Ann. Bot. Fennici 49: 319–330   ISSN 0003-3847 (print)   ISSN 1797-2442 (online)
Helsinki 30 November 2012   © Finnish Zoological and Botanical Publishing Board 2012

A preliminary assessment of *mat*K, *rbc*L and *trn*H–*psb*A as DNA barcodes for *Calamus* (Arecaceae) species in China with a note on ITS

Han-Qi Yang[1,*], Yu-Ran Dong[1], Zhi-Jia Gu[1], Ning Liang[1] & Jun-Bo Yang[2]

[1] *Research Institute of Resources Insects, Chinese Academy of Forestry, Kunming 650224, Yunnan, China (*corresponding author's e-mail: yanghanqikm@yahoo.com.cn)*
[2] *Germplasm Bank of Wild Species; Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, Yunnan, China*

Received 27 Mar. 2012, final version received 23 May 2012, accepted 29 May 2012

Yang, H. Q., Dong, Y. R., Gu, Z. J., Liang, N. & Yang, J. B. 2012: A preliminary assessment of *mat*K, *rbc*L and *trn*H–*psb*A as DNA barcodes for *Calamus* (Arecaceae) species in China with a note on ITS. — *Ann. Bot. Fennici* 49: 319–330.

Calamus is the largest genus in the palm family (Arecaceae) and contains many species of high ecological and economical value. In this study, we assessed the nuclear ribosomal internal transcribed spacer region (ITS), *mat*K, *rbc*L, *trn*H–*psb*A, as well as two combinations, *mat*K + *rbc*L and *mat*K + *rbc*L + *trn*H–*psb*A, as DNA barcodes for *Calamus* using 15 species or varieties distributed in China. ITS may exist as multiple copies in the examined *Calamus* species, and was eliminated from consideration as a possible barcode. The *trn*H–*psb*A spacer had the most variation, followed by *mat*K and *rbc*L. No separations between intraspecific variation and interspecific divergence (barcoding gaps) were found in the remaining candidate barcodes. At the species level, the discrimination rates of the candidate barcodes based on neighbor–joining (NJ) trees were significantly different: *mat*K (37.5%), *rbc*L (6.3%), *trn*H–*psb*A (56.3%), *mat*K + *rbc*L (43.8%) and *mat*K + *rbc*L + *trn*H–*psb*A (62.5%). Furthermore, the discrimination rates of *trn*H–*psb*A would improve to 91.7%, if the varieties of *C. nambariensis* and *C. yunnanensis* were treated as one species in the NJ tree. Thus, *trn*H–*psb*A may be an appropriate single DNA barcode for *Calamus* useful in the species identification.

Introduction

DNA barcodes generally refer to short DNA sequences, which can be used to rapidly and accurately identify species (Hebert *et al*. 2003). Besides species identification, DNA barcodes have also been deemed to improve or supplement traditional taxonomy based on morphological characters (Hebert & Gregory 2005). An ideal barcode must conform to at least three criteria: (1) universality (ease of amplification and sequencing), (2) sequence quality, and (3) discriminatory power (Hollingsworth *et al*. 2011).

The most successful DNA barcode so far is the mitochondrial gene cytochrome oxidase c subunit 1 (*COI*), which is widely used in animals (e.g. Hebert *et al*. 2004, Barrett & Hebert 2005). However, finding universal and consistent markers for land plants has proven difficult (Hollingsworth *et al*. 2011). As a result of a generally low nucleotide substitution rate, *COI* has low discriminatory power in plant taxa, and it is not suitable as a plant barcode (Cho *et al*. 2004, Fazekas *et al*. 2008). Many candidate plant barcodes, including the nuclear internal transcribed spacer (ITS) regions or ITS2, chloroplast intergenic spacers (e.g. *trn*H–*psb*A, *atp*F–*atp*H) and chloroplast coding regions (e.g. *mat*K, *rbc*L) have been proposed (e.g. Kress *et al*. 2005, Chase *et al*. 2007, Lahaye *et al*. 2008, Fazekas *et al*. 2008, CBOL Plant Working Group 2009). Many researchers have acknowledged that multiple markers would be required to obtain adequate species discrimination using plant DNA barcodes (Hollingsworth *et al*. 2011). Recently *mat*K, *rbc*L and the combination *mat*K + *rbc*L were suggested and employed as core plant barcodes (e.g., CBOL Plant Working Group 2009). Based on the assessment of effectiveness and universality of *mat*K, *rbc*L, *trn*H–*psb*A and ITS as barcode markers in seed plants of 141 genera from 75 families in China, China Plant BOL Group (2011) proposed ITS as one core barcode for seed plants.

Calamus is the largest genus in the palm family (Arecaceae), consisting of ca. 370 species distributed throughout the tropical and subtropical regions (Pei *et al*. 1991, Chen *et al*. 2003). Results of molecular phylogenetic analyses showed that *Calamus* was a paraphyletic genus; furthermore it and four other genera, i.e., *Daemonorops*, *Retispatha*, *Ceratolobus* and *Pogonotium*, formed a monophyletic group based on ITS and *rps*16 datasets (Baker *et al*. 2000a, 2000b). In Asia, China is the northern margin of the natural distribution of *Calamus*, and 37 species and 26 varieties are reported in the southwestern and southeastern China (Xing *et al*. 2006).

The canes of many species of *Calamus*, known as "rattan", are excellent materials for furniture. Due to overexploitation, the habitats and resources of *Calamus* in China have been dramatically reduced; therefore, it is important to conserve the species (Chen *et al*. 2003, Xing *et al*. 2006). The first step towards this goal is to distinguish the species. However, identification of *Calamus* species using morphological characters alone is, at least in China, difficult. DNA barcoding may be helpful in distinguishing these species. Several molecular phylogenetic studies based on or including plastid (including *mat*K, *rbc*L, *rps*16 and *trn*L–*trn*F) data have demonstrated low variation within the palm family (e.g. Baker *et al*. 2000a, 2000b, Asmussen *et al*. 2001, 2006), and the utility of plastid regions as DNA barcodes was thus thought to be low (e.g., Jeanson *et al*. 2011). In the palm family, *mat*K, *rbc*L, *trn*H–*psb*A and ITS2 barcode data have so far been reported for only 40 species from the tribe Caryoteae, which are distributed from mainland Asia to the western Pacific and Australia (Jeanson *et al*. 2011). In Caryoteae, these three plastid barcodes exhibited much lower species discrimination (26%–48%) than ITS2 (92%). To our knowledge, no DNA barcode data for *Calamus* (tribe Calameae) have been collected. In the present study, we assessed the utility of four frequently recommended DNA barcodes, i.e., *mat*K, *rbc*L, *trn*H–*psb*A and ITS, as well as two of their combinations *mat*K + *rbc*L and *mat*K + *rbc*L + *trn*H–*psb*A, for identifying 15 *Calamus* species and varieties collected in China, representing ca. 25% and 4% of *Calamus* diversity in China and the world, respectively.

Material and methods

Plant material

A total of 46 samples representing 15 *Calamus* species or varieties and *Plectocomia himalayana* were collected from Yunnan, China (Table 1). Because *Plectocomia* is closely related to *Calamus* (Baker *et al*. 2000a), three individuals of *P. himalayana* were used as an outgroup in the phylogenetic analyses. Two to five samples of each species were analyzed. The taxonomy of *Calamus* in this study follows Chen *et al*. (2003). Vouchers were deposited at the Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences (KUN). Young and healthy leaves were collected in the field, then immediately dried and stored in silica gel until DNA extraction.

DNA extraction, amplification and sequencing

Genomic DNA was extracted using the modi-

Table 1. Voucher information and GenBank accession numbers for the species and varieties examined in this study.

Taxon	Locality (all in Yunnan, China)	Latitude/longitude	Voucher	GenBank Accession no.		
				*mat*K	*rbc*L	*trn*H–*psb*A
Calamus bonianus	Menglun, Mengla	21°55´N/101°17´E	Yanghq0059	JQ042014	JQ042065	JQ042116
C. bonianus	Menglun, Mengla	21°55´N/101°17´E	Yanghq0060	JQ042015	JQ042066	JQ042117
C. bonianus	Menglun, Mengla	21°55´N/101°17´E	Yanghq0061	JQ042016	JQ042067	JQ042118
C. erectus	Nanbang, Yingjiang	24°42´N/97°34´E	Yanghq0024	JQ041983	JQ042034	JQ042085
C. erectus	Nanbang, Yingjiang	24°42´N/97°34´E	Yanghq0025	JQ041984	JQ042035	JQ042086
C. erectus	Nanbang, Yingjiang	24°42´N/97°34´E	Yanghq0026	JQ041985	JQ042036	JQ042087
C. gracilis	Nanbang, Yingjiang	24°44´N/97°34´E	Yanghq0021	JQ041980	JQ042031	JQ042082
C. gracilis	Nanbang, Yingjiang	24°44´N/97°34´E	Yanghq0022	JQ041981	JQ042032	JQ042083
C. guruba var. *elipsoideus*	Nanxi, Hekou	22°39´N/103°59E	Yanghq0001	JQ041966	JQ042017	JQ042068
C. guruba var. *elipsoideus*	Nanxi, Hekou	22°39´N/103°59E	Yanghq0002	JQ041967	JQ042018	JQ042069
C. guruba var. *elipsoideus*	Nanxi, Hekou	22°39´N/103°59E	Yanghq0003	JQ041968	JQ042019	JQ042070
C. henryanus	Mengla, Mengla	21°30´N/101°34´E	Yanghq0052	JQ042007	JQ042058	JQ042109
C. henryanus	Mengla, Mengla	21°30´N/101°34´E	Yanghq0053	JQ042008	JQ042059	JQ042110
C. karinensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0048	JQ042004	JQ042055	JQ042106
C. karinensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0049	JQ042005	JQ042056	JQ042107
C. karinensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0050	JQ042006	JQ042057	JQ042108
C. nambariensis var. *alpinus*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0017	JQ041977	JQ042028	JQ042079
C. nambariensis var. *alpinus*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0019	JQ041979	JQ042030	JQ042081
C. nambariensis var. *menglongensis*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0040	JQ041996	JQ042047	JQ042098
C. nambariensis var. *menglongensis*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0041	JQ041997	JQ042048	JQ042099
C. nambariensis var. *menglongensis*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0042	JQ041998	JQ042049	JQ042100
C. nambariensis var. *xishuangbannaensis*	Menglong, Jinghong	21°31´N/100°31´E	Yanghq0034	JQ041991	JQ042042	JQ042093
C. nambariensis var. *xishuangbannaensis*	Menglong, Jinghong	21°31´N/100°31´E	Yanghq0036	JQ041993	JQ042044	JQ042095
C. nambariensis var. *xishuangbannaensis*	Menglong, Jinghong	21°31´N/100°31´E	Yanghq0037	JQ041994	JQ042045	JQ042096
C. nambariensis var. *xishuangbannaensis*	Menglong, Jinghong	21°31´N/100°31´E	Yanghq0038	JQ041995	JQ042046	JQ042097
C. platyacanthus var. *longicarpus*	Nanxi, Hekou	22°39´N/103°58E	Yanghq0007	JQ041972	JQ042023	JQ042074
C. platyacanthus var. *longicarpus*	Nanxi, Hekou	22°39´N/103°58E	Yanghq0008	JQ041973	JQ042024	JQ042075
C. rhabdocladus	Nanxi, Hekou	22°39´N/103°57E	Yanghq0004	JQ041969	JQ042020	JQ042071
C. rhabdocladus	Nanxi, Hekou	22°39´N/103°57E	Yanghq0005	JQ041970	JQ042021	JQ042072
C. rhabdocladus	Nanxi, Hekou	22°39´N/103°57E	Yanghq0006	JQ041971	JQ042022	JQ042073
C. viminalis var. *fasciculatus*	Mengmian,Mengla	21°21´N/101°20´E	Yanghq0054	JQ042009	JQ042060	JQ042111
C. viminalis var. *fasciculatus*	Mengmian,Mengla	21°21´N/101°20´E	Yanghq0055	JQ042010	JQ042061	JQ042112
C. viminalis var. *fasciculatus*	Mengmian,Mengla	21°21´N/101°20´E	Yanghq0056	JQ042011	JQ042062	JQ042113
C. yunnanensis	Tongbiguan,Yingjiang	24°37´N/97°39´E	Yanghq0029	JQ041986	JQ042037	JQ042088

continued

Table 1. Continued.

Taxon	Locality (all in Yunnan, China)	Latitude/longitude	Voucher	GenBank Accession no.		
				*mat*K	*rbc*L	*trn*H–*psb*A
C. yunnanensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0030	JQ041987	JQ042038	JQ042089
C. yunnanensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0031	JQ041988	JQ042039	JQ042090
C. yunnanensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0032	JQ041989	JQ042040	JQ042091
C. yunnanensis	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0033	JQ041990	JQ042041	JQ042092
C. yunnanensis var. *densiflorus*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0046	JQ042002	JQ042053	JQ042104
C. yunnanensis var. *densiflorus*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0047	JQ042003	JQ042054	JQ042105
C. yunnanensis var. *intermedius*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0043	JQ041999	JQ042050	JQ042101
C. yunnanensis var. *intermedius*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0044	JQ042000	JQ042051	JQ042102
C. yunnanensis var. *intermedius*	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0045	JQ042001	JQ042052	JQ042103
Plectocomia himalayana	Mengjiao,Cangyuan	23°18´N/99°10´E	Yanghq0010	JQ041974	JQ042025	JQ042076
Pl. himalayana	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0011	JQ041975	JQ042026	JQ042077
Pl. himalayana	Menglong, Jinghong	21°31´N/100°30´E	Yanghq0012	JQ041976	JQ042027	JQ042078

fied CTAB method (Doyle & Doyle 1987). DNA was dissolved in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) to a final concentration of 30–60 ng l^{-1}. The PCR amplification was performed in a 25 μl reaction mixture containing 20 ng DNA, 10 mmol l^{-1} Tris-HCl (pH 8.3), 50 mmol l^{-1} KCl, 1.5 mmol l^{-1} $MgCl_2$, 200 μmol l^{-1} each dNTP, 0.4 μmol l^{-1} each primer, and 1 U Taq DNA polymerase (TaKaRa, Dalian, China). For the amplification and sequencing, we used the following primers suggested by the China Plant BOL Group (2011): ITS4 and ITS5 for ITS, including ITS1, 5.8S and ITS2 (White *et al*. 1990), 390F and 1326R for *mat*K (Cuénoud *et al*. 2002), 1F and 724R for *rbc*L (Fay *et al*. 1997), *trn*H (Tate & Simpson 2003) and *psb*A3 (Sang *et al*. 1997) for *trn*H–*psb*A. The PCR amplification conditions for *mat*K, *rbc*L, *trn*H–*psb*A and ITS were as follows: an initial predenaturation step at 94 °C for 5 min, followed by 30 cycles of 30 s at 94 °C, 30 s at 52 °C, and 1 min at 72 °C, with a final extension step of 10 min at 72 °C. For the ITS amplification, three additional annealing temperatures (51, 54 and 56 °C) were also applied. The amplification of genomic DNA was done in a PTC-100 thermocycler (Bio-Rad, Hercules, CA, USA).

The PCR products of *mat*K, *rbc*L and *trn*H–*psb*A were run on a 1.0% agarose gel in 1.0× TBE (Tris-borate-EDTA) buffer, purified using the Tiangen Midi purification Kit (Tiangen Biotech, Beijing, China) and then sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit and an Applied Biosystems ABI3730 DNA Sequencer.

Data analysis

Sequences were assembled using the SeqMan program (DNAStar Inc., Madison, Wisconsin, Burland, 2000) and aligned using CLUSTAL X (Thompson *et al*. 1997), then adjusted manually. The inter- and intraspecific variation of each barcoding region was characterized by calculating Kimura-2-parameter (K2P) distances in MEGA 4.0 (Tamura *et al*. 2007). K2P is one of the optimal models when distances are very small (Hebert *et al*. 2003). To assess the significance of intra- and interspecific divergence, the Wilcoxon signed-rank and Wilcoxon two-sample tests in SPSS 16.0 (SPSS, Chicago, IL, USA) were used. The separations between intraspecific variation and interspecific divergence ("barcoding gaps", Meyer & Paulay 2005) were gained by comparing the distributions of intra- and interspecific divergences of each candidate locus using the program TaxonDNA (Meier *et al.* 2006).

We also used TaxonDNA to analyze discrimination rates of DNA barcodes based on genetic distance (Meier *et al.* 2006). We employed three methods of this program, i.e., "Best match", "Best close match" and "All species barcodes", to ensure accurate species assignments in the datasets of *mat*K, *rbc*L, *trn*H–*psb*A, as well as the two combinations *mat*K + *rbc*L and *mat*K + *rbc*L + *trn*H–*psb*A. For the "Best match", a query is assigned the species name of its best-matching barcode sequences, regardless of how similar the query and barcode sequences are. With the "Best close match", a threshold similarity value is required to define how similar a barcode match needs to be before it can be identified. Using "All species barcodes", a query is assigned a species name only if the query is followed by all known barcodes for a particular species and only if there are at least two conspecific matches.

Tree-based methods were used to display the molecular identification results and test the monophyly of the species. Analyses using different methods may result in different trees which differ in relationships among individuals, species or genera. We, therefore, performed four methods, including neighbor-joining (NJ), unweighted pair group method with arithmetic mean (UPGMA), maximum parsimony (MP) and maximum likelihood (ML), to confirm the monophyly of species. NJ and UPGMA trees were generated using MEGA 4.0 under K2P model, and MP and ML trees were obtained with PAUP 4.0b10 (Swofford 2002) under general time reversible + I + G model assessed by ModelTest 3.7 (Posada & Crandall 1998). The resolution of species was characterized by calculating the percentage of species recovered as monophyletic based on the molecular trees. We regarded a species or variety as monophyletic only if all of its individuals grouped in a clade with more than 50% bootstrap values.

Results

PCR amplification and sequencing success

In the examined species, the *mat*K, *rbc*L and *trn*H–*psb*A regions exhibited 100% amplification and sequencing success (Table 2). Although four annealing temperatures (i.e., 51, 52, 54 and 56 °C) were used in the amplification, the PCR success rates for the ITS region were lower than 25%, and the success rate for bidirectional sequencing of ITS was zero due to strong overlapping signals in the sequencing. The poor success rates for the ITS amplification and sequencing may be due to the primer set of ITS4/ITS5 which is initially designed for fungi (White *et al.* 1990). Another probable reason is that the ITS region of the species examined in this study may have multiple divergent copies as shown in the subfamily Calamoideae including *Calamus* by Baker *et al.* (2000a). Multiple divergent ITS copies may potentially lead to misidentification in DNA barcoding due to differential sampling of divergent paralogues (Jeanson *et al.* 2011); we, therefore, abandoned the ITS region. In total 46 new sequences of *mat*K, *rbc*L and *trn*H–*psb*A were obtained from *Plectocomia himalayana* and from the 15 species or varieties of *Calamus*.

Alignment and character analysis of each locus

The aligned sequence lengths were 795 bp for *mat*K, 695 bp for *rbc*L, 1020 bp for *trn*H–*psb*A, 1490 bp for *mat*K + *rbc*L, and 2485 bp for *mat*K+ *rbc*L + *trn*H–*psb*A (Table 2). Of the three plastid barcodes, the *trn*H–*psb*A region showed the greatest number of variable sites (132) and greatest mean interspecific distance (0.0751). No intraspecific inversions were detected in the *trn*H–*psb*A dataset. There were many indels in the aligned *trn*H–*psb*A dataset, the longest comprising 258 bp in two individuals of *C. gracilis* (Table 2). The variable sites of *trn*H–*psb*A were approximately 6.8 and 21 times more than *mat*K and *rbc*L, respectively. *Mat*K had 17 variable sites, approximately 2.8 times more than *rbc*L, which had six variable sites.

Table 2. Evaluation of the four DNA loci and two combinations in 15 *Calamus* species or varieties.

	*mat*K	*rbc*L	*trn*H–*psb*A	ITS	*mat*K+ *rbc*L	*mat*K + *rbc*L + *trn*H–*psb*A
PCR success (%)	100	100	100	> 25	–	–
Sequencing success (%)	100	100	100	0	–	–
Aligned sequence length (bp)	795	695	1020	–	1490	2485
Indel length (bp)	6–9	0	1–258	–	3–6	1–149
No. informative sites/variable sites	17/17	6/6	132/132	–	23/23	94/101
Mean interspecific distance (range)	0.0040 (0–0.0129)	0.0026 (0–0.0072)	0.0709 (0–0.1708)	–	0.0033 (0–0.0082)	0.0104 (0–0.0221)
Mean intraspecific distance (range)	0.0001 (0–0.0009)	0 (0–0)	0.0001 (0–0.0020)	–	0.0001 (0–0.0005)	0.0004 (0–0.0038)

Genetic divergence analysis

The mean interspecific distances of the examined loci were much greater than the intraspecific distances in the present study (Table 2). The mean intraspecific distance was 0.0001 in the *mat*K dataset, varying from zero to 0.0009; and two varieties (*C. nambariensis* var. *alpinus* and *C. nambariensis* var. *xishuangbannaensis*) showed intraspecific variation (0.0009 and 0.0005, respectively). No species or varieties exhibited intraspecific variation in the *rbc*L dataset. The mean intraspecific distance was 0.0001 in the *trn*H–*psb*A dataset, varying from zero to 0.0020; and three species or varieties (*C. bonianus*, *C. nambariensis* var. *alpinus* and *C. nambariensis* var. *xishuangbannaensis*) showed intraspecific variation (0.0020, 0.0010 and 0.0006, respectively). The results of the Wilcoxon two-sample test indicated that the interspecific divergences for all five barcode sequences were significantly higher than the corresponding intraspecific variations. The combination *mat*K + *rbc*L + *trn*H–*psb*A had the greatest inter- *versus* intraspecific variation (Wilcoxon two-sample test: $p << 0.001$), followed by *trn*H–*psb*A and *mat*K + *rbc*L, while *mat*K had the lowest value (Table 3).

According to the results of the Wilcoxon signed-rank test, the rank order for the interspecific variation of the five candidate barcode sequences was *trn*H–*psb*A > *mat*K + *rbc*L + *trn*H–*psb*A > *mat*K > *mat*K + *rbc*L > *rbc*L. *Trn*H–*psb*A showed the highest variation among all of the candidate barcodes and their combinations (Table 4).

Monophyly test based on molecular trees

The discriminatory success of single and combined barcodes was determined by evaluating the percentage of each species or variety determined to be monophyletic using NJ, UPGMA, MP, and ML trees. Of these four molecular tree analyses, the UPGMA tree always yielded the best results, with more species resolved and higher bootstrap values. Based on the monophyletic species value of the NJ tree, the rank order of monophyletic species and varieties identification power of the candidate barcodes was: *mat*K + *rbc*L + *trn*H–*psb*A (62.5%) > *trn*H–*psb*A (56.3%) > *mat*K + *rbc*L (43.8%) > *mat*K (37.5%) > *rbc*L (6.3%) (Table 5). Furthermore, when treating the varieties of *Calamus nambariensis* (including *C. nambariensis* var. *alpinus*, *C. nambariensis* var. *menglongensis*, *C. nambariensis* var. *xishuangbannaensis*) and *C. yunnanensis* (including *C. yunnanensis*, *C. yunnanensis* var. *densiflorus*, *C. yunnanensis* var. *intermedius*) as one species, the monophyletic species value in the NJ trees improved to 100% (12/12) for the combined barcode *mat*K + *rbc*L + *trn*H–*psb*A (Fig. 1) and 91.7% (11/12) for *trn*H–*psb*A (Fig. 2). Respecive values for *mat*K, *rbc*L and *mat*K + *rbc*L were 58.3% (7/12), 8.3% (1/12) and 75% (9/12).

Barcoding gap test

The barcoding gap between intra- and interspecific distances was determined by graphing the distribution of the K2P distances for the five candidate barcode sequences (Fig. 3). We did not find any large barcoding gaps, although in the *trn*H–*psb*A dataset the distribution of intra- versus interspecific distances was considerably well separated. For all candidate barcodes, the discrimination rates based on the "Best match" of the TaxonDNA were identical to those of the "Best close match". The discrimination rates obtained with these two methods were apparently different among the candidate barcodes *mat*K (41.3%), *rbc*L (8.7%), *trn*H–*psb*A (58.7%), *mat*K + *rbc*L (47.8%) and *mat*K + *rbc*L + *trn*H–*psb*A (58.7%) (Table 6). According to the "All species

Table 3. Divergence of inter- *versus* intraspecific distances of each locus and different combinations. $p << 0.001$ in all cases.

Region	Wilcoxon two-sample test		
	#A	#B	*W*
*mat*K	9027	275	5568
*rbc*L	9075	251	5838
*trn*H–*psb*A	9129	173	6203
*mat*K + *rbc*L	9128	172	6258
*mat*K + *rbc*L + *trn*H–*psb*A	9134	179	6169

Table 4. Results of the Wilcoxon signed-rank test of interspecific divergence among loci.

W+	*W*–	Relative rank		*n*	*p* <<	Result
		W+	*W*–			
*mat*K	*rbc*L	5329	1457	116	0.001	*mat*K > *rbc*L
*mat*K	*trn*H–*psb*A	0	6759	116	0.001	*mat*K < *trn*H–*psb*A
*rbc*L	*trn*H–*psb*A	6	6873	117	0.001	*rbc*L < *trn*H–*psb*A
*mat*K + *rbc*L	*mat*K	1461	5321	116	0.001	*mat*K + *rbc*L < *mat*K
*mat*K + *rbc*L	*rbc*L	4419	1029	104	0.001	*mat*K + *rbc*L > *rbc*L
*mat*K + *rbc*L	*trn*H–*psb*A	3	6739	116	0.001	*mat*K + *rbc*L < *trn*H–*psb*A
*mat*K + *rbc*L + *trn*H–*psb*A	*mat*K	6882	5	117	0.001	*mat*K + *rbc*L + *trn*H–*psb*A > *mat*K
*mat*K + *rbc*L + *trn*H–*psb*A	*rbc*L	6548	0	114	0.001	*mat*K + *rbc*L + *trn*H–*psb*A > *rbc*L
*mat*K + *rbc*L + *trn*H–*psb*A	*trn*H–*psb*A	62	6718	116	0.001	*mat*K + *rbc*L + *trn*H–*psb*A < *trn*H–*psb*A
*mat*K + *rbc*L+ *trn*H–*psb*A	*mat*K+ *rbc*L	6889	0	117	0.001	*mat*K + *rbc*L + *trn*H–*psb*A > *mat*K + *rbc*L

Table 5. Species identification power of the DNA markers based on the tree-based methods.

Ability to discriminate	*mat*K	*rbc*L	*trn*H–*psb*A	*mat*K + *rbc*L	*mat*K + *rbc*L+ *trn*H–*psb*A
UPGMA tree	43.8% (7/16)	12.5% (2/16)	62.5% (10/16)	43.8% (7/16)	62.5% (10/16)
NJ tree	37.5% (6/16)	6.3% (1/16)	56.3% (9/16)	43.8% (7/16)	62.5% (10/16)
MP tree	37.5% (6/16)	6.3% (1/16)	56.3% (9/16)	37.5% (6/16)	62.5% (10/16)
ML tree	37.5% (6/16)	6.3% (1/16)	56.3% (9/16)	37.5% (6/16)	62.5% (10/16)

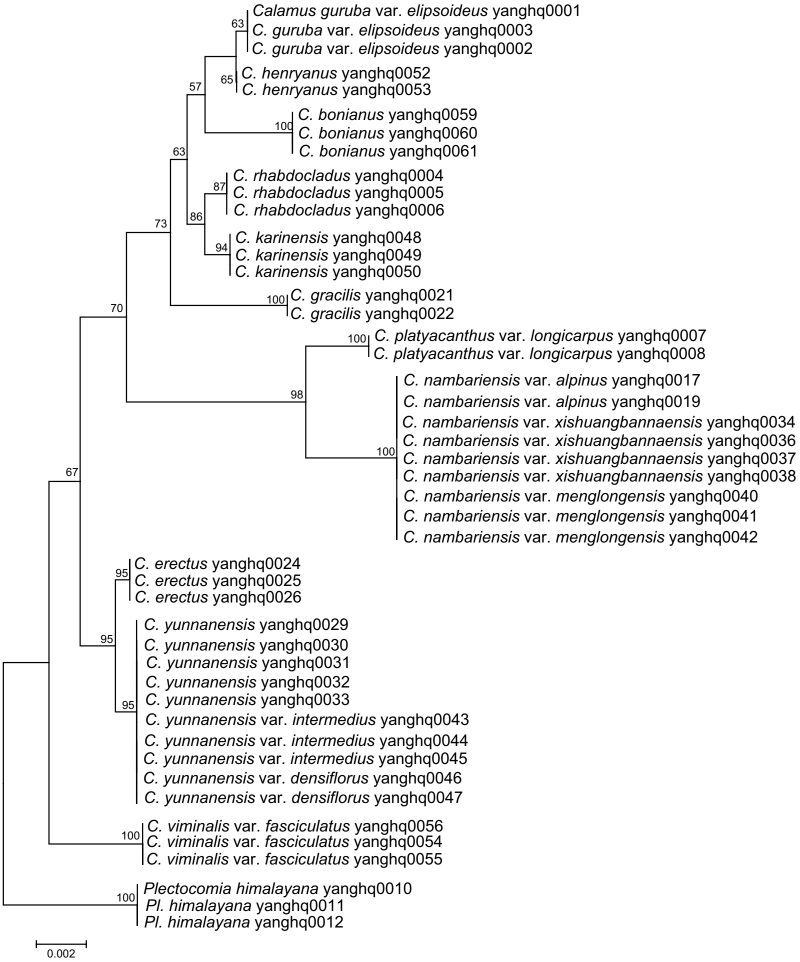
63
Calamus guruba var. elipsoideus yanghq0001
C. guruba var. elipsoideus yanghq0003
C. guruba var. elipsoideus yanghq0002
65
C. henryanus yanghq0052
C. henryanus yanghq0053
57
100
C. bonianus yanghq0059
C. bonianus yanghq0060
C. bonianus yanghq0061
63
87
C. rhabdocladus yanghq0004
C. rhabdocladus yanghq0005
C. rhabdocladus yanghq0006
86
73
94
C. karinensis yanghq0048
C. karinensis yanghq0049
C. karinensis yanghq0050
100
C. gracilis yanghq0021
C. gracilis yanghq0022
70
100
C. platyacanthus var. longicarpus yanghq0007
C. platyacanthus var. longicarpus yanghq0008
98
C. nambariensis var. alpinus yanghq0017
C. nambariensis var. alpinus yanghq0019
C. nambariensis var. xishuangbannaensis yanghq0034
C. nambariensis var. xishuangbannaensis yanghq0036
100
C. nambariensis var. xishuangbannaensis yanghq0037
C. nambariensis var. xishuangbannaensis yanghq0038
67
C. nambariensis var. menglongensis yanghq0040
C. nambariensis var. menglongensis yanghq0041
C. nambariensis var. menglongensis yanghq0042
95
C. erectus yanghq0024
C. erectus yanghq0025
C. erectus yanghq0026
C. yunnanensis yanghq0029
95
C. yunnanensis yanghq0030
C. yunnanensis yanghq0031
C. yunnanensis yanghq0032
95
C. yunnanensis yanghq0033
C. yunnanensis var. intermedius yanghq0043
C. yunnanensis var. intermedius yanghq0044
C. yunnanensis var. intermedius yanghq0045
C. yunnanensis var. densiflorus yanghq0046
C. yunnanensis var. densiflorus yanghq0047
C. viminalis var. fasciculatus yanghq0056
100
C. viminalis var. fasciculatus yanghq0054
C. viminalis var. fasciculatus yanghq0055
Plectocomia himalayana yanghq0010
100
Pl. himalayana yanghq0011
Pl. himalayana yanghq0012
0.002

Fig. 1. A taxon identification tree for 15 *Calamus* species or varieties created using the neighbor-joining (NJ) analysis of Kimura-2-parameter distances based on combined *mat*K + *rbc*L + *trn*H–*psb*A sequences. Bootstrap values (> 50%) are shown above the branches. Species names are followed by voucher numbers.

barcodes” method, all of the candidate barcodes had 78.3% discrimination rates.

Discussion and conclusion

For an appropriate DNA barcode, one of the most important criteria is universality, i.e., high PCR and sequencing success (e.g., Kress *et al.* 2005, Chase *et al.* 2007, Hollingsworth *et al.* 2011, China Plant BOL Group 2011). In our study, all of the *mat*K, *rbc*L and *trn*H–*psb*A regions performed well with 100% PCR and sequencing success. High-quality bidirectional sequences could thus be obtained easily for the *mat*K, *rbc*L and *trn*H–*psb*A loci.

ITS was proposed as a complementary marker to the core barcodes (CBOL Plant Working Group 2009) or a core barcode (China Plant BOL Group 2011). Many studies have demonstrated high variability in ITS (e.g. Kress *et al.* 2005, Sass *et al.* 2007, Liu *et al.* 2011). However in our study, ITS had poor success rates of amplification with the ITS4/ITS5 primer set, which may indicate that more universal primers for ITS as a plant DNA barcode are still needed. On the other hand, Baker *et al.* (2000a) revealed multi-copies of ITS in the calamoid palms, and their ITS sequences were proven to come from pseudogenic ITS regions (Harpke & Peterson 2008). Although in plant DNA barcoding, information from divergent putative pseudogenes can be useful for phylogenetic analyses (Razafimandimbison *et al.* 2004), additional procedures in cloning and analysis will take more time and expense. Recently ITS2 exhibited the highest

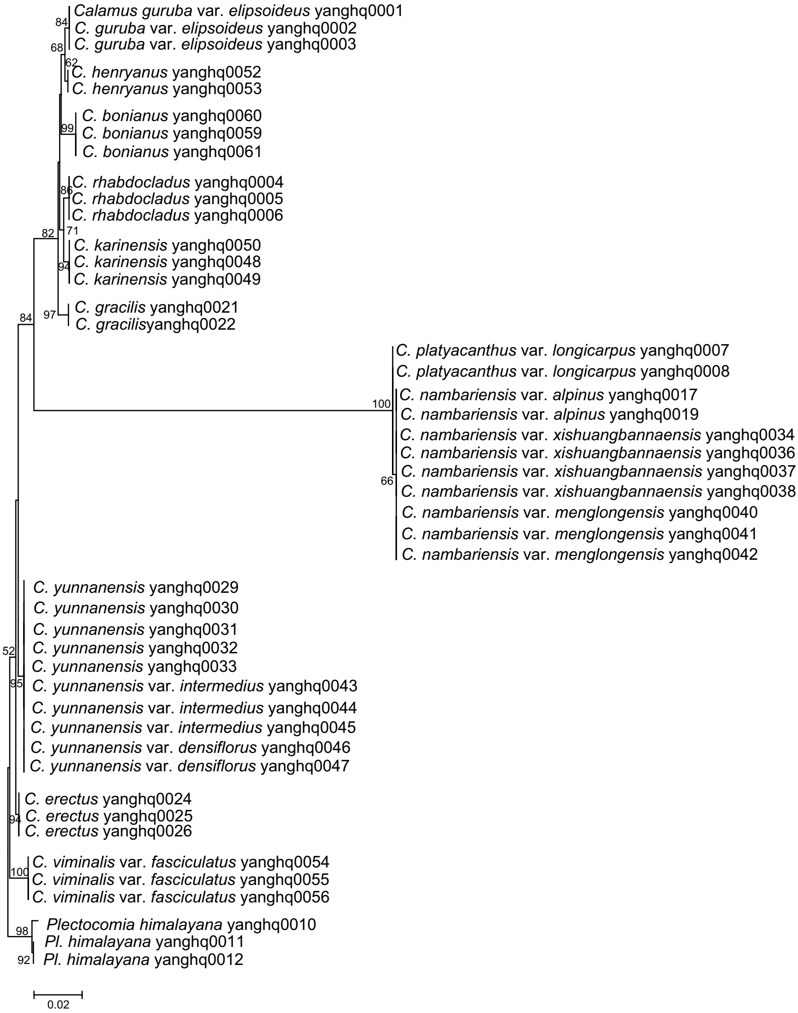
Calamus guruba var. *elipsoideus* yanghq0001
C. guruba var. *elipsoideus* yanghq0002
C. guruba var. *elipsoideus* yanghq0003
C. henryanus yanghq0052
C. henryanus yanghq0053
C. bonianus yanghq0060
C. bonianus yanghq0059
C. bonianus yanghq0061
C. rhabdocladus yanghq0004
C. rhabdocladus yanghq0005
C. rhabdocladus yanghq0006
C. karinensis yanghq0050
C. karinensis yanghq0048
C. karinensis yanghq0049
C. gracilis yanghq0021
*C. gracilis*yanghq0022
C. platyacanthus var. *longicarpus* yanghq0007
C. platyacanthus var. *longicarpus* yanghq0008
C. nambariensis var. *alpinus* yanghq0017
C. nambariensis var. *alpinus* yanghq0019
C. nambariensis var. *xishuangbannaensis* yanghq0034
C. nambariensis var. *xishuangbannaensis* yanghq0036
C. nambariensis var. *xishuangbannaensis* yanghq0037
C. nambariensis var. *xishuangbannaensis* yanghq0038
C. nambariensis var. *menglongensis* yanghq0040
C. nambariensis var. *menglongensis* yanghq0041
C. nambariensis var. *menglongensis* yanghq0042
C. yunnanensis yanghq0029
C. yunnanensis yanghq0030
C. yunnanensis yanghq0031
C. yunnanensis yanghq0032
C. yunnanensis yanghq0033
C. yunnanensis var. *intermedius* yanghq0043
C. yunnanensis var. *intermedius* yanghq0044
C. yunnanensis var. *intermedius* yanghq0045
C. yunnanensis var. *densiflorus* yanghq0046
C. yunnanensis var. *densiflorus* yanghq0047
C. erectus yanghq0024
C. erectus yanghq0025
C. erectus yanghq0026
C. viminalis var. *fasciculatus* yanghq0054
C. viminalis var. *fasciculatus* yanghq0055
C. viminalis var. *fasciculatus* yanghq0056
Plectocomia himalayana yanghq0010
Pl. himalayana yanghq0011
Pl. himalayana yanghq0012
84
68
62
99
86
82
71
94
97
84
100
66
52
95
94
100
98
92
0.02

Fig. 2. A taxon identification tree for 15 *Calamus* species or varieties created using neighbor-joining (NJ) analysis of Kimura-2-parameter distances based on *trn*H–*psb*A sequence. Bootstrap values (> 50%) are shown above the branches. Species names are followed by voucher numbers.

species discrimination (92%) in 40 species from the tribe Caryoteae of the palm family as compared with that of the *mat*K, *rbc*L and *psb*A-*trn*H loci (Jeanson *et al*. 2011). It will be expected to test ITS2 as a barcode in *Calamus* and other palm genera in the future.

Similar to the results of China Plant BOL Group (2011) and Jeanson *et al*. (2011), the two core markers — *mat*K and *rbc*L — individually exhibited low species discrimination rates. In *Calamus*, the success rates of *mat*K and *rbc*L based on the NJ tree were only 37.5% and 6.3% at the species level, respectively. Meanwhile, using "Best match" or "Best close match" of the TaxonDNA analysis, the success rates for individual identification of *mat*K and *rbc*L were 41.3% and 8.7%, respectively. Similarly, in the recent barcoding analysis of Caryoteae, the spe-

Table 6. Individual identification success rate based on the TaxonDNA analysis.

Criteria	*mat*K	*rbc*L	*trn*H–*psb*A	*mat*K + *rbc*L	*mat*K + *rbc*L + *trn*H–*psb*A
Best match	19 (41.3%)	4 (8.7%)	27 (58.7%)	22 (47.8%)	27 (58.7%)
Best close match	19 (41.3%)	4 (8.7%)	27 (58.7%)	22 (47.8%)	27 (58.7%)
All species barcodes	36 (78.3%)	36 (78.3%)	36 (78.3%)	36 (78.3%)	36 (78.3%)

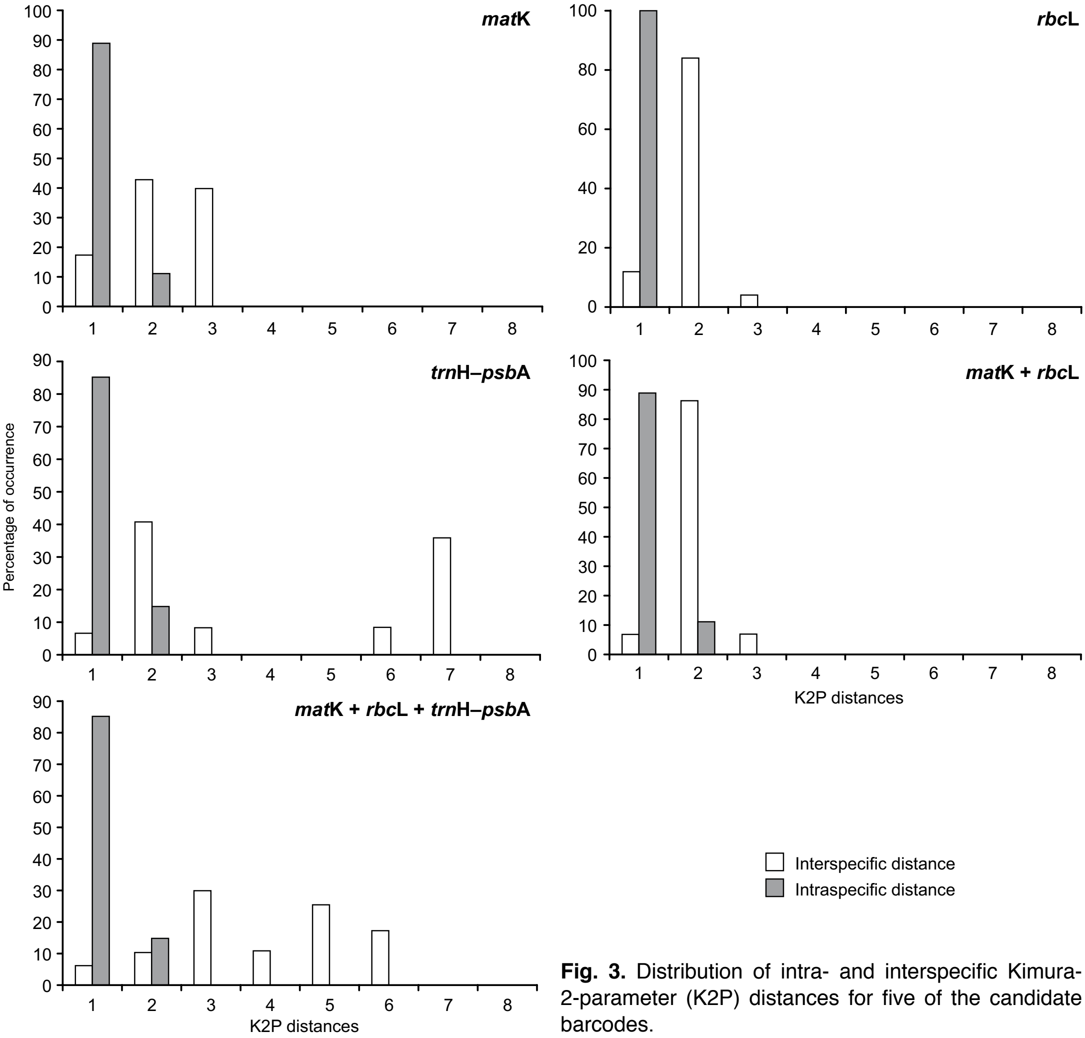

Fig. 3. Distribution of intra- and interspecific Kimura-2-parameter (K2P) distances for five of the candidate barcodes.

cies discriminations were 48% for *mat*K and 26% for *rbc*L (Jeanson *et al.* 2011). It is clear that the identification power of *mat*K and *rbc*L is significantly lower at infrageneric levels than at the generic level; meanwhile, these two plastid DNA regions have low species identification power at the species level in some plant groups such as Poales, Laurales, Dioscoreales, Apiales, and Zygophyllales (China Plant BOL Group 2011). Therefore, in *Calamus* neither *mat*K nor *rbc*L is capable of identifying closely related species.

*Trn*H–*psb*A has been suggested as a promising plant DNA barcoding marker by many studies (e.g. Lahaye *et al.* 2008, Nitta 2008, China Plant BOL Group 2011). However, one flaw of *trn*H–*psb*A as a barcode is its dramatic change in sequence lengths among different taxa and even congeneric species, caused by insertions/deletions (Kress *et al.* 2005). This can lead to difficulties in sequence alignment (Chase *et al.* 2007, CBOL Plant Working Group 2009). In our study, many indels were also found in the aligned *trn*H–*psb*A dataset, and two individuals of *C. gracilis* had an indel of 258 bp. Though *psb*A–*trn*H only had the second lowest species discrimination (37%) in the barcoding of Caryoteae (Jeanson *et al.* 2011), *trn*H–*psb*A, in the current study, it exhibited more variation than *mat*K and *rbc*L, and has a higher discrimination rate than *mat*K, *rbc*L, and even *mat*K + *rbc*L. Consequently, the *trn*H–*psb*A region has a potential to be used as a single barcode in *Calamus*. Based on the NJ tree and TaxonDNA, the combination of *mat*K + *rbc*L, a core plant barcode proposed by CBOL Plant Working Group (2009), greatly improved the species discriminating rates to 43.8% and 47.8%, respectively. Similarly, *mat*K + *rbc*L had 51.8% species discrimination in the barcoding of Caryoteae (Jeanson *et al*. 2011). As a whole, the identification power of this combination is unsatisfactory at the species level. Due to *trn*H–*psb*A, the species discrimination rates of *mat*K + *rbc*L + *trn*H–*psb*A considerably improved to 62.5% (NJ tree) and 58.7% ("Best match" of TaxonDNA). By ignoring the varieties of *C. yunnanensis* and *C. nambariensis*, its discrimination rates reach 100% (NJ tree), making it an appropriate combination barcode for *Calamus*.

In conclusion, of the regions examined in this study, the *trn*H–*psb*A region is an appropriate single barcode in *Calamus*. We consider DNA barcoding to be a useful tool to identify species within this economically and ecologically important genus. As far as we know, this is the first report contributed to DNA barcoding of *Calamus*, the largest genus of the palm family. Although considerable efforts have gone into testing barcoding markers, only 15 species or varieties collected in China were examined in the present study. For accurate species identification in *Calamus*, further studies on the species from other geographic regions and more candidate barcodes are required.

Acknowledgements

We thank an anonymous reviewer for his considerate and helpful comments; William Baker (Royal Botanic Gardens, Kew) for suggestions on an earlier draft of this manuscript; and Mao-Sheng Sun for help in material collection. This work was supported by the project of the State Forestry Administration of China (2008-4-30), the Large-scale Scientific Facilities of the Chinese Academy of Sciences (2009-LSF-GBOWS-01) and the Applied Basic Program of Yunnan (2010CD141).

References

Asmussen, C. B. & Chase, M. W. 2001: Coding and noncoding plastid DNA in palm systematics. — *American Journal of Botany* 88: 1103–1117.

Asmussen, C. B., Dransfield, J., Deickmann, V., Barfod A. S., Pintaud, J. C. & Baker, W. J. 2006: A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. — *Botanical Journal of the Linnean Society* 151: 15–38.

Baker, W. J., Hedderson, T. A. & Dransfield, J. 2000a: Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rps*16 intron sequence data. — *Molecular Phylogenetics and Evolution* 14: 195–217.

Baker, W. J., Hedderson, T. A. & Dransfield, J. 2000b: Molecular phylogenetics of *Calamus* (Palmae) and related rattan genera based on 5S nrDNA spacer sequence data. — *Molecular Phylogenetics and Evolution* 14: 218–231.

Barrett, R. D. H. & Hebert, P. D. N. 2005: Identifying spiders through DNA barcodes. — *Canadian Journal of Zoology* 83: 481–491.

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

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madrinan, S., Petersen, G., Seberg, O., Jorgsensen, T., Cameron, K. M., Carine, M., Pedersen, N., Hedderson, T. A. J., Conrad, F., Salazar, G. A., Richardson, J. E., Hollingsworth, M. L., Barraclough, T. G., Kelly, L. & Wilkinson, M. 2007: A proposal for a standardised protocol to barcode all land plants. — *Taxon* 56: 295–299.

Chen, S. Y., Pei, S. J. & Wang, K. L. 2003: *Palmer*. — In: Li, H. (ed.), *Flora of Yunnan*, vol. 14: 1–99. Science Press, Beijing.

China Plant BOL Group 2011: Comparative analysis of a large dataset indicates that internal transcribed spacer (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.

Cho, Y., Mower, J. P., Qiu, Y. L. & Palmer, J. D. 2004: Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. — *Proceedings of the National Academy of Sciences of the United States of America* 101: 17741–17746.

Cuénoud, P., Savolainen, V., Chatrou, L. W., Powell, M., Grayer, R. J. & Chase, M. W. 2002: Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *mat*K sequences. — *American Journal of Botany* 89: 132–144.

Doyle, J. & Doyle, J. 1987: A rapid DNA isolation procedure for small quantities of fresh leaf material. — *Phytochemical Bulletin* 19: 11–15.

Fay, M. F., Swensen, S. M. & Chase, M. W. 1997: Taxonomic affinities of *Medusagyne oppositefolia* (Medusagynaceae). — *Kew Bulletin* 52: 111–120.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., Percy, D. M., Hajibabaei, M. & Barrett, S. C. H. 2008: Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. — *PLoS ONE* 3(7), e2802, doi:10.1371/journal.pone.0002802.

Harpke, D. & Peterson, A. 2008: 5.8S motifs for the identification of pseudogenic ITS regions. — *Botany* 86: 300–305.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & de Waard, J. R. 2003: Biological identifications through DNA barcodes. — *Proceedings of the Royal Society B* 270: 313–321.

Hebert, P. D. N. & Gregory, T. R. 2005: The promise of DNA barcoding for taxonomy. — *System Biology* 54: 852–859.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. 2004: Identification of birds through DNA barcodes. — *PLoS Biology* 2(10), e312, doi:10.1371/journal.pbio.0020312.

Hollingsworth, P. M., Graham, S. W. & Little, D. P. 2011: Choosing and using a plant DNA barcode. — *PLoS ONE* 6(5), e19254, doi:10.1371/journal.pone.0019254.

Jeanson, M. L., Labat, J. N. & Little, D. P. 2011: DNA barcoding: a new tool for palm taxonomists? — *Annals of Botany* 108: 1445–1451.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. 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.

Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T. G. & Savolainen, V. 2008: DNA barcoding the floras of biodiversity hotspots. — *Proceedings of the National Academy of Sciences of the United States of America* 105: 2923–2928.

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.

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

Meyer, C. P. & Paulay, G. 2005: DNA barcoding: error rates based on comprehensive sampling. — *PLoS Biology* 3(12), e422, doi:10.1371/journal.pbio.0030422.

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

Pei, S. J., Chen, S. Y. & Tong, S. Q. 1991: *Flora of China*, vol. 13. — Science Press, Beijing.

Posada, D. & Crandall, K. A. 1998: ModelTest: testing the model of DNA substitution. — *Bioinformatics* 14: 817–818.

Razafimandimbison, S. G., Kellogg, E. A. & Bremer, B. 2004: Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleeae (Rubiaceae). — *Systematic Biology* 53: 177–192.

Sang, T., Crawford, D. J. & Stuessy, T. F. 1997: Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). — *American Journal of Botany* 84: 1120–1136.

Sass, C., Little, D. P., Stevenson, D. W. & Specht, C. D. 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.

Swofford, D. L. 2002: *PAUP. Phylogenetic analysis using parsimony (and other methods)*, version 4.0b10. — Sinauer, Sunderland.

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.

Tate, J. A. & Simpson, B. B. 2003: Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. — *Systematic Botany* 28: 723–737.

Thompson, J. D., Gibson, T. J., Jeanmougin, F. & Higgins, D. G. 1997: The Clustal X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. — *Nucleic Acids Research* 25: 4876–4882.

White, T. J., Bruns, T., Lee, S. & Taylor, J. W. 1990: *Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics*. — In: Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (eds.), *PCR protocols: a guide to methods and applications*: 315–322. Academic Press, New York.

Xing, Y. W., Wang, K. L. & Yang, Y. M. 2006: Floristic geography of *Calamus* (Palmae: Calamoideae) in China. — *Acta Botanica Yunnanica* 28: 461–467.