plants by Yao et al. (2010) and Chen et al. (2010) for its success in the identification of plants and animals. Our results indicated that ITS2 has the highest interspecific variability among Juglandaceae species. It is short after alignment (240 bp) with modest intraspecific distance. Based on tree building and variable site analysis, ITS2 shows low species discrimination ability.

For *trnH-psbA*, it was easy to amplify and sequence across species in Juglandaceae, as in previous studies (Kress et al., 2005; Newmaster et al., 2008; Ren et al., 2010). Although this region was easy to align, it was only 260 bp long with lower variability among Juglandaceae species, compared to ITS. So far, *trnH-psbA* was evaluated as a potential region because of its universality, ease of amplification, and alignment (Kress et al., 2005; Newmaster et al., 2008; Lahaye et al., 2008a; Ren et al., 2010). Nevertheless, *trnH-psbA* is not suitable for Juglandaceae, for this region is short and lacks sufficient variable sites to discriminate species within a genus.

### 3.4 Suggestions for plant DNA barcoding

A successful barcoding project requires a comprehensive species sampling and high rates of distinguishing species. Several DNA barcoding markers have been tested in woody and herbaceous plants with different levels of taxon sampling (Edwards et al., 2008; Lahaye et al., 2008a; Logacheva et al., 2008; Newmaster et al., 2008; Nitta, 2008; Newmaster & Ragupathy, 2009; Starr et al., 2009; Ren et al., 2010; Kelly et al., 2010). Unlike *coxI*, widely used in animal DNA barcoding, there is no such single region that has sufficient variations to identify all plant species. A tiered method for barcoding plants, suggested by Newmaster et al. (2006), provided a new insight for the development of plant DNA barcoding, and suggested that multilocus DNA barcoding is practical and applicable (Kress & Erickson, 2007). We support the tiered barcoding method, and suggest that *matK* is the priority tier DNA region to resolve the generic level, but the second-tier locus has not enough variation to resolve the species level problems and thus needs more attention.

**Acknowledgements** The authors thank: Bao-Qing REN, Xiao-Yu DONG, and Hong-Lei LI for their help with field work; Jian-Wen ZHANG and Susan L. KELLEY for providing plant tissue for this study; Hong-Lei LI, Miao SUN, and Li-Min LU for their laboratory assistance; Bao-Qing REN, Miao SUN, Ruo-Zhu LIN, Xiao-Yu DONG, Jun-Xia SU, Li-Min LU, and Jian ZHANG for revising the manuscript. This research was supported by the National Natural Science Foundation of China (Grant Nos. 30740088, 30570106, and NSFC 40830209), and the large-scale Scientific Facilities of the Chinese Academy of Sciences (Grant No. 2009-LSF-GBOWS-01).

## References

Baldwin BG. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Molecular Phylogenetics and Evolution 1: 3–16.

Barrett RDH, Hebert PDN. 2005. Identifying spiders through DNA barcodes. Canadian Journal of Zoology 83: 481–491.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences USA 31: 12794–12797.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, le Seberg O, Jørgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ,

Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56: 295–299.

Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, Luo K, Li Y, Li XW, Jia XC, Lin YL, Leon C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5: e8613.

Cuenoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. 2002. Molecular phylogenetics of the Caryophylles based on 18SrDNA, *rbcL*, *atpB* and *matK* sequences. American Journal of Botany 89: 132–144.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Dunning LT, Savolainen V. 2010. Broad-scale amplification of *matK* for DNA barcoding plants, a technical note. Botanical Journal of the Linnean Society 164: 1–9.

Edwards D, Horn A, Taylor D, Savolainen V, Hawkins JA. 2008. DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). Taxon 57: 1317–1314.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3: e2802.

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6: 959–964.

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

Hebert PDN, Cywinska A, Ball SL, Dewaard JR. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Science 270: 313–322.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Sciences USA 101: 14812–14817.

Kelly LJ, Ameka GK, Chase MW. 2010. DNA barcoding of African Podostemaceae (river-weeds): A test of proposed barcode regions. Taxon 59: 251–260.

Kress W, Erickson D. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS One 2: e508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences USA 102: 8369–8374.

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics 9: 299–306.

Lahaye R, Savolainen V, Duthoit S, Maurin O, Van Der Bank M. 2008a. A test of *psbK-psbI* and *atpF-atpH* as potential plant DNA barcodes using the flora of the Kruger National Park (South Africa) as a model system. Nature Precedings [online]. Available from: http://precedings.nature.com/documents/1896/version/1/files/npre20081896-1.pdf [accessed 10 November 2008].

Lahaye RRY, Van Der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008b. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences USA 105: 2923–2928.

Li RQ, Chen ZD, Lu AM, Soltis DE, Soltis PS, Manos PS. 2004. Phylogenetic relationships in Fagales based on DNA sequences from three genomes. International Journal of Plant Sciences 165: 311–324.

Logacheva MD, Valiejo-Roman CM, Pimenov MG. 2008. ITS phylogeny of West Asian *Heracleum* species and related taxa of Umbelliferae-Tordylieae WDJ Koch, with notes on evolution of their *trnH-psbA* sequences. Plant Systematics and Evolution 270: 139–157.

Lu AM. 1982. On the geographical distribution of the Juglandaceae. Acta Phytotaxonomica Sinica 20: 257–274 (in Chinese with English abstract).

Lu AM, Stone DE. 1999. Juglandaceae. In: Wu ZY, Raven PH eds. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 4: 277–285.

Manchester SR. 1987. The fossil history of the Juglandaceae. Monographs in Systematic Botany from the Missouri Botanical Garden 21: 1–137.

Manchester SR. 1999. Biogeographical relationships of North American Tertiary floras. Annals of the Missouri Botanical Garden 86: 472–522.

Manning WE. 1978. The classification within the Juglandaceae. Annals of the Missouri Botanical Garden 65: 1058–1087.

Manos PS, Steele KP. 1997. Phylogenetic analyses of "higher" Hamamelididae based on plastid sequence data. American Journal of Botany 84: 1407–1419.

Manos PS, Stone DE. 2001. Evolution, phylogeny, and systematics of the Juglandaceae. Annals of the Missouri Botanical Garden 88: 231–269.

Manos PS, Soltis PS, Soltis DE, Manchester SR, Oh SH, Bell CD, Dilcher DL, Stone DE. 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. Systematic Biology 56: 412–430.

Meier R, Shiyang K, Vaidya G, Ng PKC. 2006. DNA barcoding and taxonomy in *Diptera*: A tale of high intraspecific variability and low identification success. Systematic Biology 55: 715–728.

Newmaster SG, Fazekas AJ, Ragupathy S. 2006. DNA barcoding in land plants: Evaluation of *rbcL* in a multigene tiered approach. Botany 84: 335–341.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J. 2008. Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources 8: 480–490.

Newmaster SG, Ragupathy S. 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). Molecular Ecology Resources 9: 172–180.

Nitta JH. 2008. Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea. Taxon 57: 725–2E(–722).

Ren BQ, Xiang XG, Chen ZD. 2010. Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. Molecular Ecology Resources 10: 594–605.

Rubinoff D. 2006. Utility of mitochondrial DNA barcodes in species conservation. Conservation Biology 20: 1026–1033.

Sass C, Little DP, Stevenson DW, Specht CD. 2007. DNA barcoding in the cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. PLoS One 2: e1154.

Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. Systematic Biology 49: 369–381.

Smith MA, Fisher BL, Hebert PDN. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of Madagascar. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1825–1834.

Starr JR, Naczi RFC, Chouinard BN. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). Molecular Ecology Resources 9: 151–163.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876–4882.

Yao H, Song JY, Liu C, Luo K, Han JP, Li Y, Pang XH, Xu HX, Zhu YJ, Xiao PG, Chen SL. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5: e13102.

## Supplementary Material

The following supplementary material is available in the online version of this article:

**Fig. S1.** Relative distribution of interspecific p-distance from the four DNA regions.

**Fig. S2.** Neighbor-joining tree based on *matK* matrix for Juglandaceae. The stars representing genera can be identified by both tree construction and sequences analysis; cycles indicate genera can be identified only by sequence characters; the squares represent the ones that have the same sequences.

**Fig. S3.** Neighbor-joining tree based on ITS matrix for Juglandaceae individuals. The lines representing these species can be identified by tree construction.

**Table S1.** Sample and sequence information in this study.

**Table S2.** Polymerase chain reaction and sequencing primers in this study.

**Table S3.** Character-based DNA database for Juglandaceae species from the *rbcL* region.

**Table S4.** Character-based DNA database for *Carya* species from the ITS region.

**Table S5.** Character-based DNA database for *Engelhardia* species from the ITS region.

**Table S6.** Character-based DNA database for Juglandaceae species from the *trnH-psbA* region.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.