JSE Journal of Systematics and Evolution

Journal of Systematics and Evolution 49 (5): 425–437 (2011) doi: 10.1111/j.1759-6831.2011.00154.x

Research Article

DNA barcoding of *Pedicularis* L. (Orobanchaceae): Evaluating four universal barcode loci in a large and hemiparasitic genus

[1,2]Wen-Bin YU[†] [1,2]Pan-Hui HUANG[†] [3]Richard H. REE [1,2]Min-Lu LIU
[1]De-Zhu LI [1]Hong WANG*

[1](*Key Laboratory of Biodiversity and Biogeography, and Plant Germplasm and Genomics Center, Kunming Institute of Botany, Chinese Academy of Sciences,* Kunming 650204, China)

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

[3](*Department of Botany, Field Museum of Natural History,* Chicago, Illinois 60605, USA)

Abstract One application of DNA barcoding is species identification based on sequences of a short and standardized DNA region. In plants, various DNA regions, alone or in combination, have been proposed and investigated, but consensus on a universal plant barcode remains elusive. In this study, we tested the utility of four candidate barcoding regions (*rbcL*, *matK*, *trnH-psbA*, and internal transcribed spacer (ITS)) as DNA barcodes for discriminating species in a large and hemiparasitic genus *Pedicularis* (Orobanchaceae). Amplification and sequencing was successful using single primer pairs for *rbcL*, *trnH-psbA*, and ITS, whereas two primer pairs were required for *matK*. Patterns of sequence divergence commonly showed a "barcoding gap", that is, a bimodal frequency distribution of pairwise distances representing genetic diversity within and between species, respectively. Considering primer universality, ease of amplification and sequencing, and performance in discriminating species, we found the most effective single-region barcode for *Pedicularis* to be ITS, and the most effective two-region barcode to be *rbcL* + ITS. Both discriminated at least 78% of the 88 species and correctly identified at least 89% of the sequences in our sample, and were effective in placing unidentified samples in known species groups. Our results suggest that DNA barcoding has the potential to aid taxonomic research in *Pedicularis*, a species-rich cosmopolitan clade much in need of revision, as well as ecological studies in its center of diversity, the Hengduan Mountains region of China.

Key words DNA barcode, Hengduan Mountains region, ITS, *Pedicularis*, *rbcL*, species identification.

One goal of DNA barcoding is to allow individual organisms to be identified to species based on sequence analysis of a short and standardized DNA region (the "barcode") (Hebert et al., 2003a, 2003b; Kress et al., 2005). This has the potential to aid morphology-based taxonomy, as well as studies in ecology, evolution, conservation, and biodiversity management (Blaxter, 2004; Hebert & Gregory, 2005). It may also accelerate the pace of species discovery, by enabling rapid sorting of (possibly fragmentary) specimens into distinct genetic groups (Blaxter, 2004; Hebert & Gregory, 2005). In some cases, it may provide data relevant to studies of molecular phylogenetics and population genetics (Hajibabaei et al., 2007).

Key to the success of DNA barcoding is adoption of a "universal" region, easily amplified and sequenced across distantly related species, that is suitably informative, meaning that it varies much less within species than it does among species. In animals, the research community has settled on cytochrome c oxidase subunit I (COI), a short region of the mitochondrial genome (Hebert et al., 2003a, 2003b). In plants, however, low substitution rates of mitochondrial DNA have motivated a search for alternative markers (Kress et al., 2005). Various nuclear and plastid DNA regions, alone or in combination, have been suggested as DNA barcodes for plants (e.g., Kress et al., 2005; Chase et al., 2007; Kress & Erickson, 2007; Lee et al., 2007; Lahaye et al., 2008; CBOL Plant Working Group, 2009; Chen et al., 2010; Yao et al., 2010), but consensus on a universal marker remains elusive. For example, the CBOL Plant Working Group (2009) proposed a two-locus combination of *rbcL* + *matK*, whereas Chen et al. (2010) and Yao et al. (2010) proposed the internal transcribed spacer 2 (ITS2) region. These differences suggest a need for further testing of markers across more species and more taxonomic groups.

Received: 23 January 2011 Accepted: 30 May 2011

† These authors contributed equally to this work.
* Author for correspondence. E-mail: wanghong@mail.kib.ac.cn; Tel.: 86-871-5223534; Fax: 86-871-5217791.

© 2011 Institute of Botany, Chinese Academy of Sciences

In this study, we tested four DNA regions (*rbcL*, *matK*, *trnH-psbA*, and ITS) as candidates for DNA barcoding in the genus *Pedicularis* L. (Orobanchaceae), a clade of approximately 600–800 species of hemiparasitic herbs distributed mainly in temperate mountain ranges of the Northern Hemisphere (Li, 1948; Yang et al., 1998; Mill, 2001; Wang et al., 2009). It is most diverse in China, where 364 species are currently recognized (Yang et al., 2003a; Zhang et al., 2006; Yu et al., 2010), classified into 13 groups and 112 series (Tsoong, 1963; Yang et al., 1998). Most of these are endemic to the Hengduan Mountains region of northern Yunnan, western Sichuan, and eastern Xizang (Yang et al., 1998; Yang et al., 2003b; Ree, 2005). The large size and wide geographic distribution of the genus have inhibited attempts at a global systematic monograph. However, molecular phylogenetic studies (Yang et al., 2003b; Ree, 2005; Yang & Wang, 2007) have begun to provide a framework for assessing infrageneric taxonomic concepts proposed in regional treatments. Our primary motivation here was to develop a DNA barcode suitable for discriminating species of *Pedicularis* in its center of species richness and floral diversity, the Hengduan Mountains region, where many species are also valued in traditional Chinese medicine.

1 Material and methods

1.1 Study plants

Specimens of *Pedicularis* were collected from 179 localities in China, primarily in the Hengduan Mountains region (Fig. 1), and identified to species on the basis of morphological–geographical approaches. Fresh leaf tissue was preserved in silica gel for 328 individuals representing 12 groups, 51 series, and 88 species (Table 1). All species were represented by 2–8 samples, each taken from a distinct locality, with the exception of 10 rare species having 2–3 samples from a single population. An overview of vouchers and collection localities are listed in Appendix S1. All DNA samples and voucher specimens are stored at the Germplasm Bank of Wild Species and the herbarium of the Kunming Institute of Botany (KUN), respectively.

1.2 DNA isolation and sequencing

Genomic DNA was extracted from dried leaves by a modified CTAB method (Doyle & Doyle, 1987). Polymerase chain reaction (PCR) amplifications of *rbcL*, *matK*, *trnH-psbA*, and ITS were carried out in 21 μL reactions containing 1 μL genomic DNA (30–80 ng/μL), 0.5 μL each primer (1 μmol/L), 10 μL 2× Taq PCR MasterMix (Tiangen Biotech, Beijing, China) comprising 0.1 U Taq polymerase/μL, 0.5 mmol/L dNTP each, 20 mmol/L Tris-HCl (pH 8.3), 100 mmol/L KCl, and 3 mmol/L $MgCl_2$, and 9 μL ddH_2O. The PCR reactions of all primers used the following protocol: one cycle for 3 min at 94 °C, 35 cycles (45 s at 94 °C, 1 min at 53–55 °C, 1 min at 72 °C), followed by 5 min at 72 °C. Primers used in this study are listed in Appendix S2. For *matK*, PCR amplification was first attempted using primer pair *3F_KIM* and 1*R_KIM*, followed by another primer pair, *matK 5r* and *1R_KIM*.

The PCR products were visualized on 2% agarose gels in 1× TAE buffer, then purified using ExoSAP-IT (Affymetrix, Santa Clara, CA, USA). Purification was carried out in 7 μL reactions comprising 5 μL PCR product and 2 μL ExoSAP-IT. The purified PCR products were sequenced directly in both directions with the same primers used for amplification. Cycle sequencing was carried out in 6 μL reactions containing 0.2 μL purified PCR product, 1.15 μL BigDye Sequence Buffer, and 0.15 μL BigDye Reaction Mix (Applied Biosystems, Foster City, CA, USA), 0.5 μL unidirection primer (1 μmol/L), and 4 μL ddH_2O. Sequencing reactions were carried out with 30–32 cycles of: 10 s at 96 °C, 5 s at 50 °C, and 4 min at 60 °C. Products were precipitated in sodium acetate–alcohol (95%) solution (1: 24) for more than 30 min at −20 °C, washed twice with 75% alcohol and dried at 75 °C for 3—5 min. Dried products were diluted by 15 μL HiDi solution, then run on an ABI 3730 xl automated sequencer (Applied Biosystems).

Sequences were assembled and edited using SeqMan (DNAstar), and verified to specific DNA regions by BLAST searches on GenBank. Alignments were produced using MAFFT version 5.0 (Katoh et al., 2005), adjusted manually in BioEdit version 7.0 (Hall, 1999). GenBank accession numbers of all sequences are listed in Appendix S1.

1.3 Data analysis

We analyzed the sequence alignments for each DNA region alone or in combination, with each dataset representing a DNA barcode candidate. We considered a total of 10 barcodes: the four individual regions; four two-region combinations previously proposed by others (ITS and *trnH-psbA*, Kress et al., 2005; *rbcL* and *trnH-psbA*, Kress & Erickson, 2007; *matK* and *trnH-psbA*, Chase et al., 2007; *rbcL* and *matK*, CBOL Plant Working Group, 2009); and two new combinations (*rbcL* and ITS, and *matK* and ITS). For each region, we calculated the number of variable and parsimony-informative sites. For each barcode candidate, we calculated pairwise Kimura 2-parameter (K2P) distances (mean values per species) using MEGA version 4.3 (Tamura et al.,

© 2011 Institute of Botany, Chinese Academy of Sciences

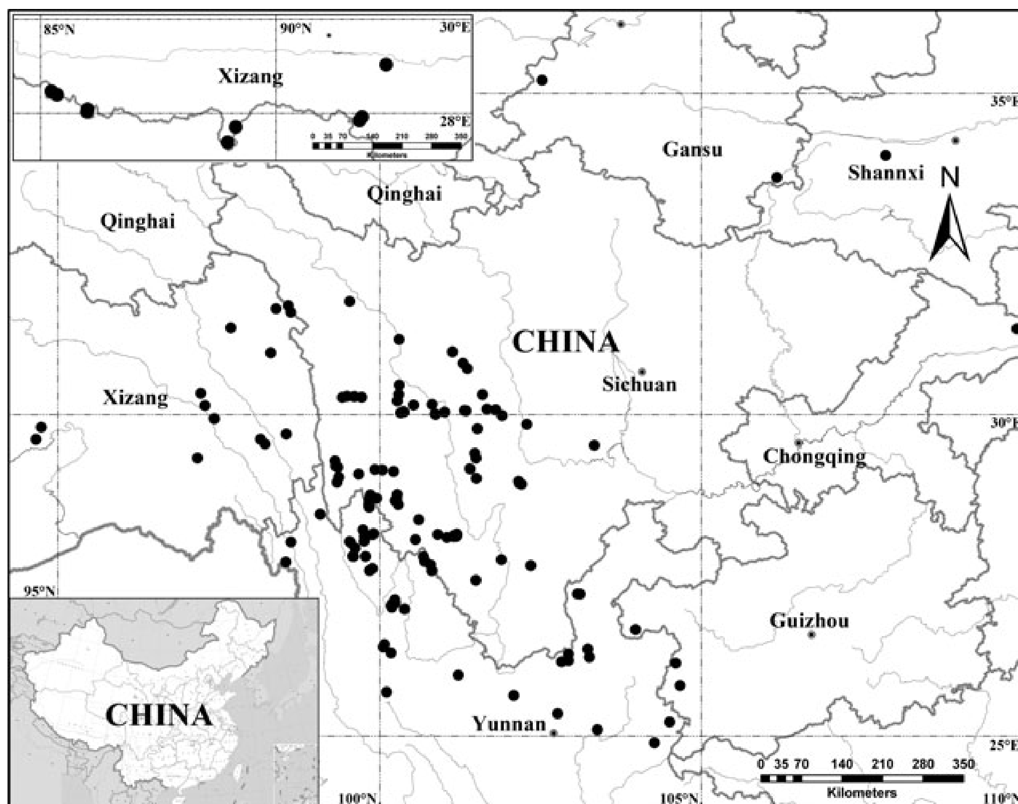

Fig. 1. Collection localities of *Pedicularis* in western China, mostly concentrated in northern Yunnan and western Sichuan. One sample of *P. resupinata* L. collected from Jilin Province, northeastern China, is not marked in this map. Detailed collections are listed in Appendix S1. The Chinese map was downloaded from the National Fundamental Geographic Information System in China (http://nfgis.nsdi.gov.cn/).

2007). Differences between intraspecific and interspecific K2P distances were tested using the two-sample Kolmogorov–Smirnov (K-S) test. Ranks of intraspecific/interspecific distance comparisons between barcode candidates were evaluated using paired-sample *t*-tests. All statistical analyses were carried out using SPSS version 13.0 (SPSS, Chicago, IL, USA). To assess a "barcoding gap" between intraspecific and interspecific distances, frequency distributions of K2P distances (0.5% categories), constructing a histogram, were obtained from TaxonDNA version 1.7 (Meier et al., 2006). This program was also used to calculate total overlap and 90% overlap between intraspecific and interspecific distances, and to find out identical sequences from different species.

Tree-based methods (Hebert et al., 2003a) and distance methods (Meier et al., 2006) were used to assess performance in species discrimination. Neighbor-joining trees (K2P distance model and pair deletion) were constructed by MEGA (Tamura et al., 2007) with bootstrap support measured from 1000 replicates. A species was considered successfully discriminated only if its samples formed an exclusive clade with bootstrap value $\geq$ 50%. In using distance methods, TaxonDNA was used to identify species based on three options, "best match", "best close match", and "all species barcode", using K2P distances model (Meier et al., 2006). The "best match" was used to find for each query its closest barcode match. If both sequences were from the same species, the identification was considered "correct", whereas mismatched names were assigned "incorrect". If multiple best matches from different species were found, the sequence was classified as "ambiguous". Before the "best close match" and the "all species barcode" were calculated, the distance threshold was calculated from the observed values of intraspecific distances for each barcode with 5% of the distances cut-off. In the "best close match", all query sequences without barcode match below the threshold were considered "no match"; "correct", "ambiguous", and "incorrect" identifications were similar to those in the "best match". For "all species barcodes", query sequences were classified as: "correct" if they were followed by all conspecific sequences; "ambiguous" when they were followed by at least one conspecific sequence; "incorrect" when they were followed by all sequences from another species;

© 2011 Institute of Botany, Chinese Academy of Sciences

Table 1 Summary for number of samples investigated (*n*) and species discrimination (bootstrap value ≥ 50%) using four DNA loci

Taxa[†]	*n*	L	K	H	I	LK	LH	LI	KH	KI	HI
Group I. *Apocladus*											
Ser. *Amplitubae*											
Pedicularis amplituba H. L. Li	2	–	99	68	97	98	71	97	99	99	99
Ser. *Oliganthae*											
P. cephalantha Franch. ex Maxim.[‡]	7	–	–	–	–	–	–	–	–	–	–
P. fengii H. L. Li	3	–	–	–	72	–	–	69	–	69	65
P. tachanensis Bonati	4	–	97	–	97	97	–	98	92	99	91
P. tahaiensis Bonati	2	–	95	61	99	88	67	99	99	99	99
Ser. *Oxycarpae*											
P. davidii Franch.[‡]	3	74	76	83	99	95	96	99	96	99	99
P. oxycarpa Franch. ex Maxim.	6	–	99	99	99	–	99	99	–	93	99
P. tibetica Franch.	4	99	91	80	99	99	99	99	98	99	99
Ser. *Rhinanthoides*											
P. rhinanthoides Shrenk[‡]	5	62	99	89	99	99	95	99	99	99	99
Group II. *Brachyphyllum*											
Ser. *Brevifoliae*											
P. alopecuros Franch. ex Maxim.	6	61	87	97	99	97	98	99	93	99	99
Ser. *Debiles*											
P. confertiflora Prain[‡]	3	–	–	–	–	–	–	–	–	–	–
P. debilis Franch. ex Maxim	4	–	79	99	99	85	99	99	99	99	99
P. maxionii Bonati	3	–	–	–	–	–	–	–	–	–	–
P. subacaulis Bonati	5	–	79	97	99	85	99	99	99	99	99
Ser. *Integrifoliae*											
P. integrifolia Hook. f.[‡]	4	90	99	99	99	99	99	99	99	99	99
Ser. *Lyratae*											
P. cymbalaria Bonati	4	–	–	–	–	–	–	–	–	–	–
P. deltoidea Franch. ex Maxim.	2	–	–	–	99	–	–	99	–	99	99
P. lutescens Franch. ex Maxim.	4	–	–	–	99	–	–	99	–	96	96
P. lyrata Prain	8	–	–	–	61	–	–	66	–	71	69
P. rizhaoensis H. P. Yang	2	–	67	84	87	62	86	86	96	95	94
Ser. *Urceolatae*											
P. urceolata P. C. Tsoong	5	64	99	97	99	99	98	99	99	99	99
Group III. *Cyathophora*											
Ser. *Cyathophyllae*											
P. cyathophylla Franch.	5	–	–	–	87	–	–	98	–	96	89
Ser. *Cyathophylloides*											
P. cyathophylloides W. Limpricht	5	–	–	–	98	–	–	70	–	80	97
Ser. *Reges*											
P. rex C. B. Clarke ex Maxim.[‡]	5	–	–	–	–	–	–	–	–	–	–
P. thamnophila (Hand.-Mazz.) H. L. Li	5	–	–	87	–	–	90	–	94	–	77
Ser. *Superbae*											
P. superba Franch. ex Maxim.[‡]	4	–	64	93	81	75	84	79	98	99	99
Group IV. *Cyclocladus*											
Ser. *Graciles*											
P. gracilis Wall. ex Benth.[‡]	5	80	96	99	99	99	99	99	99	99	99
Ser. *Longicaules*											
P. longicaulis Franch. ex Maxim.[‡]	4	–	83	52	99	82	82	99	97	99	99
Ser. *Melampyriflorae*											
P. floribunda Franch.	2	–	–	–	99	–	–	99	89	99	99
P. pseudomelampyriflora Bonati	5	–	81	99	99	81	99	99	99	99	99
Ser. *Salviiflorae*											
P. salviiflora Franch. ex Forbes & Hemsl.	3	98	99	98	99	99	99	99	99	99	99
Group V. *Dolichomiscus*											
Ser. *Axillares*											
P. axillaris Franch. ex Maxim.	5	85	99	99	99	99	97	99	99	99	99
Ser. *Batangenses*											
P. batangensis Bureau & Franch.	4	80	99	–	99	99	–	99	87	99	99
Ser. *Muscicolae*											
P. macrosiphon Franch.[‡]	3	82	65	98	98	96	99	99	99	99	99
P. sorbifolia P. C. Tsoong	3	91	99	98	98	99	99	99	99	99	99
Group VI. *Neosceptrum*											
Ser. *Lachnoglossae*											
P. lachnoglossa Hook. f.[‡]	5	65	70	95	99	89	96	99	99	99	99

Continued.

© 2011 Institute of Botany, Chinese Academy of Sciences

Table 1 Continued

Taxa[†]	*n*	L	K	H	I	LK	LH	LI	KH	KI	HI
Group VII. *Orthosiphonia*											
Ser. *Abrotanifoliae*											
P. brevilabris Franch.	4	98	87	98	99	99	99	99	99	99	99
P. densispica Franch. ex Maxim.	5	57	94	94	99	99	99	99	99	99	99
Ser. *Myriophyllae*											
P. alaschanica Maxim.[‡]	3	85	–	–	99	77	93	99	85	99	99
Ser. *Sermitorta*											
P. gyrorhycha Franch. ex Maxim.	3	–	99	99	99	99	99	99	99	99	99
Group VIII. *Pedicularis*											
Ser. *Carnosae*											
P. crenata Maxim.	4	65	51	–	–	88	66	–	66	–	–
P. nigra Vaniot ex Bonati[‡]	2	–	–	58	–	–	65	–	–	–	–
Ser. *Microphyllae*											
P. labordei Vant. ex Bonati[‡]	5	–	77	–	–	79	–	–	–	–	–
P. tenuisecta Franch. ex Maxim.[‡]	4	–	–	96	89	–	97	98	95	98	99
Ser. *Polyphyllae*											
P. gruina Franch. ex Maxim.	6	–	–	–	56	–	51	68	–	65	87
Ser. *Racemosae*											
P. resupinata L.[‡]	3	98	99	99	99	99	99	99	99	99	99
Group IX. *Polyschistophyllum*											
Ser. *Dichotomae*											
P. dichotoma Bonati[‡]	4	–	91	99	99	97	99	99	99	99	99
Group X. *Rhizophyllum*											
Ser. *Filiculae*											
P. lecomtei Bonati	3	–	–	–	74	–	–	70	–	69	64
Ser. *Flammeae*											
P. oederi Vahl[‡]	4	–	–	–	–	–	–	–	–	–	–
P. orthocoryne H. L. Li	2	68	99	56	99	99	–	99	94	99	99
Ser. *Franchetianae*											
P. mussotii Franch.	5	–	98	99	99	99	99	99	99	99	99
Ser. *Longiflorae*											
P. cranolopha Maxim.[‡]	7	93	96	99	–	99	99	90	93	89	99
P. decorissima Diels[‡]	2	–	–	–	99	–	–	99	–	99	99
P. latituba Bonati	3	95	99	99	99	99	99	99	99	99	99
P. longiflora Rudolph[‡]	5	73	80	94	98	95	98	99	99	99	99
P. sigmoidea Franch. ex Maxim.	2	–	91	–	–	92	–	–	–	–	–
P. siphonantha D. Don[‡]	7	–	–	–	–	–	–	–	–	–	–
Pedicularis sp.1	4	66	99	86	98	99	93	98	99	99	99
P. tricolor Hand.-Mazz.	3	88	99	99	99	99	99	99	99	99	99
P. variegata H. L. Li	2	94	98	99	99	99	99	99	99	94	99
Ser. *Macrorhynchae*											
P. macrorhyncha H. L. Li	2	–	97	85	99	–	91	99	99	99	99
Ser. *Megalanthae*											
P. megalantha D. Don[‡]	2	98	99	–	99	99	53	99	92	99	99
Ser. *Paucifoliatae*											
P. umbelliformis H. L. Li	2	99	99	95	99	99	99	99	99	99	99
Ser. *Pseudomacranthae*											
P. elwesii Hook. f.[‡]	4	83	99	99	99	99	99	99	99	99	99
Ser. *Pseudo-oederianae*											
P. pseudoversicolor Hand.-Mazz.	3	–	–	–	–	–	–	–	–	–	–
Ser. *Pumilliones*											
P. przewalskii Maxim.[‡]	3	79	99	99	99	99	99	99	99	99	99
Ser. *Rhynchodontae*											
P. rhynchodonta Bureau & Franch.	5	–	–	–	95	–	–	97	–	91	91
Group XI. *Sceptrum*											
Ser. *Craspedotrichae*											
P. ingens Maxim.[‡]	2	–	–	–	–	–	–	–	–	–	60
P. tongolensis Franch.	5	–	–	93	–	–	94	–	98	–	73
Ser. *Dolichocymbae*											
P. dolichocymba Hand.-Mazz.[‡]	2	–	95	99	99	99	99	99	99	99	99
Ser. *Imbricatae*											
P. clarkei Hook. f.	2	–	97	99	99	98	98	99	99	99	99

Continued.

© 2011 Institute of Botany, Chinese Academy of Sciences

Table 1 Continued

Taxa†	*n*	L	K	H	I	LK	LH	LI	KH	KI	HI
Ser. *Lasiophrydes*											
P. cinerascens Franch.	2	–	87	93	99	87	98	99	99	99	99
Ser. *Rudes*											
P. decora Franch.‡	2	–	–	–	66	–	–	68	–	71	–
P. duniana Bonati‡	3	–	–	–	67	–	–	69	–	63	–
Ser. *Subsurrectae*											
P. vialii Franch. ex Forbes & Hemsl.‡	2	82	99	99	99	99	99	99	99	99	99
Ser. *Trichoglossae*											
P. rhodotricha Maxim.	5	–	–	–	99	–	–	99	–	96	96
P. trichoglossa Hook. f.‡	2	–	–	–	98	–	–	97	–	94	93
Ser. *Tsekouenses*											
P. tsekouensis Bonati	3	74	99	99	99	99	99	99	99	99	99
Group XII. *Sigmantha*											
Ser. *Cheilanthifoliae*											
P. anas Maxim.‡	5	–	–	87	–	–	90	–	93	–	68
Ser. *Rigidae*											
P. comptoniifolia Franch. ex Maxim.‡	3	–	–	73	99	–	70	99	61	99	99
P. rigida Franch. ex Maxim.	3	–	85	99	99	86	99	99	99	99	99
Ser. *Verticillatae*											
P. glabrescens H. L. Li	4	–	–	–	99	72	–	99	–	99	93
P. kansuensis Maxim.‡	5	–	–	–	–	–	–	–	–	–	–
P. likiangensis Franch. ex Maxim.	2	94	79	–	99	97	–	99	–	99	93
P. roylei Maxim.‡	3	–	64	–	69	65	–	72	–	96	91
P. rupicola Franch. ex Maxim.	6	–	–	–	–	–	–	–	–	–	–
P. sikangensis H. L. Li	3	98	99	99	99	99	99	99	99	99	99
Pedicularis sp.2	2	77	95	–	99	98	–	99	–	99	60

†Taxonomic position in accordance with Tsoong (1963) and Yang et al. (1998). ‡Species with medicinal value (Wu et al., 1990; Wu, 1993; Zhang et al., 1994; Xie et al., 1996). –, Bootstrap value < 50%. H, *trnH-psbA*; HI, *trnH-psbA* + ITS; I, internal transcribed spacer (ITS); K, *matK*; KH, *matK* + *trnH-psbA*; KI, *matK* + ITS; L, *rbcL*; LH, *rbcL* + *trnH-psbA*; LI, *rbcL* + ITS; LK, *rbcL* + *matK*.

and "no match" when they had no match below the threshold.

2 Results

2.1 Polymerase chain reaction amplification and sequencing success

For *rbcL*, *trnH-psbA*, and ITS, all 328 samples were successfully PCR amplified and sequenced using single primer pairs. For *matK*, 259 samples were successfully PCR amplified and sequenced using the primer pair *3F_KIM* and *1R_KIM*, and 67 samples required an alternative forward primer, *matK 5r* (noted in Appendix S1). For two samples, LIDZ1504 (*Pedicularis oxycarpa* Franch. ex Maxim.) and LIDZ1580 (*P. siphonantha* var. *delavayi* (Franch.) P. C. Tsoong), PCR amplification but not sequencing was successful using the latter forward primer. Other primer pairs for *mat*K (*390F* and *1326R*, *matK xf*, and *matK 5r*) were also tested and failed in sequencing these recalcitrant samples.

2.2 Sequence quality and variation

Sequence characteristics for each region are summarized in Table 2. In general, consensus sequences could be directly assembled from reads of each DNA strand; in a few cases, manual editing of basecalls were required for *trnH-psbA* and ITS. For *trnH-psbA*, these occurred after mononucleotide repeats, whereas for ITS, ambiguous basecalls were the result of multiple superimposed peaks in chromatograms, likely resulting from incomplete concerted evolution of the tandem repeat. One such base was found in six samples (03055, HW10333, HW10169, LIDZ1187, LIDZ1007, and LIDZ1268), two were found in two samples (LIDZ1059 and HW10126), and nine were found in one sample (GLM103010). The ambiguous site was assigned using IUPAC ambiguity characters. The frequency of variable sites was lower in the two coding genes (*rbcL* and *matK*) compared to the non-coding spacers (*trnH-psbA* and ITS), and lower in the plastid loci compared to ITS (Table 2). Sequence alignment of all regions was straightforward except for *trnH-psbA*, which varied in length from 217 bp in one sample (081191) of *P. megalantha* D. Don to 589 bp in another sample (LIDZ1135) of *P. rhodotricha* Maxim. (Fig. 2). This alignment had a high frequency of indels that varied in size from 1 to 449 bp.

2.3 Variation within and between species

For all barcode candidates, mean K2P distances were significantly lower within species than between species (K-S test, $P < 0.001$, Table 3). Frequency distributions of K2P distances generally showed a

© 2011 Institute of Botany, Chinese Academy of Sciences

Table 2 Summary statistics for four candidate DNA barcodes examined from 328 individuals representing 88 *Pedicularis* species

Parameters	*rbcL*	*matK*	*trnH-psbA*	ITS
PCR success (%)	100	100	100	100
Sequencing success (%)	100	99.4	100	100
Aligned length (size, bp)	624 (624)	800 (737–764)	1195 (217–589)	655 (608–634)
GC (range,%)	43.4 (42.8–44.1)	33.8 (32.9–43.6)	23.5 (21.7–28.6)	61.2 (57.9–64.1)
Indel length (bp)	–	3–21	1–449	1–18
No. of variable/parsimony informative sites	92/88	303/252	534/428	360/320
Ratios of variable/parsimony informative sites (%)	15.0/14.0	39.0/32.0	44.7/35.8	55.9/48.9

–, No indel. GC, GC content; ITS, internal transcribed spacer; PCR, polymerase chain reaction.

"barcoding gap", separating infraspecific distances from interspecific distances (Fig. 3), although in some cases, infraspecific distances of *trnH-psbA* were well within the range of interspecific distances (e.g., in *P. likiangensis* Franch. ex Maxim., *P. rupicola* Franch. ex Maxim., *P. siphonantha* D. Don, and *Pedicularis* sp.2; Fig. 3). In the three barcode candidates, rates of total overlap between intraspcific and interspecific distances were higher than 59%, but the rates abruptly shrank to zero (but 0.49% in *trnH-psbA*) at a level of 90% overlap (Table 3). For four single DNA regions, queried with allospecific identical match varied from 2.38% (four matches) of the total intraspecific and interspecific identical matches in ITS to 47.95% (141 matches) of those in *rbcL* (Table 3).

2.4 Performance of candidate barcodes

2.4.1 Tree-based method The performance of barcode candidates at discriminating species based on the neighbor-joining monophyly criterion is summarized in Table 3, with species-specific results listed in Table 1. For all candidate barcodes, high mean interspecific K2P distances enhanced the performance of species discrimination ($r = 0.671$, $P < 0.001$), with the exception of *trnH-psbA*. Twenty-six species (29.5%) were successfully discriminated by all barcode candidates, but 10 species (11.4%) could be discriminated by none of them. The single-region barcode with highest discrimination success was ITS (78.4%), and the lowest was *rbcL* (38.6%). The two-region combination with highest discrimination success was *trnH-psbA* + ITS (81.8%), which outperformed *rbcL* + ITS and *matK* + ITS (both 79.5%), and the combinations with the lowest success were *rbcL* + *matK* and *rbcL* + *trnH-psbA* (both 61.4%). The 16 species not discriminated by ITS were clustered into 11 species groups; the 15 species not discriminated by either *rbcL* + ITS or *matK* + ITS, and the 14 species not discriminated by *trnH-psbA* + ITS, were clustered into 10 species groups (Appendix S3).

2.4.2 Distance method The distance threshold of each barcode varied from 0.8% (*rbcL*) to 3.3% (*trnH-psbA*) (Fig. 4). A specific threshold for each barcode reduced few "correct" and "incorrect" identifications in the "best match" as "no match" in the "best close match". Discrimination success for candidate barcodes, alone or in combination, varied from 53% (*rbcL*) to 94.5% (*trnH-psbA* + ITS) using the "best match", from 52.4% (*rbcL*) to 92.1% (*trnH-psbA* + ITS) using the

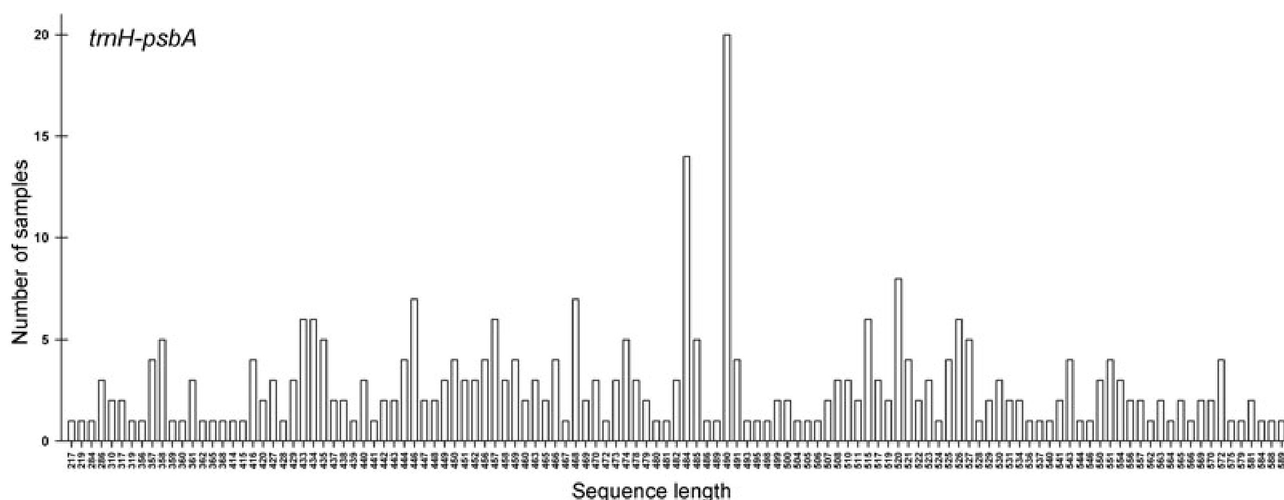

Fig. 2. Length variations of unaligned *trnH-psbA* spacer.

© 2011 Institute of Botany, Chinese Academy of Sciences

Table 3 Summary of intraspecific and interspecific Kimura 2-parameter (K2P) distances, and species discrimination of tree-based method

Barcode	Total overlap (% observation)	90% Overlap (% observation)	Queried with allospecific identical match	Mean K2P distance (%, range)†		K-S test‡		Species discrimination (BV ≥ 50%)		
				Intraspecific	Interspecific	Z-value	P-value	≥50%	≥70%	≥90%
L	0–1.46% (58.23)	0.48–0.8% (15.87)	47.95% (141)	0.08 ± 0.02 (0–0.8)^{a}	1.45 ± 0.01 (0–3.7)^{a}	8.53	<0.001	34	26	13
K	0–2.28% (23.16)	1.2–1.6% (5.02)	32.25% (70)	0.27 ± 0.04 (0–1.9)^{c}	2.96 ± 0.02 (0–5.4)^{c}	8.49	<0.001	53	48	36
H	0–9.87% (92.83)	3.24–3.33% (0.49)	20.93% (50)	0.73 ± 0.15(0–9.9)^{h}	6.86 ± 0.04 (0–18.7)^{h}	8.57	<0.001	51	47	37
I	0–4.38% (7.85)	2.63–3.7% (0.0)	2.38% (4)	0.62 ± 0.07 (0–3.0)^{g,h}	8.53 ± 0.04 (0.2–17.0)^{j}	8.99	<0.001	69	64	58
LK	0–1.77% (26.92)	0.94–1.16% (3.5)	32.53% (68)	0.18 ± 0.03 (0–1.4)^{b}	2.26 ± 0.01 (0–4.1)^{b}	8.61	<0.001	54	52	39
LH	0–4.15% (76.18)	1.59–1.79% (0)	17.15% (35)	0.35 ± 0.07 (0–4.2)^{c,d,e}	3.51 ± 0.02 (0–8.4)^{d}	8.62	<0.001	54	49	44
LI	0–2.8% (8.48)	1.38–2.45 (0)	1.35% (2)	0.32 ± 0.04 (0–1.8)^{c,d}	4.86 ± 0.02 (0.1–8.8)^{f}	9.02	<0.001	70	65	60
KH	0–4.6% (59.75)	1.99–2.22 (0)	15.0% (27)	0.44 ± 0.08 (0–4.1)^{f,g}	4.31 ± 0.02 (0.1–9.4)^{e}	8.61	<0.001	55	53	50
KI	0.07–3.13% (7.91)	1.64–2.75% (0)	0	0.40 ± 0.05 (0–2.1)^{e,f}	5.4 ± 0.02 (0.2–9.7)^{g}	9.04	<0.001	70	66	62
HI	0.08–5.18% (11.2)	2.33–3.84% (0)	0	0.63 ± 0.09 (0–4.5)^{h}	7.87 ± 0.34 (0.2–14.5)^{i}	9.46	<0.001	72	66	62

†Different superscript letter indicating significant difference among barcode candidates using paired samples *t*-test ($P < 0.05$). ‡Two-sample Kolmogorov–Smirnov (K-S) test of non-parameter tests testing difference between intraspecific and interspecific distances. BV, bootstrap value of neighbor-joining trees; H, *trnH-psbA*; HI, *trnH-psbA* + ITS; I, internal transcribed spacer (ITS); K, *matK*; KH, *matK* + *trnH-psbA*; KI, *matK* + ITS; L, *rbcL*; LH, *rbcL* + *trnH-psbA*; LI, *rbcL* + ITS; LK, *rbcL* + *matK*.

"best close match", and from 57.7% (*matK*) to 70.4% (*trnH-psbA* + ITS) using the "all species barcodes" (Fig. 4). Misidentification rates of sequences for all barcodes were lower than 10% in the "best match" and "best close match", and lower than 5% in "all species barcodes". For the "all species barcode", the false-positive identification of 67.98% was found in *rbc*L (Fig. 4), likely resulting from high rates of identical sequences from different species (Ross et al., 2008).

3 Discussion

3.1 Barcode selection for the genus *Pedicularis*

To date, a variety of plant barcode candidates have been suggested by a number of investigators (e.g., Kress et al., 2005; Chase et al., 2007; Kress & Erickson, 2007; Lee et al., 2007; Lahaye et al., 2008; CBOL Plant Working Group, 2009; Chen et al., 2010; Yao et al., 2010), with the two-locus combination of *rbcL* + *matK* receiving much of the attention. The gene *matK* is successfully used in a few genera, such as *Carex* (Starr et al., 2009; Le Clerc-Blain et al., 2010) and *Crocus* (Seberg & Petersen, 2009). However, for recently diverged species, discrimination success of *rbcL* + *matK* is likely to be limited by the low substitution rates of these coding genes. Two highly variable non-coding spacers, *trnH-psbA* and ITS/ITS2, although recommended by some researchers (e.g., Edwards et al., 2008; Song et al., 2009; Gao et al., 2010; Luo et al., 2010; Ren et al., 2010; Li et al., 2011; Liu et al., 2011), have other disadvantages. For example, in *trnH-psbA*, these include ambiguous raw sequences, difficulty in alignment, frequent inversions, and high intraspecific variation; in ITS, the problems arise from the presence of multiple and divergent copies and lack of universal primers and divergent paralogs in ITS (e.g., Kress et al., 2005; Chase et al., 2007; Kress & Erickson, 2007; Lahaye et al., 2008; CBOL Plant Working Group, 2009; Whitlock et al., 2010). The bottom line is that, so far, no candidate has been found to meet all three key criteria for an ideal barcode, that is, universality of primers, sequence quality and coverage (quality of bidirectional sequences), and high performance at species discrimination (CBOL Plant Working Group, 2009).

In the present study, four regions (*rbcL*, *matK*, *trnH-psbA*, and ITS) were evaluated as candidate DNA barcodes in *Pedicularis*. In terms of primer conservatism, the best candidates were *rbcL*, *trnH-psbA*, and ITS. In terms of sequence quality and coverage, *rbcL* and *matK* yielded clean, unambiguous reads, whereas some sequences of *trnH-psbA* and ITS had a low frequency of ambiguous basecalls and required manual

© 2011 Institute of Botany, Chinese Academy of Sciences

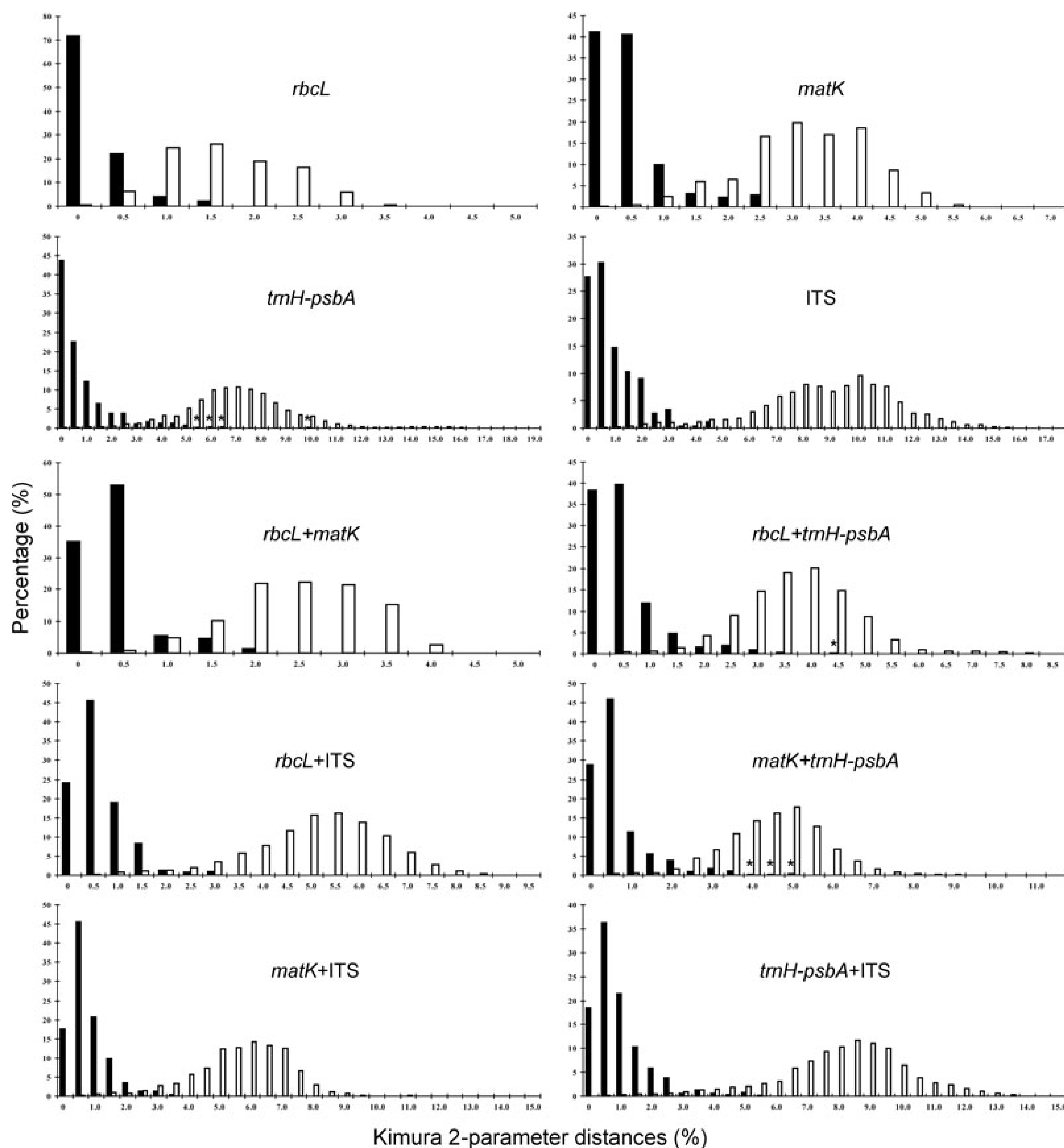

Fig. 3. Frequency distributions of intraspecific distances (black) and interspecific distances (white) that fell into 0.5% Kimura 2-parameter distance distance categories for 10 barcode candidates. *Intraspecific distances well within the range of interspecific distances in histograms of *trnH-psbA*, *rbcL* + *trnH-psbA*, and *matK* + *trnH-psbA*.

editing of basecalls. Performance at species discrimination was the highest with ITS. All things considered, it seems that ITS is the most effective single-region DNA barcode in *Pedicularis*, for which it has been used for phylogenetic studies and appears to be exceptionally variable (Yang et al., 2003b; Ree, 2005). Its adoption could benefit these efforts by providing potentially useful data.

Of the six two-region combinations, species discrimination was high when ITS was included.

© 2011 Institute of Botany, Chinese Academy of Sciences

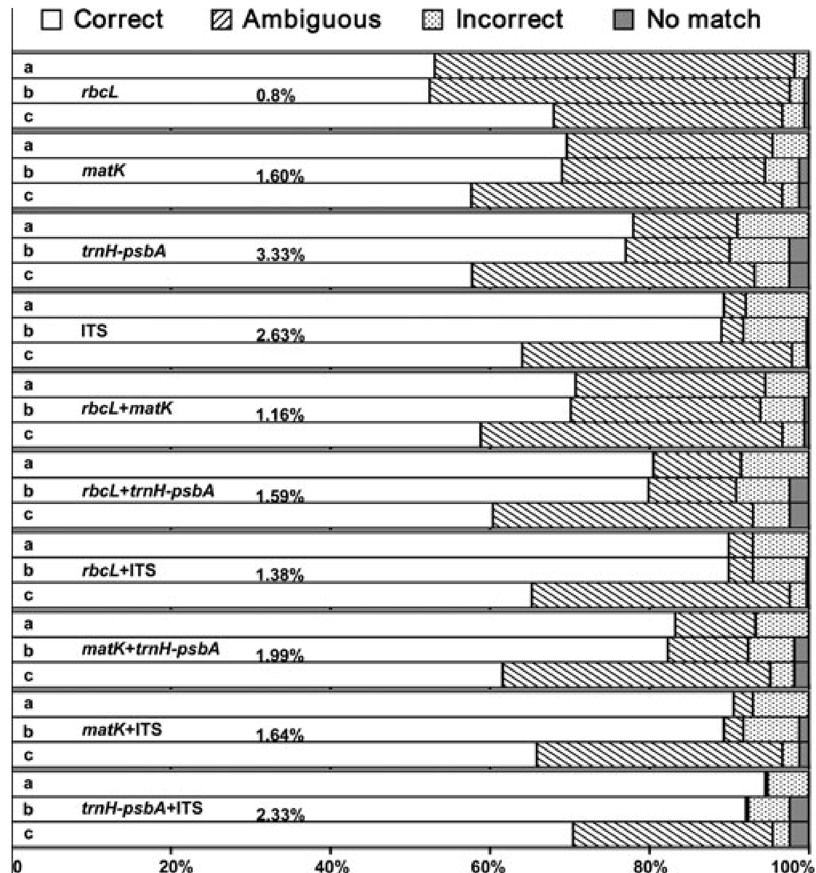

Fig. 4. Performance of 10 barcode candidates based on "best match" (a), "best close match" (b), and "all species barcode" (c). The distance threshold (showing on histograms) for "best close match" and "all species barcode" was calculated from the observed values of intraspecific distances for each barcode with 5% of the distances cut-off.

According to performance at species discrimination, *trnH-psbA* + ITS is the best among the two-region combinations. For *trnH-psbA*, however, frequent ambiguous base reads require manual editing, and its highly variable length hinders multiple alignment, and high intraspecific variation leads to misidentification. Moreover, multiple primer pairs are required for *matK*. It thus appears that *rbcL* + ITS is the best barcode when all criteria are considered. In *Pedicularis*, therefore, we suggest *rbcL* + ITS as a two-region barcode that represents both nuclear and plastid genomes. Generally speaking, plastid barcodes have received most attention in published reports (e.g., Chase et al., 2007; Kress & Erickson, 2007; Lee et al., 2007; Lahaye et al., 2008). In flowering plants, however, maternally inherited plastid loci may have higher overall rates of introgression compared to biparentally inherited nuclear loci, and for that reason may have lower accuracy at species discrimination (Petit & Excoffier, 2009). In the long term, multi-locus plant barcodes should be selected from the plastid and nuclear genomes (Chase et al., 2005).

3.2 Applications of DNA barcoding

DNA barcoding represents an effort to apply molecular data to taxonomic questions (e.g., see Tautz et al., 2003). To what extent can we expect it to help in taxonomic studies of *Pedicularis*? In this study, we have shown that clustering sequences by neighbor joining can yield moderate correspondence of monophyletic groups with morphological species determinations – the combination of *rbcL* + ITS discriminated 79.5% of species, and distance method correctly identified 89.9% of sequences in our sample. This suggests that the methods used here do reasonably well at matching barcode sequences to species.

The need for molecular tools for taxonomy is particularly apparent in the context of species identification and discovery. Like many groups of plants, specimens of *Pedicularis* can be difficult to identify to species by morphology alone, especially if flowers are absent or poorly preserved, as might be commonly the case with samples of medicinal plants. More than 60 species of *Pedicularis* have been recorded as herbal plants in traditional

© 2011 Institute of Botany, Chinese Academy of Sciences

Chinese medicine (Wu et al., 1990; Wu, 1993; Zhang et al., 1994; Xie et al., 1996). For example, in western China, *P. decora* Franch., *P. davidii* Franch., and *P. dunniana* Bonati are called "Taibai ginseng" (Xie et al., 1996). The present study includes 35 species recognized as having medicinal value (see Table 1), of which 71% were discriminated using either ITS or *rbcL* + ITS.

In the Hengduan Mountains region, where over 300 species are endemic, many new taxa have been described in the last 10 years (e.g., Yang et al., 2003a; Yu et al., 2008; Yu et al., 2010), and surely more await discovery. In this study, barcode data revealed the affinities of two new species not yet described, and suggest the existence of cryptic lineages in *P. anas* Maxim., *P. rupicola*, *P. cymbalaria* Bonati, and *P. siphonantha*. This suggests that barcoding can contribute useful data toward the discovery and description of new species.

In this study, we observed only moderate performance of DNA barcodes at discriminating species of *Pedicularis* from the Hengduan Mountains region. This is an expected result if our taxon sample contained lineages in various stages of divergence, as is likely to be the case if this biodiversity hotspot is the site of ongoing speciation. In this context, failure to discriminate species by DNA barcodes may highlight high-priority cases for further investigation using population genetic methods. For example, we are currently studying the *P. rex* C. B. Clarke ex Maxim. and *P. thamnophila* (Hand.-Mazz.) H. L. Li species group using plastid DNA locus and simple sequence repeat nuclear markers (Yu, 2011); others include ser. *Lyratae* Maxim. and the *P. cephalantha* Franch. ex Maxim. and *P. fengii* H. L. Li species groups. Detailed studies at this level promise to provide insights into the evolutionary dynamics of species divergence at shallow timescales in this biodiversity hotspot.

Acknowledgements This study is supported by grants from the Large-Scale Scientific Facilities of the Chinese Academy of Sciences (Grant No. 2009-LSF-GBOWS-01), the National Basic Research Program of China (973 Program, Grant No. 2007CB411600), and the National Natural Science Foundation of China (Grant Nos. 30970201, 30570115). We are grateful to Lian-Ming GAO, Hua-Jie HE, Jun HE, Wei JIANG, Jie LIU, Lu LU, Yang LUO, Chun-Lei XIANG, Ji-Dong YA, Qiu-Lin YANG, and Shu-Dong ZHANG for their kind help in the field or providing samples for this study; to Jun-Bo YANG, Jing YANG, Chun-Xia ZENG, and Yu-Xiao ZHANG for their assistance in molecular experiments and data analysis; and to Deren EATON for reading an earlier version of this manuscript.

© 2011 Institute of Botany, Chinese Academy of Sciences

References

Blaxter ML. 2004. The promise of a DNA taxonomy. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 359: 669–679.

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

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

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 B-Biological Sciences 360: 1889–1895.

Chen S-L, Yao H, Han J-P, Liu C, Song J-Y, Shi L-C, Zhu Y-J, Ma X-Y, Gao T, Pang X-H, Luo K, Li Y, Li X-W, Jia X-C, Lin Y-L, Leon C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. Plos One 5: e8613.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry 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.

Gao T, Yao H, Song J-Y, Zhu Y-Y, Liu C, Chen S-L. 2010. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. BMC Evolutionary Biology 10: 324.

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 Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

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

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series 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 of London Series B-Biological Sciences 270: S96–S99.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33: 511–518.

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.

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.
Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences USA 105: 2923–2928.
Le Clerc-Blain J, 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.
Lee H-L, Yi D-K, Kim J-S, Kim K-J. 2007. Development of plant DNA barcoding markers from the variable noncoding regions of chloroplast genome. In: Schindel DE, Shao K-T, Miller SE eds. Conference abstracts: Second International Barcode of Life conference. 18–20 September, 2007. Taipei: Academia Sinica. 112.
Li D-Z, Liu J-Q, Chen Z-D, Wang H, Ge X-J, Zhou S-L, Gao L-M, Fu C-X, Chen S-L. 2011. Plant DNA barcoding in China. Journal of Systematics and Evolution 49: 165–168.
Li H-L. 1948. A revision of the genus *Pedicularis* in China I. Proceedings of the Academy of Natural Sciences of Philadelphia 100: 205–378.
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.
Luo K, Chen S-L, Chen K-L, Song J-Y, Yao H, Ma X-Y, Zhu Y-J, Pang X-H, Yu H, Li X-W, Liu Z. 2010. Assessment of candidate plant DNA barcodes using the Rutaceae family. Science China Life Sciences 53: 701–708.
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.
Mill RR. 2001. Notes relating to the flora of Bhutan: XLIII. Scrophulariaceae (Pedicularis). Edinburgh Journal of Botany 58: 57–98.
Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. Trends in Ecology and Evolution 24: 386–393.
Ree RH. 2005. Phylogeny and the evolution of floral diversity in *Pedicularis* (Orobanchaceae). International Journal of Plant Sciences 166: 595–613.
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.
Ross HA, Murugan S, Li WLS. 2008. Testing the reliability of genetic methods of species identification via simulation. Systematic Biology 57: 216–230.
Seberg O, Petersen G. 2009. How many loci does it take to DNA barcode a Crocus? PLoS One 4: e4598.
Song J-Y, Yao H, Li Y, Li X-W, Lin Y-L, Liu C, Han J-P, Xie C-X, Chen S-L. 2009. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. Journal of Ethnopharmacology 124: 434–439.
Starr JR, Naczi RFC, Chouinard BN. 2009. Plant DNA barcodes and species resolution in sedges (Carex, Cyperaceae). Molecular Ecology Resources 9: 151–163.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596–1599.
Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. 2003. A plea for DNA taxonomy. Trends in Ecology and Evolution 18: 70–74.
Tsoong P-C. 1963. Scrophulariaceae (Pars II). In: Flora Reipublicae Popularis Sinacae. Beijing: Science Press. 68: 1–378.
Wang H, Yu W-B, Chen J-Q, Blackmore S. 2009. Pollen morphology in relation to floral types and pollination syndromes in *Pedicularis* (Orobanchaceae). Plant Systematics and Evolution 277: 153–162.
Whitlock BA, Hale AM, Groff PA. 2010. Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. PLoS One 5: e11533.
Wu Z-Y ed. 1993. Checklist of traditional Chinese medicine resources in Yunnan (transliteration from Chinese). Beijing: Science Press.
Wu Z-Y, Zhou T-Y, Xia P-G eds. 1990. Synopsis of Chinese medicinal plants, Vol. III (transliteration from Chinese). Shanghai: Shanghai Scientific and Technical Publishers.
Xie Z-W, Fan C-S, Zhu Z-Y eds. 1996. Compilation of Chinese herbal medicine, Vol. II (transliteration from Chinese). Beijing: People's Medical Publishing House.
Yang F-S, Hong D-Y, Wang X-Q. 2003a. A new species and a new specific synonym of *Pedicularis* (Scrophulariaceae) from the Hengduan Mountains, China. Novon 13: 363–367.
Yang F-S, Wang X-Q. 2007. Extensive length variation in the cpDNA *trnT-trnF* region of hemiparasitic *Pedicularis* and its phylogenetic implications. Plant Systematics and Evolution 264: 251–264.
Yang F-S, Wang X-Q, Hong D-Y. 2003b. Unexpected high divergence in nrDNA ITS and extensive parallelism in floral morphology of *Pedicularis* (Orobanchaceae). Plant Systematics and Evolution 240: 91–105.
Yang H-B, Holmgren NH, Mill RR. 1998. *Pedicularis* Linn. In: Wu Z-Y, Raven PH eds. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 16: 97–209.
Yao H, Song J-Y, Liu C, Luo K, Han J-P, Li Y, Pang X-H, Xu H-X, Zhu Y-J, Xiao P-G, Chen S-L. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS One 5: e13102.
Yu W-B. 2011. Species divergence and molecular identification of *Pedicularis* L. (Orobanchaceae) in the Himalaya–Hengduan Mountains region. Ph.D. Dissertation. Kunming: Kunming Institute of Botany, Chinese Academy of Sciences.
Yu W-B, Huang P-H, Li D-Z, Wang H. 2010. A new species of *Pedicularis* (Orobanchaceae) from the Hengduan Mountains, southwest China. Novon 20: 512–518.
Yu W-B, Zhang S-D, Wang H. 2008. New taxa of *Pedicularis* (Scrophulariaceae) from the Hengduan Mountains, southwestern China. Novon 18: 125–129.
Zhang H-K, Qin S-F, Li J, Chen W-D, Zhang H-Y eds. 1994. Compendium of Chinese medicinal material resources (transliteration from Chinese). Beijing: Science Press.
Zhang S-D, Wang H, Mill RR. 2006. A new species of *Pedicularis* (Scrophulariaceae) from the Yaoshan Mountain, Yunnan, China. Novon 16: 286–290.

© 2011 Institute of Botany, Chinese Academy of Sciences

Supporting Information

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

Appendix S1. Voucher information and GenBank accession numbers, indicating taxonomic position in accordance with Tsoong (1963) and Yang et al. (1998).

Appendix S2. Polymerase chain reaction primers used for amplification and sequencing in this study.

Appendix S3. Species not discriminated using internal transcribed spacer (ITS), *rbcL* + ITS (LI), *matK* + ITS (KI), or *trnH-psbA* + ITS (HI). Numbers in the cells represent bootstrap values (BV) of species groups. dis., Discriminated species (BV $\geq$ 50%, see Table 1).

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.

© 2011 Institute of Botany, Chinese Academy of Sciences