



Comparative analysis of complete plastid genome reveals powerful barcode regions for identifying wood of *Dalbergia odorifera* and *D. tonkinensis* (Leguminosae)

Qin, Ming; Zhu, Cheng-jie; Yang, Jun-bo; Vatanparast, Mohammad; Schley, Rowan; Lai, Qiang; Zhang, Dan-yan; Tu, Tie-yao; Klitgård, Bente B.; Li, Shi-jin; Zhang, Dian-xiang

Published in:
Journal of Systematics and Evolution

DOI:
[10.1111/jse.12598](https://doi.org/10.1111/jse.12598)

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Qin, M., Zhu, C., Yang, J., Vatanparast, M., Schley, R., Lai, Q., Zhang, D., Tu, T., Klitgård, B. B., Li, S., & Zhang, D. (2022). Comparative analysis of complete plastid genome reveals powerful barcode regions for identifying wood of *Dalbergia odorifera* and *D. tonkinensis* (Leguminosae). *Journal of Systematics and Evolution*, 60(1), 73-84. <https://doi.org/10.1111/jse.12598>

Research Article

Comparative analysis of complete plastid genome reveals powerful barcode regions for identifying wood of *Dalbergia odorifera* and *D. tonkinensis* (Leguminosae)

Ming Qin^{1,2} , Cheng-Jie Zhu³, Jun-Bo Yang⁴, Mohammad Vatanparast^{5,6}, Rowan Schley⁷, Qiang Lai^{1,2}, Dan-Yan Zhang⁸, Tie-Yao Tu^{1*} , Bente B. Klitgaard^{7*}, Shi-Jin Li^{1*}, and Dian-Xiang Zhang¹

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³Shenzhen Academy of Metrology and Quality Inspection, Shenzhen 518109, China

⁴Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

⁵Department of Geosciences and Natural Resource Management, University of Copenhagen, 1958 Frederiksberg, Denmark

⁶US National Herbarium, Department of Botany, Smithsonian Institution-NMNH, Washington, DC 20560, USA

⁷Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK

⁸School of Pharmaceutical Science, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

*Authors for correspondence. Tie-Yao Tu. E-mail: tutieyao@scbg.ac.cn; Bente B. Klitgaard. E-mail: B.klitgaard@kew.org;

Shi-Jin Li. E-mail: lisj@scib.ac.cn

Received 5 March 2020; Accepted 1 May 2020; Article first published online 8 May 2020

Abstract *Dalbergia odorifera* T. C. Chen (Leguminosae), a rare and endangered tree species endemic to Hainan Island of China, produces the most expensive and rarest wood in China. The wood characteristics of *D. odorifera* are remarkably similar to those of *D. tonkinensis* (a much less sought-after species from Vietnam), and the DNA from wood is often highly degraded, making it very difficult to identify the two species using anatomical features or DNA barcoding based on regular DNA markers. To solve the confusion of identifying wood reliably from the two species, we built and analyzed the plastome library of 26 samples from 18 *Dalbergia* species, of which 12 samples from eight closely related species of *D. odorifera* are newly sequenced in this study. Phylogenomic analysis suggested that the relationships among the 26 samples are mostly well resolved, and conspecific individuals from different populations of *D. odorifera* and *D. tonkinensis* clustered together. Between the plastid genomes of the two species, we identified 129 indels and 114 single nucleotide polymorphisms. By assessing a subset of 20 nucleotide polymorphisms and 10 indels using 37 population-level samples (20 samples of *D. odorifera* and 17 samples of *D. tonkinensis*), we recovered eight species-specific barcode regions that could be suitable for identifying the wood *D. odorifera* and *D. tonkinensis*. To examine their utility in wood identification, we amplified the eight DNA barcodes using six wood samples and recovered an amplification success rate of 83.3%, demonstrating a reliable method for precise wood identification of the two species.

Key words: DNA barcoding, Fabaceae, Hongmu, phylogenomics, rosewood, wood identification.

1 Introduction

Dalbergia L.f., a genus of Leguminosae (alternatively Fabaceae), comprises ca. 250 species of trees, shrubs, and lianas mainly in the pantropic region (Klitgaard & Lavin, 2005; Vatanparast et al., 2013; Bolson et al., 2015; Li, 2017). Many species within this genus are of major economic importance for their high quality wood (Bhagwat et al., 2015), such as the richly hued and costly woods of the Brazilian rosewood *D. nigra* (Vell.) Allemão ex Benth., the Indian rosewood *D. latifolia* Roxb., and the

Madagascar rosewood *D. maritima* R. Vig. Some other species, such as *D. cultrata* Graham ex Benth. and *D. odorifera* T. C. Chen, have been recognized as “Hongmu”, referring to a group of hardwood species that are red in color and are widely used in high-end Chinese furniture.

According to the updated national standard published by the Chinese government in 2017 (General administration of quality supervision, GB/T 18107-2017 Hongmu [S], 2017), Hongmu comprises 29 tree species, of which 24 belong to the legume family. Of these, 15 belong to *Dalbergia*; six

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

are from *Pterocarpus* Jacq., two are *Millettia* Wight & Arn. species, and one is a species of *Senna* Miller Gard. Five species of *Diospyros* L. belonging to the family Ebenaceae have also been recognized as Hongmu. *Dalbergia odorifera*, a rare Hongmu species endemic to Hainan Island of southern China, produces excellent quality wood with characteristic aroma, which is the most expensive found in the Chinese wood market. Since the 10th century, the wood of *D. odorifera* has been used in furniture manufacturing, and during the Ming and Qing dynasties (ca. 1600–1700) it was widely used by the nobility and royal families of China (Xu & Li, 2003). Due to its smooth, beautiful grain, and the lasting pleasant aroma it produces, there is high demand for this wood, and as such products made from *D. odorifera* have become a symbol of wealth and a lavish display of status. The heartwood of *D. odorifera* has also been used as medicine for blood disorders, ischemia, swelling, and epigastric pain in China (Chan et al., 1998; Lee et al., 2014). However, the resources of *D. odorifera* are extremely limited because of restricted distribution and the fact that it takes more than 100 years to grow to maturity. This species is nearly extinct in the wild due to deforestation and illegal logging. It was last assessed in 1998 for the IUCN Red List of Threatened Species as of “Vulnerable” status (Ban, 1998), and was ranked in the second-class category of the National Protected Plants list, issued by the Chinese government in 1999 (Yu et al., 2017). Based on the comprehensive survey of wild populations of *D. odorifera*, Yang et al. (2012) assessed *D. odorifera* as “Critically Endangered” according to the IUCN endangered level assessment standard. Only seven individuals were found in their study. According to our follow-up survey in 2016, only three wild individuals of the species were found in the National Natural Reserve of Bawangling of Hainan Island, China.

The high-quality wood and the minimal distributions as well as survival of individuals of the species have resulted in the wood price of *D. odorifera* being \$1000–1500 per kilogram (<http://www.chinatimber.org/price/60290.html>; 2019.01–2019.06). Consequently, this could cause traffickers to sell cheaper wood from similar species fraudulently (e.g., *D. tonkinensis* Prain, *D. hainanensis* Merr. & Chun, *Pterocarpus erinaceus* Poir, *Streblus asper* Lour., and *Spirostachys africana* Sond.) as *D. odorifera* (Cheng et al., 2014). Most of the wood from these counterfeit species can be distinguished from *D. odorifera* by congeneric wood anatomy. However, wood anatomy cannot distinguish the most closely related species, *D. tonkinensis* (northern to central Vietnam and Laos; Yu et al., 2016; Li, 2017), from *D. odorifera*. Recently, studies concerning wood identification of *D. odorifera* and *D. tonkinensis* have attracted the attention of several research projects in China (Guo et al., 2011; Wang et al., 2016; Ma et al., 2018; Zhang et al., 2018). Unfortunately, however, wood identification of these two species is still far from being resolved. The wood color, density, and structure of the two species are very similar (Luo et al., 2012; Yu et al., 2016), making it impossible to identify the wood of the two species by anatomical means. Moreover, it is not applicable to discriminate between woods of the two species using chemical methods or isotopic fingerprinting, due to very similar chemical compounds of these congeneric species (Marco et al., 1994; Deguilloux et al., 2002; Dev et al., 2014).

The chemical compounds found in many plants could be unstable, depending on the age of the trees or the habitat (Deguilloux et al., 2002). Yang et al. (2016) examined the volatile and fat-soluble components from *D. odorifera* and *D. tonkinensis*, respectively, by gas chromatography–mass spectrometry and found the two species can be differentiated by fat-soluble components based on their composition. However, when including more samples of different ages and places of origin, these results proved less accurate (Zhang DY, 2018, unpublished data).

The difficulties in identifying wood of *D. odorifera* and *D. tonkinensis* using anatomical and chemical methods call for new and effective techniques in order to facilitate their identification. One means by which this could be accomplished is via DNA barcoding. For example, Ohshima et al. (2001) explored the identification of six species of *Cyclobalanopsis* Oerst. (Fagaceae) using the DNA barcoding approach and examined the feasibility of the method in identifying the woods of this genus. Such identification based on DNA barcoding was adopted by many researchers due to its utility and relative accuracy. Within *Dalbergia*, in particular, there have been multiple studies examining the efficacy of DNA barcoding. For example, Hartvig et al. (2015) sequenced *rbcL*, *matK*, and internal transcribed spacer (ITS) of 95 samples covering 31 *Dalbergia* species and tested their discriminatory ability with both distance-based and model-based methods. They suggested *rbcL* + *matK* as the potential combined barcodes to identify woods within *Dalbergia*. Furthermore, Yu et al. (2017) sequenced eight DNA regions of nine endangered *Dalbergia* species and suggested ITS2 + *trnH-psbA* as the best combination barcode. These studies strongly suggested that DNA barcoding is robust in identifying distantly related species within *Dalbergia*. However, these DNA barcodes might not identify closely related species, such as the species pair *D. odorifera* and *D. tonkinensis*. Yu et al. (2016) advocated the intergenic spacer of *trnH-psbA* that contains seven single nucleotide polymorphism (SNP) sites as a barcode region for the identification of *D. odorifera* and *D. tonkinensis*. We carefully analyzed the *trnH-psbA* region of the two species and found that the seven SNP sites are locating in a 47-bp palindromic region, which is equal to a reverse complement of itself (Figs. 1A, 1B, 1D). This means that the alignment concerning the 47-bp palindromic region and the seven SNPs in Yu et al.'s (2016) study is not correct. Furthermore, the palindromic sequences can form stem-and-loop structures and are often unconserved among different individuals or even among different cells of the same plant (Cavalier, 1974). This is the case in our study because we detected both the forward strand and the reverse complementary strand of the 47-bp palindromic region in the raw reads of the same individual for both *D. odorifera* and *D. tonkinensis* (Fig. 1C). Additionally, previous studies only covered a few samples of *D. odorifera* and *D. tonkinensis*, which cannot answer whether the diagnostic sites in those studies are species-specific.

Genomic data have been used by many researchers to resolve the phylogenetic relationships of either deep or shallow recalcitrant nodes (Kumar et al., 2009; Nock et al., 2011; Yang et al., 2013; Song et al., 2015; Gamboa-Tuz et al., 2018; Wu et al., 2018; Zhao et al., 2018). It is useful for resolving

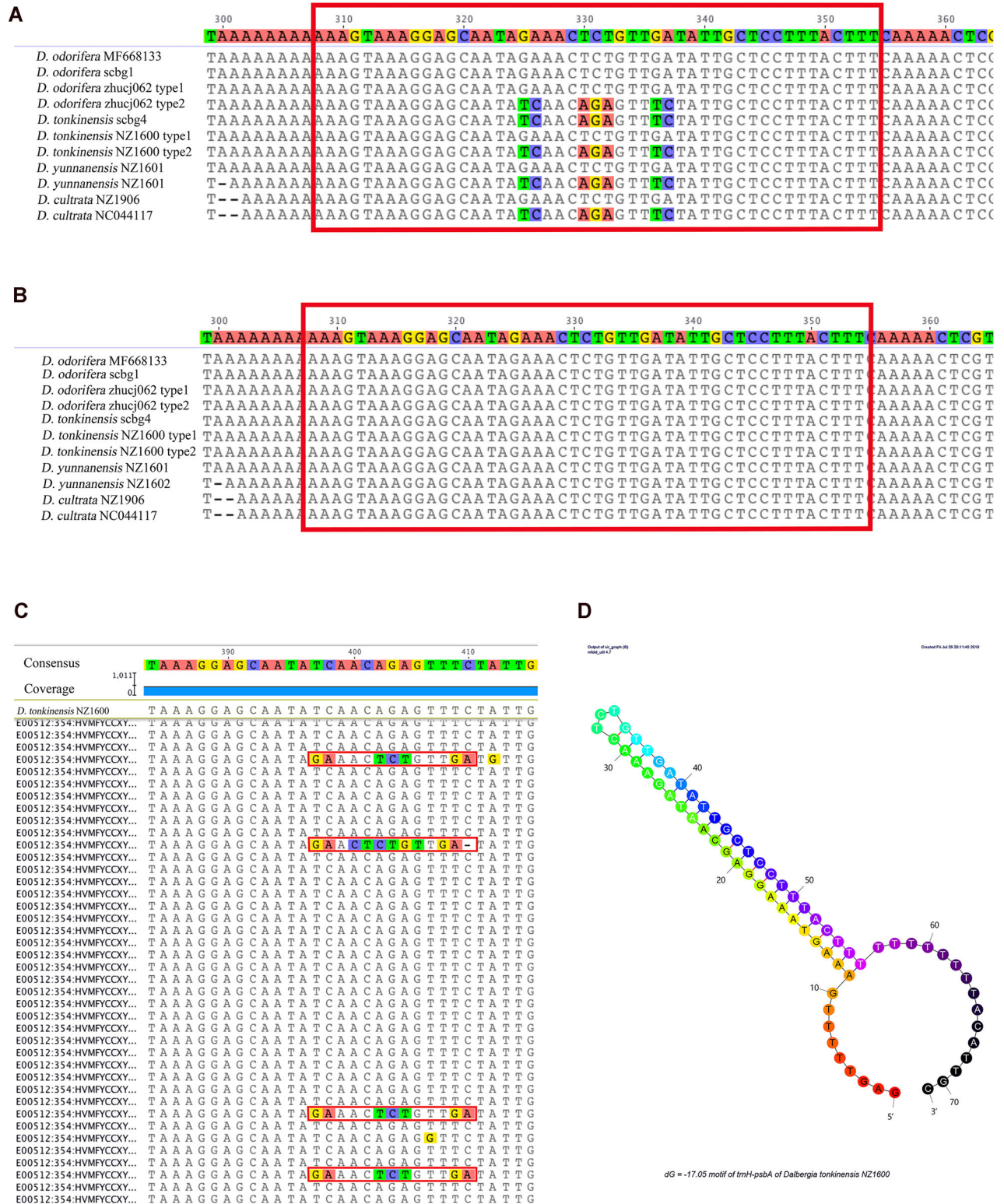


Fig.1. Continued

relationships between closely related species. Liu et al. (2019a, 2019b) developed 19 single sequence repeat markers for *D. odorifera* and *D. tonkinensis* based on transcriptome data, and they recovered genetic diversity and population structure of the species. These studies also showed the potential power of genomic data to differentiate *D. odorifera* from *D. tonkinensis* using fresh or silicon-dried plant tissues. In the case of identifying woods of *D. odorifera* and *D. tonkinensis*, however, the utility of nuclear genome data could be restricted because DNA from aged or processed wood was often severely degraded and the concentration could be too low to be enriched for sequencing. Genome data derived from the plastid could be a better choice for wood identification because most plant cells contain many more copies of the plastid genome than the nuclear genome (Fazekas et al., 2009). Song et al. (2019) identified eight mutation hotspot regions as candidate DNA barcodes by comparing the chloroplast genomes of nine species of *Dalbergia*. The length of these eight fragments ranges from 700 to 1500 bp. However, its feasibility in wood identification might not be guaranteed because it lacks of species-specific test and it is unlikely to obtain such long sequences from severely degraded wood products through Sanger sequencing. In general, shorter DNA fragments are more likely to be successfully amplified (Rachmayanti et al., 2009; Jiao et al., 2019).

An accurate identification of the two species is urgent, as currently *D. odorifera* is widely cultivated in southern China as a measure to upgrade the forestry production in the country. In Guangdong, Guangxi, and Hainan alone the total cultivation area could exceed 30 000 ha and new plantations are increasing rapidly (Xu DP, 2020, pers. comm.). The confusion in the germplasm sources could cause heavy loss to the farmers and forestry sections.

In this study, our main objectives were to: (i) screen potential barcoding regions for *D. odorifera* and *D. tonkinensis* through complete plastid genome comparative analysis; (ii) explore the most powerful and easily-amplified species-specific barcodes for the two species using population-level sampling; and (iii) verify the utility of these barcodes by testing their efficacy on wood samples.

2 Material and Methods

2.1 Sampling, DNA extraction, and sequencing

For plastid genome sequencing, we sampled fresh leaves from two individuals each of *Dalbergia odorifera*, *D. tonkinensis*, and *D. yunnanensis* Franch, in addition to other eight *Dalbergia* species. For Sanger sequencing, we used silica gel-dried leaves of 20 and 17 samples of *D. odorifera* and *D. tonkinensis*, respectively (covering their distributional range in China and Vietnam). Six wood samples of *D. odorifera* and *D. tonkinensis* were either

bought from a local seller in Hainan, China or through online shopping as a test case. The voucher information is presented in Table S1, and the specimens were deposited in the herbarium of South China Botanical Garden (IBSC).

Total DNA from leaf samples was extracted using the modified CTAB method (Doyle & Doyle, 1990). DNA from wood samples was extracted from 0.5–1 g wood powder using the Qiagen DNeasy Plant Mini Kit (Hilden, Germany), combined with N-phenacylthiazolium bromide (Yu et al., 2016).

Primers (Table S2) were designed in Primer3 version 4.1.0 (Koressaar & Remm, 2007; Untergasser et al., 2012). Due to the poor quality and severe degradation of wood DNA, we aimed at short DNA fragments with 100–250 bp in length for Sanger sequencing. The high-throughput sequencing of the plastid genomes was carried out using the Illumina HiSeq 2000 (San Diego, CA) at the Beijing Genomics Institute (Beijing, China). Sanger sequencing was performed using the ABI 3730 sequencer at the Shanghai Majorbio Bio-pharm Technology Co. (Shanghai, China).

2.2 Plastid genome assembly and annotation

The clean data of the Illumina sequencing received from Beijing Genomics Institute were directly assembled using the GetOrganelle pipeline (Bankevich et al., 2012; Langmead & Salzberg, 2012; Jin et al., 2018). Bandage version 5.6.0 (Wick et al., 2015) was used to visualize and manually correct the assembly results. We checked the sequence directions and verified the accuracy of the assemblies in Geneious version 9.0.5 (Kearse et al., 2012). The annotation of the chloroplast genomes was undertaken in Plastid Genome Annotator (Qu et al., 2019). Manual correction of start/stop codons and intron/exon boundaries was carried out in Geneious. To further verify the identified transfer RNA (tRNA) genes, the tRNAscan-SE version 1.21 program with default parameters was used to predict their corresponding structures (Schattner et al., 2005). All genome maps were drawn by OrganellarGenomeDRAW version 1.3.1 (Greiner et al., 2019). The annotated chloroplast genomes were deposited in GenBank (accession numbers see Table 1).

2.3 Phylogenetic analysis

We built a reference library using 14 plastid genome sequences generated in this study, in addition to 12 plastome sequences of *Dalbergia* and three outgroup plastome sequences of *Pterocarpus* downloaded from GenBank (Table S1). The sequence of *D. hainanensis* was not included (MF926268, Deng et al., 2018), due to the possible misidentification of the sample (see also Song et al., 2019). Based on the reference library, we undertook sequence alignment using the MAFFT version 7.0.17 (Katoh & Standley, 2013) plugin in Geneious by concatenating 77 coding genes, 4 ribosomal RNA (rRNA) genes, and 26

Fig. 1. The 47-bp palindromic region between intergenic region *trnH-psbA* of the genus of *Dalbergia*. **A**, Seven high-lighting nucleotide mutation sites reveal that this 47-bp palindromic region (the region inside the red square) is unconserved among different species or even among different individuals of the same species. **B**, The forward sequences are identical with the reverse complement of the 47-bp palindromic regions. **C**, Two types of reads detected in the raw next generation sequencing data of the same sample (*Dalbergia tonkinensis*, voucher NZ1600). **D**, The palindromic region is capable of forming stem-and-loop structures, which is lethal to the carrier and replicon (inviability), and unstable among different individuals or even among different cells of the same plant (instability) (Zuker, 2003).

intergenic spacers. Glaring alignment errors were adjusted by hand. We inferred phylogenetic relationships for species within the reference library by constructing a maximum likelihood tree using IQ-TREE version 1.6.11 (Trifinopoulos et al., 2016). Branch supports were estimated using standard bootstraps with 100 iterations, and a minimum correlation coefficient of 0.99.

2.4 Species-specific population test and the utility of selected barcodes in wood product identification

Based on the alignment of complete plastomes from the reference library, potential insertion–deletion (indel) and SNP sites, which were flanked by regions suitable for designing polymerase chain reaction (PCR) primers and were effective in the identification of *D. odorifera* and *D. tonkinensis*, were selected as barcode regions. The primers were designed using Primer3. We sequenced 37 population-level samples (20 samples of *D. odorifera* and 17 samples of *D. tonkinensis*) to test whether the selected SNP and indel sites were species-specific. The species-specific variable sites were considered to be powerful barcodes for identifying *D. odorifera* and *D. tonkinensis*. We built the maximum likelihood tree using IQ-TREE version 1.6.11 (Trifinopoulos et al., 2016) based on the concatenated alignment of the species-specific variable sites of the 37 population-level samples. Branch supports were estimated using standard bootstraps with 100 iterations, and a minimum correlation coefficient of 0.99. Additionally, we used six wood product samples of *D. odorifera* and *D. tonkinensis* to test the validity of identification of these species-specific variable sites. The sequences of the 37 population-level samples are provided in Doc. S1.

3 Results

3.1 Features of the *Dalbergia* plastid genomes

The 14 complete plastid genome sequences newly generated in this study range from 155 823 to 156 314 bp in length. The total GC-content was 36.1% for all plastomes except *D. cearensis*

Ducke (36%). The 14 plastid genomes each contained 77 coding genes, 30 tRNA genes, and 4 rRNA genes. In general, the plastid genomes of the *Dalbergia* species analyzed in this study were similar in terms of genome size, gene content, gene order, introns, intergenic spacers, and GC-content. A summary of the structure of the plastid genomes is shown in Table 1. The gene map is shown in Fig. 2.

3.2 Single nucleotide polymorphism and indel mutations between plastid genomes of *D. odorifera* and *D. tonkinensis*

We detected 129 indels ranging from 1 to 22 bp in size (105 in intergenic regions, 22 in introns, and 2 in coding regions) among the four plastid genomes generated in this study of *D. odorifera* and *D. tonkinensis*, of which 66% were single base-pair. The longest indel was 22 bp in length, which was in the intergenic region of *psbA-trnK-UUU*. In the coding region of *ndhF*, we found an 8-bp insertion in the *ndhF* gene of *D. tonkinensis* (voucher: Lishijin 3616), an advanced stop codon, and a five-amino difference when compared to the homologous region of *D. odorifera* (vouchers: scbg1 and zhucj062). A 10-bp insertion in the *ndhF* gene of *D. tonkinensis* (voucher: Lishijin 3616) resulted in a delayed stop codon and a loss of five amino acids when compared to the same region of *D. tonkinensis* (voucher: scbg4).

We found 114 SNPs between *D. odorifera* and *D. tonkinensis*, with a transversion to transition ratio of 2.56:1. Of these, 69 SNPs (58 transversions and 11 transitions) were in the non-coding regions and 45 SNPs (24 transversions and 21 transitions) were from coding regions. We found 28 synonymous and 17 non-synonymous mutations in the coding regions between the two species.

3.3 Phylogenetic relationships and species-specific barcode sites

We built a reference library based on 26 plastid genomes of *Dalbergia* and three outgroup plastid genomes of *Pterocarpus*. The phylogenetic relationships among all species were well resolved (Fig. 3) in the maximum likelihood tree. The sister relationship between *D. odorifera* and

Table 1 Summary of complete chloroplast genomes generated in this study

Species	Total (bp)	GC%	LSC (bp)	IR (bp)	SSC (bp)	Total genes	Protein coding genes	tRNA	rRNA	GenBank accession No.
<i>Dalbergia odorifera</i>	156 049	36.1	85 807	25 702	18 838	111	77	30	4	MN202180
<i>D. odorifera</i>	156 031	36.1	85 781	25 702	18 846	111	77	30	4	MN202181
<i>D. tonkinensis</i>	156 090	36.1	85 765	25 720	18 886	111	77	30	4	MN202179
<i>D. tonkinensis</i>	156 045	36.1	85 825	25 720	18 780	111	77	30	4	MN202182
<i>D. yunnanensis</i>	155 823	36.1	85 698	25 715	18 695	111	77	30	4	MN202183
<i>D. yunnanensis</i>	156 154	36.1	85 946	25 706	18 796	111	77	30	4	MN202184
<i>D. sissoo</i>	155 863	36.1	85 501	25 681	19 000	111	77	30	4	MN936016
<i>D. frutescens</i>	156 105	36.1	85 289	25 719	19 378	111	77	30	4	MN936017
<i>D. obovata</i>	156 012	36.1	85 405	25 719	19 169	111	77	30	4	MN936018
<i>D. cearensis</i>	156 314	36.0	85 509	25 729	19 347	111	77	30	4	MN936019
<i>D. chlorocarpa</i>	155 745	36.1	85 646	25 663	18 773	111	77	30	4	MN936020
<i>D. armata</i>	156 246	36.1	85 435	25 739	19 333	111	77	30	4	MN936021
<i>D. martinii</i>	157 123	36.1	85 673	25 706	20 038	111	77	30	4	MN936022
<i>D. vietnamensis</i>	156 088	36.1	85 818	25 723	18 824	111	77	30	4	MN936023

IR, inverted repeat; LSC, large single copy; rRNA, ribosomal RNA; SSC, small single copy; tRNA, transfer RNA.

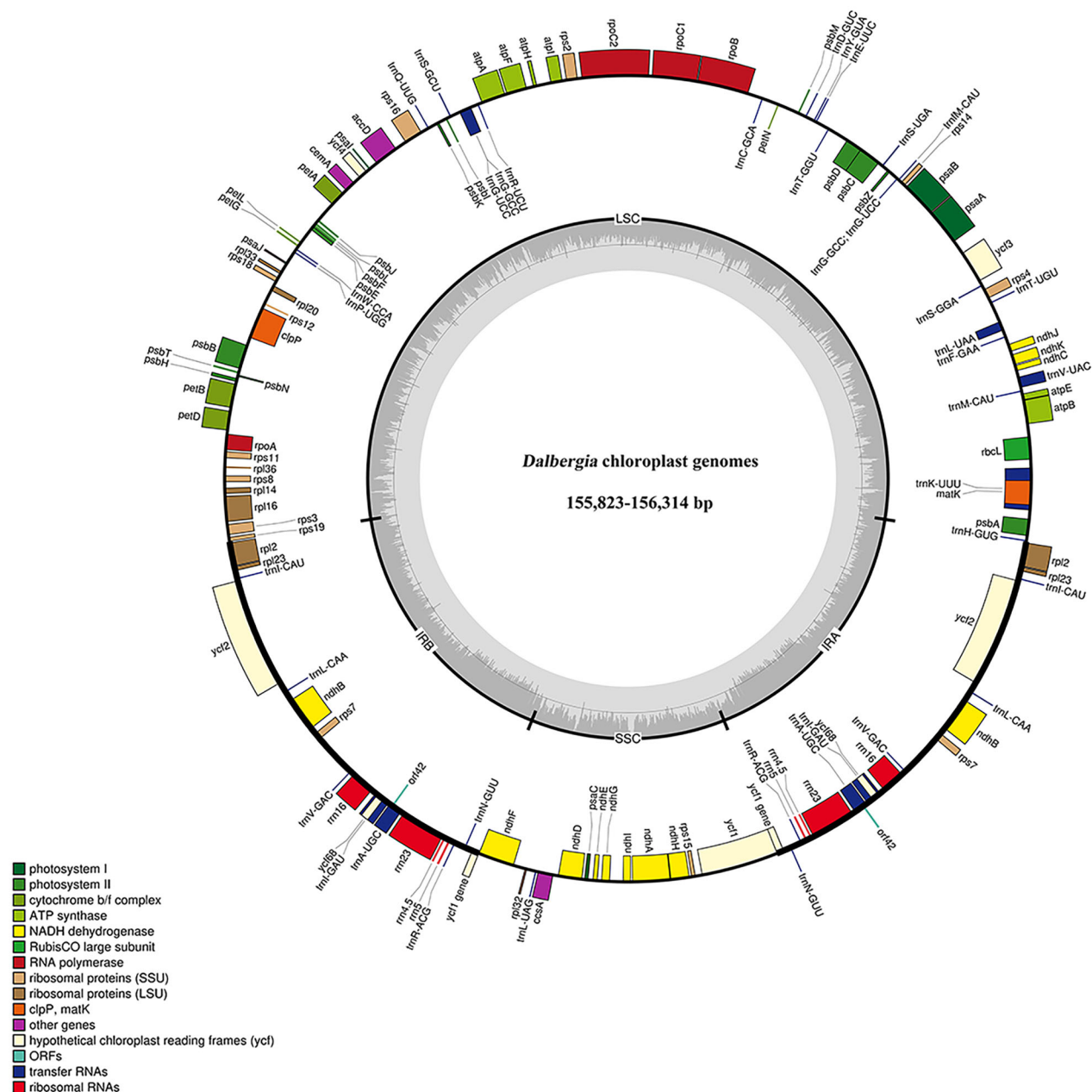
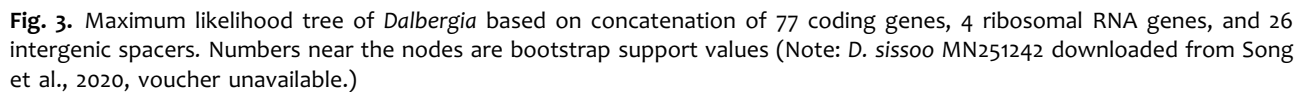


Fig. 2. Gene map of the plastid genomes of *Dalbergia* generated in this study. The genes outside the circle are transcribed clockwise, whereas those inside are transcribed counterclockwise. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRA and IRB) are indicated.

D. tonkinensis was guaranteed in both the phylogenomic analysis of 26 *Dalbergia* samples at species level (Fig. 3) and the analysis based on eight barcoding regions using the population-level sampling (Fig. 4). A sister relationship between the clade of *D. odorifera*–*D. tonkinensis* and *D. yunnanensis*–*D. vietnamsis* P.H.Hô and Niyomdham is recovered with a bootstrap value of 100 (Fig. 3). Morphologically, the latter two species are most comparable to *D. odorifera*–*D. tonkinensis* in the genus. The intraspecific relationships within *D. odorifera*–*D. tonkinensis* were poorly resolved and polytomies were found for both species. *Dalbergia sissoo* Roxb. ex DC. newly sequenced in this study does not cluster with the one published by Song et al.

(2019), which forms a clade with *D. cochinchinensis* Pierre ex Laness. and *D. hupeana* Hance. The identification of *D. sissoo* and *D. hupeana* in the study of Song et al. (2019) needs to be reexamined because: (i) the two species are comparable in morphology to the clade of *D. odorifera* but very different from *D. cochinchinensis*; and (ii) we found living trees that could be mislabeled as *D. sissoo* and *D. hupeana* in Jianfengling Nature Reserve where the two species were sampled by Song et al. (2019).

Comparative analysis of the 26 plastid genomes of *Dalbergia* revealed that 20 SNPs and 10 indels flanked with sequences suitable for primer design were considered



Combined indel and SNP DNA barcodes using comparative plastid genome analysis has proved to have significant potential for species identification in other taxa. For example, Huang et al. (2014) found 15 highly variable non-coding regions with more than 1.5% sequence divergence, which led to successful phylogenetic reconstruction and identification of species in the genus *Camellia* L. Similar work was carried out on the medicinal herb *Panax* L. (Dong et al., 2014), which revealed *rps16*, *ycf1a*, and *ycf1b* as the best mini-barcodes for identification of *Panax ginseng* C. A. Mey.

Among the 129 indels found within the plastid genomes of *D. odorifera* and *D. tonkinensis*, 105 indels were found in intergenic spacers, 22 were from introns, and 2 in coding regions. Some highly variable regions, such as the intergenic spacers *trnP_UGG-psaJ*, *rps11-rpl36*, *trnN_GUU-ycf1*, *ndhG-ndhI*, *psbA-trnK_UUU*, *trnV_UAC-ndhC*, *trnC_GCC-psbZ*, *trnG_UCC-trnS_UCC*, contain multiple indel sites. Compared to gene regions that show biological functions, the intergenic spacers and gene introns appear to be more variable among and within species, as has been reported by many authors (Timme et al., 2007; Liu et al., 2017; Wu et al., 2018). In this study, we found the intergenic spacer *ndhK-ndhJ* and the intron region *trnG_UCC* are the best indel fragments to distinguish *D. odorifera* and *D. tonkinensis*. These two regions have also been shown to be very efficient at distinguishing between different, closely related species belonging to multiple taxonomic groups, including bacteria and species from several different plant families (Kaneko et al., 2000; Iwasaki et al., 2012; Panero et al., 2014). The mutations in these two regions are species-specific and conserved among

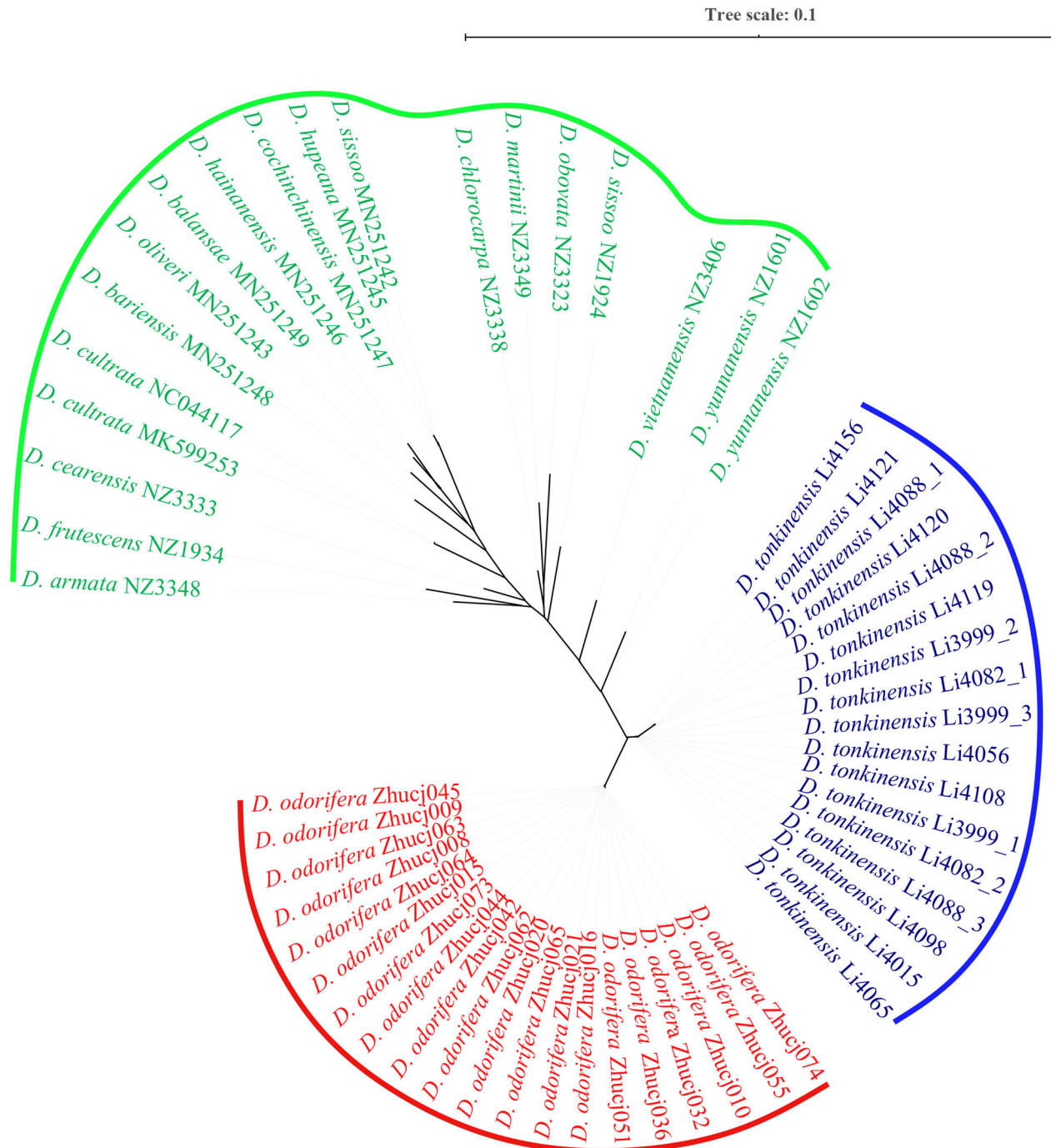


Fig. 4. Maximum likelihood tree based on eight barcode regions by sampling *Dalbergia odorifera* and *D. tonkinensis* at population level strongly suggested that conspecific individuals of the two species clustered together.

different individuals within *D. odorifera* and *D. tonkinensis*, as shown in our phylogenetic result using 37 population-level samples (Fig. 4).

The 114 nucleotide substitutions occurring in the plastid genomes of *D. odorifera* and *D. tonkinensis* suggest that such SNP events are much less common in this genus than in other genera such as *Panax*, *Machilus* Rumph. ex Nees and *Citrus* L. (Dong et al., 2014; Su et al., 2014; Song et al., 2015).

Several SNP sites were proved to be species-specific between *D. odorifera* and *D. tonkinensis* in this study. The promising gene, *ycf1*, which has shown great potential as a plastid DNA barcode for land plants (Cai et al., 2012; Dong et al., 2015), contained three species-specific SNP sites between *D. odorifera* and *D. tonkinensis*. In the *atpI* gene, we found one species-specific SNP which resulted in an amino acid change from valine into isoleucine between the two

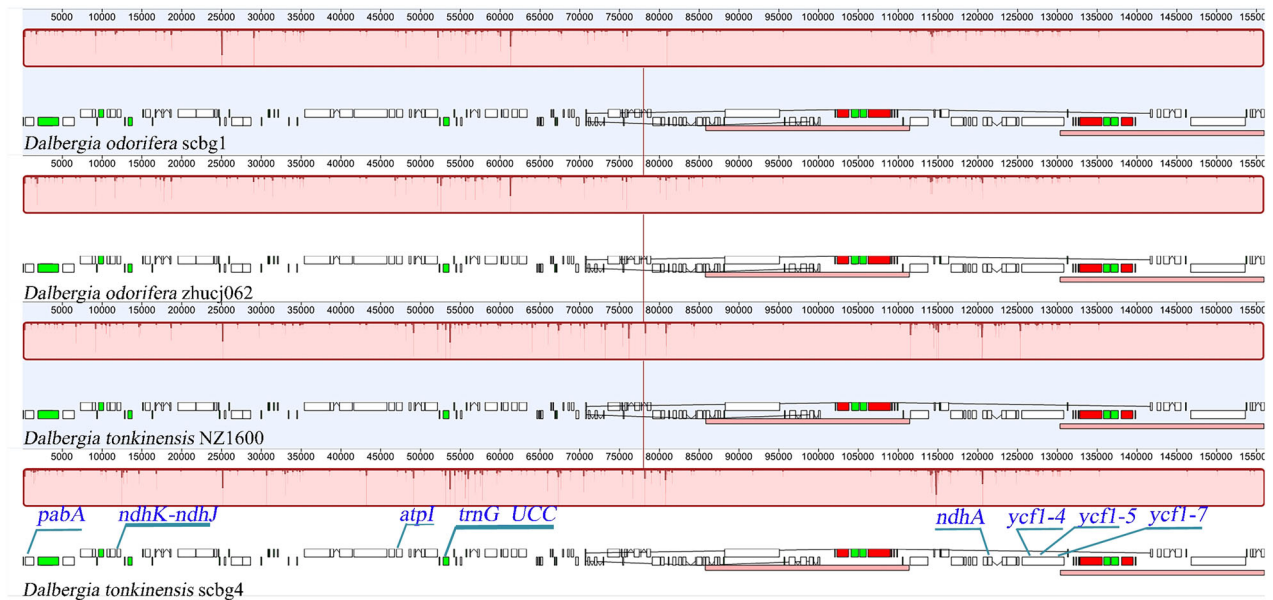


Fig. 5. Location of eight barcoding fragments in the whole plastid genomes of *Dalbergia odorifera* and *D. tonkinensis*. The regions were ordered based on their position in the plastid genomes.

species. The *psbA* and *ndhA* genes were also highly effective in distinguishing *D. odorifera* from *D. tonkinensis*. Even though there is only one single mutation in each gene, they prove to be highly stable within species and with no interspecies overlap among the 37 population-level samples.

To find a single-barcoding gene for all plants is a challenging task due to the inherent inaccuracies of many barcode regions (Rubinoff et al., 2006). For our study, we combined two indel regions (*ndhK-ndhJ* and *trnG_UCC*) and six SNP regions (*psbA*, *atpI*, *ndhA*, *ycf1-4*, *ycf1-5*, and *ycf1-7*) to distinguish *D. odorifera* from *D. tonkinensis* robustly. Single-copy nuclear genes were excluded from the present study because it is particularly difficult to obtain nuclear DNA from timbers or wood products, from which DNA quality is very poor and the content is often extremely low (Fazekas et al., 2009).

The major obstacle for DNA barcoding of wood could be the difficulties in attaining high-quality DNA from commercial woods and wood products (Yu et al., 2017). Therefore, we have avoided amplifying long sequence regions and designed primers to amplify eight smaller regions in 100–250 bp for all the 37 population-level samples. From our results (Fig. 4), it is evident that these eight regions are highly discriminatory and that the mutations in these eight regions are species-specific as, in this plot, *D. tonkinensis* and *D. odorifera* accessions fell into distinctly separated groups. Furthermore, the eight barcode regions performed well in the wood product test as well. Five of the six wood samples were successfully amplified and identified, with a success rate of 83.3%.

To conclude, this study sequenced 14 complete plastid genomes for the species of *Dalbergia* using Illumina next generation sequencing and built a reference library of *Dalbergia* using 26 complete plastid genomes. Based on the

reference library, 10 indels and 20 SNP sites that were flanked by regions suitable for designing PCR primers were selected to test whether they were species-specific among 37 population-level samples of *D. odorifera* and *D. tonkinensis*. Our work showed that two indel regions (*ndhK-ndhJ* and *trnG_UCC*) and six SNP regions (*psbA*, *atpI*, *ndhA*, *ycf1-4*, *ycf1-5*, and *ycf1-7*) are species-specific. We also found that these eight species-specific barcodes are effective at species identification even when using DNA extracted from wood products. Five of the six wood samples were successfully amplified and identified, with a success rate of 83.3%. The eight barcodes proved to be very successful in the identification of *D. tonkinensis* and *D. odorifera*. We recommend the use of these eight barcodes for species identification and wood product identification of these two species.

A significant breakthrough could be made for many existing wood identification problems through the rapid development of high-throughput sequencing technology. DNA extracted from wood is often serially degraded and fragmented, and analysis rendering PCR amplification from such samples is very difficult, which has been a pervasive problem in the identification of timber species. By sequencing herbarium specimens up to 80 years old from various flowering plant families, Zeng et al. (2018) reported that genome skimming could be used to generate genomic information using as little as 500 pg of degraded starting DNA. We can enrich the degraded DNA extracted from wood by using commercial kits despite the low concentration and poor quality of wood-derived DNA. These enriched DNA libraries can then be sequenced using a genome-skimming approach. Beyond *D. odorifera*, another 14 Hongmu species of *Dalbergia* meet the weak identification problem as well. Based on the reference library of 26 *Dalbergia* plastid

genomes, we will extend current research by including more *Dalbergia* species collected from different areas throughout the distributional range of the genus.

Our results, based on multiple samples, clearly indicate that *D. odorifera* and *D. tonkinensis* represent two independent evolutionary lineages. In order to reveal their taxonomic relationship, it is necessary to study the differences in morphological characters between *D. odorifera* and *D. tonkinensis* in a large population scale. Whether the distinction between them is at the level of species or subspecies is worth discussing. However, it would be less wise to merge the two independent evolutionary lineages, whose wood products possesses huge differences in price, as one taxonomic entity, as the practice could cause damage to the forestry session, and bring about confusion in forming conservation strategies.

Acknowledgements

We would like to thank Lianxuan Zhou, Xianqiang Li, and Kai Zhang from South China Botanical Garden, Chinese Academy of Sciences for the materials collection. We would also like to thank Zhirong Zhang and Hongtao Li from the Kunming Institute of Botany, Chinese Academy of Sciences for helping in the experiment works and data analyses of next generation sequencing. Moreover, thanks to Oscar Alejandro Perez Escobar from Royal Botanic Gardens, Kew for providing useful suggestions on the data analysis. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA13020501), the National Natural Science Foundation of China (Grant No. 31670193), Publicity and Education Base Construction of Guangdong Forestry Science & Technology (Grant No. 2019-08), and Science and Technology Planning Project of Guangdong Province (Grant No. 2019B030316020).

References

- Ban NT. 1998. *Dalbergia odorifera*. The IUCN red list of threatened species 1998:e.T32398A9698077 [online]. Available from <http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32398A9698077.en>
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.
- Bhagwat RM, Dholakia BB, Kadoo NY, Balasundaran M, Gupta VS. 2015. Two new potential barcodes to discriminate *Dalbergia* species. *PLoS One* 10: e0142965.
- Bolson M, Smidt ED, Brotto ML, Silva-Pereira V. 2015. ITS and *trnH-psbA* as efficient DNA barcodes to identify threatened commercial woody angiosperms from southern Brazilian Atlantic rainforests. *PLoS One* 10: e0143049.
- Cai ZM, Zhang YX, Zhang LN, Gao LM, Li DZ. 2012. Testing four candidate barcode markers in temperate woody bamboos (Poaceae: Bambusoideae). *Journal of Systematics and Evolution* 50: 527–539.
- Cavalier T. 1974. Palindromic base sequences and replication of eukaryote chromosome ends. *Nature* 250: 467–470.
- Chan SC, Chang YS, Wang JP, Chen SC, Kuo SC. 1998. Three new flavonoids and antiallergic, anti-inflammatory constituents from the heartwood of *Dalbergia odorifera*. *Planta Medica* 64: 153–158.
- Cheng HT, Liu SW, Li M. 2014. *Precious hardwood species and wood recognition*. Beijing: China Agricultural Science Technology Press.
- Deguiloux MF, Pemonge MH, Petit RJ. 2002. Novel perspectives in wood certification and forensics: Dry wood as a source of DNA. *Proceedings of the Royal Society B: Biological Sciences* 269: 1039–1046.
- Deng CY, Xin GL, Zhang JQ, Zhao DM. 2018. Characterization of the complete chloroplast genome of *Dalbergia hainanensis* (Leguminosae), a vulnerable endangered legume endemic to China. *Conservation Genetics Resources* 11: 105–108.
- Dev SA, Muralidharan EM, Sujanalal P, Balasundaran M. 2014. Identification of market adulterants in East Indian sandalwood using DNA barcoding. *Annals of Forest Science* 71: 517–522.
- Dong WP, Liu H, Xu C, Zuo YJ, Chen ZJ, Zhou SL. 2014. A chloroplast genomic strategy for designing taxon specific DNA mini-barcode: A case study on ginsengs. *BMC Genetics* 15: 138.
- Dong WP, Xu C, Li CH, Sun JH, Zuo YJ, Shi S, Cheng T, Guo JJ, Zhou SL. 2015. Ycf1, the most promising plastid DNA barcode of land plants. *Scientific Reports* 5: 8348.
- Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.
- Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster SG, Hajibabaei M, Husband BC. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources* 9: 130–139.
- Gamboa-Tuz SD, Pereira-Santana A, Zhao T, Schranz ME, Castano E, Rodriguez-Zapata LC. 2018. New insights into the phylogeny of the TMBIM superfamily across the tree of life: Comparative genomics and synteny networks reveal independent evolution of the BI and LFG families in plants. *Molecular Phylogenetics and Evolution* 126: 266–278.
- General administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. 2017. National standard of the People's Republic of China –Hongmu. GB/T 18107-2017 Hongmu (S).
- Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47: W59–W64.
- Guo X, Zheng LH, Chai M, Zhu PJ, Chen ZX. 2011. Study on chemistry composition of essential oil from the seeds of *Dalbergia odorifera* T. Chen extracted by two methods. *Science and Technology of Food Industry* 10: 95–98.
- Hartvig I, Czako M, Kjaer ED, Nielsen LR, Theilade I. 2015. The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). *PLoS One* 10: e0138231.
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evolutionary and Biology* 14: 151.
- Iwasaki T, Aoki K, Seo A, Murakami N. 2012. Comparative phylogeography of four component species of deciduous broad-leaved forests in Japan based on chloroplast DNA variation. *Journal of Plant Research* 125: 207–221.

- Jiao LC, Lu Y, He T, Li JN, Yin YF. 2019. A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three *Pterocarpus* species. *Planta* 250: 95–104.
- Jin JJ, Yu WB, Yang JB, Song Y, Yi TS, Li DZ. 2018. GetOrganelle: A simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv*.
- Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, Watanabe A, Idesawa K, Ishikawa A, Kawashima K. 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* 7: 331–338.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.
- Klitgård BB, Lavin M. 2005. *Dalbergieae s.l.* In: Lewis G, Schrire B, Mackinder B, Lock M eds. *Legumes of the world*. London: Royal Botanical Garden, Kew. 307–335.
- Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289–1291.
- Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC. 2009. Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. *BMC Plant Biology* 9: 131.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.
- Lee DS, Kim KS, Ko W, Li B, Keo S, Jeong GS, Oh H, Kim YC. 2014. The neoflavonoid latifolin isolated from MeOH extract of *Dalbergia odorifera* attenuates inflammatory responses by inhibiting NF- κ B activation via Nrf2-mediated heme oxygenase-1 expression. *Phytotherapy Research* 28: 1216–1223.
- Li SJ. 2017. *Dalbergia in Asia*. Beijing: Science Press.
- Liu FM, Hong Z, Yang ZJ, Zhang NN, Liu XJ, Xu DP. 2019a. De novo transcriptome analysis of *Dalbergia odorifera* and transferability of SSR markers developed from the transcriptome. *Forests* 10: 98.
- Liu FM, Zhang NN, Liu XJ, Yang ZJ, Jia HY, Xu DP. 2019b. Genetic diversity and population structure analysis of *Dalbergia odorifera* germplasm and development of a core collection using microsatellite markers. *Genes* 10: 281.
- Liu LX, Li R, Worth JRP, Li X, Li P, Cameron KM, Fu CX. 2017. The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myricaceae): Implications for understanding the evolution of Fagales. *Frontiers in Plant Science* 8: 968.
- Luo ZF, Zhang XF, Pan B, Lu BY, Ruan ZY, Wu B. 2012. Anatomy structure and physical and mechanical properties of *Dalbergia tonkinensis*. *Journal of Anhui Agricultural University* 39: 493–496.
- Ma RK, Chen X, Li YJ, Fu YL. 2018. The identification of *Dalbergia odorifera* and *Dalbergia benthami* Prain on the basis of PY-GC-MS analysis. *China Forest Products Industry* 10: 37–41.
- Marco J, Artajona J, Larrechi MS, Rius FX. 1994. Relationship between geographical origin and chemical-composition of wood for oak barrels. *American Journal of Enology and Viticulture* 45: 192–200.
- Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ. 2011. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal* 9: 328–333.
- Ohyama M, Baba K, Itoh T. 2001. Wood identification of Japanese *Cyclobalanopsis* species (Fagaceae) based on DNA polymorphism of the intergenic spacer between *trnT* and *trnL* 5' exon. *Journal of Wood Science* 47: 81–86.
- Panero JL, Freire SE, Espinar LA, Crozier BS, Barboza GE, Cantero JJ. 2014. Resolution of deep nodes yields an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Molecular Phylogenetics and Evolution* 80: 43–53.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15: 50.
- Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R. 2009. DNA from processed and unprocessed wood: Factors influencing the isolation success. *Forensic Science International Genetics* 3: 185–192.
- Rubinoff D, Cameron S, Will K. 2006. Are plant DNA barcodes a search for the Holy Grail? *Trends in Ecology and Evolution* 21: 1–2.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* 33: W686–W689.
- Song Y, Dong WP, Liu B, Xu C, Yao X, Gao J, Corlett RT. 2015. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Frontiers in Plant Science* 6: 662.
- Song Y, Zhang YJ, Xu J, Li WM, Li MF. 2019. Characterization of the complete chloroplast genome sequence of *Dalbergia* species and its phylogenetic implications. *Scientific Reports* 9: 20401.
- Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH. 2014. Complete chloroplast genome sequence of Omani Lime (*Citrus aurantifolia*) and comparative analysis within the Rosids. *PLoS One* 9: e113049.
- Timme RE, Kuehl JV, Boore JL, Jansen RK. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *American Journal of Botany* 94: 302–312.
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* 44: W232–W235.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research* 40: e115.
- Vatanparast M, Klitgård BB, Adema FACB, Pennington RT, Yahara T, Kajita T. 2013. First molecular phylogeny of the pantropical genus *Dalbergia*: Implications for infrageneric circumscription and biogeography. *South African Journal of Botany* 89: 143–149.
- Wang F, Wang SN, Huang AM, Zhou Q. 2016. Distinction of *Dalbergia odorifera* confused species by FTIR, 2nd Derivative IR, 2D-IR spectroscopy. *Spectroscopy and Spectral Analysis* S1: 210–211.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* 31: 3350–3352.
- Wu Y, Liu F, Yang DG, Li W, Zhou XJ, Pei XY, Liu YG, He KL, Zhang WS, Ren ZY, Zhou KH, Ma XF, Li ZH. 2018. Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny. *Frontiers in Plant Science* 9: 376.
- Xu JY, Li WH. 2003. Identification of rosewood. *China Timber* 5: 8–11.
- Yang DH, Chen F, Song XQ, Luo GY, Xu Q. 2012. Floristic composition and community characteristics of *Dalbergia odorifera* T. Chen, an

- endemic species to Hainan island. *Chinese Journal of Tropical Agriculture* 32: 110–115.
- Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. 2013. Complete chloroplast genome of the genus *Cymbidium*: Lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolution and Biology* 13: 84.
- Yang L, Fang CR, Zhang J, Yu HX, Wang Z, Zhu JL, Yang WM, Chai ZL. 2016. The identification of *Dalbergia odorifera* and *D. tonkinensis* by gas chromatography mass spectrometry. *Journal of Nanjing Forestry University* 40: 97–103.
- Yu M, Jiao LC, Guo J, Wiedenhoef AC, He T, Jiang XM, Yin YF. 2017. DNA barcoding of vouchered xylarium wood specimens of nine endangered *Dalbergia* species. *Planta* 246: 1165–1176.
- Yu M, Liu K, Zhou L, Zhao L, Liu SQ. 2016. Testing three proposed DNA barcodes for the wood identification of *Dalbergia odorifera* T. Chen and *Dalbergia tonkinensis* Prain. *Holzfor-schung* 70: 127–136.
- Zeng CX, Hollingsworth PM, Jing Y, He ZS, Zhang ZR, Li DZ, Yang JB. 2018. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* 14: 43.
- Zhang LH, Zhou DS, Guo CY, Liao WT, Zeng Y. 2018. Identification and analysis of the volatile oil from *Dalbergia odorifera* T. Chen and other plants in *Dalbergia* genus by GC-MS. *Journal of Guangdong Pharmaceutical University* 5: 579–585.
- Zhao ML, Song Y, Ni J, Yao X, Tan YH, Xu ZF. 2018. Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). *Scientific Reports* 8: 8844.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406–15.

Supplementary Material

The following supplementary material is available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12598/supinfo:Supplementary information>.

Table S1. Materials information for the chloroplast genome samples and the population samples in this study.

Table S2. Primer pairs for species-specific population test.

Doc. S1. The sequences matrix of Fig. 4.