Research Article

Testing four proposed barcoding markers for the identification of species within *Ligustrum* L. (Oleaceae)

[1,2]Jing GU   [2]Jun-Xia SU   [2]Ruo-Zhu LIN   [2]Rui-Qi LI   [1]Pei-Gen XIAO*

[1](*Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College,* Beijing 100193, China)
[2](*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences,* Beijing 100093, China)

Abstract   DNA barcoding is a biological technique that uses short and standardized genes or DNA regions to facilitate species identification. DNA barcoding has been used successfully in several animal and plant groups. *Ligustrum* (Oleaceae) species occur widely throughout the world and are used as medicinal plants in China. Therefore, the accurate identification of species in this genus is necessary. Four potential DNA barcodes, namely the nuclear ribosomal internal transcribed spacer (ITS) and three chloroplast (cp) DNA regions (*rbcL*, *matK*, and *trnH–psbA*), were used to differentiate species within *Ligustrum*. BLAST, character-based method, tree-based methods and TAXONDNA analysis were used to investigate the molecular identification capabilities of the chosen markers for discriminating 92 samples representing 20 species of this genus. The results showed that the ITS sequences have the most variable information, followed by *trnH–psbA*, *matK*, and *rbcL*. All sequences of the four regions correctly identified the species at the genus level using BLAST alignment. At the species level, the discriminating power of *rbcL*, *matK*, *trnH–psbA*, and ITS based on neighbor-joining (NJ) trees was 36.8%, 38.9%, 77.8%, and 80%, respectively. Using character-based and maximum parsimony (MP) tree methods together, the discriminating ability of *trnH–psbA* increased to 88.9%. All species could be differentiated using ITS when combining the NJ tree method with character-based or MP tree methods. Overall, the results indicate that DNA barcoding is an effective molecular identification method for *Ligustrum* species. We propose the nuclear ribosomal ITS as a plant barcode for plant identification and *trnH–psbA* as a candidate barcode sequence.
Key words   DNA barcoding, internal transcribed spacer (ITS), *Ligustrum*, *matK*, *trnH–psbA*, *rbcL*.

DNA barcoding is a molecular technique that uses short and standardized DNA sequences known as "DNA barcodes" to provide rapid, accurate, and automatable species identification (Hebert et al., 2003a, b; Hebert & Gregory, 2005). DNA barcoding was developed as a new approach to remedy the limitations of morphology-based identification systems (Hebert et al., 2003a; Hebert & Gregory, 2005). In animals, the mitochondrial gene cytochrome *c* oxidase subunit 1 (*CO1*) has been established as an effective DNA barcode in several animal groups (Hebert et al., 2003a, 2004; Barrett & Hebert, 2005; Hogg & Hebert, 2004) and can be searched via the Canadian Barcode of Life (http://www.bolnet.ca, accessed 01 December 2006) and the Consortium for the Barcode of Life (CBOL) (http://www.barcoding.si.edu, accessed 01 April 2007; Newmaster et al., 2007). In plants, although many studies have demonstrated that DNA barcoding is effective for species identification

(Newmaster & Ragupathy, 2009b; Moniz & Kaczmarska, 2010), a barcode for plant barcoding similar to *CO1* being used in animals remains elusive (Kress et al., 2005).

Previous plant barcoding studies may be classified into four categories: (i) those looking for universal and consistent makers for large-scale land plants (e.g. Chase et al., 2005, 2007; Kress et al., 2005; Cowan et al., 2006; Newmaster et al., 2006; Presting, 2006; Kress & Erickson, 2007; Sass et al., 2007; Erickson et al., 2008; Fazekas et al., 2008; Lahaye et al., 2008a, 2008b; Devey et al., 2009; Ford et al., 2009; Chen et al., 2010); (ii) those testing the identification power of different markers in a single family or genus (e.g. Hymenophyllaceae: Nitta, 2008; *Compsoneura* Warb.: Newmaster et al., 2007; *Heracleum* L.: Logacheva et al., 2008; *Aspalathus* L.: Edwards et al., 2008; *Acacia* Mill.: Newmaster & Ragupathy, 2009b; *Carex* L.: Starr et al., 2009; *Crocus* L.: Seberg & Petersen, 2009; *Alnus* Mill.: Ren et al., 2010); (iii) those assessing the analysis methods and sampling strategies adopted by DNA barcoding studies (Little & Stevenson, 2007; Erickson et al., 2008;

Bergmann et al., 2009); and (iv) those using one or several candidate markers in practical applications, such as distinguishing invasive from non-invasive species (Van de Wiel et al., 2009) or identifying poisonous plants (Bruni et al., 2010), species of economic importance (Newmaster & Ragupathy, 2009a), or medicinal species (Midgley & Turnbull, 2003; He et al., 2010). The present plant barcoding study of *Ligustrum* species belongs to categories (ii) and (iv).

The genus *Ligustrum* L. (Oleaceae) contains 37–50 species, mostly native to Asia, with a few species distributed in Europe, Australia, America, and North Africa (Green, 1990; Starr et al., 2003; Qin, 2009). *Ligustrum* species, usually called "privet", have long been cultivated in many areas of the world as hedge plants and street trees (Starr et al., 2003). Among these species, *L. sinense*, *L. lucidum*, *L. japonicum*, and *L. robustum* are described as invasive species in some areas (Morris et al., 2002; Milne & Abbott, 2004; Ferreras et al., 2008), whereas *L. obtusifolium* is recommended as a potentially suitable species for landfill remediation (Kim & Lee, 2005). According to the *Flora of China* (Parasyringa, 1996), there are 27 *Ligustrum* species in China. Fruits of *L. lucidum* with the English name Fructus Ligustri Lucidi are used as a valued tonic in traditional Chinese medicine, and the leaves and twigs of *L. robustum* subsp. *chinense* and its relatives are used as herbal tea substitutions in southwestern China. However, some *Ligustrum* species, such as *L. lucidum*, *L. japonicum*, and *L. vulgare*, are reported to be poisonous (Bruni et al., 2010). Accurate identification of different species in this genus is important, but *Ligustrum* species are difficult to discriminate using morphological characters (Milne & Abbott, 2004). A few molecular phylogenetic studies have included *Ligustrum*, such as studies of Oleaceae based on plastid DNA *trnL–F* and *matK* gene sequences (Kim & Kim, 2010), plastid DNA (*trnL–trnF*, *matK*, *trnT–trnL*, and *trnS–trnG*) and nuclear functional internal transcribed spacer (ITS) 1 sequences (Besnard et al., 2009), and nuclear ribosomal ITS and the external transcribed spacer (ETS) regions (Li et al., 2002).

An appropriate DNA region or combination of DNA regions used in plant barcoding should be routinely amplifiable with universal primers, easily sequenced via single-pass sequencing, appropriate in length (300–800 bp; Kress et al., 2005), variable enough to separate closely related species, and exhibiting less variability within species (Chase et al., 2005; Cowan et al., 2006; Newmaster et al., 2007). A number of candidate regions have been investigated. The plastid gene *rbcL* and the nuclear ribosomal ITS have been assessed for their ability to identify both *Moraea* and *Protea* using

the BLAST procedure (Altschul et al., 1990) and both DNA regions performed well (Chase et al., 2005). The nuclear ITS region and the plastid *trnH–psbA* intergenic spacer have been proposed as potential DNA barcoding regions for flowering plants (Kress et al., 2005). The plastid gene *rbcL* has been proposed as a core DNA barcoding gene to discriminate plant species at the genus level in a tiered approach wherein a highly variable locus can be implemented if necessary (Newmaster et al., 2006). Chase et al. (2007) indicated that low levels of variation in plastid DNA made three regions necessary and proposed two combinations, namely *rpoC1*, *rpoB*, and *matK*, and *rpoC1*, *matK*, and *psbA–trnH*, as usable markers for land plants. Kress & Erickson (2007) proposed the combination of *trnH–psbA* and *rbcL* as a two-locus global DNA barcode for land plants. Newmaster et al. (2007) tested seven chloroplast loci for barcoding in *Compsoneura* and two regions (*matK* and *trnH–psbA*) showed promise as plant barcodes in nutmeg. Sass et al. (2007) tested eight DNA regions to discriminate cycads (Cycadaceae) and found that the ITS contains enough variability to identify most samples to the species level and is promising as a barcoding region. Fazekas et al. (2008) used eight chloroplast (cp) DNA regions (*rbcL*, *matK*, *rpoC1*, *rpoB*, *23S* rDNA, *trnH–psbA*, *atpF–atpH*, and *psbK–psbI*) and the mitochondrial *CO1* gene to identify species and found that the resolution of a single locus ranged from 7% (*23S* rDNA) to 59% (*trnH–psbA*). Three chloroplast DNA regions (*rbcL*, *trnSGG*, and *trnH–psbA*) were tested in Hymenophyllaceae and the results showed that *trnH–psbA* had the greatest possibility using as a marker for DNA-based identification (Nitta, 2008). The Consortium for the Barcode of Life (CBOL) Plant Working Group compared the performance of seven candidate plastid DNA regions and recommended the combination of *rbcL* and *matK* as the core barcode for land plants (CBOL, 2009). Chen et al. (2010) proposed that ITS2 could serve as a novel universal barcode for plant identification. Morejón et al. (2010) suggested multigene combinations as DNA barcodes for the identification of species in the Cycadales. Ren et al. (2010) demonstrated that the combination of *trnH–psbA* and ITS was the best option for identifying species in *Alnus* (Betulaceae).

In the present study, the four frequently recommended DNA barcodes, namely *rbcL*, *matK*, *trnH–psbA*, and ITS, were used to identify *Ligustrum* species. The aims of the study were: (i) to test the universality of the four DNA barcode loci in *Ligustrum* (Oleaceae); and (ii) to estimate the discrimination power of the four potential DNA barcodes using four evaluation criteria (BLAST, character-based methods, tree-based methods and TAXONDNA). As additional objectives, we

aimed to demonstrate DNA barcoding as an effective molecular technique for species identification and that it is practicable to combine plant DNA barcoding with traditional morphology-based taxonomy.

1   Material and methods

1.1   Plant samples

Eighty-two fresh samples representing 18 species of *Ligustrum* were collected from various sites in China. Two species native to Japan and South Korea (*L. japonicum* and *L. ovalifolium*) and two cultivated species from North America (*L. quihoui* and *L. ovalifolium*) were also included in the study. Fifteen sequences of six species, including the widely distributed European species *L. vulgare*, were downloaded from GenBank. Detailed information on the sampled taxa is provided in Table S1 available as supplementary material to this paper.

1.2   DNA extraction, amplification, and sequencing

Genomic DNA was extracted from silica gel-dried leaves using the 2× cetyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987). Amplification of DNA regions was performed using standard PCR. The primers used in the present study are listed in Table S2. The PCR cycling and sequencing conditions were according to Kress et al. (2005) and Sass et al. (2007). The PCR products were run on a 1.0% agarose gel in 1.0× TBE (Tris-borate-EDTA) buffer, purified using the Tiangen Midi purification Kit (Tiangen Biotech, Beijing, China) and then sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit and an Applied Biosystems ABI3730 DNA Sequencer.

1.3   DNA barcode analysis

Bidirectional sequences were assembled with Sequencer v. 4.6 and aligned using ClustalX (Thompson et al., 1997). The alignments were adjusted manually in BioEdit v. 7.0 (Hall, 1999). The GenBank accession numbers of newly determined sequences are given in Table S1. Pairwise Kimura 2-parameter (K2P) distances for all four DNA regions were calculated in MEGA v. 4.0 (Kumar et al., 2008) to estimate intra- and interspecific divergence. BLAST, sequence character-based methods, tree-based methods and TAXONDNA analysis were used to evaluate the discrimination power of the candidate markers.

The BLAST (Altschul et al., 1990) procedure was used to search the National Center for Biotechnology Information (NCBI) nucleotide database (http://www.ncbi.nlm.nih.gov/BLAST/, accessed 01 September 2006) for sequences that give the best alignments to all or part of a query sequence. In gymnosperms, BLAST was found to give accurate identification at the generic level (Little & Stevenson, 2007), but not at the species level, partly because the top hits are often not the closest phylogenetic relatives (Koski & Golding, 2001). BLAST was used to evaluate the generic-level identification power of the four markers in the present study.

The sequence character-based method (Rach et al., 2008), unlike distance-based analysis, which compares the sequences as whole units and constructs trees on the basis of overall similarity, searches for diagnostic discrete characters or combinations of characters (Bergmann et al., 2009). The information from each site of the four DNA regions was treated as a character to distinguish the taxa from each other.

Two tree-based methods were used to display the molecular identification results and test the monophyly of species. Neighbor-joining (NJ) trees were generated using MEGA v. 4.0 with the K2P model (Kumar et al., 2008). For maximum parsimony (MP) trees, heuristic searching was performed in PAUP v. 4.0b10 (Swofford, 2002). Bootstrap support was calculated using parsimony with 1 000 random addition replicates to determine statistical support for clusters and clades.

TAXONDNA is a program developed on the basis of genetic distance for analyzing identification rates of DNA barcodes (Meier et al., 2006). "Best Match", "Best Close Match", and "All Species Barcodes" are three commonly used criteria. With "Best Match", a query is assigned the species name of its best-matching barcode sequences, regardless of how similar the query and barcode sequences are. With "Best Close Match", a threshold similarity value is required to define how similar a barcode match needs to be before it can be identified. "All Species Barcodes" is the most rigorous application for identifying queries. A query is assigned a species name only if the query is followed by all known barcodes for a particular species and only if there are at least two conspecific matches (Meier et al., 2006). All three strategies were used to ensure the accuracy of the species assignments based on *rbcL*, *matK*, *trnH–psbA*, and ITS in the present study.

2   Results

2.1   PCR amplification and sequencing

The primers for the four DNA regions are universally applicable for the samples used in the present study. The success rate for PCR amplification for *rbcL*, *trnH–psbA*, ITS, and *matK* was 100%, 100%, 98%, and 96%, respectively. The success rate for bidirectional

Table 1　Evaluation of the four DNA loci

	rbcL	*matK*	*trnH–psbA*	ITS
Universal primer	Yes	Yes	Yes	Yes
PCR success (%)	100	96	100	98
Sequencing success (%)	100	100	98	95
Aligned sequence length (bp)	706	863	497	626
Indel length (bp)	0	6	2–14	1
No. informative sites/variable sites	12/17	13/19	55/57	80/90
No. samples (individuals)	82	81	79	84
Mean interspecific distance (range)	0.0036 (0–0.0110)	0.0004 (0–0.012)	0.031 (0–0.079)	0.032 (0–0.070)
Mean intraspecific distance (range)	0.0001 (0–0.0020)	0.0000 (0–0.0000)	0.00015 (0–0.0030)	0.0002 (0–0.0040)

ITS, internal transcribed spacer.

sequencing was the highest for *rbcL* (100%), followed by *trnH–psbA* (98%), *matK* (98%), and ITS (95%). New sequences generated in the present study include 78 *rbcL* sequences for 18 species, 75 *matK* sequences for 17 species, 77 *trnH–psbA* sequences for 17 species, and 74 ITS sequences for 17 species. In total, 304 newly generated sequences were submitted to GenBank (Table 1).

2.2　Alignment and variability

The length of the aligned *rbcL* sequences was 706 bp with 12 informative sites and 17 variable sites dispersed across the alignment. The aligned *matK* ma-trix is 863 bp long, with 13 informative sites and 19 variable sites. In the *trnH–psbA* matrix, the sequences are 497 bp in length, with 55 informative sites, 57 variable sites, and six 2–14 bp indels. For the ITS matrix, the aligned sequences are 646 bp long: the ITS1 region is 1–242 bp long, 5.8S is 243–401 bp long, and ITS2 is 402–626 bp long. The distribution of the 80 informative sites and 90 variable sites is relatively concentrated in ITS1 and ITS2 (Table 1; Fig. 1). The mean interspecific distances of the chosen loci were much greater than the intraspecific distances in the present study according to the tests of the four regions (Table 1). The distribution of the inter- and intraspecific distances for the DNA regions is shown in Fig. 2.
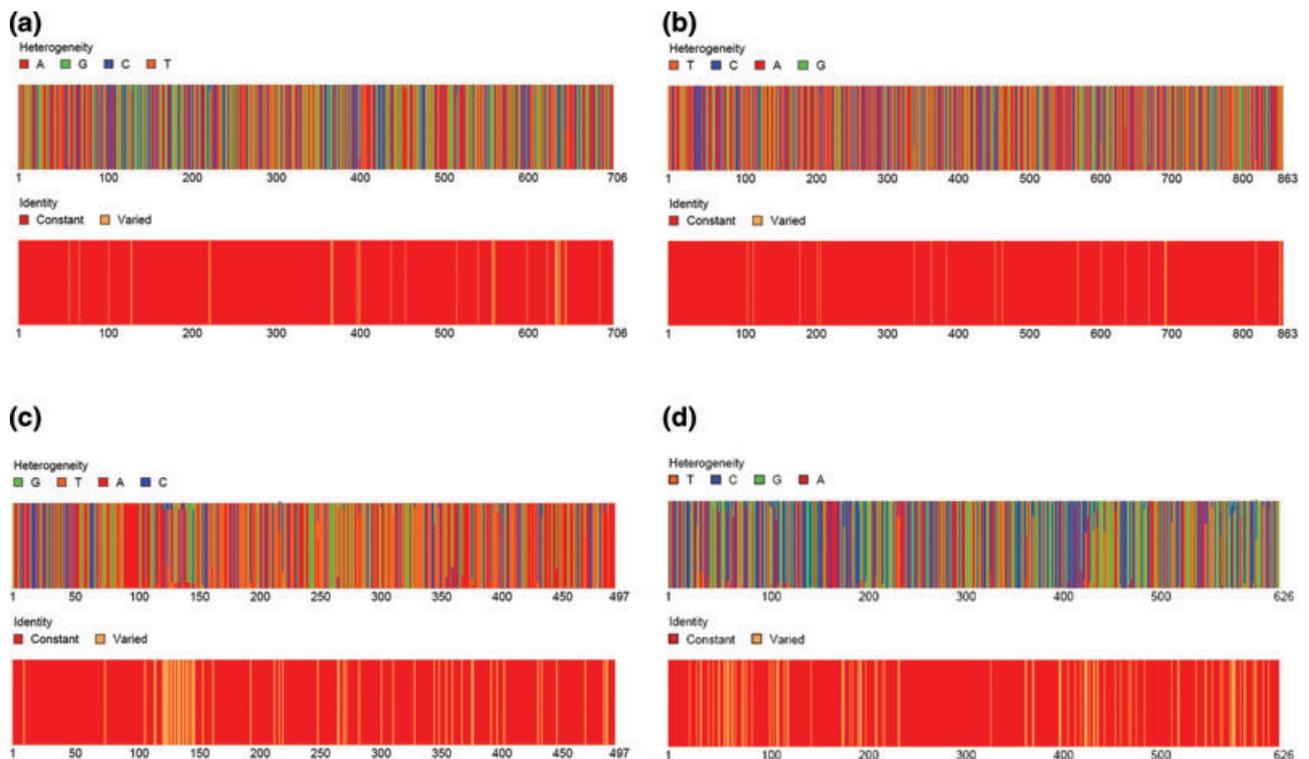

Fig. 1.　Nucleotide composition and variability distribution of four DNA loci. (a) *rbcL*, (b) *matK*, (c) *trnH–psbA* and (d) internal transcribed spacer (ITS).
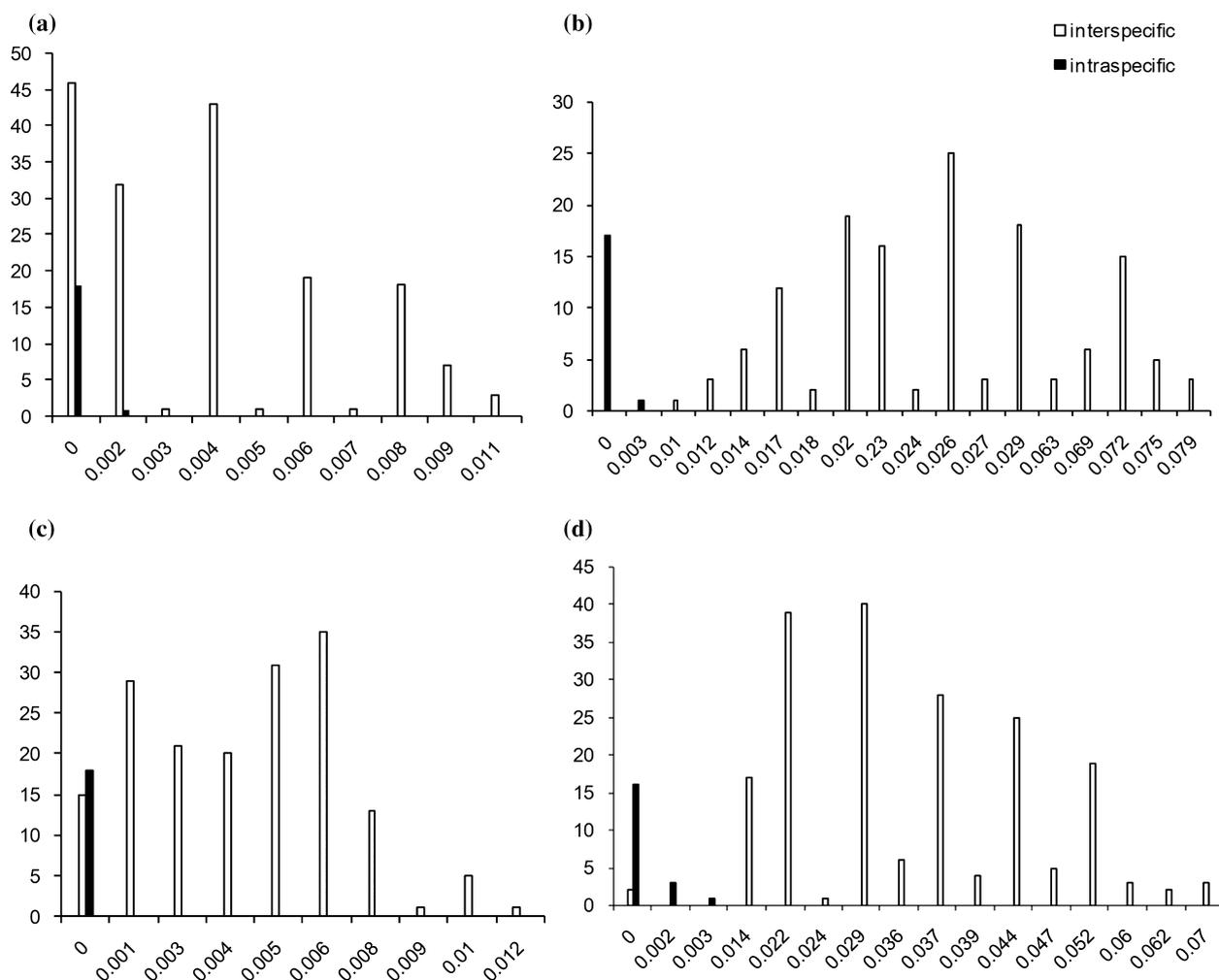
(a)

(b)

□ interspecific

■ intraspecific

(c)

(d)

Fig. 2. Relative distribution of the inter- and intraspecific Kimura 2-parameter (K2P) distances for four DNA loci. (a) *rbcL*, (b) *trnH–psbA*, (c) *matK* and (d) internal transcribed spacer (ITS).

2.3 Species identification

2.3.1 *rbcL* Each of the 82 *rbcL* sequences from the 19 species investigated in the present study was processed by BLAST in GenBank and all matched a species in this genus with the highest total score. According to the character-based method, there are 13 stable variable sites within the matrix (Table S3). Five species have unique character states that differentiate them from the other species: *L. compactum* (position 606: A), *L. gracile* (position 461: G), *L. henryi* (position 228: C),

L. japonicum (position 444: T), and *L. strongylohyllum* (position 61: C; position 73: C; position 108: G). Two species have unique character combinations that distinguish them from the other species: *L. ibota* var. *microphyllum* (position 135: G; position 403: A; position 548: T) and *L. robustum* subsp. *chinense* (position 403: T; position 652: A). The results based on the NJ or MP tree methods are consistent with those obtained by the character-based method and seven of the 19 species could be distinguished (Table 2).

Table 2 Species identification power of the DNA markers based on BLAST, character-based, neighbor-joining tree and maximum parsimony tree methods

Ability to discriminate	*rbcL*	*matK*	*trnH–psbA*	ITS
BLAST (genus level)	100%(19/19)	100%(18/18)	100%(18/18)	100%(20/20)
Character-based method (species level)	36.8%(7/19)	44.4%(8/18)	88.9%(16/18)	100%(20/20)
NJ tree (species level)	36.8%(7/19)	44.4%(8/18)	77.8%(14/18)	80%(16/20)
MP tree (species level)	36.8%(7/19)	44.4%(8/18)	77.8%(14/18)	75%(15/20)

ITS, internal transcribed spacer; NJ, neighbor-joining; MP, maximum parsimony.

2.3.2   *matK*   BLAST in GenBank matched all 78 *matK* sequences to *Ligustrum* species. In the sequence character-based data, the matrix includes 15 variable sites and a 6-bp indel for 14 species, excluding *L. robustum* subsp. *chinense* (Table S4). Seven species, namely *L. confusum* (position 371: C), *L. delavayanum* (position 471: G), *L. lucidum* (position 211: C), *L. robustum* subsp. *chinense* (position 677: A), *L. sempervirens* (position 392: T), *L. strongylohyllum* (position 186: T; position 577: C), and *L. vulgare* (position 112: C; position 121: A), have unique character states that differentiate them from the other species. *L. henryi* has a unique character combination (position 644: G and position 828: T) that distinguishes it from the other species. The results are the same with those obtained from the NJ or MP tree methods and eight of 18 species can be distinguished (Table 2).

2.3.3   *rbcL* and *matK*   The combination of *rbcL* and *matK* greatly improved the ability to identify species in the *Ligustrum* genus. Nine species could be identified by combining *rbcL* and *matK*, including five species that could only be identified with *matK* (*L. confusum*, *L. delavayanum*, *L. lucidum*, *L. sempervirens*, and *L. vulgare*) and four species that could only be identified by *rbcL* (*L. compactum*, *L. gracile*, *L. ibota* var. *microphyllum*, and *L. japonicum*). In total, 12 species could be differentiated from the others for an identification success rate of 63.2% (12/19).

2.3.4   *trnH–psbA*   BLAST searches in GenBank with the 78 *trnH–psbA* sequences of 18 species matched all the sequences with *Ligustrum* species. At the species level, 52 diagnostic sites and five indels were found in the sequence-based alignment database (Table S4). Eleven species have unique sites by which they can be differentiated: *L. confusum*, *L. delavayanum*, *L. henryi*, *L. japonicum*, *L. lucidum*, *L. obtusifolium* subsp. *suave*, *L. ovalifolium*, *L. sempervirens*, *L. strongylohyllum*, *L. vulgare*, and *L. xingrenens*. Being variable at one site (position 166: G) and constant at another site (position 253: A), *L. gracile* is distinct from the other species; similarly, *L. sinense* var. *sinense* is distinguishable by a variable site (position 221: C) and a constant site (position 77: G; position 269: G). *L. robustum* subsp. *chinense* could be distinguished by a unique combination of variable sites (position 253: C; position 274: C; position 349: G; position 380: A; position 407: G; and position 435: C). In addition, indels (positions 344–349 and 352–357) differentiate *L. ibota* var. *microphyllum* and *L. acutissimum*, respectively, from the other species. Of the total 18 species above, 14 could be easily distinguished in the NJ tree, forming 14 monophyletic groups with high support values obtained. Of the remaining four

species, *L. ibota* var. *microphyllum* and *L. acutissimum* could be distinguished by combining the NJ tree method with either the sequence character-based method or the MP tree method; *L. quihoui* and *L. expansum* could not be distinguished even with all three methods (Table 2, Fig. 3).

2.3.5   Internal transcribed spacer   All 20 species in the present study could be matched to the right genus by BLAST searching GenBank with their ITS sequences. A degree of variation exists in the ITS alignment region, with 79 diagnostic sites and three 1-bp indels for species identification (Table S6). Based on the unique diagnostic sites of the aligned ITS matrix, 12 species could be differentiated, including *L. compactum* (position 582: A), *L. delavayanum* (position 104: C; position 124: T; position 195: T), *L. japonicum* (position 418: G), *L. lucidum* (position 428: T), *L. massalongianum* (position 223: A), *L. obtusifolium* subsp. *suave* (position 81: G), *L. ovalifolium* (position 42: A; position 61: T; position 87: G; position 433: T), *L. robustum* subsp. *chinense* (position 601: T), *L. sempervirens* (position 123:T; position 221:T), *L. sinense* var. *sinense* (position 47: T; position 52: T; position 58: T; position 121: A; position 190: A; position 203: T), *L. vulgare* (position 481: A; position 524: T; position 542: T; position 569: C), and *L. xingrenense* (position 214: G). Moreover, *L. acutissimum*, *L. confusum*, *L. expansum*, *L. gracile*, *L. henryi*, *L. ibota* var. *microphyllum*, and *L. quihoui* can be distinguished from the other species by unique site combinations. Being variable at one site (position 601: T) but without other variable sites distinguishes *L. strongylohyllum* (Table S6). Thus, all 20 species could be distinguished by the sequence-based method. In the NJ tree, 16 of the 20 species above are identified as monophyletic groups and the other four species form two paraphyletic groups: *L. ibota* var. *microphyllum* + *L. acutissimum* and *L. quihoui* + *L. expansum* (Fig. 4). These two groups are distinct in the MP tree.

2.4   TAXONDNA analysis

Based on the "Best Match" and "Best Close Match" strategies of the TAXONDNA method, the success rate of species identification for ITS and *trnH–psbA* was 86.74% and 79.48%, respectively. However, the success rate for both *rbcL* and *matK* were <50%. According to the "All Species Barcodes" strategy, *trnH–psbA* had the highest percentage (91.2%) and there were no obvious differences among the three loci: *rbcL* 72.28%, *matK* 75.94%, and ITS 75.90% (Table 3). This indicates that *trnH–psbA* meets the rigorous standards for identifying queries accurately.
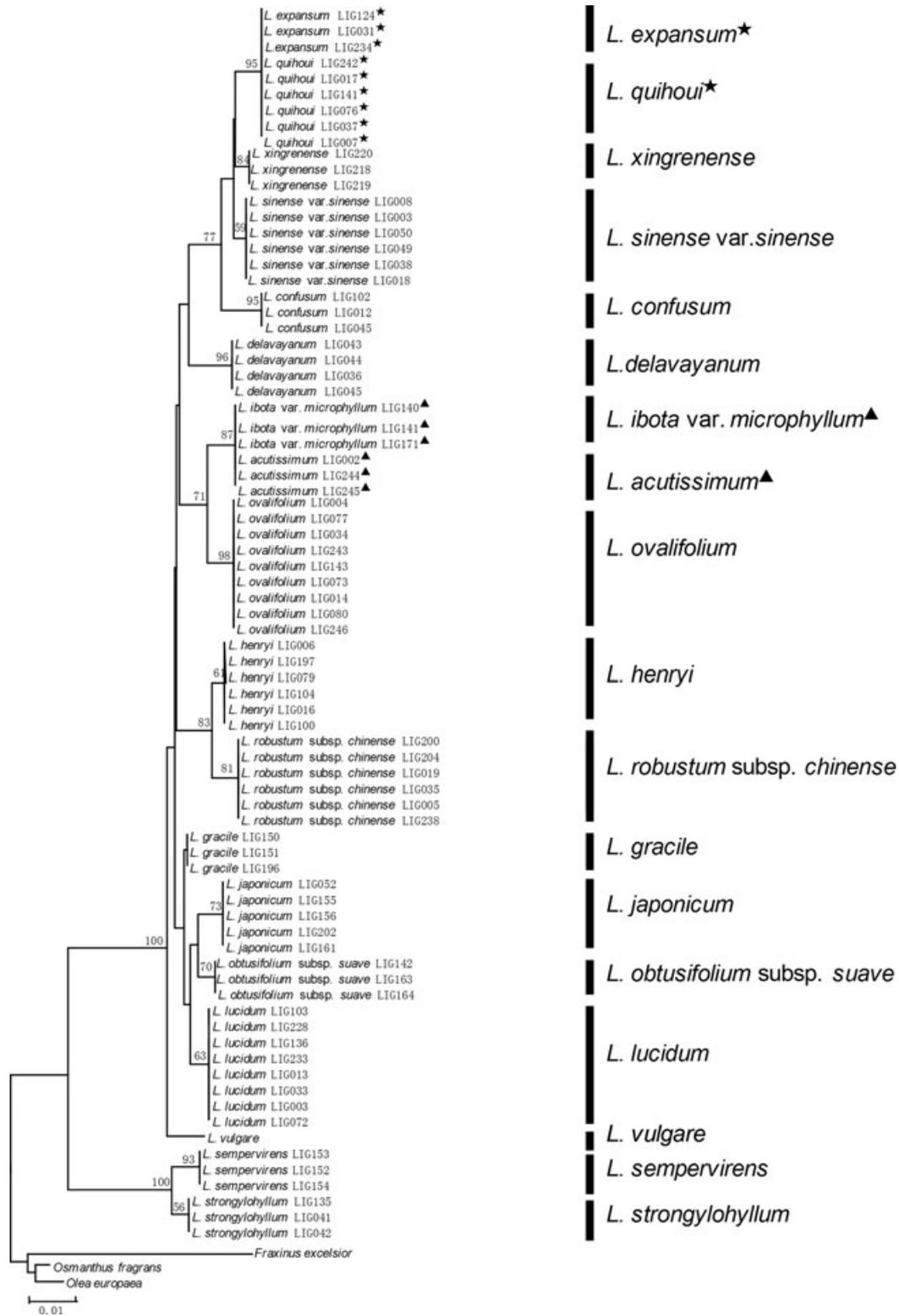
Fig. 3. A taxon identification tree generated using neighbor-joining analysis of Kimura 2-parameter distances showing patterns of *trnH–psbA* sequence divergence for 18 *Ligustrum* species. Bootstrap values (>50%) are shown above the relevant branches. DNA numbers follow the species names. Paraphyletic species that could be distinguished by the sequence character-based method or maximum parsimony tree are marked with pentacles. Paraphyletic species that could not be distinguished are triangles.
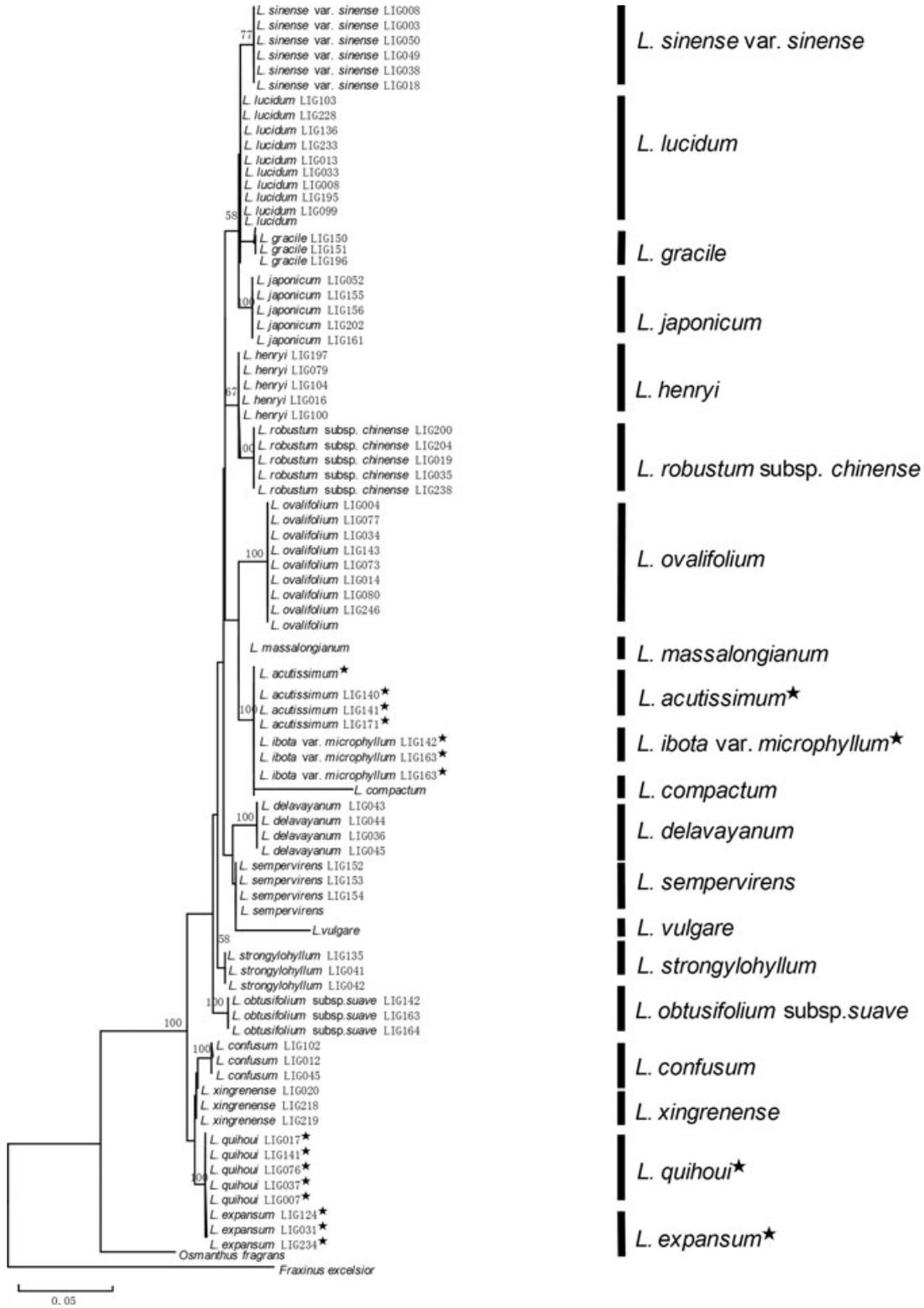
© 2011 Institute of Botany, Chinese Academy of Sciences

Fig. 4.   A taxon identification tree generated using neighbor-joining analysis of Kimura 2-parameter distances showing patterns of internal transcribed spacer (ITS) sequence divergence for 20 *Ligustrum* species. Bootstrap values (>50%) are shown above the relevant branches. DNA numbers follow the species name. Paraphyletic species that could be distinguished by the sequence character-based method or maximum parsimony tree are marked with pentacles.

Table 3   Species identification success rate based on TAXONDNA analysis

Criteria	*rbcL*	*matK*	*trnH–psbA*	ITS
Best match (%)	34 (40.96%)	32 (40.5%)	62 (79.48%)	72 (86.74%)
Best close match (%)	34 (40.96%)	32 (40.5%)	62 (79.48%)	72 (86.74%)
All species barcodes (%)	60 (72.28%)	60 (75.94%)	71 (91.02%)	63 (75.90%)

ITS, internal transcribed spacer.

3.   Discussion

3.1   Proposed loci

In the present study, all four potential barcodes were universal and could identify the *Ligustrum* species sampled correctly at the genus level through BLAST searching in GenBank. For identification at the species level, the nuclear ribosomal ITS showed promise as a barcoding region because it contains the most variations of the barcoding regions tested and was able to differentiate all species involved. The mean interspecific distance was more than 10-fold greater than the genetic distance within species. The length of the sequence alignment was 626 bp, which was suitable for barcoding, and all taxa were successfully amplified and sequenced using universal primers. The superiority of ITS for identifying species due to more variability has been demonstrated previously and ITS has been proposed as a plant barcode in other plant barcoding studies (e.g. *Moraea* and *Protea*: Altschul et al., 1990; Kress et al., 2005; cycads (Cycadaceae): Sass et al., 2007; and *Alnus* Mill. (Betulaceae): Ren et al., 2010). However, ITS is not generally accepted as a barcode. Previous studies have shown that the universal primers fail in gymnosperms, ferns, and mosses (Kress & Erickson, 2007). There have been difficulties with bidirectional sequencing with some samples and ITS has not always displayed sufficient variability for identification at the species level (Edwards et al., 2008). Another flaw of ITS is that intraspecific taxon pairs, such as those of *Aspalathus* L. (Fabaceae), show divergence in barcoding studies (Edwards et al., 2008). These problems are not universal and were not obvious in the present study, although ITS was found to have lower PCR amplification and sequence success rates than the other three loci (Table 1). The only sample (Voucher no. 23359; Table S1) that failed to be amplified was the sole degraded sample tested in the present study. Bidirectional sequencing problems for three other samples (Voucher nos. HYLIHE002, OLO411, and QYLIRO05; Table S1) were solved through cloning. Recent studies have shown that primers nested within the ITS region or in the partial variable region ITS2, which have been proposed as novel barcodes (Chen et al., 2010), could be universal and facilitate bidirectional sequencing in specific groups such as Bryophyta (Liu et al., 2010).

Many studies have suggested *trnH–psbA* as a promising marker for DNA barcoding, such as in *Compsoneura* (Newmaster et al. 2007), orchid (Lahaye et al., 2008b), and filmy ferns (Nitta, 2008). In the present study, *trnH–psbA* has a median length of 497 bp in the dataset, which is suitable for DNA barcoding. However, in some plant groups, the *trnH–psbA* spacer is exceedingly short (<300 bp: Kress et al., 2005), although it reaches 1 000 bp in orchids (Lahaye et al., 2008a), some monocots (Chase et al., 2007), and conifers (Hollingsworth et al., 2009). This shortness can lead to problems obtaining bidirectional sequences without using taxon-specific internal sequencing primers (CBOL, 2009) and alignment difficulties (Chase et al., 2007). In the present study, *trnH–psbA* was found to contain fewer variations than ITS, and two species were not discriminated. However, *trnH–psbA* can identify species with a relatively high success rate (88.9%), high interspecies variation, and low intraspecies variation, and the clusters shown in the NJ tree clearly demonstrate the relationships of the different species. Furthermore, combining nuclear DNA with cpDNA markers may be advantageous for discerning hybrid species due to their different patterns of inheritance (Ren et al., 2010). Thus, *trnH–psbA* is proposed as a candidate DNA barcode.

Both *rbcL* and *matK* showed low identification power (<50%) based on sequence character-based or tree-based methods at the species level in the present study. Using TAXONDNA analysis with "Best Match" or "Best Close Match" criteria, the success rate of *rbcL* and *matK* was 40.96% and 40.5%, respectively, suggesting that neither *rbcL* nor *matK* is sufficient to identify closely related species. The combination of *rbcL* and *matK*, which has been recommended as a core barcode for land plants (CBOL, 2009), greatly improved the identification success rate to 63.2%. However, the identification power of the combination was still lower than that of ITS or *trnH–psbA* at the species level. Thus, *rbcL* and *matK* are not suitable as plant barcodes based on the results of the present study.

3.2   Combination with traditional taxonomy

To make sure that the accuracy of each sequence is verified against other conspecifics and that the range of variation within a species is included, more than two individuals from each species should be included (Ren

et al., 2010). In the present study, most samples were collected from various natural habitats according to the records of specimens conserved in herbaria. Specimens with flowers or fruits were collected to enable strict morphological identification. This greatly decreases errors generated by incorrectly labeled samples.

The two original purposes of DNA barcoding have been discussed and debated, namely species "discovery" and species identification (Desalle, 2006). For species "discovery", when an unknown sample does not return a close match to known species in the barcode library and the barcode sequence does not qualify to designate the unknown sample as a new species, only expert taxonomists can resolve the relationship (Hajibabaei et al., 2007). For species identification, a means of identifying unknown organisms by querying an existing database in a barcoding study can only be established by the delimitations of taxonomy (Desalle, 2006). Furthermore, using morphological information helps reduce incorrect molecular inferences (Haase et al. 2007; Song et al., 2008). In the present study, two paraphyletic groups (*L. ibota* var. *microphyllum+L. acutissimum* and *L. quihoui+L. expansum*) shown in the NJ tree of the ITS region could only be differentiated by character-based methods, which treat each variable as a character to distinguish taxa from each other, or in combination with an MP tree, which uses different arithmetic from the NJ method. The relationships among these species indicated by the barcoding analysis can only be solved combining it with traditional taxonomic methods.

When viewed in the context of the traditional taxonomic framework, DNA barcoding provides a rapid diagnostic method for sorting specimens into genetically divergent groups (Hajibabaei et al., 2007). When used as a supplement to other taxonomic datasets, DNA barcoding can help with the process of delimiting species boundaries and identifying samples in cases where morphological features are missing. In the present study, the DNA sequence database of the species studied and the relationships between species reflected by the DNA barcode loci were found to provide information for the further taxonomic study of *Ligustrum* species.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. doi: 10.1006/jmbi.1990.9999.

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

Bergmann T, Hadrys H, Breves G, Schierwater B. 2009. Character-based DNA barcoding: a superior tool for species classification. Berliner und Münchener Tierärztliche Wochenschrift 122, Heft 11/12, Seiten: 446–460.

Besnard G, Casas RR, Christin PA, Vargas P. 2009. Phylogenetics of *Olea* (Oleaceae) based on plastid and nuclear ribosomal DNA sequences: tertiary climatic shifts and lineage differentiation times. Annals of Botany 104: 143–160.

Bruni I, Mattia FD, Galimberti A, Galasso G, Banfi E, Casiraghi M, Labra M. 2010. Identification of poisonous plants by DNA barcoding approach. International Journal of Legal Medicine 124: 595–603.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy Sciences, USA 106: 12794–12797.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Seberg O, Jørgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56: 295–299.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society of London B: Biological Sciences 360: 1889–1895.

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.

Cowan RS, Chase MW, Kress WJ, Savolainen V. 2006. 300 000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55: 611–616.

Desalle R. 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conservation Biology 20:1545–1547.

Devey DS, Chase MW, Clarkson JJ. 2009. A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. Taxon 58: 7–15.

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

Edwards D, Horn A, Taylor D, Savolain V, Hawkins JA. 2008. DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). Taxon 57: 1317–1327.

Erickson DL, Spouge J, Resch A, Weigt LA, Kress WJ. 2008. DNA barcoding in land plants: Developing standards to quantify and maximize success. Taxon 57: 1304–1316.

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.

Ferreras AE, Torres C, Galetto L. 2008. Fruit removal of an invasive exotic species (*Ligustrum lucidum*) in a fragmented landscape. Journal of Arid Environments 72: 1573–1580.

Ford CS, Ayres KL, Toomey N, Van Alphen Stahl J, Kelly LJ, Wikström N, Hollingsworth PM, Duff RJ, Hoot SB, Cowan RS, Chase MW, Wilkinson MJ. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. Botanical Journal of the Linnean Society 159: 1–11.

Green PS. 1990. *Ligustrum* (Oleaceae) in southern India. Kew Bulletin 45: 693–696.

Haase M, Wilke T, Mildner P. 2007. Identifying species of *Bythinella* (Caenogastropoda: Rissooidea): a plea for an integrative approach. Zootaxa 1563: 1–16.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23: 167–172.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

He J, Wong KL, Shaw PC, Wang H, Li DZ. 2010. Identification of the medicinal plants in *Aconitum* L. by DNA barcoding technique. Planta Medica 76: 1622–1628.

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. System Biology 54: 852–859.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003a. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270: 313–321.

Hebert PDN, Ratnasingham S, deWaard JR. 2003b. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society London B (Suppl.) 270: S96–S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. PLoS Biology 2: e312.

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular Ecology Resources 9: 439–457.

Hogg ID, Hebert PDN. 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. Canadian Journal of Zoology 82: 749–754.

Kim DK, Kim JH. 2010. Molecular phylogeny of tribe Forsythieae (Oleaceae) based on nuclear ribosomal DNA internal transcribed spacers and plastid DNA *trnL-F* and *matK* gene sequences. Journal of Plant Research (doi: 10.1007/s10265-010-0383-9)

Kim KD, Lee EJ. 2005. Potential tree species for use in the restoration of unsanitary landfills. Environmental Management 36: 1–14.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plant: The coding *rbcL* gene complements the noncoding *psbA–trnH* 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.

Koski LB, Golding GB. 2001. The closest BLAST hit is often not the nearest neighbor. Journal of Molecular Evolution 52: 540–542.

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 RRY, 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 [avaliable at the website: http://hdl.handle.net/10101/npre.2008.1896.1].

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 JH, Alexander JH, Zhang DL. 2002. Paraphyletic *Syringa* (Oleaceae): Evidence from sequences of nuclear ribosomal DNA ITS and ETS regions. Systematic Botany 27(3): 592–597.

Little DP, Stevenson DW. 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics 23: 1–21.

Liu Y, Yan HF, Cao T, Ge XJ. 2010. Evaluation of 10 plant barcodes in Bryophyta (Mosses). Journal of Systematics and Evolution 48: 36–46.

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

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

Midgley SJ, Turnbull JW. 2003. Domestication and use of Australian acacias: case studies of five important species. Australian Systematic Botany 16: 89–102.

Milne RI, Abbott RJ. 2004. Geographic origin and taxonomic status of the invasive Privet, *Ligustrum robustum* (Oleaceae), in the Mascarene Islands, determined by chloroplast DNA and RAPDs. Heredity 92: 78–87.

Moniz MB, Kaczmarska I. 2010. Barcoding of diatoms: nuclear encoded ITS revisited. Protist 161: 7–34.

Morejón FN, Silvac FV, Astorgaa JG, Stevensond DW, Vovidese AP, Sosaf V. 2010. A character-based approach in the

Mexican cycads supports diverse multigene combinations for DNA barcoding. Cladistics 26: 1–15.

Morris LL, Walck JL, Hidayati SN. 2002. Growth and reproduction of the invasive *Ligustrum sinense* and native *Forestiera ligustrina* (Oleaceae): implications for the invasion and persistence of a non-native shrub. International Journal of Plant Sciences 163: 1001–1010.

Newmaster SG, Ragupathy S. 2009a. Ethnobotany genomics: use of DNA barcoding to explore cryptic diversity in economically important plants. Indian Journal of Science and Technology 2: 1–7.

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

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

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J. 2007. Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources 8: 480–490.

Nitta JH. 2008. Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea. Taxon 57: 725–736.

Parasyringa WWS. 1996. *Ligustrum* L. In: Wu ZY, Raven PH eds. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 15: 299–307.

Presting GG. 2006. Identification of conserved regions in the plastid genome: implications for DNA barcoding and biological function. Canadian Journal of Botany 84: 1434–1443.

Qin XK. 2009. A New System of *Ligustrum* (Oleaceae) Acta Botanica Yunnanica 31: 97–116.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H. 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proceedings of the Royal Society of London B: Biological Sciences 275: 237–247.

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

Sass C, Littlle DP, Stevenson DW, Specht CD. 2007. DNA barcoding in the Cycadales: testing the potential of proposed barcoding marker for species identification of cycads. PLoS One 2: e1154.

Seberg O, Petersen G. 2009. How many loci does it take to DNA barcode a crocus? PLoS One 4: e4598.

Song HJ, Buhay JE, Whiting MF, Crandall KA. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proceedings of the National Academy of Sciences USA 36: 13486–13491.

Starr F, Starr K, Loope L. 2003. *Ligustrum* spp.: Privet, Oleaceae. Maui: United States Geological Survey-Biological Resources Division, Haleakala Field Station. Avaliable at the website http://www.hear.org/pier/pdf/pohreports/ligustrum_spp.pdf.

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

Swofford DL. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), version 4.0b10. Sunderland: Sinauer.

Thompson JD, Gibson TJ, Jeanmougin F, Higgins DG. 1997. The Clustal X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876–4882.

Van de Wiel CCM, Van der Schoot J, Van Valkenburg JLCH, Duistermaat H, Smulders MJM. 2009. DNA barcoding discriminates the noxious invasive plant species, floating pennywort (*Hydrocotyle ranunculoides* L.f.), from non-invasive relatives. Molecular Ecology Resources 9:1086–1091.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Sample details and sequence information.

Table S2. PCR and sequencing primers.

Table S3. A character-based DNA database for the *rbcL* region of *Ligustrum* species.

Table S4. A character-based DNA database for the *matK* region of *Ligustrum* species.

Table S5. A character-based DNA database for the *trnH–psbA* region of *Ligustrum* species.

Table S6. A character-based DNA database for the internal transcribed spacer (ITS) region of *Ligustrum* species.

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.