Journal of Systematics and Evolution 49 (3): 169–175 (2011) doi: 10.1111/j.1759-6831.2011.00128.x

Research Article

High universality of *matK* primers for barcoding gymnosperms

[1,2]Yan LI [1]Lian-Ming GAO* [1,3,4]Ram C. POUDEL [1,3]De-Zhu Li [5]Alan FORREST

[1](*Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences*, Kunming 650204, China)
[2](*Institute of Alpine Economic Plants, Yunnan Academy of Agricultural Sciences*, Lijiang 674100, China)
[3](*Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences*, Kunming 650204, China)
[4](*The Graduate University of Chinese Academy of Sciences*, Beijing 100049, China)
[5](*Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK*)

Abstract DNA barcoding is a tool to provide rapid and accurate taxonomic identification using a standard DNA region. A two-marker combination of *matK*+*rbcL* was formally proposed as the core barcode for land plants by the Consortium for the Barcode of Life Plant Working Group. However, there are currently no barcoding primers for *matK* showing high universality in gymnosperms. We used 57 gymnosperm species representing 40 genera, 11 families and four subclasses to evaluate the universality of nine candidate *matK* primers and one *rbcL* primer in this study. Primer (1F/724R) of *rbcL* is proposed here as a universal primer for gymnosperms due to high universality. One of the nine candidate *matK* primers (Gym_F1A/Gym_R1A) is proposed as the best "universal" *matK* primer for gymnosperms because of high polymerase chain reaction success and routine generation of high quality bidirectional sequences. A specific *matK* primer for *Ephedra* was newly designed in this study, which performed well on the sampled species. The primers proposed here for *rbcL* and *matK* can be easily and successfully amplified for most gymnosperms.

Key words DNA barcoding, *matK*, gymnosperm, primer universality.

DNA barcoding is a technique to provide rapid and accurate taxonomic identification using a specific DNA region (Hebert & Gregory, 2005). It has become a useful tool for species identification (Kress & Erickson, 2007; Erickson et al., 2008), and discovering new or cryptic species (Newmaster & Ragupathy, 2009; Valentini et al., 2009). The mitochondrial cytochrome c oxidase subunit 1 (*CO1*) has been used as a barcode for species identification in many animal groups (Hebert et al., 2003). However, the low substitution rate of *CO1* makes it unsuitable for barcoding plants, and has led to the search for alternative barcoding regions (Kress et al., 2005; Cowan et al., 2006; Fazekas et al., 2008; Hollingsworth et al., 2009). Recently, a two-marker combination of *matK*+*rbcL* was recommended as the core barcode for land plants by the Consortium for the Barcode of Life (CBOL) (CBOL Plant Working Group, 2009).

Universality of primers for polymerase chain reaction (PCR) and sequencing is one of the most important criteria for DNA barcoding (Chase et al., 2007; Kress & Erickson, 2007; Ford et al., 2009; Hollingsworth et al., 2009). In the core barcode, the primers for *rbcL* show a high level of universality in land plants (CBOL Plant Working Group, 2009; Kress et al., 2009). However, universality of *matK* primers was reported to be low in some studies (Sass et al., 2007; Fazekas et al., 2008; Ford et al., 2009; Kress et al., 2009), which may limit its use as a barcode. For example, 69% success recovery for *matK* in a floristic study was reported (Kress et al., 2009). Improvement of *matK* primers for barcoding can increase universality in angiosperms, and was a key reason for its recommendation as part of the core barcode (Lahaye et al., 2008; CBOL Plant Working Group, 2009). Although high levels of nucleotide substitutions in *matK* make it a good barcode marker in terms of species identification, it is difficult to routinely amplify and sequence across divergent lineages and as such more studies to develop "universal" *matK* primers are necessary (Kress & Erickson, 2007; Hollingsworth, 2008; CBOL Plant Working Group, 2009; Ford et al., 2009). Recently, order-specific *matK* primers for angiosperms were suggested by Dunning & Savolainen (2010). Nevertheless, there are still significant challenges that need to be overcome for high-throughput floristic studies using this DNA region for barcoding plants.

Gymnosperms are a group of seed-bearing plants that consist of four subclasses, Cycadidae, Ginkgoidae, Gnetidae, and Pinidae, representing 8 orders, 12 families, 83 genera, and approximately 990 species (Christenhusz et al., 2011). Gymnosperms are all woody trees, shrubs or lianas. They grow throughout most of the

Received: 23 December 2010 Accepted: 23 February 2011
* Author for correspondence. E-mail: gaolm@mail.kib.ac.cn; Tel.: 86-871-5223505; Fax: 86-871-5217791.

world, and are the dominant vegetation in many colder and arctic regions (Judd et al., 2007). Gymnosperms have major economic uses: many are important wood resources and ornamental plants. In China, there are about 250 gymnosperm species in 34 genera and 10 families (Wu & Raven, 1999). For gymnosperms, there is no universal specific *matK* primer for DNA barcoding that has performed well to date. Several *matK* primers proposed for barcoding land plants show low levels of PCR success in gymnosperms (CBOL Plant Working Group, 2009; Ford et al., 2009; Liu et al. 2011). For example, the PCR success rate is less than 50% in gymnosperms based on data compiled from several laboratories (CBOL Plant Working Group, 2009), and only 33% PCR success was obtained with the best *matK* primer pair for sampled gymnosperm species by Ford et al. (2009). The fact that no specific *matK* primers have been developed as barcodes for gymnosperms may explain the low amplification of this region. Primer development to improve the overall success rate of *matK* amplification is essential for its application as part of the universal plant barcode (CBOL Plant Working Group, 2009). However, it is a challenge to develop a universal *matK* primer for the whole plant kingdom because of an extensive variation within the DNA region (Kress & Erickson 2007). Further primer development efforts are required in non-angiosperms for *matK* (CBOL Plant Working Group, 2009).

In this study, we used nine candidate *matK* primers and one *rbcL* primer to evaluate the PCR and sequencing universality for gymnosperms. Six of the nine candidate *matK* primers are recently designed with five as yet unpublished, and the remaining three were proposed as "universal" primers for barcoding land plants. We selected 57 species, representing 40 genera of 11 families that correspond to all four subclasses of gymnosperm for trial in this study. Our aim here is to evaluate the universality of the *matK* primers for gymnosperms.

1 Material and methods

1.1 Sampling strategy

A total of 57 species representing 40 genera, 11 families, and 8 orders were selected in this study (Table 1), covering approximately 47% of genera, 92% of families, 88% of orders, and 100% of the subclasses of gymnosperms. These samples represent all the genera and families occurring in China. To represent the wider diversity of larger genera (those containing >20 species occurring in China), more than one species per genus were included in the analysis. Voucher specimens were deposited in the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (KUN). Leaf material was collected in the field or from Kunming Botanic Garden, and immediately dried in silica gel until laboratory analysis.

1.2 Candidate primers screening strategy

The nine candidate *matK* primer pairs and a single *rbcL* primer pair are outlined in Table 2. Six of the nine *matK* primer pairs are recently designed and have not been tested systematically, and the remaining three *matK* primer pairs are those most commonly used for barcoding land plants. We adopted two steps to test the universality of candidate primers. First, we selected 14 species representing the 11 families for PCR amplification with all candidate *matK* primers and one *rbcL* primer as control. For the large families Pinaceae and Cupressaceae, two samples from each were included in this family-level assessment (see Table 1). Second, candidate *matK* primers showing more than 50% PCR success were used for further assessment on all remaining samples (five of the nine *matK* primer pairs met this requirement).

1.3 DNA extraction, amplification, and sequencing

Genomic DNA was extracted from silica-dried leaf material using the CTAB method as described by Doyle & Doyle (1987). The genomic DNA was dissolved in TE buffer (10 mmol/L Tris-HCl, pH 8.0, 1 mmol/L EDTA) to a final concentration of 40–50 ng/μL to avoid any variation in PCR success due to DNA concentration differences.

All PCR reactions were carried out on a GeneAmp PCR System 9700 thermal cycler (Perkin Elmer, Foster City, CA, USA) in a total volume of 20 μL, which contained 10 μL 2 × Taq PCR Master Mix (0.1 U Taq polymerase/μL, 0.5 mmol/L each dNTP, 20 mmol/L Tris-HCl (pH 8.3), 100 mmol/L KCl, 3 mmol/L $MgCl_2$; Tiangen Biotech, Beijing, China), 0.2 μL bovine serum albumin (10 μg/μL), 0.5 μL each of the forward and reverse primers (10 μmol/L), and 1 μL template DNA. Negative controls were run alongside all PCR reactions. The *matK* thermocycling profile was: 94°C for 3 min, 35 cycles of 94°C for 30 s, 52°C for 30 s, 72°C for 45 s, with a final extension of 72°C for 10 min. The *rbcL* thermocycling profile was: 94°C for 1 min, 30 cycles of 94°C for 50 s, 52°C for 1 min, 72°C for 1 min, with a final extension of 72°C for 10 min. The PCR products were visualized on 1% Tris–acetate–EDTA agarose gel stained with ethidium bromide alongside a GeneRuler 100 bp DNA ladder (Fermentas, Glen Burnie, MD, USA). The PCR products were purified using ExoSAP-IT (GE Healthcare, Cleveland, OH, USA). Sequencing reactions were carried out in a total volume of 6 μL containing 0.2 μL

Table 1 Vouchers of the species sampled in this study and polymerase chain reaction success for the five candidate *matK* primers

Subclass	Family	Species	Voucher	Primer				
				Gym_F1A Gym_R1A	Gym_F2A Gym_R2A	Gym_F1B Gym_R1B	Gym_F2B Gym_R2B	PKF4 PKR1
Cycadidae	Cycadaceae	***Cycas guizhouensis***†	Glm-06032	++	++	++	+	+
Cycadidae	Cycadaceae	*Cycas micholitzii*	Ly-012	++	++	++	+	++
Cycadidae	Zamiaceae	***Encephalartos lehmannii***†	Ly-014	++	++	++	+	+
Ginkgoidae	Ginkgoaceae	*Ginkgo biloba*†	Ly-028	−	+	+	+	+
Gnetidae	Ephedraceae	*Ephedra gerardiana*	090937	−	−	−	−	−
Gnetidae	Ephedraceae	*Ephedra intermedia*	Sunh-zx-1688	−	−	−	−	−
Gnetidae	Ephedraceae	*Ephedra likiangensis*†	Ly-024	−	−	−	−	−
Gnetidae	Gnetaceae	*Gnetum pendulum*†	GBOWS0801	++	−	−	++	++
Pinidae	Araucariaceae	***Araucaria bidwillii***	Ly-015	++	−	++	+	++
Pinidae	Araucariaceae	*Araucaria cunninghamii*†	Ly-016	++	−	++	+	++
Pinidae	Araucariaceae	*Araucaria heterophylla*	Ly-004	++	−	++	+	++
Pinidae	Cupressaceae	*Calocedrus macrolepis*	Glm-103074	++	−	++	+	−
Pinidae	Cupressaceae	*Chamaecyparis formosensis*	Glm-103065	++	++	++	−	+
Pinidae	Cupressaceae	*Cunninghamia lanceolata*	Ly-007	++	++	++	+	++
Pinidae	Cupressaceae	*Cupressus duclouxiana*	Ly-011	++	−	++	+	+
Pinidae	Cupressaceae	*Cryptomeria japonica*	Glm-103142	++	−	++	+	++
Pinidae	Cupressaceae	*Fokienia hodginsii*	Ly-006	++	++	++	++	++
Pinidae	Cupressaceae	***Glyptostrobus pensilis***	Ly-005	++	++	++	++	++
Pinidae	Cupressaceae	*Juniperus pingii* var. *wilsonii*	Ly-026	++	−	++	−	−
Pinidae	Cupressaceae	*Juniperus squamata*†	Glm-082126	++	−	++	−	−
Pinidae	Cupressaceae	*Metasequoia glyptostroboides*†	Ly-002	++	++	++	+	++
Pinidae	Cupressaceae	*Platycladus orientalis*	Ly-020	++	−	++	+	−
Pinidae	Cupressaceae	*Sequoia sempervirens*	Ly-021	++	++	++	+	+
Pinidae	Cupressaceae	*Taiwania cryptomerioides*	Glm-103046	++	++	++	+	++
Pinidae	Cupressaceae	*Taiwania flousiana*	Glm-092411	++	++	++	++	++
Pinidae	Cupressaceae	*Taxodium distichum*	Ly-009	++	+	++	+	++
Pinidae	Cupressaceae	*Thujopsis dolabrata*	Ly-019	++	++	++	+	++
Pinidae	Cupressaceae	***Thuja sutchuenensis***	Ly-018	++	++	++	++	++
Pinidae	Pinaceae	*Abies georgei*†	Glm-102904	++	++	++	++	++
Pinidae	Pinaceae	*Cathaya argyrophylla*	Ly-017	++	+	++	++	++
Pinidae	Pinaceae	*Cedrus deodara*	Ly-029	++	++	++	++	++
Pinidae	Pinaceae	*Keteleeria davidiana*	Ly-001	++	+	++	+	++
Pinidae	Pinaceae	*Keteleeria evelyniana*	Ly-003	++	+	++	++	++
Pinidae	Pinaceae	*Keteleeria fortunei*	Ly-010	++	+	++	+	++
Pinidae	Pinaceae	*Larix potaninii var. chinensis*	Glm-06081	++	−	++	++	++
Pinidae	Pinaceae	***Larix himalaica***	Glm-081604	++	−	++	+	++
Pinidae	Pinaceae	*Picea likiangensis*	Glm-08912	++	−	++	+	++
Pinidae	Pinaceae	*Picea smithiana*	Glm-081533	++	−	++	+	++
Pinidae	Pinaceae	*Picea spinulosa*	Glm-081883	++	−	++	++	++
Pinidae	Pinaceae	*Pinus armandii*	Ly-027	++	−	++	++	++
Pinidae	Pinaceae	*Pinus wallichiana*	81471	++	−	++	+	++
Pinidae	Pinaceae	*Pinus yunnanensis*†	Ly-025	++	+	++	++	++
Pinidae	Pinaceae	*Pseudolarix amabilis*	Ly-008	++	+	++	++	++
Pinidae	Pinaceae	*Pseudotsuga sinensis*	Glm-103063	++	−	++	−	+
Pinidae	Pinaceae	*Tsuga dumosa*	Glm-081797	++	−	++	−	++
Pinidae	Podocarpaceae	*Dacrycarpus imbricatus*†	Ly-023	++	++	++	++	+
Pinidae	Podocarpaceae	***Nageia fleuryi***	Ly-022	++	++	++	++	++
Pinidae	Podocarpaceae	*Nageia nagi*	Ly-013	++	++	++	++	−
Pinidae	Podocarpaceae	*Podocarpus macrophyllus*	GBOWS0294	++	++	++	++	−
Pinidae	Podocarpaceae	*Podocarpus neriifolius*	Glm06211	++	++	++	+	−
Pinidae	Sciadopityaceae	*Sciadopitys verticillata*†	Glm-092241	++	+	++	+	+
Pinidae	Taxaceae	*Amentotaxus argotaenia*	182	++	++	++	++	+
Pinidae	Taxaceae	*Cephalotaxus mannii*	ZSD001	++	++	++	++	++
Pinidae	Taxaceae	***Cephalotaxus wilsoniana***†	Glm-103119	++	++	++	++	++
Pinidae	Taxaceae	*Pseudotaxus chienii*	Glm-07549	++	++	++	++	++
Pinidae	Taxaceae	*Taxus wallichiana*†	RC1289	++	++	++	+	++
Pinidae	Taxaceae	***Torreya fargesii* var. *yunnanensis***	GLM-092567-18	++	++	++	++	++

Species in bold were selected for sequencing to evaluate sequence quality and coverage in this study. †Species selected for evaluating PCR success using the nine candidate *matK* primer sets at family-level. −, no band; +,weak band; ++, strong band.

Table 2 Polymerase chain reaction primers used for screening universal primers for gymnosperms in this study

Gene	Primer name	Direction	Sequence (5′–3′)	Source
matK	Gym_R1A	f	5′- TCA YCC GGA RAT TTT GGT TCG-3′	This study, designed by Alan Forrest
	Gym_F1A	r	5′- ATY GYR CTT TTA TGT TTA CAR GC-3′	This study, designed by Alan Forrest
	Gym_R2A	f	5′- ACY TTT CGY YRC TGG ATC CAA G-3′	This study, designed by Alan Forrest
	Gym_F2A	r	5′- GTT TTA GCR CAT GRR AGT CGA AG-3′	This study, designed by Alan Forrest
	Gym_R1B	f	5′- TCA TCC RGA AAT TTT GGT KCG-3′	This study, designed by Alan Forrest
	Gym_F1B	r	5′- ATM GTA CTT TTA TGT TTA CAR GC-3′	This study, designed by Alan Forrest
	Gym_R2B	f	5′- AYY TTT CGT CGC TGG ATC CGA G-3′	This study, designed by Alan Forrest
	Gym_F2B	r	5′- GTT TTA GCR CAT GRW ART RRA AG-3′	This study, designed by Alan Forrest
	NY552F	f	5′- CTG GAT YCA AGA TGC TCC TT-3′	Damon Little, unpublished
	NY1150R	r	5′- GGT CTT TGA GAA GAA CGG AGA-3′	Damon Little, unpublished
	PKF4	f	5′- CCC TAT TCT ATT CAY CCN GA-3′	Fazekas et al (2008)
	PKR1	r	5′- CGT ATC GTG CTT TTR TGY TT-3′	Fazekas et al (2008)
	5R	f	5′-GTT CTA GCA CAA GAA AGT CG-3′	Ford et al. (2009)
	XF	r	5′-TAA TTT ACG ATC AAT TCA TTC-3′	Ford et al. (2009)
	3F_KIM	f	5′-CGT ACA GTA CTT TTG TGT TTA CGA G-3′	Ki-Joong Kim, unpublished
	1R_KIM	r	5′-ACC CAG TCC ATC TGG AAA TCT TGG TTC-3′	Ki-Joong Kim, unpublished
	1326R	f	5′-TCT AGC ACA CGA AAG TCG AAG T-3′	Cuénoud et al. 2002
	390F	r	5′-CGA TCT ATT CAT TCA ATA TTT C-3′	Cuénoud et al. 2002
rbcL	1F	f	5′-ATG TCA CCA CAA ACA GAA AC-3′	Fay et al. 1997
	724R	r	5′-TCG CAT GAT CCT GCA GTA GC-3′	Fay et al. 1997

f, forward; r, reverse. Damon Little is based at New York Botanic Garden, New York. Ki-Joong Kim is based at Korea University, Seoul.

purified PCR product, 0.15 μL BigDye terminator sequencing mixture (V3.1), 1.2 μL sequence buffer, and 1.4 μL μmol/L primer. The cycle sequencing profile was 32 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The purified sequencing products were run on an ABI 3730xl automated sequencer (Applied Biosystems, Foster City, CA, USA).

1.4 Sequence quality and coverage analysis

Sequences were trimmed using a window size of 20 bp, with segments with 2 bp showing <20 Quality Value (QV, an established metric for determining quality sequencing data) trimmed. Sequence quality was assessed using Sequencing Analysis 5.3.1 software (Applied Biosystems). The sequence quality recovered was defined such that both the forward and reverse reads should have high length of read: after trimming to an average QV $\geq$ 20 in a 20-bp window, the post-trim lengths should be >50% of the original read length and the assembled contigs should have >50% overlap in the alignment of the forward and reverse reads with <1% low-quality bases (<20 QV) and <1% internal gaps and substitutions when aligning the forward and reverse reads. The percentage of bases with >20 QV, >30 QV, and >40 QV was determined using Sequencher 4.9 (Gene Codes, Ann Arbor, MI, USA). Error probabilities for consensus sequences were calculated after base pair calling using Phred as implemented in Aligner 5.3.6 (CodonCode, Dedham, MA, USA).

2 Results

2.1 Primer universality

The universality of primer for *rbcL* (1F/724R) performed well with 100% PCR success on all the sampled 14 species at family level (Table 1). The universality for the nine *matK* primers showed a much more variable PCR success rate (0.0%–92.9%) on the 14 selected species in the family level assessment. Five of the nine candidate *matK* primers (Gym_F1A/Gym_R1A, Gym_F2A/Gym_R2A, Gym_F1B/Gym_R1B, Gym_F2B/Gym_R2B, and PKF4/PKR1) showed high PCR universality at genus level (50.0%–92.9%) and family level (75%–100%), respectively. The remaining four *matK* primers (3F_KIM/1R_KIM, 390F/1326R, XF/5R, and NY552F/NY1150R) showed low PCR success (0.0%–14.3%). These four primers with low PCR success were rejected for further evaluation on the expanded sample set.

On all 57 sampled species, *rbcL* performed well with 100% PCR universality (Table 1). For the five candidate *matK* primers used for further evaluation, Gym_F1A/Gym_R1A showed the highest PCR success with 94.7%, followed by Gym_F1B/Gym_R1B with 93.0%, Gym_F2B/Gym_R2B with 80.7%, PKF4/PKR1 with 77.2%, and Gym_F2A/Gym_R2A with 57.9 (Table 1). However, three primers Gym_F2A/Gym_R2A, Gym_F2B/Gym_R2B, and PKF4/PKR1, yielded weak PCR bands of 25.7%, 53.1%, and 21.3%, respectively, for the sampled 57 species. All the sampled species of family Ephedraceae failed to amplify using these five *matK* primers (Table 1).

2.2 Sequence quality and coverage

To assess the quality of generated sequence traces for *matK* primers Gym_F1A/Gym_R1A and Gym_F1B/Gym_R1B, which showed high PCR success and strong PCR bands, we selected nine species

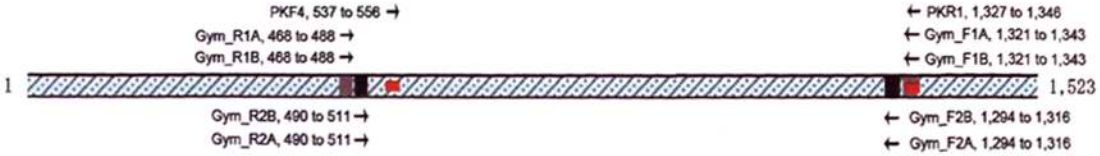

Fig. 1. Position and orientation of primers for *matK* of *Cephalotaxus fortunei* as template (AF228109). Grey, black, and red rectangles indicate the position in the *matK* gene of primer pair Gym_R1A/Gym_F1A and Gym_R1B/Gym_F1B, Gym_R2A/Gym_F2A and Gym_R2B/Gym_F2B, and PKF4/PKR1, respectively.

for sequencing, representing the nine sampled families except for Ephedraceae, Gnetaceae, and Ginkgoaceae due to PCR failure. All the generated 36 bidirectional sequence traces showed high sequence quality with few low-quality bases trimmed out. The amplicon sizes of the two primer pairs were 876 bp including primer sites. Length of read for sequence traces generated by primer Gym_F1A/Gym_R1A ranged from 799 to 829 bp with high QV (≥20), and 779–816 bp with high QV (≥20) for the sequences generated by primer Gym_F1B/Gym_R1B. There were 93.7%–97.2% and 82.8%–92.5% of bases (QV >30), and 91.3%–96.2% and 83.7%–90.4% of bases (QV >40) for primer pairs Gym_F1A/Gym_R1A and Gym_F1B/Gym_R1B, respectively. On the whole, the sequence quality of Gym_F1A/Gym_R1A is a little better than that of Gym_F1B/Gym_R1B. The assembled contigs showed >70% overlap in the alignment of the forward and reverse reads without low-quality bases among generated sequences.

2.3 Position and orientation of *matK* primers

Position and orientation of the five candidate primers for *matK*, with *Cephalotaxus fortunei* as template, is illustrated in Fig. 1. Primer Gym_F1A/Gym_R1A and Gym_F1B/Gym_R1B located at the same position with a length of 832 bp between primer sets. Primers Gym_F2A/Gym_R2A and Gym_F2B/Gym_R2B also share an identical position with a length of 782 bp between primer sets (Fig. 1). Forward primer PKF4 started at position 537 bp and the reverse primer PKR4 at 1327 bp with a length of 770 bp between primer sets (Fig. 1).

3 Discussion

Primer universality is the most important criterion for DNA barcoding in the first instance (Chase et al., 2007; Kress & Erickson, 2007; Ford et al., 2009; Hollingsworth et al., 2009). *rbcL*+*matK* has been recommended as the plant core barcode based on its recoverability, sequence quality, and level of species discrimination (CBOL Plant Working Group, 2009). High PCR amplification success of *rbcL* for gymnosperms has been reported in some studies (CBOL Plant Working Group, 2009; Liu et al., 2011), whereas PCR universality of *matK* was relatively low in gymnosperms. In this study, PCR universality of *rbcL* here performed best with 100% PCR success for all the 57 sampled species, representing 11 of 12 families of gymnosperms. So we proposed primer 1F/724R as the "universal" *rbcL* primer for barcoding gymnosperms.

The highly variable *matK* region has lower PCR amplification success than the more conserved *rbcL* gene (Kress & Erickson, 2007; CBOL Plant Working Group, 2009). It is currently not possible to amplify *matK* with a single universal primer pair across the whole plant kingdom to facilitate high throughout, rapid, automated, and cost-effective species identification (Kress & Erickson, 2007; Dunning & Savolainen, 2010). Three *matK* primer sets, namely 390F/1326R (Cuénoud et al., 2002), XF/5R (Ford et al., 2009), and 3F_KIM/1R_KIM (K.-J. Kim, Korea University, Seoul, unpubl. data, also see Hollingsworth, 2008) were proposed as currently the best "universal" *matK* primers for land plant (http://www.barcoding.si.edu/PDF/Informationonbarcodeloci.pdf). Although these primers performed well for most groups of angiosperm, they showed low PCR success for sampled species of gymnosperm in this study, indicating that these three *matK* primers are unsuitable as "universal" *matK* primers for gymnosperms.

Two primers of the nine candidate *matK* primers, Gym_F1A/Gym_R1A and Gym_F1B/Gym_R1B, showed high levels of PCR universality, high sequence quality and coverage, and could also yield strong PCR products for sampled species. Of the two *matK* primer pairs, primer Gym_F1A/Gym_R1A, performed better, successfully amplifying all the 11 sampled families and 39 out of the 40 sampled genera (97.5%). Only species of *Ephedra* (Gnetales: Ephedraceae) failed to amplify (Table 1). For primer pair Gym_F1B/Gym_R1B, 37 of 40 genera (92.5%) were amplified successfully, but species of two sampled families (Ephedraceae

and Ginkgoaceae) could not be amplified in this study (Table 1). Based on assessments of primer universality and sequence quality and coverage, we propose that primer Gym_F1A/Gym_R1A currently represents the best universal primer for *matK* in gymnosperms.

An ideal DNA barcode should satisfy the criterion of an appropriately short sequence length (300–800 bp) to facilitate DNA extraction and amplification (Kress et al., 2005). Position and orientation of the five candidate *matK* primers showing high PCR success is given in Fig. 1. The length of the proposed *matK* barcode region for primer Gym_F1A/Gym_R1A is 832 bp long using the *Cephalotaxus fortunei matK* sequence (AF228109) as a reference (excluding primer sequence). The size of this product is amenable to the acquisition of bidirectional sequence reads with a single primer pair, thus meeting the criterion of DNA barcode length.

As none of the candidate *matK* primers amplified samples of *Ephedra*, a specific *matK* primer for *Ephedra* (Eph_F: 5′-TCATTCAGAGCTGTTAGTTAG-3′, Eph_R: 5′-ATCGTACTTTTATGCTTACAGGC-3′) was newly designed in this study. This primer performed very well on the three sampled species of *Ephedra* with 100% PCR success. This primer is located at a similar position on the *matK* gene as primer Gym_F1A/Gym_R1A, ensuring that the length of the barcode region for gymnosperms is similar and easy to align. The primers proposed in this study for core barcode *rbcL* and *matK* can be easily and successfully amplified for all the genera of gymnosperms in China. We hope our work will facilitate DNA barcoding of gymnosperms worldwide as well.

Acknowledgements We are grateful to Dr. Chun-Xia ZENG, Mr. Jun-Bo YANG, Mrs Jing YANG, and other colleagues for laboratory help or providing samples. We also thank Prof. Peter HOLLINGSWORTH for his constructive comments on the manuscript. This study was supported by the Research Fund for the Large-Scale Scientific Facilities of Chinese Academy of Sciences (Grant No. 2009-LSF-GBOWS-01), the West Light Programme of the Chinese Academy of Sciences (Grant No. 9223111W1) and the Talent Project of Yunnan Province, China (Grant No. 2008PY064). Alan FORREST is grateful to the Gordon and Betty Moore Foundation (#2535) for funding which supported the development of *matK* primers described in this paper.

References

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, Madriñán S, Petersen G, Seberg O, Jørgsensen 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.

Christenhusz MJM, Gardner MF, Mill RR, Reveal JL, Chase MW. 2011. A new classification and linear sequence of extant gymnosperms. Phytotaxa 19: 55–70.

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

Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. 2002. Molecular phylogenetics of *Caryophyllales* based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *mat*K DNA 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.

Dunning LT, Savolainen V. 2010. Broad-scale amplification of *mat*K for DNA barcoding plants, a technical note. Botanical Journal of the Linnean Society 164: 1–9.

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

Fay MF, Swensen SM, Chase MW. 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). Kew Bulletin 52: 111–120.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3: e2802.

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

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

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

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

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

Judd WS, Campbell CC, Kellogg EA, Stevens PF, Donoghue MJ. 2007. Plant systematics: A phylogenetic approach. 3rd edition. Sunderland: Sinauer Associates.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: The coding *rbc*L gene complements the non-coding *trn*H-*psb*A spacer region. PLoS One 2: e508.

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences USA 106: 18621–18626.

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.

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.

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

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

Valentini A, Pompanon F, Taberlet P. 2009. DNA barcoding for ecologists. Trends in Ecology & Evolution 24: 110–117.

Wu ZY, Raven PH. 1999. Flora of China. Vol. 4. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press.

High universality of matK primers for barcoding gymnosperms

作者： Yan LI， Lian-Ming GAO， RAM C.POUDEL， De-Zhu Li， Alan FORREST

作者单位： Yan LI(Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China;Institute of Alpine Economic Plants, Yunnan Academy of Agricultural Sciences, Lijiang 674100, China)， Lian-Ming GAO(Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China)， RAM C.POUDEL,De-Zhu Li(Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China;Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, C)， Alan FORREST(Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK)

刊名： 植物分类学报 ISTIC SCI PKU

英文刊名： JOURNAL OF SYSTEMATICS AND EVOLUTION

年，卷(期)： 2011,49(3)