MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) **12**, 267–275 doi: 10.1111/j.1755-0998.2011.03095.x

Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae)

JUN-BO YANG,*¶ YI-PING WANG,†¶ MICHAEL MÖLLER,‡ LIAN-MING GAO* and DING WU†§
*Key Laboratory of Biodiversity and Biogeography, and Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650204, China, †Jiangxi Agricultural University, Nanchang 330045, China, ‡Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, Scotland, UK, §Jingdezhen College, Jingdezhen 333000, China

Abstract

DNA barcoding is a technique to identify species by using standardized DNA sequences. In this study, a total of 105 samples, representing 30 *Parnassia* species, were collected to test the effectiveness of four proposed DNA barcodes (*rbc*L, *mat*K, *trn*H-*psb*A and ITS) for species identification. Our results demonstrated that all four candidate DNA markers have a maximum level of primer universality and sequencing success. As a single DNA marker, the ITS region provided the highest species resolution with 86.7%, followed by *trn*H-*psb*A with 73.3%. The combination of the core barcode regions, *mat*K+*rbc*L, gave the lowest species identification success (63.3%) among any combination of multiple markers and was found unsuitable as DNA barcode for *Parnassia*. The combination of ITS+*trn*H-*psb*A achieved the highest species discrimination with 90.0% resolution (27 of 30 sampled species), equal to the four-marker combination and higher than any two or three marker combination including *rbc*L or *mat*K. Therefore, *mat*K and *rbc*L should not be used as DNA barcodes for the species identification of *Parnassia*. Based on the overall performance, the combination of ITS+*trn*H-*psb*A is proposed as the most suitable DNA barcode for identifying *Parnassia* species. DNA barcoding is a useful technique and provides a reliable and effective mean for the discrimination of *Parnassia* species, and in combination with morphology-based taxonomy, will be a robust approach for tackling taxonomically complex groups. In the light of our findings, we found among the three species not identified a possible cryptic speciation event in *Parnassia*.

Keywords: cryptic species, DNA barcoding, *Parnassia*, species identification, taxonomy

Received 15 July 2011; revision received 18 October 2011; accepted 25 October 2011

Introduction

Parnassia L. (Parnassiaceae), a genus of perennial herbs, is predominantly distributed in arctic and temperate zones of the Northern Hemisphere (Phillips 1982; Hultgård 1987; Gu & Hultgård 2001; Simmons 2004). Specimens of this genus are easily recognized by their basal leaf rosettes, generally long peduncles, single-flowed inflorescences, pentamerous flowers, commissural stigmas and the presence of a staminodium opposite each petal (Gu & Hultgård 2001). *Parnassia* includes about 50–70 species, depending on the taxonomy followed (Phillips 1982; Gu & Hultgård 2001; Wu *et al.* 2003; Simmons 2004). *Parnassia palustris* L. is the most widely distributed species of the genus, with populations in North America, Europe and Asia, reaching Morocco in the south (Korta 1972). However, most of the species have a limited distribution range, and more than 30 species are confined to China and the Himalayan region (Hultgård 1987). The East Himalaya–Hengduan Mountain region is one important centre of diversification and high endemism of *Parnassia* (Handel-Mazzetti 1941; Phillips 1982; Ku 1987; Wu *et al.* 2003; Simmons 2004).

The taxonomic status of *Parnassia* and its infrageneric classification have been widely discussed (Drude 1875; Franchet 1897; Engler 1930; Handel-Mazzetti 1941; Phillips 1982; Ku 1987; Wu *et al.* 2003). However, taxonomic treatments of *Parnassia* species are very controversial because of their often restricted geographic distribution, and the treatment of species in regional floras (e.g., Hooker & Thomson 1858; Handel-Mazzetti 1941; Wien 1966; Nasir & Ali 1972; Phillips 1982; Grierson 1987; Ku 1995). The complexity of the taxonomy and the limited morphological variation of this genus make species identification difficult, especially for some closely related species, such as *Parnassia wightiana*, *P. delavayi* and *P. mysorensis*.

Correspondence: Lian-Ming Gao, Fax: 86-871-5217791;
E-mail: gaolm@mail.kib.ac.cn
Ding Wu, Fax: 86-798-8386194; E-mail: parnassia@sohu.com
¶These authors contributed equally to this work.

DNA barcoding is a new biological tool to achieve accurate, rapid and automatable species identification without morphological knowledge, by using short DNA regions (Hebert *et al.* 2003; Savolainen *et al.* 2005). Combining DNA sequences with existing morphological characters can fasten the rate of identification and classification of species (Smith *et al.* 2005; Will *et al.* 2005; DeSalle 2006; Hajibabaei *et al.* 2007). A two-marker combination of *rbcL*+*mat*K was recommended for barcoding land plants by the Consortium for the DNA Barcode of Life (CBOL) Plant Working Group (2009). The nuclear ribosomal internal transcribed spacer (ITS) and plastid *trn*H-*psb*A region were suggested to be tested as complementary plant barcoding regions at the Third International Barcoding of Life Conference (Hollingsworth *et al.* 2011).

Here, we selected four candidate barcodes (i.e. *mat*K, *rbc*L, *trn*H-*psb*A and ITS) for 30 *Parnassia* species. Our aims were (i) to examine the effectiveness of these four regions as barcodes for *Parnassia*; (ii) to evaluate the congruence of traditional taxonomic treatments for some closely related species based on morphological data, with DNA barcoding results.

Materials and methods

Plant materials

A total of 105 samples were collected in this study, representing 30 species of *Parnassia*. There were at least two individuals sampled for each species from different populations, and more individuals for widespread species to fully represent their distribution ranges (Table S1, Supporting information). All corresponding vouchers were deposited in the Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences (KUN). Materials of nine specimens of five species occurring in North America and Central Asia were sampled from the Herbarium of the Royal Botanic Garden Edinburgh (E). Species concepts followed Ku (1987) and Phillips (1982). *Lepuropetalon spathulatum* was selected as outgroup for tree-based analysis (Zhang & Simmons 2006).

DNA extraction, PCR amplification and sequencing

Total DNA was extracted from silica-gel dried-leaf material and herbarium specimens using the CTAB procedure (Doyle & Doyle 1987). Polymerase chain reaction (PCR) amplifications were performed in a 20-μL reaction mixture containing 1× *Taq* buffer [50 mM $(NH4)_2SO_4$; 75 mM Tris–HCl (pH 8.3); 50 mM KCl; 0.001% gelatin]; 2.5 mM $MgCl_2$, 0.4 mM of dNTPs, 0.5 μM of each primer, 1.0 U of *Taq* DNA Polymerase (TaKaRa Biotechnology Co. Ltd., Dalian, China), and 1 μL of genomic DNA (25–30 ng). For *matK*, the addition of 4% DMSO and 0.2 μL 0.1 mg/mL BSA was used for PCR amplification. The primer information and thermocycling conditions for the four markers used in this study are listed in Table 1. Purified PCR products were sequenced in both directions with the PCR primers on an ABI 3730 DNA Sequencer (Applied Biosystems). The newly acquired DNA sequences have been deposited in GenBank and their accession numbers provided in Table S1 (Supporting information).

Data analysis

Sequences for each region were aligned with CLUSTAL X (Thompson *et al.* 1997) and then manually adjusted in BIOEDT v.7 (Hall 1999). The genetic pairwise distance for each marker was calculated using MEGA4 and the Kimura two-parameter (K2P) distance model (Tamura *et al.* 2007). Additionally, inter- and intra-specific genetic divergences of the four candidate DNA regions were analysed by Wilcoxon signed-rank tests (Meyer & Paulay 2005). To test whether accurate species assignments can be made among the samples on single marker and combinations of markers, we used the 'best match' and 'best close match' functions of the program TaxonDNA (Meier *et al.* 2006). Furthermore, to evaluate whether individual samples of a species clustered in species-specific monophyletic clades, neighbour-joining (NJ) trees were constructed based on single marker and all possible combinations of the four candidate markers in MEGA4, with pairwise deletion based on the K2P distance model. Bootstrap values (BP) were calculated over 10 000 replications.

Table 1 Primers and reaction condition used in this study

DNA region	Primer pairs	Primer sequences (5′–3′)	Thermocycling conditions
rbcL	rbcLa_f	ATGTCACCACAAACAGAGACTAAAGC	95 °C 4 min; [35 cycles: 94 °C 50 s; 52 °C 1 min; 72 °C 80 s]; 72 °C 10 min
	724R	TCGCATGTACCTGCAGTAGC	
matK	Xf	TAATTTACGATCAATTCATTC	95 °C 4 min; [35 cycles: 94 °C 50 s; 52 °C 1 min; 72 °C 80 s]; 72 °C 10 min
	5r	GTTCTAGCACAAGAAAGTCG	
trnH-psbA	*trnH*	ACTGCCTTGATCCACTTGGC	95 °C 4 min; [35 cycles: 94 °C 30 s; 55 °C 45 s; 72 °C 1 min]; 72 °C 10 min
	psbA	CGAAGCTCCATCTACAAATGG	
ITS	ITS5	GGAAGTAAAAGTCGTAACAAGG	95 °C 4 min; [35 cycles: 94 °C 50 s; 55 °C 1 min; 72 °C 80 s]; 72 °C 10 min
	ITS4	TCCTCCGCTTATTGATATGC	

Results

Variation of barcoding markers

All *Parnassia* samples were successfully amplified and sequenced using universal primer pairs for the four DNA barcoding regions (Table 1). A total number of 420 sequences were obtained from the 30 sampled *Parnassia* species. The variability of the four DNA markers for all examined samples is summarized in Table 2. ITS showed the highest interspecific sequence divergence (9.18%), followed by *trn*H-*psb*A (5.84%) and *mat*K (1.88%). *rbc*L had the lowest interspecific (0.74%) and intraspecific divergence (0.08%). The highest intraspecific sequence distance was recorded for ITS (0.71%), followed by *trn*H-*psb*A (0.53%). ITS and *trn*H-*psb*A showed a much number of variable sites and length variation than *rbc*L and *mat*K. *rbc*L was most highly conserved with fewest variable sites and lacking indels (Table 2).

Assessment for barcoding gaps

We analysed the interspecific and intraspecific genetic divergences of the four DNA regions with Wilcoxon signed-rank tests. In these, ITS exhibited the highest divergence and *rbc*L showed the lowest at the interspecific level (Table 3). At the intraspecific level, the lowest divergence was recorded for *rbc*L, while there were no significant differences of intraspecific sequence divergences between ITS and *trn*H-*psb*A (Table 4). This was also supported by similar mean intraspecific sequence distances for ITS and *trn*H-*psb*A, with 0.53% and 0.71%, respectively (Table 2).

We did not find any distinct barcoding gap in the distributions of divergences for any marker, especially in the core barcodes *rbc*L and *mat*K (Fig. 1). These results demonstrated that there was a larger range distribution of inter- and intra-specific distances for ITS and *trn*H-*psb*A than for *rbc*L or *mat*K (Tables 3 and 4), indicative of the higher sequence variation among individuals/species for ITS and *trn*H-*psb*A with higher species resolution.

Applicability for species discrimination

TaxonDNA and a tree-based method (NJ) were used to perform the species identification of *Parnassia* (Table 5, Fig. 2). Based on the 'best match' model, the species discriminatory power of the four DNA regions was 33.33% (*rbc*L), 63.80% (*mat*K), 94.28% (*trn*H-*psb*A) and 95.23% (ITS), respectively (Table 5). Species discrimination for combinations of two-marker ranged from 71.42% to 98.09%, with the core barcode *rbc*L+*mat*K providing the lowest species discrimination (71.42%). The remaining two-marker combinations showed relatively high levels of species identification (94.28–96.19%), with ITS+*trn*H-*psb*A being the combination with the highest success rate.

Table 2 The comparisons of variability of the four DNA markers

DNA region	*rbcL*	*matK*	*trnH-psbA*	ITS
Universality to primer	Yes	Yes	Yes	Yes
Percentage PCR success	100%	100%	100%	100%
Percentage sequencing success	100%	100%	100%	100%
Aligned sequence length (bp)	615	738	870	796
Indels (length, bp)	0	1 (6)	24 (1–347)	18 (1–11)
No. information sites/variable sites	28/33	105/123	235/406	330/356
Distribution of variable sites	Di & S	Di & S	I & D	I & D
No. sampled species (individuals)	30 (105)	30 (105)	30 (105)	30 (105)
Mean interspecific distance, %	0.74	1.88	5.84	9.18
Mean intraspecific distance, %	0.08	0.24	0.53	0.71

Di, dispersive; S, sparse; I, intensive; D, dense.

Table 3 Wilcoxon signed-rank tests of interspecific divergence among DNA markers

W+	W−	Relative ranks		*n*	*P* value	Result
*rbc*L	*mat*K	W+ = 754.50	W− = 94 075.50	435	$\leq 9.854 \times 10^{-71}$	*rbc*L ≪ *mat*K
*rbc*L	*trn*H-*psb*A	W+ = 0	W− = 94 830	435	$\leq 5.6 \times 10^{-73}$	*rbc*L ≪ *trn*H-*psb*A
*rbc*L	ITS	W+ = 0	W− = 94 830	435	$\leq 5.6 \times 10^{-73}$	*rbc*L ≪ ITS
*mat*K	*trn*H-*psb*A	W+ = 33.50	W− = 94 796.50	435	$\leq 7.057 \times 10^{-73}$	*mat*K ≪ *trn*H-*psb*A
*mat*K	ITS	W+ = 8	W− = 94 822	435	$\leq 5.918 \times 10^{-73}$	*mat*K ≪ ITS
*trn*H-*psb*A	ITS	W+ = 3870	W− = 90 091	435	1.821×10^{-61}	*trn*H-*psb*A < ITS

Table 4 Wilcoxon signed-rank tests of intraspecific divergences among DNA markers

W+	W−	Relative ranks		*n*	*P* value	Result
*rbc*L	*mat*K	W+ = 4	W− = 116	15	$\leq 4.273 \times 10^{-4}$	*rbc*L<*mat*K
*rbc*L	*trn*H-*psbA*	W+ = 14	W− = 157	18	$\leq 8.392 \times 10^{-4}$	*rbc*L<*trn*H-*psb*A
*rbc*L	ITS	W+ = 11	W− = 242	22	$\leq 1.889 \times 10^{-4}$	*rbc*L<ITS
*mat*K	*trn*H-*psb*A	W+ = 30	W− = 141	18	≤0.01387	*mat*K<*trn*H-*psb*A
*mat*K	ITS	W+ = 25	W− = 251	23	$\leq 6.224 \times 10^{-4}$	*mat*K<ITS
*trn*H-*psb*A	ITS	W+ = 107	W− = 193	24	≤0.2246	*trn*H-*psb*A=ITS

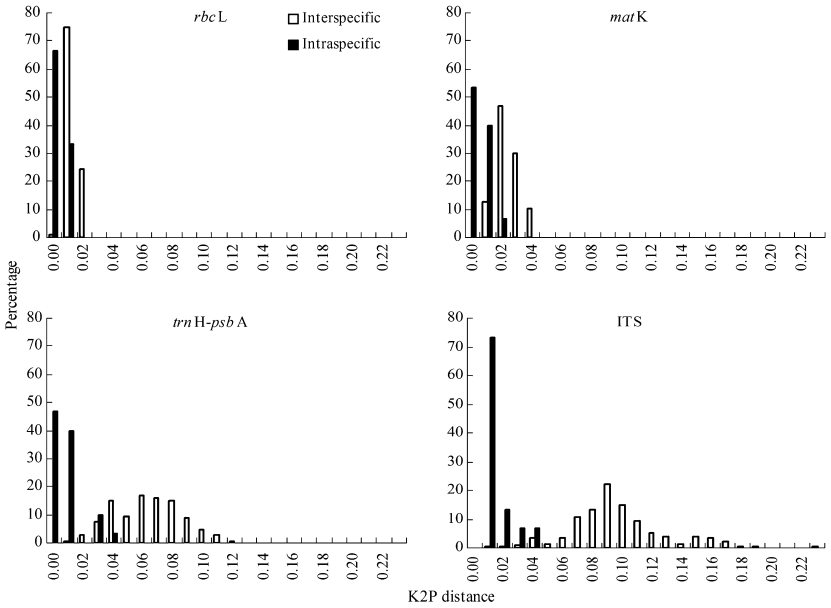

Fig. 1 Relative distribution of interspecific and intraspecific distances for the four DNA barcoding markers of *Parnassia*. *x*-axes relate to K2P distances arranged in intervals, and the y-axes correspond to the percentage of occurrences.

All two-marker combinations provided higher resolution than single markers, except for *rbc*L that showed little contribution to increase the discrimination rate. For the three-marker combinations, the core barcode (*rbc*L+*mat*K) plus ITS or *trn*H-*psb*A provided 98.09% and 96.16% species discrimination, respectively. Species discrimination of the four DNA markers combined was 98.09%, which was equal to the two-marker combination of ITS+*trn*H-*psb*A.

In the tree-based analysis, ITS provided the highest species discrimination (86.7%), followed by *trn*H-*psb*A (73.3%), *mat*K (63.3%) and *rbc*L (43.3%). ITS combined with *rbc*L, *mat*K, and *rbc*L+*mat*K provided the same ability for species discrimination as ITS alone (Tabel 6). A combination of ITS+*trn*H-*psb*A provided the highest species identification (90%) among all combinations (Fig. 2). Individuals for 27 of the 30 sampled *Parnassia* species formed monophyletic clade in the NJ tree, and most of the monophyletic species had high BP of over 90%. The samples of two species, *P. wightiana* and *P. yunnanensis*, did not fell in monophyletic clades. In addition, a monophyletic *P. viridiflora* fell among samples of *P. trinervis* (Fig. 2).

Discussion

Evaluation of the potential barcodes for Parnassia

An ideal DNA barcode should be routinely retrievable with a single primer pair, be amenable to bidirectional sequencing with little requirement for manual editing of sequence traces and provide maximal discrimination among species (CBOL Plant Working Group 2009). The single primer pairs for each of the four barcodes tested here performed without fail, as all samples of *Parnassia* were successfully amplified and sequenced (Table 2). This indicated a very high universality for all the four DNA regions used here.

Species discriminatory power is another important criterion for a DNA barcode (Hebert *et al*., 2004; Kress &

Table 5 Identification success based on the 'best match' and 'best close match' function of the program TaxonDNA and neighbour-joining phylogenetic analysis

	Best match			Best close match					
DNA region	Successfully identified (%)	Ambiguous (%)	Misidentified (%)	Successfully identified (%)	Ambiguous (%)	Misidentified (%)	No match (%)	Threshold (%)	Percentage of species monophyly (%)
*rbc*L	33.33	64.76	1.90	33.33	64.76	1.90	0.00	0.65	43.3
*mat*K	63.80	28.57	7.61	62.85	28.57	7.61	0.95	1.62	63.3
*trn*H-*psb*A	94.28	0.00	5.71	94.28	0.00	5.71	0.00	23.90	73.3
ITS	95.23	2.85	1.90	95.23	2.85	1.90	0.00	6.03	86.7
*rbc*L+*mat*K	71.42	21.90	6.66	70.47	21.906	6.66	0.95	1.10	63.3
*rbc*L+ITS	95.23	2.85	1.90	95.23	2.85	1.90	0.00	3.54	86.7
*rbc*L+*trn*H-*psb*A	94.28	0.00	5.71	94.28	0.00	5.71	0.00	14.14	70.0
*mat*K+ITS	95.23	3.80	0.95	95.23	3.80	0.95	0.00	3.58	86.7
*mat*K+*trn*H-*psb*A	96.19	0.00	3.80	96.19	0.00	3.80	0.00	13.74	80.0
*trn*H-*psb*A+ITS	98.09	0.00	1.90	98.09	0.00	1.90	0.00	14.70	90.0
*rbc*L+*mat*K+ITS	96.19	2.85	0.95	96.19	2.85	0.95	0.00	2.65	86.7
*rbc*L+*mat*K +*trn*H-*psb*A	96.19	0.00	3.80	96.19	0.00	3.80	0.00	10.07	76.7
*rbc*L+*mat*K+*trn*H -*psb*A+ITS	98.09	0.00	1.90	98.09	0.00	1.90	0.00	8.38	90.0

Erickson, 2007; CBOL Plant Working Group 2009). An ideal DNA barcode should provide high ability of species discrimination and identification (Kress *et al.* 2005; Lahaye *et al.* 2008; Hollingsworth *et al.* 2009) and exhibit a 'barcode gap' between intraspecific divergence and interspecific divergence (Meyer & Paulay 2005). Here, we found no distinct barcoding gap for *rbc*L and *mat*K in *Parnassia*. However, ITS and *trn*H-*psb*A exhibited a relatively well-defined gap between intraspecific and interspecific divergences (Fig. 1). At the single DNA barcode level, the ITS region showed the highest genetic distance between and within species (9.18% and 0.71%, respectively). It also provided the highest species resolution in this study, with a total of 26 of the 30 sampled species (86.7%) successfully identified. A high level of species discrimination of ITS was also reported in other groups (Ren *et al.* 2010; Liu *et al.* 2011). ITS was initially proposed as DNA barcode for plants because of its high sequence divergence (Kress *et al.* 2005). However, it was considered too problematic and was rejected from incorporation into the core plant barcode because of the potential pitfalls of incomplete lineage sorting and inhomogeneous concerted evolution, divergent paralogous copies within individuals and pseudogenes (Alvarez & Wendel 2003; Chase *et al.* 2007; Starr *et al.* 2009; Hollingsworth *et al.* 2011). Though, in a recent study on a large data set, it was found that ITS was unproblematic in the majority of samples analysed and proposed to be incorporated into the core barcode for seed plants (China Plant BOL Group 2011).

In this study, these potential drawbacks of ITS were not found in *Parnassia*. Based on the high level of universality in PCR and sequencing aspects, and high species discrimination ability, ITS performed well as a single barcode for species identification in *Parnassia*. *trn*H-*psb*A is one of the more variable plastid spacer regions that have been widely used as plant barcodes, which distinguished the largest number of plant species for barcoding purposes (Kress *et al.* 2005; Kress & Erickson 2007; Fazekas *et al.* 2008). In this study, *trn*H-*psb*A distinguished 73.3% of the sampled species and thus showed a high potential as barcode for *Parnassia*. Of the two coding DNA regions, *rbc*L performed poorest in species discrimination with less than half the species (43.3%) successfully identified. Consequently, *rbc*L is not a good choice for barcoding *Parnassia* species. *rbc*L has high universality and sequence quality, but relatively low discriminating levels in angiosperms (CBOL Plant Working Group 2009). Though, it shows high discrimination power at family and genus level (Kress & Erickson 2007). With 19 of 30 (63.3%) sampled species of *Parnassia* distinguished by *mat*K in our study, this region performed lower than ITS or *trn*H-*psb*A. As a single barcode, *mat*K is thus not a good DNA marker among the four candidate DNA barcodes tested for *Parnassia*.

Combinations of barcoding markers often perform better than single candidate DNA barcodes, and several combinations of markers were proposed in the past (see Pennisi 2007; Hollingsworth 2008). A combination of *mat*K+*rbc*L has been officially proposed as the core barcode for land plants (CBOL Plant Working Group 2009). However, this combination only provided 63.3% species identification (19 of 30 species), the lowest resolution

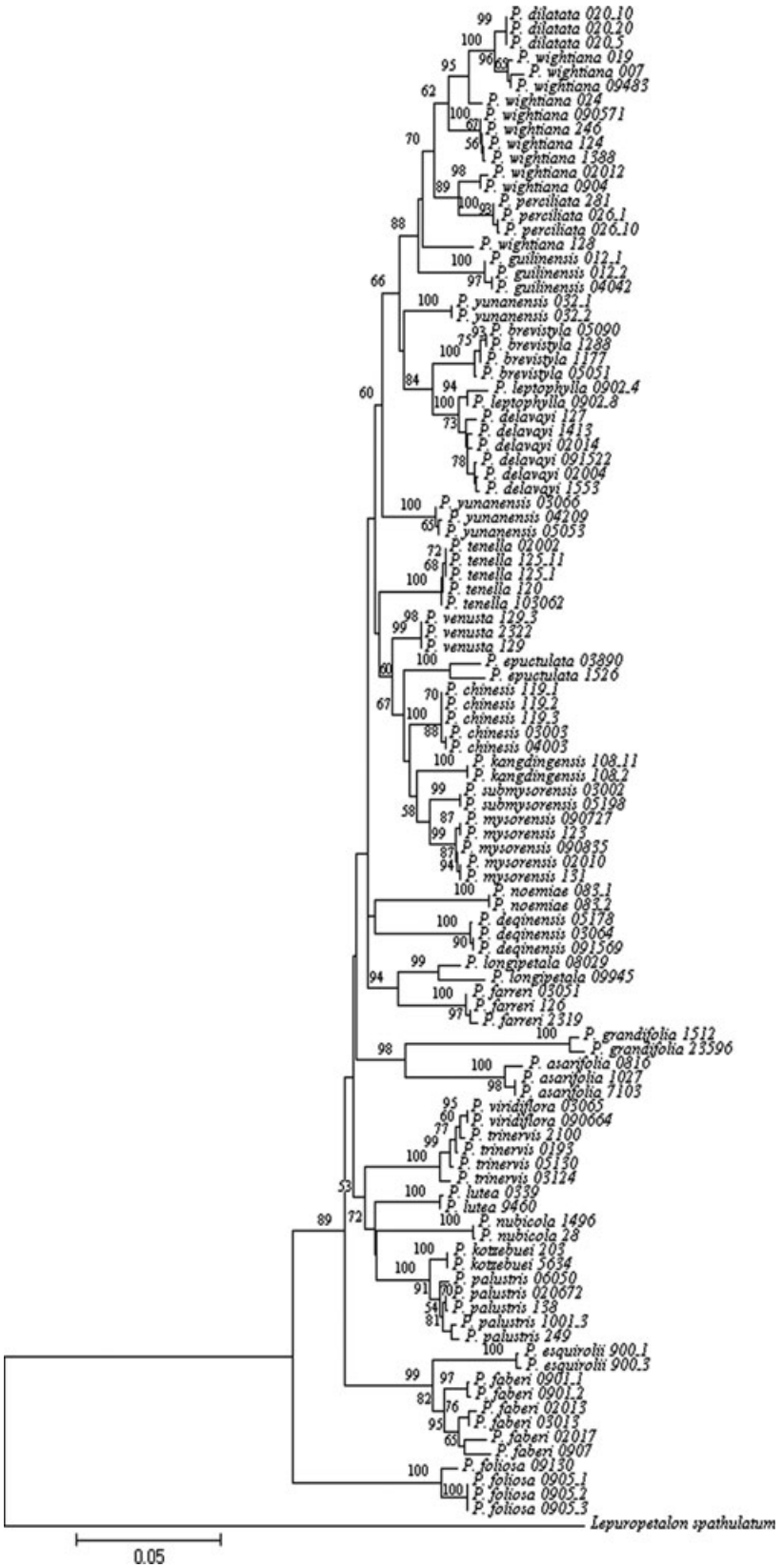

Fig. 2 Neighbour-joining tree based on the combination of ITS+*trn*H-*psb*A sequences with the Kimura 2-parameter distance model. Bootstrap values (>50%) are shown above the relevant branches.

among all combinations in our study. This result indicated that the core barcode proposed by the CBOL Plant working Group is not the best choice for barcoding *Parnassia* species and needs to be supplemented with additional markers. Combinations of the two core barcodes plus ITS (*rbc*L+*mat*K+ITS) and *trn*H-*psb*A (*rbc*L+*mat*K+*trn*H-*psb*A), respectively, did not increase the species resolution significantly (Table 5). The four-marker combination showed the highest species resolution (90.0%), equalling the two-marker combination of ITS+*trn*H-*psb*A (Table 5).

Considering the costs and the ultimate goal of high throughput species discrimination, the two-marker combination of ITS+*trn*H-*psb*A is the best choice as a barcode for species identification in *Parnassia*.

Taxonomy and species identification

Traditionally, biological specimens have been identified using morphological features, such as habit, leaf and flower shape, size and colour, etc. In morphologically and *ergo* taxonomically difficult groups, it sometime requires an experienced professional taxonomist years to deal with species identification (Li *et al.* 2011). Furthermore, if specimens are damaged or lack sufficient diagnostic characters, even the specialists may be unable to make identifications (Zhou *et al.* 2009). DNA barcoding as a new tool in the taxonomist's toolbox (http://www.barcodeoflife.org/), however, can only solve the latter problems. In our opinion, in cases of taxonomic uncertainty because of morphological complexity, however, DNA barcoding cannot substitute taxonomy. It can only be used to test a taxonomic scenario. *Parnassia* is a good test case because its taxonomy is difficult because of lack of species differentiating characters (Drude 1875; Franchet 1897; Engler 1930; Handel-Mazzetti 1941; Phillips 1982; Ku 1987; Wu *et al*. 2003). In barcoding terms, if all samples for a given species are assigned to a monophyletic clade, a species can be regarded as successfully identified (Hebert *et al*. 2004; Liu *et al.* 2011). This was the case for all but three of the 30 species of *Parnassia* included, based on the combination of just two markers, ITS+*trn*H-*psb*A (Table 5), suggesting that DNA barcoding can be an effective tool for the discriminating species of *Parnassia*.

Multiple individuals (2–11) per species were collected in this study, and the NJ tree illustrates the 27 species of *Parnassia* that were well identified, with often high clade support (>90% for 25 species) (Fig. 2). These species confirmed by DNA barcoding are characterized by reliable morphological features and geographic distributions. For example, *P. palustris* is widely distributed in the temperature zone of the northern hemisphere (Simmons 2004). This species is characterized by cauline sessile leaves, one leaf near the middle, usually with 3 to numerous basal leaves, ovate or long ovate; branched staminodes, divided into (7–) 9–21 filiform rays with globose glands at the apex. In this study, the five individuals of this species sampled, covering its entire distribution range (from North America, through Europe eastward to North Asia) clustered in a monophyletic clade in the NJ tree, even though they were collected from a very wide distribution range.

The three species included here, which did not form species-specific monophyletic clades, were *P. wightiana*, *P. yunnanensis* and *P. trinervis* (Fig. 2). The distribution of *P. wightiana* is centred in the Himalayan region, extending from southern India to western China and to central and southern China (Wu *et al.* 2003). This species showed a high level of intraspecific genetic variation. Among the 11 samples of *P. wightiana*, for instance, fell a clade of three samples of *P. dilatata* and a clade with the three samples of *P. perciliata*. There was a geographic aspect to the clustering, with four *P. wightiana* individuals from Guizhou-Guangxi-Hunan region forming a sister clade to *P. dilatata* from Guangxi, and two *P. wightiana* individuals collected from NE Yunnan-SW Sichuan clustered as sister to samples of *P. perciliata* collected from Chongqing (Fig. 2). The seeming polyphyly of *P. wightiana* might be explained by incomplete lineage sorting, hybridization or gene introgression between the species. Another, more likely explanation may be a case of polypatric speciation, with two geographically restricted species, *P. dilatata* in Guangxi and *P. perciliata* in Chongqing, having evolved independently from a widespread ancestor, *P. wightiana*. To fully understand this situation, however, further more detailed morphological and molecular studies on this species group are needed.

A similar situation was found for *P. trinervis* and *P. viridiflora*. Here, samples of *P. viridiflora* nested within the *P. trinervis* clade (Fig. 2). These two species are very similar in morphological features; *P. trinervis* and *P. viridiflora* are differentiated only by white vs. green petals, basal leaf blade usually cuneate at the base vs. usually subcordate at the base (Gu & Hultgård 2001). These are somewhat variable characteristics, and the two species could be recognized at intraspecific level. *Parnassia viridiflora* possibly derived from a *P. trinervis* population. Because of the small morphological differences between the two species, and before any taxonomic changes can be proposed, further population samples included for study to test the robustness of the molecular data are needed; the branches supporting the *P. trinervis* grades leading to *P. viridiflora* are not highly supported (60% and 77%, Fig. 2). If our data here are confirmed, it would be more reasonable to merge *P. viridiflora* into *P. trinervis* or to reduce it to a variety of *P. trinervis*.

A different situation was found for samples of *P. yunnanensis*; accessions of this species were divided

into two distinct clades, indicating two distinct lineages. The samples from Northwest Yunnan (Deqin and Shangri-la) and the samples collected from West Sichuan (Kangding) each formed clades with 100% BP. The samples cover local geographic ranges, and more samples from intervening population are needed to conclude whether this is a case of cryptic speciation.

A successful barcoding project requires comprehensive species sampling and should facilitate high rate of species identification. DNA barcoding can be used for species identification. It can provide a deeper understanding of biodiversity at large and species boundaries, the latter only in support, not in place of morphological taxonomic approaches. The inclusion of multiple samples from different population covering the entire distribution range for species is an important prerequisite for plant barcoding projects, especially for closely related species or species with a complex morphology. Here, we found that DNA barcoding is a useful technique and the combination of just two markers, ITS+*trn*H-*psb*A, provided a reliable and highly efficient and effective mean for the discrimination of *Parnassia* species.

Acknowledgements

This study was funded by the Large-scale Scientific Facilities of the Chinese Academy of Sciences (grant number: 2009-LSF-GBOWS-01), the National Science Foundation of China (30670144), the Chinese Academy of Sciences grant (grant number: KSCX2-YW-N-0807, KSCX2-EW-Z-2) and the talent project of the Yunnan province (grant number: 2008YP064). The Royal Botanic Garden Edinburgh is funded by the Rural and Environment Science and Analytical Services division (RESAS) in the Scottish Government. We are grateful to Dr. Liu Jin-Mei, Dr. Jin Xiao-Hua, Dr. Yue Ji-Pei, Dr. Zhang Shu-Dong, Dr. Chen Jia-Hui, Dr. Ren Zong-Xin, Dr. Li Rong, Dr. Pan Cheng, Mr. Cai Jie, Mr. Yu Wen-Bin and Mr. Zhang Zhi-Rong for their help in field and laboratory work, and discussion on the project, and/or provision of plant materials. We also thank the keepers of E, K and US for permission to access their collections. Thanks also to Dr. Jin Qiao-Jun from the Max-Planck-Institute of Terrestrial Microbiology, Germany, for revising an earlier version of this manuscript.

References

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

CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.

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

China Plant BOL Group (2011) Comparative analysis of a large dataset indicates that ITS should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 19641–19646.

DeSalle R (2006) Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology*, **20**, 1545–1547.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemical Bulletin*, **19**, 11–15.

Drude O (1875) Über die Blüthengestaltung und die Verwandtschaftsverhältnisse des Genus Parnassia, nebst einer systematischen Revision seiner Arten. *Linnaea*, **39**, 239–324.

Engler A (1930) Saxifragaceae. In: *Die Natürlichen Pflanzenfamillien* (eds Engler A, Prantl K), vol. 18(a), pp. 178–182. W Englmann, Leipzig.

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.

Franchet A (1897) Les Parnassia de l'Asie Orientale. *Bulletin de la Société royale de Botanique de Belgique*, **44**, 244–263.

Grierson AJC (1987) Parnassiaceae. In: *Flora of Bhutan* (eds Grierson AJC, Long DG), vol. 1, pp. 515–517. Royal Botanic Garden, Edinburgh.

Gu C, Hultgård UM (2001) *Parnassia*. In: *Flora of China* (eds Wu C-Y, Raven PH), vol. 8, pp. 95–118. Science Press, Beijing; Missouri Bot. Garden Press, St Louis, MO.

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.

Handel-Mazzetti H (1941) Die Chinesischen Parnassia-Arten. *Oesterreichische Botanische Zeitschrift*, **90**, 127–136.

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series B: Biological Science*, **270**, 313–321.

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

Hollingsworth PM (2008) DNA barcoding plants in biodiversity hot spots: Progress and outstanding questions. *Heredity*, **101**, 1–2.

Hollingsworth ML, Clark A, 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.

Hooker JD, Thomson T (1858) Praecursores ad Floram Indicam. *Botanical Journal of the Linnean Society*, **2**, 54.

Hultgård UM (1987) *Parnassia palustris* L. in Scandinavia. *Acat Universitatis Upsaliensis Symbolae Botanicae Upsalienses*, **28**, 1–128.

Korta J (1972) Anatomical analysis of *Parnassia palustris* L. *Acta Biologica Cracoviensia*, **15**, 31–37.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plant: the coding *rbcL* gene complements the noncoding *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 of the United States of America*, **102**, 8369–8374.

Ku TC (1987) A revision of the genus *Parnassia* (Saxifragaceae) in China. *Bulletin of Botanical Research*, **7**, 1–59.

Ku TC (1995) Parnassioideae Engl. In: *Flora Reipublica Popularis Sinicae* (eds Lu L, Hwang S), vol. 35, pp. 1–66. Science Press, Beijing.

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G (2008) DNA barcoding the floras of biodiversity hotspots. *Proceeding of the National Academy of Science of the United States of America*, **105**, 2923–2928.

Li DZ, Liu JQ, Chen ZD *et al.* (2011) Plant DNA barcoding in China. *Journal of Systematics and Evolution*, **49**, 165–168.

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

Meier RS, Kwong S, 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.

Nasir E, Ali SI (1972) *Parnassiaceae*. In: *Flora of West Pakistan* (ed. Stewaet RR), pp. 339–340. Department of Botany-Taxonomy, University of Karachi, Karachi.

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

Phillips RB (1982) *Systematics of Parnassia (Parnassiaceae): generic overview and revision of North American taxa*. PhD Thesis, University of California, Berkerley, CA.

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

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **360**, 1805–1811.

Simmons MP (2004) Parnassiaceae. In: *The Families and Genera of Vascular Plants* (ed. Kubitzki K), pp. 291–296. Springer Verlag, Berlin.

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

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

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.

Thompson JD, Gibson TJ, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.

Wien ES (1966) Parnassiaceae. In: *Flora Iranica* (ed. Rechinger KH), pp. 1–4. Lfg. Cont. 20.

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, **54**, 844–851.

Wu ZY, Lu AM, Tang YC, Chen ZD, Li DZ (2003) *The Families and Genera of Angiosperms in China, a Comprehensive Analysis*. Science Press, Beijing, pp. 649–654.

Zhang LB, Simmons MP (2006) Phylogeny and delimitation of the Celastrales inferred from nuclear and plastid genes. *Systematic Botany*, **31**, 122–137.

Zhou X, Adamowicz SJ, Jacobus LM, Dewalt RE, Hebert PDN (2009) Towards a comprehensive barcode library for arctic life-Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology*, **6**, 30.

Data accessibility

DNA sequences: GenBank accession numbers are given in Table S1 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Samples of Parnassia included in the present barcoding study, with collection and voucher information and GenBank accession numbers.

Please note: Wiley-Blackwell is 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.