

Application of DNA barcoding markers to the identification of *Hopea* species

Nguyen T.P. Trang¹, N.M. Duc¹, N.V. Sinh² and L. Triest³

¹Department of Molecular Systematics and Conservation Genetics, Institute of Ecology and Biological Resources (IBER), Vietnam Academy of Science and Technology, Cau Giay, Hanoi, Vietnam ²Department of Plant Ecology, IBER, Cau Giay, Hanoi, Vietnam ³Research Group Plant Biology and Nature Management, Vrije Universiteit Brussel, Brussels, Belgium

Corresponding author: Nguyen T.P. Trang E-mail: ntptrang@iebr.ac.vn

Genet. Mol. Res. 14 (3): 9181-9190 (2015) Received January 7, 2015 Accepted May 8, 2015 Published August 7, 2015 DOI http://dx.doi.org/10.4238/2015.August.7.28

ABSTRACT. Hopea chinensis (synonym: H. hongayensis) (Dipterocarpaceae) is a threatened species found so far in only two locations: Quang Ninh (Vietnam) and Guangxi (China). The species shares many morphological characteristics with H. mollissima and the two species are often confused. To overcome this problem of identification and to investigate the genetic relationships of *Hopea* species with other Dipterocarp species, we sequenced three candidate DNA barcodes for the chloroplast markers rbcL, trnH-psbA, and matK. These markers were used separately and in different combinations to determine whether they could establish an accurate and effective identification system for H. chinensis in Quang Ninh (Vietnam). Our analyses indicated that two of the candidate DNA barcodes, matK and rbcL, performed best. We also generated a neighbor-joining phylogenetic tree and confirmed the presence of four Hopea species (H. odorata, H. hainanensis, H. mollissima, and H. chinensis) in nature reserves and natural parks of Vietnam. These species showed a close relationship with an average

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

genetic distance of 0.0045; both matK and rbcL separated all species, but their use in combination gave higher bootstrap values. The matK region was found to provide the most reliable barcode for the identification of the most closely related Dipterocarp species. Our study provides a means to identify rare *Hopea* species non-ambiguously and to support the protection of this decreasing natural genetic resource.

Key words: *Hopea chinensis*; Chloroplast DNA; Genetic relationship; Dipterocarpaceae

INTRODUCTION

The tree family Dipterocarpaceae is widespread in the tropics and is very important for its timber; Dipterocarp species represent a large part of the international timber market and play an important role in many countries, particularly those of South East Asia (Appanah and Turnbull, 1998). In addition to timber, Dipterocarp species also yield many other valuable products such as essential oils, balsam, resins, camphor, and tannins (Ban, 2003). In Vietnam, more than 40 Dipterocarp species are known; these fall into 6 genera (*Anisoptera, Hopea, Shorea, Parashorea, Vatica*, and *Dipterocarpus*) that are mostly native and endemic (Nghia, 2005).

Hopea chinensis Hand-Mazz (synonym *H. hongayensis* Tardieu) is a medium-sized tree reaching 30-50 m in height and 40-70 cm in diameter at maturity; it was first identified in Quang Ninh Province (Vietnam) by Tardieu in 1942. Due to over logging and habitat loss, this species is now threatened; in China, for example, it is now only found in a narrow range in Guangxi Province (Flora of China). The stem bark of *H. chinensis* contains many useful biologically active substances, such as immunosuppressive polyphenols and acetylcholine esterase inhibitors, and novel polyphenols (hopeachinols and diptoindonesin) have been isolated from ethanol extracts (Ge et al., 2010; Yan et al., 2012). The durable wood is used for making boats and furniture, and for building bridges. Globally, *H. chinensis* is listed as Critically Endangered under IUCN (2014) (http://www.iucnredlist.org/details/32356/0) criteria. In Vietnam, due to habitat reduction and over-exploitation, this species is listed as endangered (MOST and VAST, 2007).

Identifying *H. chinensis* individuals in the wild is problematical as the species shares many morphological characteristics with H. mollissima C.Y. Wu (Ashton, 1998; Nghia, 2005). One possible means to overcome this difficulty would be to establish a reliable DNA barcode. DNA barcodes have been developed using the sequences of specific regions of chloroplast DNA, and they have rapidly become an important tool for species identification. The current recommendation is the use of a two-marker combination of chloroplast rbcL and matK as the core plant barcode, supplemented with the more variable trnH-psbA markers (CBOL Plant Working Group, 2009). In the present study, our objectives were 1) to test all three recommended DNA marker regions for use in identifying *H. chinensis* and for elucidating the relationship of this species with three other Hopea species (H. mollissima, H. hainanensis Merr. et Chun., and H. odorata Roxb.) and 2) to measure the relative genetic distances in the genus Hopea and in Dipterocarp species in Vietnam [Dipterocarpus intricatus Dyer., D. dyeri Pierre, D. tuberculatus Roxb., D. costatus C.F. Gaertn. and Parashorea chinensis (H.C. Wang) H. Zhu]. Estimation of the genetic distances of closely related species in the genus Hopea and of less closely related Dipterocarp species will enable us to evaluate the efficiency of three potential DNA barcodes to distinguish H. chinensis and H. mollissima.

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

MATERIAL AND METHODS

Plant materials

H. mollissima and *H. chinensis* share many morphological characteristics: at maturity, they are medium to large trees, 20-40 m in height, 40-60 cm in diameter; the bark is greyishbrown. In mature trees, barks is splitted into large pieces causing circular scars on the stem; the leaves are simple and alternate, and the leaf blades are ovate, 10-20 cm long, 3-6 cm wide, with a round base and a pointed apex; the flowers are small, with 5 sepals and 5 petals (Figures 1 and 2). In an attempt to avoid identity confusion, morphological characteristics that show some differences between the two species are generally employed (Table 1).

In the present study, ten leaves or bark samples were obtained from nine Dipterocarp species from eight different areas in Vietnam (Table 2 and Figure 3).

The collected leaves and bark samples were dried immediately on silica gel in the field and then transported to the Institute of Ecology and Biological Resources (IEBR) for DNA extraction. Herbarium specimens were also collected simultaneously for the confirmation of species identification by botanists; these specimens were stored at the IEBR, Vietnam Academy of Science and Technology (VAST).



Figure 1. Similar morphological characteristics of *Hopea mollissima* (A) and *H. chinensis* (B) make it difficult to distinguish the two species.



Figure 2. Hopea mollissima (A) and H. chinensis (B) have flowers with similar appearances.

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

©FUNPEC-RP www.funpecrp.com.br

Nguyen T.P. Trang et al.

 Table 1. Morphological characteristics that can be used to differentiate Hopea chinensis and H. mollissima.
Characteristic H. chinensis H. mollissima Leaf Upper surface green and shiny, Upper surface covered with scattered lower surface light color; stellate hairs; lower surface tomentose, light brown when dry; both surfaces glabrous especially on veins Petiole Dark brown when dry Horizontally split Fruit Acutely ovoid Globose

Table 2. Locations used to collect the nine Dipterocarp species in Vietnam.

	Species	Collection place	Province	Altitude	Coordinates
1	Dipterocarpus intricatus	Binh Chau Nature Reserve	Ba Ria - Vung Tau	100 m	10°28'N-107°35'E
2	Dipterocarpus dyeri	Tan Cuu, Vinh Cuu	Dong Nai	129 m	11°12'N-107°09'E
3	Dipterocarpus costatus	Bu Gia Map National Park	Binh Phuoc	130 m	10°56'N-106°59'E
4	Dipterocarpus tuberculatus	Yok-Don National Park	Dak Lak	150 m	12°47'N-107°35'E
5	Parashorea chinensis	Cuc Phuong National Park	Ninh Binh	150 m	20°19'N-105°36'E
6	Hopea mollissima	Khe Ro Nature Reserve	Bac Giang	130 m	21°09'N-21°13'E
7	Hopea odorata	Ben En National Park	Thanh Hoa	100 m	19°35'N-105°30'E
8	Hopea chinensis	Ba Mun Island, Bai Tu Long National Park	Quang Ninh	120 m	21°02'N-107°35'E
9	Hopea chinensis	Cai Lim Island, Bai Tu Long National Park	Quang Ninh	150 m	21°06'N-107°33'E
10	Hopea hainanensis	Ben En National Park	Thanh Hoa	100 m	19°35'N-105°30'E



Figure 3. Location of the study sites of the nine Dipterocarp species.

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

DNA extraction

Total DNA was extracted from the samples using the modified CTAB method of Doyle and Doyle (1987). Liquid nitrogen was added to each sample (about 100 mg), which was then ground by hand. Total DNA yield and purity were assessed using a spectrophotometer and were then visualized on 1% agarose gels. Stock DNA was diluted to a concentration of 10 ng/ μ L.

DNA amplification and sequencing

PCR was performed in a 40- μ L reaction volume containing 4 μ L PCR 10X buffer, 1 μ L 25 mM dNTP, 1 μ L 20 μ M of each primer, 1 μ L 25 mM MgCl₂, 1 μ L BSA, 1 μ L 2.5 U Taq DNA polymerase and approximately 50 ng genomic DNA. The rbcL, matK, and trnH-psbA genes from Dipterocarp samples were amplified using a standard protocol (Table 3).

Table 3. Sequence information of the three candidate genes.								
DNA name	Foward primer	Reverse primer	Tm (°C)	Length PCR product (bp)	Reference			
matK	CGATCTATTCATTCAATATTTC	TCTAGCACACGAAAGTCGAAGT	50	900	Shaw et al. (2005)			
rbcLa	TCTAGCACACGAAAGTCGAAGT	CTTCGGCACAAAATACGAAACG ATCTCTCCA	56	700	Hasebe et al. (1994)			
rbcLc	TGAAAACGTGAATTCCCAACC GTTTATGCG	GCAGCAGCTAGTTCCGGGCTCCA	56	700	Hasebe et al. (1994)			
trnH-psbA	GTTATGCATGAACGTAATGCTC	CGCGCATGGTGGATTCACAATCC	48	300	Kress and Erickson (2007)			

The amplification conditions for the two rbcL fragments (rbcLa and rbcLc), each approximately 700 bp in length, and for the 300-bp fragment of trnH-psbA, were 95°C for 5 min; followed by 30 cycles of 95°C for 30 s, 48°C for 30 s, and 72°C for 1 min. The reaction was completed by a 10-min extension and hold at 4°C. The amplification conditions for the 900-bp matK fragment were: denaturation at 95°C for 5 min; 30 cycles of 40 s at 95°C, 40 s at 52°C and 1 min at 72°C, followed by extension at 72°C for 10 min and hold at 4°C. The identities of the PCR products were verified by electrophoresis on 0.8% agarose gels. All PCR products were purified using a QIA quick PCR purification kit (Qiagen, Germany). The purified PCR products were sequenced in both directions with the same primers as for PCR using a Bigdye terminator v3.1 cycle sequencing kit (Applied Biosystems) on an ABI 3100 capillary sequencer following the manufacturer instructions.

DNA analysis

Bidirectional DNA sequences of each fragment were assembled using the ChromasPro software (Technelysium) and then aligned by Mega 5.0 (Tamura et al., 2011). The p-distance of rbcL, trnH-psbA, and matK was calculated using MEGA 5.0 in order to evaluate intra-specific and inter-specific divergence. Neighbor-joining (NJ) trees based on p-distance were constructed using MEGA 5.0 to provide a graphical representation of genetic divergence among species (Tamura et al., 2011).

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

RESULTS

PCR and sequencing success

The efficiency of PCR amplification for matK, rbcL, and trnH-psbA was 100, 85.7, and 92.8%, respectively. The three candidate DNA barcoding markers were fully sequenced. The matK and rbcL sequencing results from the nine tested Dipterocarp species were submitted to GenBank with the accession Nos. from KM 267143 to KM 267152 for rbcL and the accession Nos. from KJ 611230 to KJ 611232, KJ 611235, and from KJ 611237 to KJ611241 for matK (Table 4).

	Species	Accession No.			
		matK	rbcL		
1.	Hopea chinensis Hand-Mazz	KJ611239	KM267146		
	(synonym H. hongayensis Tardieu)				
2.	Hopea mollissima C.Y. Wu	KJ611237	KM267145		
3.	Hopea odorata Roxb.	KJ611238	KM267144		
4.	Hopea hainanensis Merr et Chun	KJ611240	KM267147		
5.	Parashorea chinensis (H.C. Wang) H. Zhu	KJ611235	KM267143		
6.	Dipterocarpus tuberculatus Roxb.	KJ611232	KM267148		
7.	Dipterocarpus costatus C.F. Gaertn.	KJ611241	KM267152		
8.	Dipterocarpus dyeri Pierre	KJ611231	KM267151		
9.	Dipterocarpus intricatus Dyer.	KJ611230	KM267150		

Alignment and variability

For the nine species, an aligned matK sequence of 829 bp was obtained; this contained 60 variable sites, of which 30 were informative parsimony sites. The aligned rbcL sequence was 1336 bp with 52 variable sites, of which 42 showed informative parsimony. The aligned sequence of trnH-psbA was 195 bp long with 6 informative parsimony sites (Table 5).

Table 5. Sequence differences in the matK-rbcL region between Hopea chinensis and other Dipterocarp species.										
	Position									
Species	10	20 *	30 *	40 *	50 *	60 *	70 *	80 *	90 *	100
H. chinensis CCGACAGGCGCCGCCCTCCCATAATCCTGCACAGGGACGCATAAAAAAAGGGCTTATCACCTCATTTACGGGTTTATACTGACG4					GTAGAACAC	ACATAGCG				
H. chinensis										
H. mollissima	.ACC									
H. odorata										
H. hainanensis	GC.AG									
P. chinens	G.C. AT. C.A. GCCTGT. CTGCTGC.CC. G.G. CCGT. CT. A.CG.G.AT. C. GAC. GT.A.G. GA.T. TT.									
D. tuberculatus	GCTAT.TTCGT.C.ATGCGTT.TGCTG.TC.T.A.C.GGG.CGT.CTTG.ATC.CGG.TG.CCGACAGTTAGG.G.TCATT.A.									
D. intricatus	GCTAT.TTCGT.C.ATGCC.GTT.TGCTGCTCCT.A.C.GCG.CGT.CTTG.ATC.CGG.TG.CCGACAGTTAGG.G.TC.TAT.T.A									
D. costatus	GCTAT.TTCGTTATGCC.GTT.TGCTGCTCCT.A.C.GCG.CGT.CTGGGCTG.ATC.CGG.TG.CCGACAGTTAGG.G.TC.TATT.A									
D. dyeri	GCTAT.TTCG	T.C.ATGC	GTT.TGCTGC	TC.T.A.C.G	GG.CGT.CT.	.TG.ATC.CG	G.TG.CCG	ACAGTTAGG.	G.TCAT.	T

Comparison of Hopea chinensis and H. mollissima

In addition to using morphological characteristics to distinguish the two *Hopea* species, we also made use of barcoding markers based on the chloroplast genome and construction of an NJ tree. DNA barcoding can provide accurate, rapid, and automated species identifica-

tion using a short fragment of genomic DNA, and has been widely used for the authentication of plant species. Identifying single or multiple regions that can be used non-ambiguously for barcoding has been an important research focus. The recent favored candidates are the matK, rbcL, trnH-psbA, rpoC1, and ycf5 sequences of the chloroplast genome (Kress and Erickson, 2007). A two-locus combination of matK and rbcL was recommended as a plant barcode by the CBOL Plant Working Group (2009). In our study, alignment of the matK sequence identified 60 variable sites, 30 of which were informative. We also found that rbcL possessed 52 variable sites with 42 informative sites. The nucleotide sequences for matK and rbcL gave similar genetic distance values between *H. chinensis* and *H. mollissima* species of approximately 0.002. However, the amplification efficiency obtained for the longer rbcL sequence (1336 bp) was 85.7% compared to 100% for the shorter matK sequence (821 bp). The aligned trnH-psbA sequence was 195 bp long and did not show any differences between *H. chinensis* and *H. mollissima* (Figure 4).



Figure 4. Species identification based on p-distances using each DNA barcode separately and in different combinations.

Comparison of Hopea and other Dipterocarp species

Our phylogenetic analysis of the nine Dipterocarp species using an NJ tree based on matK and rbcL sequences identified the two major groups of Valvate-Dipterocarp and Impricate-Shorea. The combination of rbcL and matK sequences gave the highest bootstrap values (92) for the paraphyletic clade of tribe Shorea (*Hopea* and *Shorea*) (Figure 5). In the first major group, *H. mollissima* has a very close relationship with a group comprised of *H. chinensis*, *H. odorata*, and *H. hainanensis*. *H. mollissima* and *H. chinensis* showed the closest genetic relationship with a genetic distance of 0.002.

matK and rbcL have been recommended as a potential core barcode for land plants (CBOL Plant Working Group, 2009) because they possess richly variable and informative sites. In our study, a combination of matK and rbcL identified 14 different nucleotides at positions 48, 62, 64, 229, 630, 1465, 1479, 1633, 1688, 1822, 1838, 1875, 1932, and 1953. This analysis clearly confirmed that *H. mollissima* and *H. chinensis* are two different species even though they share many similar morphological characteristics.

Genetics and Molecular Research 14 (3): 9181-9190 (2015)



Figure 5. Phylogenetic trees based on matK and matK+rbcL gene sequences in the plant samples.

DISCUSSION

To date, species of the plant family Dipterocarpaceae have been difficult to classify. Previous studies have shown that the Dipterocarpaceae is a complex family with many inconsistent taxonomic results with regard to the identification of species and genera. Yamazaki et al. (1999) used matK, trnH-psbA, and rbcL sequences to analyze the genetic relationships between Dipterocarpaceae species in Southeast Asia and found that *Hopea* is closest to *Neobalanocarpus heimii* (*Shorea*). However, phylogenetic analysis of Dipterocarpaceae species based on sequencing of the genomic regions trnL-trnF, matK, and trnL indicated that *Hopea* was closest to *Neobalanocarpus heimii* and independent of *Shorea* (Gamage et al., 2006). Analyses of the Dipterocarpaceae suggests that there is a low level of genetic variation among species in this family (Yamazaki et al., 1999; Kamiya et al., 2005) Therefore, it is important to use a combination of gene markers for identifying species and to support the determinations made by standard morphological methods. In this study, we performed an analysis with three gene sequences, matK, rbcL, and trnH-psbA, and our results using pairs of sequences reinforce this conclusion.

Although the efficiency of amplification of trnH-psbA was very high (92.8%), the short length of the sequence and its lack of polymorphisms indicated that it was unsuitable as a DNA marker for Dipterocarpaceae identification. This sequence did not identify nucleotide differences between *H. mollissima* and *H. chinensis*, or between *D. intricatus* and *D. costatus*. Thus, even in combination with other sequences it would be uninformative in terms of species identification. trnH-psbA has been shown to be a good barcoding site in other studies because

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

it is one of the most rapidly evolving spacer sequences in chloroplast DNA and also has a 75-bp conserved fragment at the end; additionally, it is universal and has high amplification success (Yao et al., 2009; Luo et al., 2010).

The CBOL Plant Working Group (2009) compared the success rates of rpoB, rpoC1, rbcL, matK, trnH-psbA, atpF-atpH, and psbK-psbI in 550 species belonging to angiosperms, gymnosperms, and algae. They found that matK + rbcL successfully discriminated 72% of the species and proposed use of this sequence as the standard barcode for land plants. Our results clearly confirm that the combination of matK and rbcL regions greatly improved the discriminatory power of DNA barcoding in the Dipterocarpaceae.

Our phylogenetic analysis of nine species gave higher bootstrap values for genetic relationships when based on a combination of matK and rbcL sequences than that obtained when using only the matK gene. Thus, although only the matK gene of the three molecular markers tested here gave effective classification in Dipterocarp species when used on its own, more reliable results were obtained by a combined analysis using matK and rbcL.

According to the CBOL Plant Working Group (2009) standards, an ideal barcode should be a short segment that allows easy amplification and possesses sufficient variation among sequences to distinguish between species. For bisexual species, such as Dipterocarps, chloroplast DNA markers are the best choice for identification. Our study confirmed that use of barcoding, especially of trnH-psbA, could yield valuable information on PCR product content and enable the identification of their compositions. The use of matK on its own or of matK combined with rbcL was shown to be suitable to barcode all of tested Dipterocarpaceae species, whereas trnH-psbA could not be used alone for this purpose.

ACKNOWLEDGMENTS

Research supported by the Vietnam National Foundation for Science and Technology Development (#NAFOSTED) under grant #106-NN.06-2013.08 and BAS42 (Basisfinanciering, Vrije Universiteit Brussels).

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Appanah S and Turnbull JM (1998). A review of Dipterocarps: taxonomy, ecology and silviculture. CIFOR, Bogor.

Ashton P (1998). *Hopea chinensis*. IUCN Red List of Threatened Species. Version 2012. International Union for Conservation of Nature.

Ban NT (2003). Checklist of plant species of Vietnam (in Vietnamese). Vol. 2. Agricultural Publishing House, Hanoi.

- CBOL Plant Working Group (2009). A DNA barcode for land plants. Proc. Natl. Acad. Sci. U. S. A. 106: 12794-12797.
- Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.
- Gamage DT, de Silva MP, Inomata N, Yamakazi T, et al. (2006). Comprehensive molecular phylogeny of the sub-family Dipterocarpoideae (Dipterocarpaceae) based on chloroplast DNA sequences. *Genes Genet. Syst.* 81: 1-12.
- Ge HM, Yang WH, Shen Y, Jiang N, et al. (2010). Immunosuppressive resveratrol aneuploids from *Hopea chinensis*. *Chemistry* 16: 6338-6345.

Hasebe M, Omori T, Nakazawa M, Sano T, et al. (1994). rbcL gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. *Proc. Natl. Acad. Sci. U. S. A.* 91: 5730-5734.

Genetics and Molecular Research 14 (3): 9181-9190 (2015)

IUCN (2014). IUCN Red List of Threatened Species: version 2014.3.

- Kamiya K, Harada K, Tachida H and Ashton PS (2005). Phylogeny of PgiC gene in *Shorea* and its closely related genera (Dipterocarpaceae), the dominant trees in Southeast Asian tropical rain forests. *Am. J. Bot.* 92: 775-788.
- Kress WJ and Erickson DL (2007). A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS One* 2: e508.
- Luo K, Chen S, Chen K, Song J, et al. (2010). Assessment of candidate plant DNA barcodes using the Rutaceae family. Sci. China Life Sci. 53: 701-708.
- MOST (Ministry of Science and Technology) and VAST (Vietnamese Academy of Science and Technology) (2007). Vietnam Red Data Book. Part II. Plants. Khoa Hoc Tu Nhien & Cong Nghe, Vietnam.

Nghia NH (2005). Dipterocarps of Vietnam. Agriculture Publishing House, 72-78.

- Shaw J, Lickey EB, Beck JT, Farmer SB, et al. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92: 142-166.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA 5.0: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony method. *Mol. Biol. Evol.* 28: 2731-2739.
- Yamazaki T, Suchitra C, Harada K, Clyde MM, et al. (1999). Molecular phylogeny of Dipterocarpaceae in Southeast Asia. In: The Tokyo International Forum on Conservation and Sustainable Use of Tropical Bioresources. Nov. 9-10, 1998, 121-131.
- Yan T, Wang T, Wei W, Jiang N, et al. (2012). Polyphenolic acetylcholinesterase inhibitors from *Hopea chinensis*. *Planta Med.* 78: 1015-1019.
- Yao H, Song JY, Ma XY, Liu C, et al. (2009). Identification of *Dendrobium* species by a candidate DNA barcode sequence: the chloroplast psbA-trnH intergenic region. *Planta Med.* 75: 667-669.

Genetics and Molecular Research 14 (3): 9181-9190 (2015)