

Nucleotide Diversity of a Nuclear and Four Chloroplast DNA Regions in Rare Tropical Wood Species of *Dalbergia* in Vietnam: A DNA Barcode Identifying Utility

Dinh Thi Phong^{*1}, Duong Van Tang¹, Vu Thi Thu Hien¹, Nguyen Dang Ton², Nong Van Hai²

¹Vietnam National Museum of Nature (VNMN), Vietnam Academy of Science and Technology (VAST),
18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam.

²Institute of Genome Research (IGR), Vietnam Academy of Science and Technology (VAST),
18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam.

*Corresponding author: E-mail: dinhthiphong@hotmail.com;

ABSTRACT--- Five nucleotide regions (a nuclear ribosomal DNA internal transcribed sequence (ITS) and four chloroplast DNA fragments (*trnL* intron, *matK*-designed, *matK*-barcode and *psbA-trnH* spacer) were used for the assessment nucleotide diversity for the eight woody tree species in genus *Dalbergia* (*Dalbergia oliveri* Gamble ex Prain, *Dalbergia cochinchinensis* Pierre, *Dalbergia nigrescens* Kurz, *Dalbergia tonkinensis* Prain, *Dalbergia assamica* Benth, *Dalbergia entadoides* Pierre ex Prain, *Dalbergia hancei* Benth and *Dalbergia dialoides* Pierre Niyomdham) of Vietnam. The number of variable sites for individual loci ranged from 13 (*psbA-trnH* spacer) to 111 (ITS). The highest divergence region was ITS (7.4 %), followed by *psbA-trnH* spacer (2.2 %), *trnL* intron (1.5 %), *matK*-designed (1.2 %) and *matK*-barcode (1.3 %). Analyses based on threshold genetic distances, Phylogenetic trees-based and nucleotide character differences methods showed that only nuclear ITS region is the best nucleotide diversity of rare woody tree species in *Dalbergia* genus. Therefore, it can be used as an identification of eight *Dalbergia* species.

Keywords --- *Dalbergia*, nucleotide regions, phylogeny, rare, woody trees

1. INTRODUCTION

Dalbergia is a large genus of Fabaceae family, subfamily Faboideae. They are mainly distributed in the tropics and subtropics with variable forms from shrubs to timber trees. The number of species recognized by various authorities has fluctuated, however according to the Global Biodiversity Information Facility (GBIF), there are about 300 species of this genus ([http://data.gbif.org/tutorial/datauseagreement/GBIF Data Sharing Agreement/dalbergia](http://data.gbif.org/tutorial/datauseagreement/GBIF>Data-Sharing-Agreement/dalbergia)). In Vietnam, 27 species of *Dalbergia* have been described (Dang and Nguyen 2007), including five valuable woody trees (*D. oliveri*, *D. cochinchinensis*, *D. nigrescens*, *D. tonkinensis*, *D. assamica*). These have all been overexploited all because of their economic and commercial values. The Vietnam Red List classified *D. cochinchinensis* into group EN (Endangered) A1acd; *D. tonkinensis* into group VU (Vulnerable) A1acd and *D. oliveri* into group EN A1acd. Those are considered as endangered species groups facing a risk of extinction (Dang and Nguyen 2007). The International Union for Conservation of Nature (IUCN) reported that the natural populations of *D. oliveri* and *D. cochinchinensis* are disappearing and only limited numbers are found in the remaining forest fragments in the southern part of Vietnam (IUCN 1998).

The development of a reference database of DNA sequences for plant species has been difficult because of a lack of suitable DNA regions that can serve as a barcode (CBOL Plant Working Group 2009; Cowan et al. 2006). It is most desirable to have a single locus or a few loci that have highly conserved universal primer sequences, while at the same time having sufficient nucleotide variation to identify species. Although the difficulties of plant barcoding have been debated (Chase et al. 2005; Cowan et al. 2006; Pennisi 2007), detailed studies have demonstrated the utility of barcoding as an effective tool for identification (Newmaster et al. 2008; Lahaye et al. 2008; Fazekas et al. 2008). However, until now the study barcode for wood species of *Dalbergia* has not done yet. Only some *Dalbergia* species are interested to the population diversity using markers SSR, RAPD and some gene region for conservation purpose (Buzatti et al, 2012; Ashraf et al., 2010; Ribeiro et al., 2007, 2011).

The plastid loci have been considered as the most promising candidate barcodes for plant species (Chase et al. 2005; Kress and Erickson 2005; Lahaye et al. 2008) due to their low rates of nucleotide substitution in contrast to the high rates of chromosomal rearrangements within the plant mitochondrial genome (Palmer and Herbon 1988) and extensive gene duplication in the nuclear genome (Alvarez and Wendel 2003). Candidate barcodes from the plastid genome include both slowly evolving coding regions (*rbcL*, *rpoB*, *matK*) and more rapidly evolving loci (*rps2*, *psbA-trnH* spacer, *trnT-trnF*). However, when used alone, the more conservative loci may not possess enough nucleotide variation to discriminate among closely related species, and more variable loci may be problematic because of homoplasy. To overcome these problems, a multilocus DNA barcoding system has been suggested by some authors including Chase et al. (2005); Fazekas et al. (2008) and CBOL (The Consortium for the Barcode of Life) Plant Working Group (2009). The slowly evolving loci delineate individuals in families, genera, or groups within genera and the more rapidly evolving loci differentiate species within those higher groups (Newmaster et al. 2006; Kress and Erickson 2007). The most recent suggestions came from a group of 52 scientists using the multilocus barcode including *matK* and *rbcL* (CBOL Plant Working Group 2009).

In addition, application of DNA barcoding requires analytical methods that accurately assign query sequences of unknown taxonomic identity to species based on the sequences contained in the reference database. Consequently, the utility of DNA barcoding primarily rests on the ability to discriminate among closely related species (i.e., congeners), which has only begun to be tested recently (Newmaster et al. 2008; Fazekas et al. 2009; Hollingsworth et al. 2009).

In order to conserve genetic diversity, DNA database of *Dalbergia* species in Vietnam should be established. Recognizing this need, we conducted a study aimed at developing suitable nucleotide regions for identifying eight rare woody trees species in genus *Dalbergia* of Vietnam. This information will aid in the knowing whether DNA barcoding can serve as a tool for field botanists to identify the taxonomy species that are challenging to classify based on morphology.

2. MATERIAL AND METHODS

2.1. Sampling materials

Leaf samples or wood pieces of *D. oliveri*, *D. cochinchinensis*, *D. nigrescens*, *D. tonkinensis*, *D. assamica*, *D. entadoides*, *D. hancei* and *D. dialoides* (five samples for each species) were collected from natural stands grown across the country, including from the Yordon National Park (Dak Lak province), Cuc Phuong National Park (Ninh Binh province), Phong Nha-Ke Bang National Park (Quang Binh province), Cobia Nature Reserve (Thuan Chau, Son La province) to some suburb areas of Hanoi. Yordon National Park is the largest of Vietnam's nature preserves and one of seven internationally recognized important Centers of Plant Diversity in Vietnam. This park encompasses over 1,000 km² and extends from eastern border to Cambodia into northern Dak Lak and southern Gia Lai provinces in Vietnam. Most of the topography of this park is flat, with approximately 200 m above the sea level. Cuc Phuong is the oldest national park, located only 120 km southwest of Hanoi and nestled between the provinces of Ninh Binh, Hoa Binh and Thanh Hoa. Cuc Phuong is situated in the foothills of the northern Truong Son Mountains. The park consists of verdant karst mountains and lush valleys. Elevation varies from 150 m to 656 m at the summit of May Bac Mountain (Silver Cloud Mountain). The area of this park is about 220,00 km². Phong Nha - Ke Bang National Park is located in the middle of the Truong Son Mountain Range 500 km from Hanoi, and close to the Vietnam - Laos border, just several kilometers to the west. Phong Nha - Ke Bang National Park is one of the world's two largest limestone regions. The park covers a total area of 857.54 km² which are divided into three zones: a "strictly protected zone" (648.94 km²), an "ecological recovery zone" (174.49 km²), and an "administrative service zone" (34.11 km²). Cobia Nature Reserve is located in Thuan Chau district, Son La province. The proposed nature reserve is centered on Mount Cobia, a 1,800 m peak. The area of this park about 19,253 hectares.

The collection places of the eight species used in this study are shown in Figure 1. The numbers of individual samples and conservation status are shown in Table 1. The samples were stored at -20°C until used.

2.2. Total DNA extraction

Total genomic DNA was isolated from leaves and wood specimens using the method described by Porebski et al. (1997) and DNA purification by Genomic DNA Purification Kit (Code #512 Fermentas). The concentration of DNA was determined by UV visible spectrophotometer (UVS 2700, Labomed, USA).

2.3. Selecting nucleotide regions

Previous DNA barcoding analyses of molecular data (White et al. 1990; Ribeiro et al. 2007; Lavin et al. 2000; Sang et al. 1997; Tate et al. 2003; Fazekas et al. 2012) suggested that several DNA regions are suitable for barcoding plants. Based on these studies, we chose five regions to examine. Those include ITS, *trnL* intron, *psbA-trnH* spacer, and two overlapping region of *matK*: the "*matK*-barcode used by Fazekas et al. (2012) and *matK*-designed" prepared for this study (Fig. 2) for barcoding eight *Dalbergia* species of Vietnam. Primer sequences for the regions of ITS (ITS1-f/ITS4-r), *trnL* intron (*trnL*-f/*trnL*-r), *matK*-designed (*matK*-f/*matK*-r), *matK*-barcode (*kim1R*/*kim3F* of Ki-Joong Kim as in the

reference of Fazekas et al. 2012) and *psbA-trnH* spacer (*psbA-f/trnH-r*) were used for the study. The information of primer pairs and the map of *matK*-designed and *matK*-barcode is presented in Table 2 and in Fig. 2.

2.4. Amplification and sequencing

DNA was amplified in 0.025 cm³ reaction mixture containing 1x PCR buffer (Fermentas, EP0701), 0.2 mM of each dNTP, 0.25 μM of each primer, 2.5 mM MgCl₂, 2.5 U of DreamTaqTM (Fermentas, EP0701) and 10 ~ 20 ng of template DNA. PCR products were amplified on a GeneAmp[®] PCR System 9700, (Applied Biosystems, USA). The temperature cycle was 3 min at 95°C, followed by 35 cycles of 30 s at 95 °C, 30 s at 58 °C and 1 min at 72 °C. The PCR was terminated by a final extension step of 72 °C for 10 min. For the partial *trnL* region, the PCR mixture and reaction conditions were the same as described above except the primer annealing temperature was 52 °C. The purified PCR products were sequenced using ABI PRISM[®] 3100 Genetic Analyzer (Applied Biosystems, USA). The bigDye terminator 3.1 cycle sequencing kit was used followed the manufacturer's protocol.

2.5. Data analysis

For each PCR sequence, forward and reverse sequences were aligned and traces examined using CodonCode Aligner (Codon Code Corporation). Sequences were aligned and trimmed using the ClustalW algorithm in MEGA v. 4.0.2 (Tamura et al. 2007) with gap opening penalty 15, gap extension penalty 7.5, IUB weight matrix, transition weight 0.5 and delay divergent cut-off 50. MEGA v. 4.0.2 was also used to calculate the proportion of sites differing between pairs of sequences. These are quoted in the text as percentage differences. In the phylogenetic inference method, gaps were not considered as character states. The phylogenetic trees were reconstructed using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) analysis and bootstrap test of phylogeny with 1000 replications (Saitou and Nei 1987). The sequence distances were calculated using nucleotide substitution model: Maximum Composite Likelihood (Tamura et al. 2004). The gaps and missing data were deleted. Substitutions include the transition and transversions. Pattern among lineages was homogeneous and rates among sites were uniform.

3. RESULTS

3.1. Character analysis of gene regions

All five selected regions of genomes (ITS, *trnL* intron, *matK*-designed, *matK*-barcode and *psbA-trnH* spacer) were amplified for eight species with the same annealing temperature 56°C. The DNA fragments obtained are in size 700 bp, 510 bp, 750 bp, 850 bp and 450 bp, respectively. We were also able to extract DNA from wood product used in furniture that was PCR amplified with these primer pairs. Successful DNA extraction, amplification and sequencing from used materials are useful to develop a forensic identification method for pieces of wood. The cpDNA regions and nuclear ITS contained variable numbers of single bp (base pair) substitutions and varying numbers of insertions/deletions (indels) that were variable in length (Table 3). No intraspecific variability was observed for the occurrence of substitutions with cpDNA regions, but variation was observed in sequences of the nuclear ITS region within one species (data not shown). Comparison of the fragment nucleotides across the eight species indicated that nuclear ITS region has the highest sequence divergence values.

The number of variable sites for individual loci among all species ranged from 111 (ITS) to 13 (*psbA-trnH*). Interspecies nucleotide difference was only observed in the ITS region. The nucleotide diversity within species is quite small compared with the diversity between species. The mean sequence divergence of ITS regions of the eight species was 7.4 %, while it was 2.2 %, 1.5 %, 1.2 % and 1.3 % for *psbA-trnH* spacer, *trnL* intron, *matK*-designed, *matK*-barcode regions, respectively. The highest sequence divergence value across eight species based on the mean nucleotide diversity was 7.4 % for ITS. We also checked the divergence to distinguish species on eight species, ITS is the most powerful, with 100 % discriminated species pairs (eight species made up of 28 pairs). Other regions do not have a sufficient power to distinguish all species (Table 3).

3.2. Nucleotide diversity

The percentage differentiation based on nucleotide diversity among *Dalbergia* species ranged from 1.1 % (*D. oliveri* and *D. hancei*) to 10.7 % (*D. entadoides*) for ITS gene, from 0.0 % (*D. dialoides* and *D. entadoides*) to 2.8 % (*D. cochinchinensis*) for *trnL* intron gene, from 0.0 % (*D. dialoides* and *D. tonkinensis* and *D. entadoides* or *D. hancei*, and *D. oliveri*) to 4.1 % (*D. nigrescens*) for *psbA-trnH* spacer, from 0.0 % (*D. entadoides* and *D. dialoides* or *D. hancei* and *D. oliveri*) to 2.0 % (*D. cochinchinensis*) for *matK*-designed and from 0.0 % (*D. entadoides* and *D. dialoides* or *D. hancei* and *D. oliveri*) to 1.4 % (*D. cochinchinensis*) for *matK*-barcode gene (Table 3).

3.3. Phylogeny based on gene regions

Because individuals of the same species have the identical sequence, only one individual of each species was selected to reconstruct phylogenetic tree from five DNA regions of eight *Dalbergia* species, species were separated supported by bootstrap values > 50 %. Analysis of gene regions indicates that variation in bp substitutions within the nuclear ITS

region could separate eight species of *Dalbergia* (Fig. 3a). The other regions were not able to separate eight species (Fig. 3d, Fig. 3c, Fig. 3d and Fig. 3e). However, the phylogenetic trees from two gene *matK*-designe and *matK*-barcorde for eight species were showed similar (Fig. 3d and 3e). Ability of separating of *trnL* intron regions was 96 % of total species pairs and it was 92.9 % with *matK*-designed and *matK*-barcode and only 85.7 % with *psbA-trnH* spacer (Table 3).

Taken together, nuclear ITS region has the best nucleotide diversity of rare woody tree species in *Dalbergia* genus, therefore, can be used as an identification of eight *Dalbergia* species.

Total 120 nucleotide sequences of five DNA or gene regions have been obtained for eight *Dalbergia* species during this study and have been deposited in Genbank/EBI databases, including 120 accession numbers (FR854093 - FR854192), (HG313770 - HG313781) and (HG326215 – HG326222) available online.

4. DISCUSSION

The ITS region, *trnL*, *matK*-designed, *matK*-barcode and *psbA-trnH* spacer have been shown to provide utility as a DNA barcode in many studies, however limitation still exists. For example, the nuclear ITS region cannot be amplified for some taxa, the *psbA-trnH* spacer has many indels leading to difficulties in alignment and *matK* and *trnL* gene has low variation (Taberlet et al. 2006; Newmaster et al. 2009; Liu et al. 2010; Barrer et al. 2000; Pirie et al. 2007). To deal with the problems, this study was conducted to find a suitable region to precisely distinguish eight woody tree species of Vietnam. Our results showed that ITS sequence is the most promising testing region. Another advantage of the ITS region is that it can be amplified in two smaller fragments (ITS1 and ITS2) and therefore is especially useful for degraded samples (Starr et al. 2009; James and Neel 2010; Sylvain and Bremer 2002). This DNA testing tool could overcome some limitations of traditional morphological identification method, which usually require training in plant morphology, access to suitable identification keys (in books, floras, or scientific papers) and suitable plant material (e.g. reproductive organs).

The *psbA-trnH* intergenic regions was proposed as DNA barcoding in previous studies (Song et al. 2009, Hao et al. 2010). This marker was tested in the study for its utility. The sequences of eight tree species were subjected to phylogenetic analyses, and a UPGMA tree generated by MEGA4 were shown in Fig. 3c. Among eight species, the phylogenetic relationship was not resolved, such as that between *D. oliveri* and *D. hancei* with a bootstrap value of 74 %, or that among *D. tonkinensis*, *D. entadoides* and *D. dialoides* with a bootstfrap value of 96 %. Previously, it has been shown the topology of the *psbA-trnH* spacer tree may reflect the suitability of this marker for DNA barcoding in Taxaceae but it is not suitable for Cephalotaxaceae (Hao et al. 2010). Our results obtained in this study of the eight *Dalbergia* genus are similar.

The *matK* gene is currently proposed as DNA barcode. This is given by many previous studies (Peter et al. 2009). The limitation preventing *matK* from becoming a DNA barcode is its inability to separate closely sister species, and as such has not yet been demonstrated to pass the test for a successful plant barcode. In this study, UPGMA trees of *matK*-designed and *matK*-barcode were not able to separate four species, such as *D. entadoides* and *D. dialoides* or *D. oliveri* and *D. hancei* with a bootstrap values of 90 % and 96 %, respectively (Fig. 3d and 3e).

Of the plastid loci tested, *psbA-trnH* spacer, *trnL* intron, *matK*-barcode are the only regions that meet all requirements identified by the Consortium for the Barcode of Life of being suitable for DNA barcoding. Those loci were routinely retrievable with single primer pair, easy to obtain bidirectional sequence reads, and provided maximal discrimination among plant species (CBOL Plant Working Group, 2009). While *psbA-trnH* spacer (Julian et al. 2009; Steven and Ragupathy 2009, Hao et al. 2010) and *matK*-barcode (Fazekas et al. 2012) have been proposed as DNA barcoding, our study reconfirms the appropriate use of those gene regions only for identification purpose to discriminate the three *Dalbergia* species of Vietnam, including *D. assamica* with a bootstrap value of 66 %, *D. cochinchinensis* and *D. nigrescens* with a bootstrap value of 81 %), (Fig. 3e).

The study on origins of Taxaceae and Cephalotaxaceae showed that the nuclear ITS region and cpDNA data formed a similar topologies of phylogenetic relationship, but the interspecies resolution in the nuclear ITS region was lower than cpDNA data (Hao et al. 2008). However, our study provides clear evidence that the ITS is the most appropriate region to distinguish eight *Dalbergia species* with a > 50 % bootstrap values (Fig. 3a). These loci have been eliminated as potential barcode loci in the previous studies, because of the lack of universal primers (either published or with the potential to be developed based on available information) and lack of success using existing primers. This is probably due to difficulties associated with of amplifying genes with low numbers of copies in degraded samples and the frequent need to clone PCR products before sequencing. Our research did not face these difficulties. The *psbA-trnH* intergenic spacer, *trnL* or *matK* gene is not powerful enough to distinguish all pair species of eight species. For example, *trnL* intron gene cannot separate *D. entadoides* and *D. dialoides* with a 74 % bootstrap value (Fig. 3b), Two gene *matK*-designed (Fig. 3d) and *matK*-barcode (Fig. 3e) cannot separate two pair species *D. entadoides*, *dialoides* (with bootstrap values of 90 % and 78 %, respectively) and *D. oliveri*, *D. hancei* (with bootstrap values of 96 % and 97 %, respectively) the *psbA-trnH* intergenic spacer (Fig. 3c) also cannot separate three species *D. tonkinensis*, *D. entadoides* and *D. dialoides* with a 71 % bootstrap

values. We suggest that the ITS is the best option for a DNA barcode sequence that has interspecific variation of *Dalbergia* genus in Vietnam.

7. ACKNOWLEDGEMENTS

This research was funded by the National Foundation for Science and Technology Development (NAFOSTED) of Vietnam. The authors gratefully acknowledge the assistance of Yordon National Park (Dak Lak), Cuc Phuong National Park (Ninh Binh), Copia Nature Reserve (Sown La) and Phong Nha - Ke Bang National Park (Quang Binh) from Vietnam for providing research samples.

8. REFERENCES

- [1] I. Alvarez, J. F. Wendel, “Ribosomal ITS sequences and plant phylogenetic inference”. *Mol. Phylogenet Evol.*, vol. 29, pp. 417-434, 2003.
- [2] M. Ashraf, A. B. Mumtaz, R. Riasat, S. Tabassum, “A molecular study of genetic diversity in Shisham (*Dalbergia sisso*) plantation of NWFP”, *Pakistan. Pak J Bot*, vol. 42, no. 1, pp. 79-88, 2010.
- [3] F. T. Barrer, A. Culham, C. E. Pkhurst, M. Gibby, “Mitochondrial and chloroplast DNA-based phylogeny of pelargonium (Geraniaceae)”, *Amer J Bot*, vol. 87, pp. 727-734, 2000.
- [4] O. R. S. Buzatti, R. A. Ribeiro, J. P. L. Filho, M. B. Iovato, “Fine-scale spatial genetic structure of *Dalbergia nigra* (Fabaceae), a threatened and endemic tree of the Brazilian Atlantic Forest”, *Gen Mol Bio*, vol. 35, no. 4, pp. 838-846, 2012.
- [5] CBOL Plant Working Group, “A DNA barcode for land plants”, *Proc Natl Acad Sci USA*, vol. 106, pp. 12794-12797, 2009.
- [6] M. W. Chase, N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, V. Savolainen, “Land plants and DNA barcodes: short-term and long-term goals”, *Biol Sci*, vol. 360, no. 1889-1895, 2005.
- [7] R. S. Cowan, M. W. Chase, W. J. Kress, V. Savolainen, “300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants”, *Taxon*, vol. 55, no. 611-616, 2006.
- [8] N. T. Dang, T. B. Nguyen, *Vietnam Red List*, Publishing House for Science and Technology Hanoi, 2007.
- [9] A. J. Fazekas, K. S. Burgess, P. R. Kesanakurthi, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, M. Hajibabaei, S. C. H. Barrett, “Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well”, *PloS ONE*, vol. 3, pp. 2802 – 2901, 2008.
- [10] A. J. Fazekas, P. R. Kesanakurthi, K. S. Burgess, D. M. Percy, S. W. Graham, S. C. H. Barrett, S. G. Newmaster, M. Hajibabaei, B. C. Husband, “Are plant species inherently harder to discriminate than animal species using DNA barcoding markers”, *Mol Ecol Res*, vol. 9, pp. 130-139, 2009.
- [11] A. J. Fazekas, L. Kuzmina, S. G. Newmaster, P. M. Hollingsworth, “DNA Barcoding methods for land plants. In: Kress WJ, Erickson DL (ed) *DNA Barcodes: Methods and Protocols*”, *Meth Mol Biol*, vol. 858, pp. 223-252, 2012.
- [12] D. C. Hao, S. L. Chen, P. G. Xiao, “Sequence characteristics and divergent evolution of the chloroplast *psbA-trnH* noncoding region in gymnosperms”, *J App Genet*, vol. 51, pp. 259-273, 2010.
- [13] D. C. Hao, P. G. Xiao, B. Huang, G. B. Ge, L. Yang, “Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences”, *Pl Syst Evol*, vol. 276, pp. 89-104, 2008.
- [14] P. M. Hollingsworth, L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, V. D. M. Bank, “A DNA barcode for land plants”, *Proc Natl Acad Sci USA*, vol. 106, pp. 12794-12797, 2009.
- [15] <http://data.gbif.org/tutorial/datauseagreement/> GBIF Data Sharing Agreement/dalbergia
- [16] IUCN (1998) *The World List of Threatened Trees*. World Conservation Press, UK
- [17] B. P. James, M. C. Neel, “An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae)”, *Amer J Bot*, vol. 97, pp. 1391-1406, 2010.
- [18] R. S. Julian, F. C. N. Robert, N. C. Btianna, “Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae)”, *Mol Ecol Res*, vol. 9, pp. 151-163, 2009.
- [19] W. J. Kress, D. L. Erickson, “A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region”, *PloS ONE*, vol. 2, pp. 508-610, 2007.

- [20] W. J. Kress, K. J. Wurdack, E. A. Zimmer, L. A. Weigt, D. H. Janzen, “Use of DNA barcodes to identify flowering plants”, *Proc Natl Acad Sci USA*, vol. 102, pp. 8369-8374, 2005.
- [21] R. Lahaye, M. van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough, V. Savolainen, “DNA barcoding the floras of biodiversity hotspots”, *Proc Natl Acad Sci USA*, vol. 105, pp. 2923-2928, 2008.
- [22] M. Lavin, M. Thulin, J. N. Labat, R. T. Pennington, “Africa, the odd man out: molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise”, *Syst Bot*, vol. 25, pp. 449-467, 2000.
- [23] Y. Liu, H. F. Yan, X. J. Ge, “Evaluation of 10 plant barcodes in Bryophyta (Mosses)”, *J Syst Evol*, vol. 48, pp. 36-46, 2010.
- [24] S. G. Newmaster, A. J. Fazekas, S. Ragupathy, “DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach”, *Can J Bot*, vol. 84, pp. 335-341, 2006.
- [25] Newmaster SG, Fazekas AJ, Steeves AD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. *Mol Ecol Res* 8: 480-490
- [26] S. G. Newmaster, S. Ragupathy, “Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae)”, *Mol Ecol Res*, vol. 9, pp. 172-180, 2009.
- [27] J. D. Palmer, L. A. Herbon, “Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence”, *J Mol Evol*, vol. 28, pp. 87-97, 1988.
- [28] E. Pennisi, “Taxonomy. Wanted: a barcode for plants”, *Sciences*, vol. 318, pp. 190-191, 2007.
- [29] M. H. Peter, L. F. Laura, L. S. John, “A DNA barcode for land plants”, *Proc Natl Acad Sci USA*, vol. 106, pp. 12794-12797, 2009.
- [30] M. D. Pirie, M. P. B. Vargas, M. Botermans, F. T. Bakker, L. W. Chatrou, “Ancient paralogy in the cpDNA *trnL-F* region in Annonaceae: implications for plant molecular systematics”, *Amer J Bot*, vol. 94, pp. 1003-1016, 2007.
- [31] S. Porebski, L. G. Bailey, B. R. Baum, “Modification of a CTAB DNA extraction protocol for plants containing high polysaccharides and polyphenol component”, *Pl Mol Biol Rep*, vol. 15, pp. 8-15, 1997.
- [32] R. A. Ribeiro, M. Lavin, J. P. Lemos-Filho, F. C. V. Mendonca, F. R. Santos, M. B. Lovato, “The Genus *Machaerium* (Leguminosae) is More Closely Related to *Aeschynomene* Sect. *Ochopodium* than to *Dalbergia*: Inferences from Combined Sequence Data”, *Syst Bot*, vol. 32, pp. 762-771, 2007.
- [33] R. A. Ribeiro, J. P. Lemos-Filho, A. C. S. Ramos, M. B. Lavato, “Phylogeography of the endangered rosewood *Dalbergia nigra* (Fabaceae): insight into the evolutionary history and conservation of the Brazilia Atlantic Forest”, *Heredity*, vol. 106, pp. 46-47, 2011.
- [34] N. Saitou, M. Nei “The neighbor-joining method: A new method for reconstructing phylogenetic trees”, *Mol Ecol Res*, vol. 4, pp. 406-425, 1987.
- [35] T. Sang, D. J. Crawford, T. Stuessy “Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae)”, *Amer J Bot*, vol. 84, pp. 1120-1136, 2009.
- [36] J. Song, H. Yao, Y. Li, X. Li, Y. Lin, C. Liu, J. Han, C. Xie, S. Chen, “Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique”, *J Ethnop*, vol. 124, pp. 434-439, 2009.
- [37] J. R. Starr, F. C. N. Robert, C. Briannan, “Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae)”, *Mol Ecol Res*, vol. 9, pp. 151-163, 2009.
- [38] G. N. Steven, S. Ragupathy, “Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae)”, *Mol Ecol Res*, vol. 9, pp. 172-180, 2009.
- [39] G. R. Sylvain, B. Bremer, “Phylogeny and classification of Naucleaeae S.l (Rubiaceae) inferred from molecular (ITS, rBCL and trnT-F) and morphological data”, *Amer J Bot*, vol. 86, pp. 1027-1047, 2002.
- [40] P. Taberlet, C. Eric, P. Francois, G. Ludovic, M. Christian, V. Alice, V. Thierry, G. Cortthier, B. Christian, W. Eske, “Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding”, *Nucleic Acids Res*, vol. 35, pp. 14–20, 2006.
- [41] K. Tamura, J. Dudley, M. Nei, S. Kumar, “MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0”, *Mol Biol Evol*, vol. 24, pp. 1596-1599, 2007.

- [42] K. Tamura, M. Nei, S. Kumar, “Prospects for inferring very large phylogenies by using the neighbor-joining method”, Proc Natl Acad Sci USA, vol. 101, pp. 11030-11035, 2004.
- [43] J. A. Tate, B. B. Simpson BB, “Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species”, Syst Bot, vol. 28, pp. 723-737, 2003.
- [44] T. J. White, T. Bruns, S. Lee, D. J. Taylor DJ, “Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Gelfand D, Sminsky J, White T (ed) PCR protocols: a guide to methods and applications”, Academic Press, San Diego, California, USA, pp. 315-322, 1990.



Figure 1: Location of sampling the eight *Dalbergia* species in the Vietnam. *D. assamica* collected from Ha Noi and Cuc Phuong National Park (Ninh Binh); *D. conchinchinensis* collected from Yordon National Park (Dak Lak) and KBang (Gia Lai); *D. nigrescens* collected from Copia Nature Reserve, (Son La) and Yordon National Park (Dak Lak); *D. oliveri* collected from Yordon National Park (Dak Lak); *D. tonkinensis* collected from Hanoi suburbs and Phong Nha - Ke Bang National Park (Quang Binh); *D. hancei*, *D. dialoides* and *D. entadioides* collected from Phong Nha - Ke Bang National Park (Quang Binh).

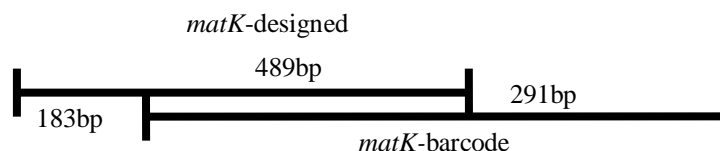


Figure 2: Map of *matK*-designed and *matK*-barcode. The *matK* gene regions overlapped in 489 bp. The *matK*-barcode was reported by Fazekas et al. (2012). The *matK*-designed prepared for this study was 183 bp upstream of overlapping region.

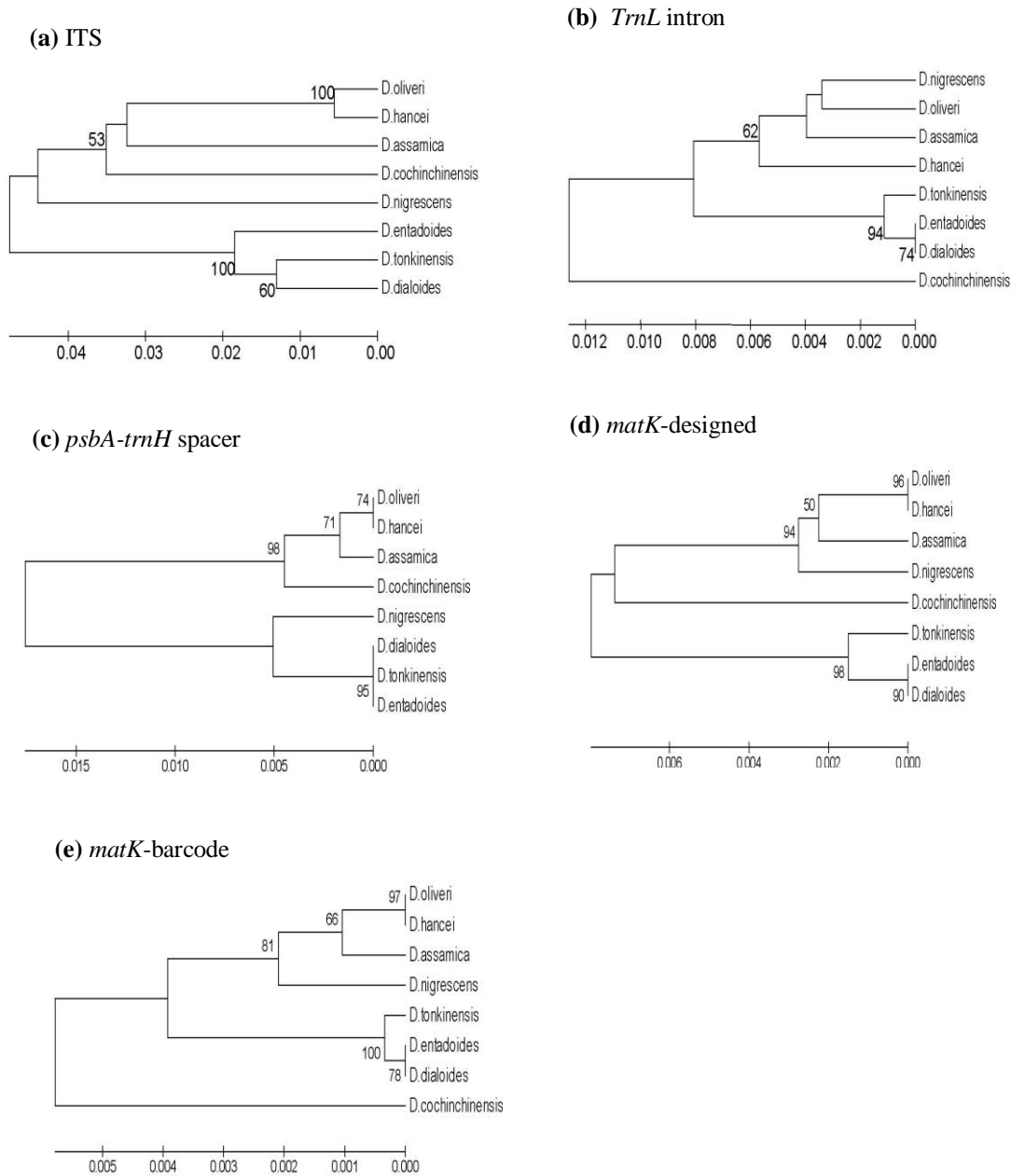


Figure 3: Phylogenetic tree reconstruction using UPGMA method for five studied DNA regions, ITS (a), *TrnL* intron (b), *psbA-trnH* spacer (c), *matK*-designed (d), *matK*-barcode (e). Only DNA fragment of sequences 650 bp for ITS, 460 bp for *trnL* intron, 325 bp for *psbA-trnH* spacer, 672 bp for *matK*-designed and 780 bp for *matK*-barcode were used for alignment. Numbers above branches indicate bootstrap values > 50 %.

Table 1: The number of conspecific, collected locations of genotypes and conservation status for the eight *Dalbergia* species employed

No.	Scientific name	Number samples	Selected location of samples	Conservation status*
1	<i>D. assamica</i>	05	Ha Noi suburbs and Cuc Phuong National Park (Ninh Binh)	Endemic to Vietnam
2	<i>D. conchinchinensis</i>		Yordon National Park (Dak Lak province) and KBang (Gia Lai province)	EN (A1acd) in Vietnam Red list and IUCN Red List
3	<i>D. nigrescens</i>	05	Copia Nature Reserve, Thuan Chau (Son La) Yordon National Park (Dak Lak)	Endemic to Vietnam
4	<i>D. oliveri</i>	05	Yordon National Park (Dak Lak province)	EN (A1acd) in Vietnam Red List and IUCN Red List
5	<i>D. tonkinensis</i>	05	Hanoi suburbs and Phong Nha - Ke Bang National Park (Quang Binh province)	VU (A1acd) in Vietnam Red List and IUCN Red List
6	<i>D. hancei</i>	05	Phong Nha - Ke Bang National Park (Quang Binh province)	Scientific value
7	<i>D. dialoides</i>	05	Phong Nha - Ke Bang National Park (Quang Binh province)	Scientific value
8	<i>D. entadioides</i>	05	Phong Nha - Ke Bang National Park (Quang Binh province)	Scientific value

* Dang and Nguyen 2007; IUCN 1998.

Table 2: List of primer pairs used in this study

Primer Names	Gene names	Primer sequences (5'– 3')	Expected size (bp)	Origin
ITS1-f ITS4-r	nrITS	TCCGTAGGTGAACCTGCGG TCCTCCGCTTATTGATATGC	700	White et al (1990)
trnL-f trnL-r	trnL intron, cpDNA	GTGATAACTTTCAAATTCAGAG CTCACGATTTCTTAAGTCGA	510	Designed based on tRNA - Leu sequence of <i>D. sissoo</i> (EF451118) (Ribeiro et al. 2007)
matK-f matK-r	matK gene, cpDNA	TCAGAGGGTTTTGCCGTCGTC CTCACGATTTCTTAAGTCGA	750	Designed based on matK sequence of <i>D. congestiflora</i> (AF142696) (Lavin et al. 2000)
Kim3F Kim1R	matK gene, cpDNA	CGTACAGTACTTTTGTGTTTACGAG ACCCAGTCCATCTGGAAATCTTGTTTC	850	Ki-Joong Kim (Fazekas et al. 2012)
psbA-f trnH-r	psbA – trnH spacer, cpDNA	GTTATGCATGAACGTAATGCTC CGCGCATGGTGGATTCACAATCC	450	Sang et al. (1997) Tate et al. (2003)

Table 3: Summary of characteristics of 5 DNA barcodes evolution and ability of distinguishing species for each region

	ITS	<i>trnL</i> intron	<i>matK</i>- designed	<i>matK</i>- barcode	<i>psbA-trnH</i> spacer
Aligned length	650	460	672	780	325
Characters of indels	12	15	3	5	25
Conservative characters	534	438	649	749	310
Variable characters	111	19	20	27	13
Parsimony informative characters	65	6	11	12	11
Singleton characters	46	13	9	15	2
Mean nucleotide diversity of eight species (%)	7.4	1.5	1.2	1.3	2.2
Nucleotide diversity among species (%)	1.1- 0.7	0-2.8	0- 2.0	0-1.4	0- 4.1
The power to distinguish species (%)	100	96.0	92.9	92.9	85.7