MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) **12**, 98–108 doi: 10.1111/j.1755-0998.2011.03076.x

Testing DNA barcoding in closely related groups of *Lysimachia* L. (Myrsinaceae)

CAI-YUN ZHANG,*† FENG-YING WANG,‡ HAI-FEI YAN,* GANG HAO,§ CHI-MING HU* and XUE-JUN GE*

**Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China, †Graduate University of the Chinese Academy of Sciences, Beijing 100049, China, ‡Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, Shanghai Chenshan Botanical Garden, Shanghai 201602, China, §School of Life Sciences, South China Agricultural University, Guangzhou 510642, China*

Abstract

It has been suggested that *rbcL* and *matK* are the core barcodes in plants, but they are not powerful enough to distinguish between closely related plant groups. Additional barcodes need to be evaluated to improve the level of discrimination between plant species. Because of their well-studied taxonomy and extreme diversity, we used Chinese *Lysimachia* (Myrsinaceae) species to test the performance of core barcodes (*rbcL* and *matK*) and two additional candidate barcodes (*trnH-psbA* and the nuclear ribosomal ITS); 97 accessions from four subgenus representing 34 putative *Lysimachia* species were included in this study. And many closely related species pairs in subgen. *Lysimachia* were covered to detect their discriminatory power. The inefficiency of *rbcL* and *matK* alone or combined in closely related plant groups was validated in this study. *TrnH-psbA* combined with *rbcL* + *matK* did not yet perform well in *Lysimachia* groups. In contrast, ITS, alone or combined with *rbcL* and/or *matK*, revealed high resolving ability in *Lysimachia*. We support ITS as a supplementary barcode on the basis of core barcode *rbcL* and *matK*. Besides, this study also illustrates several mistakes or underlying evolutionary events in *Lysimachia* detected by DNA barcoding.

Keywords: DNA barcoding, internal transcribed spacer, *Lysimachia*, species identification

Received 7 May 2011; revision received 17 August 2011; accepted 26 August 2011

Introduction

DNA barcoding has been established as a fast and reliable way of identifying species by analysing a portion of their genetic sequence. This technology has been used in a broad range of applications including creating rapid biodiversity inventories (Ebihara *et al.* 2010; Kress *et al.* 2010), identifying cryptic species (Lahaye *et al.* 2008; Newmaster & Ragupathy 2009; Ragupathy *et al.* 2009) and detecting misidentifications (Pryer *et al.* 2010). However, barcoding depends on having a reliable reference database of DNA sequences for each species. Unlike the situation with animals, the identification of suitable plant DNA barcodes has been the subject of extensive debate, largely because of the inherently slow rate of nucleotide evolution in plants and their ability to undergo complex evolutionary processes such as hybridization and polyploidy (Rieseberg *et al.* 2006; Fazekas *et al.* 2009). Consequently, no single plant barcode sequence is accepted to be universally applicable. Candidate loci including *rbcL*, *matK*, *rpoB*, *rpoC*, *trnH-psbA*, *psbK-psbI* and *atpF-atpH* have been examined individually, but none exhibit sufficient nucleotide variation by themselves to distinguish between all species. It has therefore been suggested that multilocus barcodes might be more useful in plant barcoding systems. Chase *et al.* (2005) and Newmaster *et al.* (2006) suggested a stepwise multilocus approach for identifying species (their approach has also been referred to as a traffic light approach or a tiered approach), and other researchers have suggested several candidate barcode combinations (Chase *et al.* 2007; Kress & Erickson 2007; Pennisi 2007). The most authoritative suggestion, which was published in PNAS by the CBOL Plant Working Group, involves using portions of the coding genes *rbcL* and *matK* as the core barcodes (CBOL Plant Working Group 2009). However, because coding regions always evolve slowly, these two core barcodes are not expected to perform well in closely related plant species (Spooner 2009; Starr *et al.* 2009; Ren *et al.* 2010). Despite these warnings, few studies have examined the performance of these barcodes in discriminating between closely related

Correspondence: Hai-Fei Yan, Fax: 862037252551;
E-mail: yanhaifei@scbg.ac.cn

species (Pettengill & Neel 2010). It would therefore be desirable to examine the performance of these barcodes in a wider range of species than has been done to date, and to evaluate the use of other loci in unique plant groups, as was proposed at the third International Barcode of Life Conference (Mexico 7–13 November 2009).

Noncoding regions exhibit higher rates of substitution than coding regions and are attractive candidates for use in plant barcoding systems (CBOL Plant Working Group 2009). Of the many noncoding plastid regions, *trnH-psbA* is considered to be the most promising candidate for use as a third barcode in plants because it evolves rapidly (Shaw *et al.* 2005; Kress & Erickson 2007; CBOL Plant Working Group 2009). In addition, the internal transcribed spacer (ITS) is a promising barcode because it is commonly used in phylogenetic studies and there is a large amount of ITS data in GenBank that can be used as reference material (Chase *et al.* 2005; Kress *et al.* 2005). The ITS region possesses sufficient variation and relatively high discrimination power for many groups (Sass *et al.* 2007; Edwards *et al.* 2008; Ren *et al.* 2010; Muellner *et al.* 2011). Recently, Chen *et al.* (2010) suggested that ITS2, a substitute of ITS, could be used as a universal DNA barcode across plants and animals because of its relatively high discriminatory power. However, ITS has not yet been comprehensively evaluated in a sufficiently wide range of plant groups. Therefore, there is an urgent need to evaluate the performance of the two most rapidly evolving candidate barcodes (*trnH-psbA* and ITS) in different plant groups.

Lysimachia L. is cosmopolitan, but is most common in the Northern Hemisphere, and includes approximately 180 species across the world (Hu & Kelso 1996). *Lysimachia* is one of the most intensively studied taxonomic groups in China; its taxonomy was entirely revised by Chen *et al.* (1989) and re-appraised by Hu & Kelso (1996). China is considered to be the diversity centre of this genus, being home to approximately 132 species (Hu & Kelso 1996). The Chinese members of this genus are divided into five subgenera, subg. *Idiophyton* Hand.-Mazz., subg. *Lysimachia*, subg. *Palladia* Hand.-Mazz., subg. *Heterostylandra* Chen et C. M. Hu and subg. *Naumburgia* Klatt (Chen *et al.* 1989). The first three subgenera are extremely diverse and each comprises more than 30 species, while the other two are monotypic. Subg. *Lysimachia* consists of 58 species in China, but approximately half are restricted to southwest China (Chen & Hu 1979; Chen *et al.* 1989). The relationships between species in each of the three diverse subgenera are ambiguous, which indicates that the genus might have experienced rapid evolution in China (Hao *et al.* 2004; Anderberg *et al.* 2007). In addition, many species of this genus have medical uses, particularly in Asia. For example, *Lysimachia christinae* Hance is a famous medicinal herb that is widely used to cure calculus, and *Lysimachia capillipes* Hemsl. can be used to treat influenza (Chen & Hu 1979). Given the potential uses of these species, their well-studied taxonomy and their high diversity in small areas, this genus is an excellent test case for DNA barcoding. In this study, we collected several closely related species in *Lysimachia* and aimed to (i) evaluate the performance of the core barcodes *rbcL* and *matK*; (ii) test the level of species discrimination achievable when using *trnH-psbA* and ITS as barcodes in *Lysimachia*; (iii) identify which locus or combination of loci is most suitable for distinguishing between species within the genus *Lysimachia*; and (iv) find out what useful information will be added to the taxonomy of the genus provided by DNA barcoding.

Materials and methods

Taxon sampling

Intraspecific sampling is particularly important when testing the utility of barcoding regions, so more than two accessions per species were selected in this study. In total, 97 individuals representing 34 putative *Lysimachia* species were included. The corresponding voucher specimens were deposited at the herbaria of the South China Botanical Garden (IBSC) and/or Kunming Institute of Botany (KUN) (Appendix S1, Supporting Information). Using the classification system described by Chen *et al.* (1989), 20 species were from the most diverse subg. *Lysimachia*, while six and seven species belonged to subg. *Idiophyton* and subg. *Palladia*, respectively. Moreover, *Lysimachia crispidens* Hemsl., representing the monotypic subg. *Heterotylandra*, was also included in our study. Several pairs of sister species from the *Lysimachia* subgenus were also included, namely *Lysimachia christinae* and *Lysimachia dextrosiflora* X. P. Zhang, X. H. Guo et J. W. Shao, *Lysimachia hemsleyi* Franch. and *Lysimachia erosipetala* Chen et C. M. Hu, and *Lysimachia hemsleyana* Maxim. and *Lysimachia chekiangensis* C. C. Wu. These species are all particularly difficult to distinguish in the field (Chen *et al.* 1989). *Ardisia verbascifolia* Mez was used as an outgroup for the tree-based analysis.

PCR and sequencing

Total genomic DNA was extracted from silica-dried plant leaves using a modified CTAB protocol (Doyle & Doyle 1987), and then target DNA regions including the two core barcodes (*rbcL* and *matK*), *trnH-psbA* and ITS were amplified with common DNA barcoding primers (Appendix S2, Supporting Information). Specifically, *rbcL* was amplified and sequenced using primers *rbcL*a_f (Kress & Erickson 2007) and 724R (Fay *et al.* 1997). Two primer pairs, 3F-KIM/1R-KIM and 3F-KIM/XF, were

used to amplify the *matK* region for all species in this study (Ford *et al.* 2009; K. J. K. Kim unpublished). *trnH-psbA* was amplified using the universal primers provided by Sang *et al.* (1997) and Tate & Simpson (2003). The ITS4 and ITS 5HP primer pair was used to amplify the ITS region (White *et al.* 1990; Hershkovitz & Zimmer 1996). For any failed amplification, we also used other unique plant ITS primer sets to detect the cause of the failure (Wen & Zimmer 1996). PCR mixtures (20 μL) each contained approximately 10 ng of template DNA, 1× PCR buffer ($MgCl_2$ free; TaKaRa), 0.2 μM of each primer, 1.5 mM of $MgCl_2$, 0.2 mM of each dNTP and 1 U *Taq* EX polymerase (TaKaRa). Polymerase chain reaction (PCR) amplification was performed using a PTC-200 thermocycler (Bio-Rad). Amplicons were purified using a DNA gel cleaning Kit (TaKaRa), and bidirectional sequencing reactions were carried out by Invitrogen Trading Shanghai Co., Ltd. All sequences were deposited in GenBank (Appendix S1, Supporting Information).

Data analyses

Raw bidirectional sequences were checked against the original trace files and verified by a BLASTN search on GenBank, then assembled by SeqMan™, one of the programs in the LASERGENE software package (DNASTAR, Inc.). Alignment data sets for all regions were generated by MUSCLE (Edgar 2004) and then inspected visually using Se-Al 2.0a11 (Rambaut 2002). We evaluated four single-locus barcodes and all possible combinations of them using three different methods.

Genetic distance–based method. An ideal barcode should exhibit high interspecific, but low intraspecific divergence, a so-called barcoding gap. Successful discrimination was confirmed if the minimum K2P-distance involving a species was larger than its maximum intraspecific distance (CBOL Plant Working Group 2009; Hollingsworth *et al.* 2009). Therefore, we calculated pairwise genetic distances based on the Kimura 2-parameter (K2P) nucleotide substitution model obtained from MEGA 4.0 (Tamura *et al.* 2007). We then estimated the presence of any barcoding gaps, which were estimated whether the intraspecies distances were larger than interspecies distances for each pair of regions and for all possible combinations (Hollingsworth *et al.* 2009; CBOL Plant Working Group 2009).

Tree-based method. Genetic distance overlap occurs when intraspecific variation in parts of the tree exceeds interspecific divergence in other parts of the tree, even though some species are reciprocally monophyletic to all others in tree analysis (Meyer & Paulay 2005). Thus, tree-based analyses provide a convenient method for evaluating discriminatory performance by calculating the proportion of monophyletic species, which may increase the resolving power than that provided by distance method. Neighbour-joining (NJ) trees were constructed for each individual barcode and the different barcode combinations by PAUP* version 4b10 based on a K2P distance model (Swofford 2003). Relative support for the branches of the NJ tree was assessed via 1000 bootstrap replicates. Only species with multiple individuals forming a monophyletic clade in NJ trees with a bootstrap value above 60% were considered to be successfully identified.

DNA sequence similarity–based method. A sequence comparison identification approach was conducted with TaxonDNA based on K2P distances (Meier *et al.* 2006). The 'best match' and the 'best close match' option were used to determine whether the query sequence matched correctly. In 'best match' analysis, each query is to find out its closest barcode match. If sequences were from the same species, then the species is classified as 'correct', whereas mismatched names were counted as 'incorrect'. Several equally good best matches from different species were considered 'ambiguous'. The 'best close match' was more rigorous, as it depended on a 95% pairwise distance threshold, as calculated by the 'pairwise summary' function (Meier *et al.* 2006). All queries whose pairwise genetic distances between query and reference sequence above the threshold value were classified as 'no match'. And the remaining queries that match below the threshold could be assigned to correct, incorrect and ambiguous respectively, according to the outline above for 'best match' analysis. These tests were only used for species that were represented by at least two individuals in the experimental set.

Results

PCR success and sequence characteristics

Successful amplification and high-quality sequences were obtained for *rbcL* across all 97 individuals. For *matK*, 96.9% (94/97) of the sequences were successfully amplified and sequenced using 3F_KIM and 1R_KIM, and the three failed sequences were recovered using a newly developed primer pair, XF/3F_KIM (Appendices S1 and S2, Supporting Information). Sequence alignment was reliable for *rbcL* and *matK*, as no indels existed in the two regions. In contrast, we met many difficulties in amplification, sequencing and alignment for *trnH-psbA* and ITS; 20.6% (20/97) ambiguous sequences of *trnH-psbA* resulted from the presence of two mononucleotide repeats, which all mainly composed of poly A/T and required us to carefully check and manually edit. Moreover, *trnH-psbA* was extremely difficult to align correctly

owing to its high level of length variation (ranging from 378 to 473 bp); further alignment was conducted visually using Se-Al 2.0a11, and 23 indels were found in the *trnH-psbA* data matrix.

The amplification success rate attained for ITS was the lowest of all loci examined in this study across the 97 individuals (92.8%). A total of seven accessions, belonging to *Lysimachia clethroides* Duby (two accessions), *Lysimachia hemsleyi* (one accession), *Lysimachia fistulosa* var. *wulingensis* Chen et C. M. Hu (one accession), *Lysimachia heterogenea* Klatt (one accession), *Lysimachia grammica* Hance (one accession) and *Lysimachia pentaletala* Bunge (one accession), did not yield clear and unambiguous sequences. However, we could obtain more than two unambiguous ITS sequences from each of the extant *Lysimachia* species, except *L. grammica* and *L. pentaletala*. These two species individually were represented by single ITS sequence (Appendix S1, Supporting Information) and consequently excluded from following analyses when ITS alone or combined with plastid regions. These failed ITS sequences were caused by double peaks in the whole electropherograms of both strands and could not be improved despite trying many amplification conditions and primer sets (Wen & Zimmer 1996). The presence of several indels for ITS did not affect the quality of alignment. For convenience, variable characters and the mean intra- and interspecific distances of all four regions are shown in Table 1.

Species discrimination

The three different analysis methods afforded slightly different results, but the DNA sequence similarity–based method showed the highest level of species discrimination, followed by the NJ tree method (Tables 2 and 3). Methods were not the subject of this study, thus not

Table 1 Success rates for PCR and sequencing, and sequence characteristics of the four candidate barcodes

	rbcL	*matK*	*trnH-psbA*	ITS
Length of candidate barcodes (bp)	648	763	378–473	675–688
Alignment length of candidate barcodes (bp)	648	763	579	712
Number of test samples	97	97	97	97
Success rate of PCR and sequencing	100%	100%	100%	92.8%
Number of variable nucleotide sites	39	130	120	319
Mean intraspecific distance (K2P)	0.0005	0.00122	0.002484	0.0121
Mean interspecific distance (K2P)	0.0107	0.024793	0.039842	0.119186

Table 2 Identification success rates for single-locus and multilocus combinations of the four candidate barcodes

		Best match			Best close match				
Candidate barcodes	*N*	Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect	No match	Threshold, %
rbcL	97	37 (38.1)	57 (58.8)	3 (3.1)	37 (38.1)	57 (58.8)	3 (3.1)	0 (0.0)	0.30
matK	97	59 (60.8)	33 (34.0)	5 (5.2)	58 (59.8)	32 (33.0)	5 (5.2)	2 (2.1)	0.68
trnH-psbA	97	55 (56.7)	37 (38.1)	5 (5.2)	55 (56.7)	37 (38.1)	5 (5.2)	0 (0.0)	0.94
ITS	88	83 (94.3)	0 (0.0)	5 (5.7)	82 (93.2)	0 (0.0)	5 (5.7)	1 (1.1)	4.40
rbcL + *matK*	97	59 (60.8)	31 (32.0)	7 (7.2)	58 (59.8)	31 (32.0)	6 (6.2)	2 (2.1)	0.43
rbcL + *trnH-psbA*	97	68 (70.1)	22 (22.7)	7 (7.2)	67 (69.1)	22 (22.7)	7 (7.2)	1 (1.0)	0.46
rbcL + ITS	88	82 (93.2)	1 (1.1)	5 (5.7)	81 (92.0)	1 (1.1)	5 (5.7)	1 (1.1)	2.25
matK + *trnH-psbA*	97	76 (78.3)	13 (13.4)	8 (8.2)	75 (77.3)	13 (13.4)	6 (6.2)	3 (3.1)	0.58
matK + ITS	88	82 (93.2)	0 (0.0)	6 (6.8)	82 (93.2)	0 (0.0)	6 (6.8)	0 (0.0)	2.26
trnH-psbA + ITS	88	85 (96.6)	0 (0.0)	3 (3.4)	84 (95.5)	0 (0.0)	3 (3.4)	1 (1.1)	2.86
rbcL + *matK* + *trnH-psbA*	97	79 (81.4)	12 (12.4)	6 (6.2)	77 (79.4)	12 (12.4)	5 (5.2)	3 (3.1)	0.49
rbcL + *matK* + ITS	88	82 (93.2)	0 (0.0)	6 (6.8)	82 (93.2)	0 (0.0)	6 (6.8)	0 (0.0)	1.60
rbcL + *trnH-psbA* + ITS	88	83 (94.3)	1 (1.1)	4 (4.5)	82 (93.2)	1 (1.1)	4 (4.5)	1 (1.1)	1.87
matK + *trnH-psbA* + ITS	88	84 (95.5)	1 (1.1)	3 (3.4)	84 (95.5)	1 (1.1)	3 (3.4)	0 (0.0)	1.86
rbcL + *matK* + *trnH-psbA* + ITS	88	84 (95.5)	1 (1.1)	3 (3.4)	84 (95.5)	1 (1.1)	3 (3.4)	0 (0.0)	1.40

Number in bracket was calculated by dividing each item by all tested sample, indicating percentage of successful resolution in each candidate barcode.

Table 3 Identification success rates obtained using NJ tree and distance methods for the four candidate barcodes (individually and in combination)

	NJ tree method*	Distance method†
rbcL	29.4% (10/34)	26.5% (9/34)
matK	55.9% (19/34)	55.9% (19/34)
trnH-psbA	61.8% (21/34)	50.0% (17/34)
ITS	84.4% (27/32)	68.8% (22/32)
rbcL + *matK*	58.8% (20/34)	47.1% (16/34)
rbcL + *trnH-psbA*	64.7% (22/34)	64.7% (22/34)
rbcL + ITS	87.5% (28/32)	71.9% (23/32)
matK + *trnH-psbA*	67.6% (23/34)	64.7% (22/34)
matK + ITS	87.5% (28/32)	75.0% (24/32)
trnH-psbA + ITS	93.8% (30/32)	84.4% (27/32)
rbcL + *matK* + *trnH-psbA*	73.5% (25/34)	67.6% (23/34)
rbcL + *matK* + ITS	87.5% (28/32)	78.1% (25/32)
rbcL + *trnH-psbA* + ITS	93.8% (30/32)	84.4% (27/32)
matK + *trnH-psbA* + ITS	93.8% (30/32)	81.3% (26/32)
rbcL + *matK* + *trnH-psbA* + ITS	93.8% (30/32)	78.1% (25/32)

*Based on the proportion of monophyletic species with >60% bootstrapping.
†Based on the proportion of species with a minimum interspecific distance larger than its maximum intraspecific distance.

discussed hereafter. In the single-locus analysis, the lowest discriminatory power was found for *rbcL* (26.5–38.1%), followed by *matK* and *trnH-psbA* (55.9–60.8% and 50.0–56.7%, respectively), while ITS provided the highest discrimination rate (68.8–94.3%). Of the two-locus combinations, the core barcodes (*rbcL* + *matK*) had the lowest discriminatory power (47.1–60.8%) (Tables 2 and 3), showing this combination to be unsuitable for distinguishing closely related species, such as *Lysimachia christinae* and *Lysimachia dextrosiflora*, *L. hemsleyi* and *L. erosipetala*, and *Lysimachia hemsleyana* and *Lysimachia chekiangensis*, were completely indistinguishable. Slightly better levels of discrimination levels were attained with the three-locus combinations than with the two-locus combinations. Within this group, the resolving power of three-locus combinations *rbcL* + *matK* + *trnH-psbA* was lower than that of combinations containing *rbcL* + *matK* + ITS (Tables 2 and 3).

The NJ tree method is the only one of the analytical methods used that generates a graphical representation of the results, which is useful in determining the power of a given locus combination to discriminate between *Lysimachia* species, especially when they are closely related. At the subgenus level, the rate of discrimination achieved with the chloroplast candidates was 50%. Specifically, the subg. *Idiophyton* and *Lysimachia* could be successfully distinguished using any of the single chloroplast barcodes, individually (with the exception of *rbcL*) or in combination, while these cpDNA loci (alone or combined) were not sufficient for distinguishing between subg. *Palladia* and subg. *Heterostylandra* (Fig. 1b). The resolving performance at subgenus level obtained with *rbcL* + *matK* + ITS was significantly better (Fig. 1c). Actually, all four subgenera could be readily distinguished by using ITS in combination with any of the chloroplast loci (data not shown).

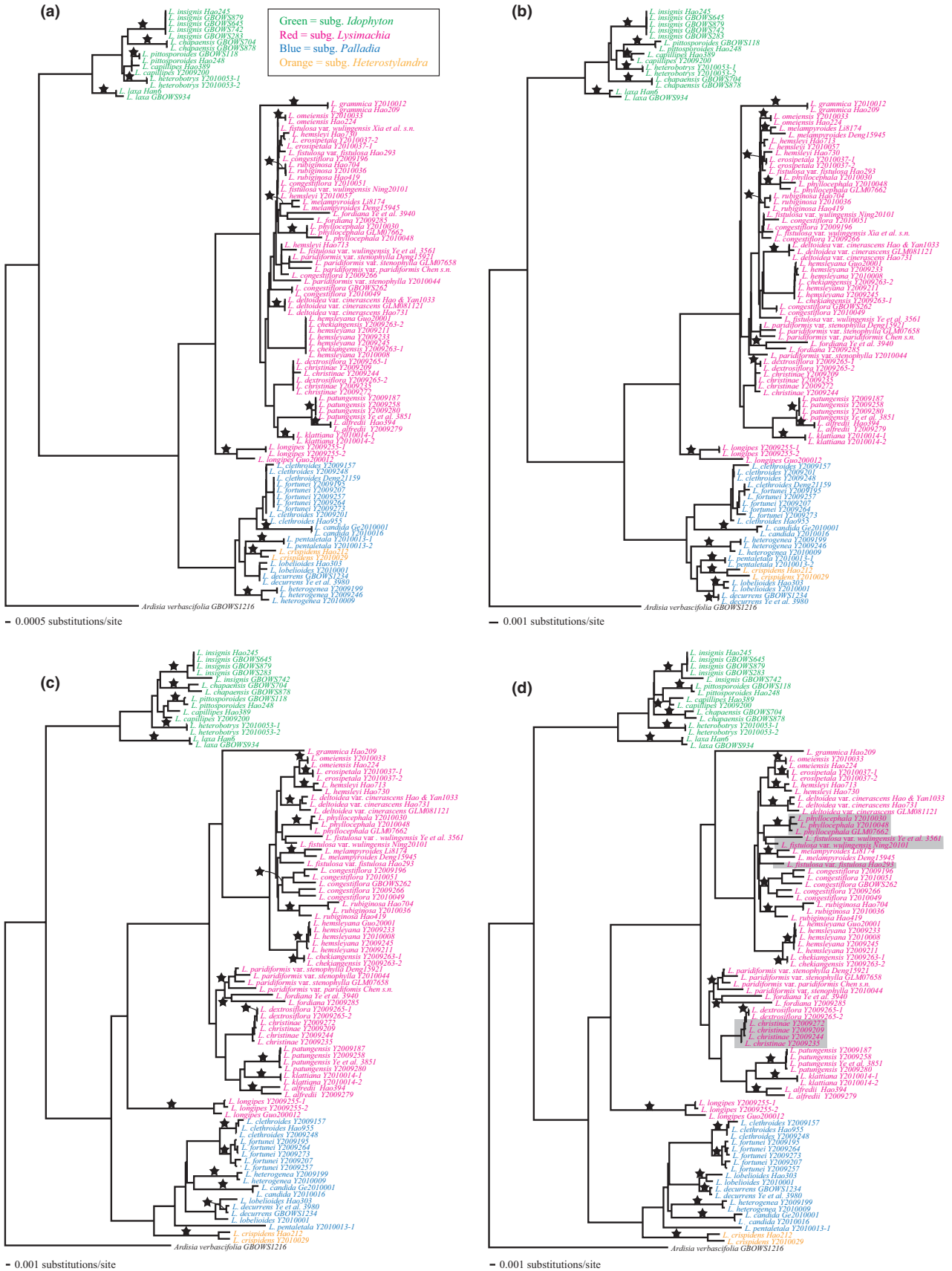
At the species level, the majority of species in the subg. *Palladia* and subg. *Idiophyton* could be identified using *matK* and/or *trnH-psbA* sequences, either individually or in combination with *rbcL*. Furthermore, *Lysimachia crispidens* of subg. *Heterostylandra* was also readily distinguished by any chloroplast candidate barcode. However, closely related species pairs in subg. *Lysimachia* were much more difficult to distinguish on the basis of chloroplast loci alone. Specifically, no more than 50% of the species examined could be identified by single *matK*, *trnH-psbA* and *rbcL* + *matK* sequences, and using *rbcL* + *matK* or *rbcL* + *matK* + *trnH-psbA* only increased this to 65% (Fig. 1a,b). In contrast to the relatively poor performance achieved with chloroplast loci at the subgenus level, ITS was sufficiently variable to discriminate between most of the closely related species. Of the seven species that were indistinguishable using candidate chloroplast barcodes (Fig. 1b), five species, *L. hemsleyi*, *L. hemsleyana*, *L. chekiangensis*, *Lysimachia congestiflora* Hemsl. and *Lysimachia pardiformis* Franch., could be successfully resolved using ITS, either alone or in combination with *rbcL* + *matK* (Fig. 1c). Only two species, *L. christinae* and *L. fistulosa* Hand.-Mazz., could not be distinguished using any of the barcode combinations (e.g. Fig. 1d).

Discussion

The low rate of discrimination obtained using rbcL and matK in Lysimachia

The discrimination ability of core DNA barcode *matK* and *rbcL* at genus level has been successfully verified in previous floristic inventories involving situations where only a limited number of closely related species co-occur

Fig. 1 Neighbour-joining tree for *Lysimachia* generated using (a) the *rbcL* + *matK* combination, (b) the *rbcL* + *matK* + *trnH-psbA* combination, (c) *rbcL* + *matK* + ITS and (d) all four loci examined in this study. Asterisks along branches indicate monophyletic species with bootstrap values above 60%. Accessions are suffixed by voucher numbers, and corresponding subgenera are colour-coded. Unresolved species and those whose identification was ambiguous when analysed using all four loci (d) are highlighted with grey shadowing.



in a single floristic region or ecological community (Lahaye *et al.* 2008; Kress *et al.* 2009; Kelly *et al.* 2010). However, when *rbcL* and *matK* have been used to distinguish between closely related species or for an exhaustively sampled genus, the rate of species discrimination dropped significantly. For example, *rbcL* could only identify 45% of Eurasian yew species (*Taxus* L., Taxaceae) (Liu *et al.* 2011) and only 10% of *Alnus* species (Ren *et al.* 2010). In this study, although *rbcL* was easy to amplify and sequence, as reported in other studies, it exhibited the lowest rate of species discrimination (25.7–32.3%) of all four candidate barcodes. The *matK* locus is considered to be a more promising and variable barcode than the more slowly evolving *rbcL* (Lahaye *et al.* 2008; Hollingsworth *et al.* 2009; Starr *et al.* 2009; Kelly *et al.* 2010). However, its primer universality often limits its use as barcode. Recently, many new designed primers and better amplification strategies are being developed (e.g. Dunning & Savolainen 2010; Li *et al.* 2011). Among the primer pairs published, 3F_KIM/1R_KIM designed by Kim (unpublished) is the best currently universal primer pair (Hollingsworth *et al.* 2011). The primer pair is successful in *Lysimachia* owing to its universality in all extant *Lysimachia* species and a total of 96.7% PCR and sequencing success. Just three accessions belonging to three different species had to be recovered by adding a new primer set (XF/3F_KIM). As a substitution of 1R_KIM, XF has been proved its utility in *Lysimachia* and *Primula* (Yan *et al.* 2011) but needs further studies in other plant groups. Given the high amplification and sequencing success of 3F_KIM/1R_KIM, we recommend that the primer set should be used in priority in *Lysimachia*, and multiple primers strategy for *matK* is necessary at present (Dunning & Savolainen 2010). Besides its primer limitation, *matK* has been shown to give very low identification rates when used in closely related groups such as *Alnus* (Ren *et al.* 2010) and *Solanum* sect. *Petota* (Spooner 2009). The performance of *matK* has often been overestimated because it is rarely tested against sister species. When this locus was used in more closely related groups in *Carex*, the discrimination rate was 50–60% (Starr *et al.* 2009). This value increased significantly to 95% in the identification of Canadian Arctic Archipelago *Carex* (Clerc-Blain *et al.* 2010).

Unsurprisingly, on the basis of the NJ tree analysis, only 19 of the 34 *Lysimachia* species examined in this work could be differentiated successfully when using *matK* alone (Table 3). Furthermore, the rates of discrimination obtained using the core *rbcL* + *matK* combination (47.1–60.82%) were the worst of all six of the possible two-locus combinations examined. For instance, within the *Lysimachia* subgenus, the NJ tree obtained using *rbcL* + *matK* indicated that only 50% of the species could be distinguished, and members of several complex groups in this subgenus, such as *Lysimachia christinae* and *Lysimachia dextrosiflora*, *Lysimachia hemsleyi* and *Lysimachia erosipetala*, and *Lysimachia hemsleyana* and *Lysimachia chekiangensis*, were completely indistinguishable. Although these two barcodes have good sequence quality and are easy to amplify and sequence, their poor performance in resolving *Lysimachia* species has largely prevented their widespread use as a barcode in the genus *Lysimachia*. The inherently slow evolution of chloroplast genes is the main cause for their poor performance in closely related plant groups, indicating that it will be necessary to focus on more rapidly evolving regions of the chloroplast genome or nuclear genome such as intergenic spacers and introns.

On the resolving power of trnH-psbA and ITS in Lysimachia

trnH-psbA has been the preferred barcode locus in many studies because of its high variability (Lahaye *et al.* 2008; Newmaster *et al.* 2008; Nitta 2008; Ren *et al.* 2010), universal primer pairs and relatively short length (Shaw *et al.* 2005). However, a major problem of the region was prevalence of homopolymer runs, which make it difficult to obtain high-quality sequences (Shaw *et al.* 2005; Chase *et al.* 2007; Fazekas *et al.* 2008; Devey *et al.* 2009; CBOL Plant Working Group 2009). Sequencing problems caused by mononucleotide repeats are also occurred in *Lysimachia*, and we found two microsatellite stretches in *trnH-psbA* sequences. One main repeat region located at the 3′ end and consequently caused 20.6% ambiguous sequences. Although microsatellite itself in chloroplast is a powerful tool for population genetics and always associated with mutational hotspots in the chloroplast genome (Weising *et al.* 2005), it provided little information in the genus, as only three species could be readily distinguished by microsatellites alone (data not shown). Moreover, the limited variation at this locus made it impossible to distinguish between certain *Lysimachia* species when using *trnH-psbA* alone or in combination with the core barcodes *rbcL* and/or *matK* (Tables 2 and 3). For instance, the combination *rbcL* + *matK* + *trnH-psbA* could not distinguish seven species in subg. *Lysimachia*, while five of them could be successfully resolved by *rbcL* + *matK* + ITS (Fig. 1). This locus therefore has only limited utility as the third barcode in this genus.

There is little doubt that faster evolving regions in plants exist in nuclear genome, which will enable to increase resolving power when they are used as barcodes. Low-copy nuclear regions were considered as the promising barcodes in future. However, there is still much work to do before they become acceptable barcodes regarding the lack of primer universality and difficulties in amplifying from even moderately

degraded DNA samples (Chase *et al.* 2005; Hollingsworth *et al.* 2011). For example, several low-copy nuclear markers (such as *Leafy*, *Adh* and GBSSI) have been successfully applied into many plant phylogenetic studies, but failed in the genus *Lysimachia*. In contrast to low-copy nuclear regions, the multicopy character of ITS makes it easy to be retrieved from degraded material (such as herbarium specimens) or low-quality DNA (Mort & Crawford 2004; Kress *et al.* 2005), which were validated in this study. In addition, ITS has been successfully used in phylogenic reconstruction for *Lysimachia* and provided more informative characters against other plastid region (*trnL-trnF*) (Hao *et al.* 2004), which indicates ITS will be more suitable as the barcode than low-copy nuclear and plastid regions in *Lysimachia*.

Although ITS was proposed as a universal DNA barcode at the early stages of plant barcoding (Chase *et al.* 2005; Kress *et al.* 2005), its utility is still a subject of debate (Sass *et al.* 2007; Edwards *et al.* 2008; Spooner 2009; Ren *et al.* 2010). ITS has some drawbacks that largely prevent its widespread use: it is prone to contamination from fungal epiphytes or endophytes (Zhang *et al.* 1997; Mayol & Rosselló 2001), exhibits incomplete concerted evolution (Álvarez & Wendel 2003; Nieto Feliner & Rosselló 2007) and can be difficult to amplify in some groups (e.g. gymnosperms, ferns and mosses) (Kress & Erickson 2007). In this study, we failed to obtain seven readable ITS sequences from six *Lysimachia* species and deleted two species (*Lysimachia grammica* and *Lysimachia pentaletala*) from further analyses. However, more ITS sequences for these failed species could be retrieved when increasing sampling.

Recent theoretical and empirical studies suggested that markers experiencing high rates of gene flow should better differentiate species (Du *et al.* 2009; Petit & Excoffier 2009; Zhou *et al.* 2010). As shown earlier, the chloroplast barcodes *rbcL* and *matK* and candidate barcode *trnH-psbA* did not perform well. However, when used by itself or in combination with core barcodes (*rbcL* and/or *matK*), ITS was capable of distinguishing between most of the closely related *Lysimachia* species examined, including *L. hemsleyi*, *L. hemsleyana*, *L. chekiangensis*, *Lysimachia congestiflora* and *Lysimachia pardiformis* (Tables 2 and 3; Fig. 1c,d). This suggests that ITS exhibits suitable inter- and intraspecific divergence in *Lysimachia*. Compared with the performance of *rbcL* + *matK* and *rbcL* + *matK* + *trnH-psbA*, the combination of *rbcL* + *matK* + ITS exhibited the best resolving possibility. Considering its high level of species resolution and relatively successful amplification and sequencing, we support the opinion of ITS as a supplementary barcode on the basis of core barcode *matK* and *rbcL* in *Lysimachia*. However, the performance should be evaluated in extensive trials in different plant groups.

Implications of DNA barcoding for the current taxonomy of Lysimachia

Despite the poor performance of DNA barcoding in some plant complexes (e.g. Spooner 2009), DNA barcoding does have utility in many plant groups, e.g. *Acacia* (Newmaster & Ragupathy 2009), *Agalinis* (Pettengill & Neel 2010), *Alnus* (Ren *et al.* 2010) and *Tolpis* (Mort *et al.* 2010). The results obtained in this study provide further evidence for the efficiency of DNA barcoding. The candidate barcodes could effectively distinguish between all four subgenera and between most of the *Lysimachia* species, even those belonging to complex groups within this subgenus. In addition to facilitating rapid identification by nonspecialists, DNA barcoding may be a useful tool for taxonomy (Cowan *et al.* 2006). This method can help assign ambiguous specimens to known species in cases involving immature, partial or damaged specimens and may be useful in identifying potential cryptic species and providing supplementary data for species delimitation studies (Schindel & Miller 2005).

In previous DNA barcoding studies on a single genus, morphologically ambiguous specimens could be sorted by DNA barcoding (e.g. Mort *et al.* 2010). In the present study, one accession (GLM-07662) initially identified as *L. hemsleyi* always clustered with *Lysimachia phyllocephala* Hand.-Mazz. (Fig. 1a–d). Confusion often occurs between *L. phyllocephala* var. *polycephala* Chen et C. M. Hu and *L. hemsleyi* in their early blooming stages, because both are characterized by nearly glabrescent and terminal inflorescence. The assignment of ambiguous specimens to *L. phyllocephala* by DNA barcoding was supported by the morphological characteristics of the plants' pollen, which are important in the classification of *Lysimachia* (Bennell & Hu 1983). The pollen of *L. hemsleyi* was coarsely reticulate (Appendix S3, G; unpublished results from our laboratories, Supporting Information), while that of *L. phyllocephala* (Y2010046) and *L. hemsleyi* (Y2010057) exhibited a finely reticulate exine sculpture (Appendix S3, C and I, Supporting Information).

In addition to facilitating the identification of ambiguous specimens, DNA barcoding could also help in resolving taxonomic uncertainties. For example, Edwards *et al.* (2008) suggested that five *Aspalathus* species with identical sequences should be treated as one species according to Van Steenis' definition of the concept (1957). However, analysis of these plants' NJ trees (Fig. 1d) showed that *L. dextrosiflora* (Y2009265) was always clustered with *L. christinae*. *Lysimachia christinae* is a widespread and polymorphic species that exhibits continuous variation in many of its characteristics, including the form of the calyx lobes (Hu & Kelso 1996). The morphology of the putative new species *L. dextrosiflora* was similar to *L. christinae* in most respects, but differed in that its

corolla lobes are right-contorted and in its limited distribution range (Shao *et al.* 2006). Our analysis suggests that *L. dextrosiflora* should be treated as a subspecies of *L. christinae*.

Studies on various animals, including butterflies (Hebert *et al.* 2004; Burns *et al.* 2008) and fig wasps (Xiao *et al.* 2010), have established that DNA barcoding is useful in identifying cryptic species. This has also proven to be the case in plants (e.g. Lahaye *et al.* 2008; Newmaster & Ragupathy 2009; Ragupathy *et al.* 2009; Starr *et al.* 2009). In this study, one *Lysimachia fistulosa* accession (Hao 293) did not cluster with the other two accessions of this species (Ye *et al.* 3561 and Ning 20101) (Fig. 1d). On the basis of their morphological characteristics, the former was identified as *L. fistulosa* var. *fistulosa*, while the latter was considered to be *L. fistulosa* var. *wulingensis* Chen et C. M. Hu. These two varieties were different with respect to their glabrous or pubescent stems, and their geographic ranges did not overlap (Hu & Kelso 1996). DNA barcoding divergence and the morphological differences provide strong evidence for the existence of cryptic species in *L. fistulosa*, which warrants further study.

Conclusion

Poor performance of two core barcodes *rbcL* and/or *matK* in resolving *Lysimachia* species indicates that additional barcodes should be used in this complex group. The success of ITS in distinguishing between members of closely related *Lysimachia* species highlights its potential utility for DNA barcoding in plants. Divergence of paralogous (incomplete concerted evolution) and the failure to find universal primers in many plant groups (Kress & Erickson 2007) make ITS difficult to apply into a broad range of plant species, but ITS may be the best nuclear locus at present as a local barcode when low resolving levels of core barcodes (*rbcL* and *matK*) are encountered in several complex groups (Chase *et al.* 2007). Therefore, we recommend that the concise combination *rbcL* + *matK* + ITS is the best choice in *Lysimachia*. Besides the success in *Lysimachia*, the ambiguous result in several species may indicate the underlying misidentification, improper taxonomical treatment and cryptic speciation in the genus. Given the relative success and implication in taxonomy, we propose that our ability to identify plant species will be largely enhanced by a reliable barcoding reference database with a thorough study of plant groups including a detailed analysis of morphological characters.

Acknowledgements

We thank Drs Lian-Ming Gao, Yu-Min Shui and Yun-Fei Deng for kindly providing plant tissue for this study. This research was supported by the Research Fund for the Large-scale Scientific Facilities of the Chinese Academy of Sciences (2009-LSF-GBOWS-01), the Major Innovation Program of the Chinese Academy of Sciences (KSCX2-YW-N-0807), the National Natural Science Foundation of China (Grant Nos. 30900160 and 31070189), The Ministry of Science and Technology of the People's Republic of China (MOST Grant Nos.2011FY120200) and Guangdong Natural Science Foundation (Nos.S2011010001178).

References

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–434.

Anderberg AA, Manns U, Källersjö M (2007) Phylogeny and floral evolution of the *Lysimachieae* (Ericales, Myrsinaceae): evidence from *ndhF* sequence data. *Willdenowia*, **37**, 407–421.

Bennell AP, Hu C-M (1983) The pollen morphology and taxonomy of *Lysimachia*. *Notes from the Royal Botanical Garden, Edinburgh*, **40**, 425–458.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservación Guanacaste, Costa Rica. *Proceeding of the National Academy of Sciences of the United States of America*, **105**, 6350–6355.

CBOL Plant Working Group (2009) DNA barcoding in land plants. *Proceedings of the National Academy of Science of the United States of America*, **106**, 12794–12797.

Chase MW, Salamin N, Wilkinson M *et al.* (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1889–1895.

Chase MW, Cowan RS, Hollingsworth PM *et al.* (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295–299.

Chen F-H, Hu C-M (1979) Taxonomic and phytogeographic studies on Chinese species of *Lysimachia*. *Acta Phytotaxonomica Sinica*, **17**, 21–53.

Chen F-H, Hu C-M, Fang Y-I, Cheng C-Z (1989) Primulaceae. In: *Flora Reipublicae Popularis Sinicae* vol. **59**, (eds Chen F-H & Hu C-M), pp. 3–133. Science Press, Beijing.

Chen S-L, Yao H, Han J-P *et al.* (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*, **5**, e8613 doi:10.1371/journal.pone.0008613.

Clerc-Blain JL, Starr JR, Bull RD, Saarela JM (2010) A regional approach to plant DNA barcoding provides high species resolution of sedges (*Carex* and *Kobresia*, Cyperaceae) in the Canadian Arctic Archipelago. *Molecular Ecology Resources*, **10**, 69–91.

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.

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.

Du F-K, Petit RJ, Liu J-Q (2009) More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other conifers. *Molecular Ecology*, **18**, 1396–1407.

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.

Ebihara A, Nitta JH, Ito M (2010) Molecular species identification with rich floristic sampling: DNA barcoding the pteridophyte flora of Japan. *PLoS ONE*, **5**, e15136 doi:10.1371/journal.pone.0015136.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

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

Fay MF, Swensen SM, Chase MW (1997) Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin*, **2**, 111–120.

Fazekas AJ, Burgess KS, Kesanakurti PR *et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome: discriminate plant species equally well. *PLoS ONE*, **3**, e2802 doi:10.1371/journal.pone.0002802.

Fazekas AJ, Kesanakurti PR, Burgess KS *et al.* (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, **9**, S130–S139.

Ford CS, Ayres KL, Toomey N *et al.* (2009) Selection of candidate coding DNA barcoding regions for use on land plants. *Botanical Journal of Linnean Society*, **159**, 1–11.

Hao G, Yuan Y-M, Hu C-M, Ge X-J, Zhao N-X (2004) Molecular phylogeny of *Lysimachia* (Myrsinaceae) based on chloroplast *trnL-F* and nuclear ribosomal ITS sequences. *Molecular Phylogenetics and Evolution*, **31**, 323–339.

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 Science of the United States of America*, **101**, 14812–14817.

Hershkovitz MA, Zimmer EA (1996) Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research*, **24**, 2857–2867.

Hollingsworth ML, Clark AA, Forrest LL *et al.* (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.

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS ONE*, **6**, e19254 doi:10.1371/journal.pone.0019254.

Hu C-M, Kelso S (1996) Primulaceae. In: *Flora of China*, vol. **15** (eds Wu Z-Y & Raven PH), pp. 39–189. Science Press, Beijing, and Missouri Botanical Garden Press, St Louis, Missouri.

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 WJ, Erickson DL (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 doi:10.1371/journal.pone.0000508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceeding of the National Academy of Sciences of the United States of America*, **102**, 8369–8374.

Kress WJ, Erickson DL, Jones FA *et al.* (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceeding of the National Academy of Sciences of the United States of America*, **106**, 18621–18626.

Kress WJ, Erickson DL, Swenson NG *et al.* (2010) Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican Forest Dynamics Plot. *PLoS ONE*, **5**, e15409 doi:10.1371/journal.pone.0015409.

Lahaye R, van der Bank M, Bogarin D *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceeding of the National Academy of Sciences of the United States of America*, **105**, 2923–2928.

Li Y, Gao L-M, Poudel RC, Li D-Z, Forrest A (2011) High universality of matK primers for barcoding gymnosperms. *Journal of Systematics and Evolution*, **49**, 169–175.

Liu J, Möller M, Gao L-M, Zhang D-Q, Li D-Z (2011) DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources*, **11**, 89–100.

Mayol M, Rosselló JA (2001) Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution*, **19**, 167–176.

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

Meyer CP, Paulay G (2005) DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422 doi:10.1371/journal.pbio.0030422.

Mort ME, Crawford DJ (2004) The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon*, **53**, 257–261.

Mort ME, Crawford DJ, Archibald JK, O'Leary TR, Santos-Guerra A (2010) Plant DNA barcoding: a test using Macaronesian taxa of *Tolpis* (Asteraceae). *Taxon*, **59**, 581–587.

Muellner AN, Schaefer H, Lahaye R (2011) Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). *Molecular Ecology Resources*, **11**, 450–460.

Newmaster SG, Ragupathy S (2009) 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 the 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 (2008) Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources*, **8**, 480–490.

Nieto Feliner G, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution*, **44**, 911–919.

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

Pennisi E (2007) Taxonomy. Wanted: a barcode for plants. *Science*, **318**, 190–191.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Pettengill JB, Neel MC (2010) An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae). *American Journal of Botany*, **97**, 1391–1406.

Pryer KM, Schuettpelz E, Huiet L *et al.* (2010) DNA barcoding exposes a case of mistaken identity in the fern horticultural trade. *Molecular Ecology Resources*, **10**, 979–985.

Ragupathy S, Newmaster SG, Murugesan M, Balasubramaniam V (2009) DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. *Molecular Ecology Resources*, **9**, 164–171.

Rambaut A (2002) Se-al version2.0a11. http://tree.bio.ed.ac.uk/software/seal/.

Ren B-Q, Xiang X-G, Chen Z-D (2010) Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. *Molecular Ecology Resources*, **10**, 594–605.

Rieseberg LH, Wood TE, Baack EJ (2006) The nature of plant species. *Nature*, **440**, 524–527.

Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, **84**, 1120–1136.

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 doi:10.1371/journal.pone.0001154.

Schindel DE, Miller SE (2005) DNA barcoding a useful tool for taxonomists. *Nature*, **435**, 17.

Shao J-W, Zhang X-P, Guo X-H (2006) *Lysimachia dextrorsiflora* X. P. Zhang, X. H. Guo & J. W. Shao, a new species of Primulaceae from China. *Acta Phytotaxonomica Sinica*, **44**, 589–594.

Shaw J, Lickey EB, Beck JT *et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.

Spooner DM (2009) DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *American Journal of Botany*, **96**, 1177–1189.

Starr JR, Naczi RFC, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources*, **9**, S151–S163.

Swofford DL (2003) *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0b10*. Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–159.

Tate JA, Simpson BB (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany*, **28**, 723–737.

Weising K, Nybom H, Wolff K, Kahl G (2005) *DNA Fingerprinting in Plants—Principles, Methods, and Applications*, 2nd edn. CRC Press, Boca Raton, Florida.

Wen J, Zimmer EA (1996) Phylogeny and biogeography of *Panax* L. (the Ginseng Genus, Araliaceae): inferences from ITS sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution*, **6**, 167–177.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* (eds Innis M, Gelfand D, Sninsky J & White T), pp. 315–322. Academic Press, San Diego, California.

Xiao J-H, Wang N-X, Li Y-W *et al.* (2010) Molecular approaches to identify cryptic species and polymorphic species within a complex community of fig wasps. *PLoS ONE*, **5**, e15067 doi:10.1371/journal.pone.0015067.

Yan H-F, Hao G, Hu C-M, Ge X-J (2011) DNA barcoding in closely related species: a case study of *Primula* L. sect. *Proliferae* Pax (Primulaceae) in China. *Journal of Systematics and Evolution*, **49**, 225–236.

Zhang W-P, Wendel JF, Clark LG (1997) Bamboozled again! Inadvertent isolation of fungal rDNA sequences from Bamboos (Poaceae: Bambusoideae). *Molecular Phylogenetics and Evolution*, **8**, 205–217.

Zhou Y-F, Abbott RJ, Jiang Z-Y *et al.* (2010) Gene flow and species delimitation: a case study of two pine species with overlapping distributions in southeast China. *Evolution*, **64**, 2342–2352.

Data Accessibility

Individual sampling location and haplotype data available in Appendix S1 (Supporting Information).

Supporting Information

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

Appendix S1 Taxon, voucher, Genbank accession numbers (*rbcL*, *matK*, *trnH-psbA*, ITS; en-dash denotes no sequence).

Appendix S2 Primers used in the amplification of the barcode regions.

Appendix S3 Pollen micro-morphology of three individuals of *Lysimachia*.

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