

# SCIENTIFIC REPORTS



OPEN

## DNA Barcode Authentication and Library Development for the Wood of Six Commercial *Pterocarpus* Species: the Critical Role of Xylarium Specimens

Lichao Jiao<sup>1,2</sup>, Min Yu<sup>1,2</sup>, Alex C. Wiedenhoef<sup>3,4,5,6</sup>, Tuo He<sup>1,2</sup>, Jianing Li<sup>7</sup>, Bo Liu<sup>1,2</sup>, Xiaomei Jiang<sup>1,2</sup> & Yafang Yin<sup>1,2,4</sup>

DNA barcoding has been proposed as a useful tool for forensic wood identification and development of a reliable DNA reference library is an essential first step. Xylaria (wood collections) are potentially enormous data repositories if DNA information could be extracted from wood specimens. In this study, 31 xylarium wood specimens and 8 leaf specimens of six important commercial species of *Pterocarpus* were selected to investigate the reliability of DNA barcodes for authentication at the species level and to determine the feasibility of building wood DNA barcode reference libraries from xylarium specimens. Four DNA barcodes (ITS2, *matK*, *ndhF-rpl32* and *rbcl*) and their combination were tested to evaluate their discrimination ability for *Pterocarpus* species with both TaxonDNA and tree-based analytical methods. The results indicated that the combination barcode of *matK* + *ndhF-rpl32* + ITS2 yielded the best discrimination for the *Pterocarpus* species studied. The mini-barcode *ndhF-rpl32* (167–173 bps) performed well distinguishing *P. santalinus* from its wood anatomically inseparable species *P. tinctorius*. Results from this study verified not only the feasibility of building DNA barcode libraries using xylarium wood specimens, but the importance of using wood rather than leaves as the source tissue, when wood is the botanical material to be identified.

Increasing concern about and demand for biodiversity conservation world-wide and substantial declines in biological diversity at various spatial, temporal and biological scales<sup>1</sup> are driving the need for species identification for forensics. For forest systems, illegal logging and the illegal timber trade are major problems domestically and internationally, threatening not just individual species, but entire ecosystems. Illegal logging is both a consumer- and producer-country driven phenomenon, and international efforts to respond to the problem consist of the enactment of laws to prohibit or limit the trade in illegally sourced timber. Broad international trade restrictions are imposed primarily through the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), which lists species in three appendices according to the degree of protection required<sup>2</sup>.

In recent years, several consumer countries and regions have also taken action to reduce the trade in forest products derived from illegally logged sources<sup>3</sup>. The United States amended the Lacey Act in 2008, which makes it unlawful to import into the United States any plant (or plant product) that was illegally harvested. In Australia, the Illegal Logging Prohibition Act (2012) was enacted to restrict the trade of illegally logged timber. The European Union's E.U. Timber Regulation (EUTR) came into effect in 2013, prohibiting illegally sourced

<sup>1</sup>Department of Wood Anatomy and Utilization, Chinese Research Institute of Wood Industry, Chinese Academy of Forestry, Beijing, 100091, China. <sup>2</sup>Wood Collections (WOODPEDIA), Chinese Academy of Forestry, Beijing, 100091, China. <sup>3</sup>Center for Wood Anatomy Research, USDA Forest Service, Forest Products Laboratory, Madison, WI, 53726, USA. <sup>4</sup>Department of Botany, University of Wisconsin, Madison, WI, 53706, USA. <sup>5</sup>Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN, 47907, USA. <sup>6</sup>Ciências Biológicas (Botânica), Univesida de Estadual Paulista – Botucatu, Botucatu, São Paulo, Brazil. <sup>7</sup>Rubber Research Institute, Chinese Academy of Tropical Agricultural Science, Hainan, 571737, China. Lichao Jiao and Min Yu contributed equally to this work. Correspondence and requests for materials should be addressed to Y.Y. (email: [yafang@caf.ac.cn](mailto:yafang@caf.ac.cn))

timber and timber products in the EU market. These legislative actions and subsequent enforcement of these laws demonstrate the urgent global attention on forest species protection. Enforcement actions to date in the United States have largely focused on ebony and rosewood from Madagascar (Dept. of Justice 2012), Eurasian hardwoods (Dept. of Justice 2016), and tropical woods from Peru (Dept. of Justice 2017) indicating that wood forensic methods, including DNA barcoding reference libraries, for valuable woods from anywhere in the world could play a critical role in law enforcement for forest protection.

*Pterocarpus* Jacq., is a pantropical genus in the family Leguminosae, containing approx. 70 species<sup>4</sup>. The timber of *Pterocarpus* is globally valued for its beauty, wood quality, medicinal properties and even valuable bioactive compounds. This high value and increase in demand for the timber has led to illegal and excessive logging resulting in threat to wild *Pterocarpus* populations. In 1995, CITES listed *P. santalinus* under Appendix II to regulate trade in logs, wood chips and unprocessed broken material<sup>5</sup>. *P. erinaceus* was added to Appendix II at the 17th Meeting of the CITES in 2017. Concurrently, the International Union for Conservation of Nature (IUCN) also listed *P. santalinus* and *P. zenkeri* as endangered, *P. indicus* and *P. marsupium* as vulnerable, and *P. angolensis* as near threatened<sup>6</sup>. In China, the species *P. indicus* was listed in the second-class category of the National List of Local Protected Flora issued by the Chinese Government in 1999<sup>7</sup>. Among the *Pterocarpus* species, *P. santalinus*, endemic to the Southern parts of Eastern Ghats of India especially in Andhra Pradesh, is known for its characteristic color, texture, quality and the medical value of its timber, which makes it of particular economic importance, especially in China. In recent years, the wood from *P. tinctorius* (non-CITES) mostly distributed in Central and Southern Africa, appeared on the international lumber market as a substitute for *P. santalinus* (CITES App. II). Its macroscopic wood properties, e.g. color, grain, density, and its anatomical structure are very similar to that of *P. santalinus*. Due to the great difference in economic value, *P. tinctorius* has often been treated as an adulterant of *P. santalinus* in the timber market. Thus, developing accurate species-level identification for *Pterocarpus* wood is significant for natural resource protection and global trade monitoring.

Traditional wood identification relies on diagnostic anatomical features, either macroscopic or microscopic but rarely can provide a precise discrimination of wood at the species level, which limits the enforcement of CITES regulations and related laws. Moreover, traditional wood identification requires expert taxonomic and anatomical knowledge that takes years to gain. To overcome such limitations, recent advances in molecular diagnostic tools for plants have the capacity to improve upon traditional methods of species identification.

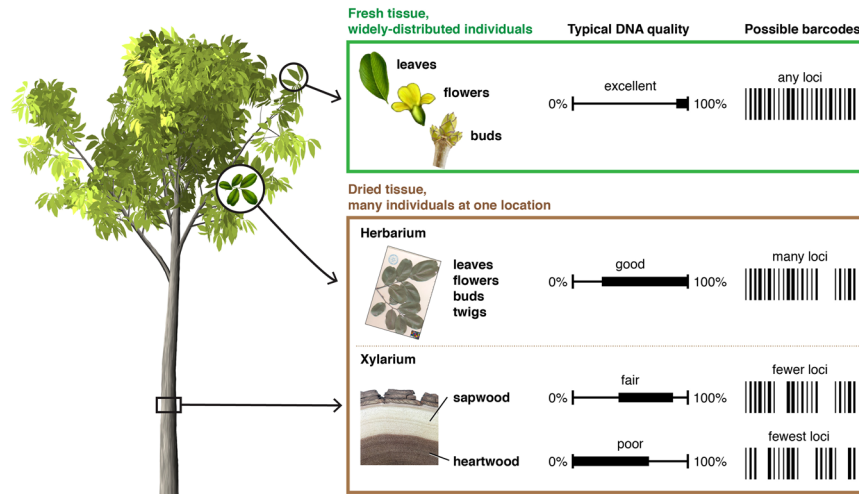
For the last decade, DNA barcoding has been the subject of extensive research and application as an accurate and convenient tool for species identification<sup>8–12</sup>. DNA barcoding is a genetic approach based on a short DNA sequence from a standard part of a genome – in animals this is typically a region of the cytochrome c oxidase subunit 1 (CO1) mitochondrial region. In plants, mitochondrial mutation rates are too slow for species-level identification, so plastid and nuclear regions are typically chosen as barcodes<sup>13,14</sup>. The Consortium for the Barcode of Life (CBOL) proposed a combination of both the chloroplast DNA (cpDNA) ribulose-bisphosphate carboxylase (*rbcL*) gene and maturase K (*matK*) genes as the core DNA barcodes for plants. Chen *et al.*<sup>15</sup> proposed that the ITS2 region could be potentially used as a standard DNA barcode, especially for identifying medicinal plants and their closely related species<sup>15</sup>. Additionally, the *ndhF-rpl32* intergenic spacer in the short single copy region of the chloroplast genome, which was noted as highly variable<sup>16</sup> by Timme *et al.*<sup>16</sup>, has also been used for phylogenetic studies<sup>17,18</sup>. DNA barcodes are established tools for identifying herbal medicinal materials, in quality control, and in forensic science<sup>10,19,20</sup>. Additionally, a number of studies relying upon DNA barcoding have verified the utility and potential for wood species identification<sup>11,12,21–23</sup>.

Despite the desirability of using DNA barcoding broadly in plant forensics, the lack of a reliable DNA barcoding reference library is the main barrier to its application for the next few years<sup>24–28</sup>. To generate such a reference library, access to correctly identified specimens of the species of interest is required. If these specimens are living individuals, extracting DNA of sufficient quality and quantity is routine, but for a widely-distributed taxon would involve significant expense and time to travel and sample across the taxon's range. An alternative to this approach is to sample from botanical collections like herbaria<sup>28</sup>, where specimens of many taxa are gathered in one place, but from which high quality DNA may not be available. It is one step more complicated to develop a reference library for DNA barcoding of wood, because DNA extraction from wood is not necessarily as simple or direct as from other plant parts that can be collected and analyzed in the living state (Fig. 1). Because wood is a botanically poor source of DNA even prior to industrial processing, developing DNA reference libraries for wood discrimination is most sensibly done from scientifically collected wood specimens from xylaria – this ensures that extraction protocols, chosen barcodes, and developed methods are directly applicable to wood as a commercial product. There are approx. 180 xylaria containing on the order of 1.5 million wood specimens in the world<sup>29</sup>. Historically, xylaria played an important role in the development of wood science and priority forestry programs, as resources supporting timber trade, law enforcement, archaeology, and conservation and restoration of architectural wood heritage. Xylaria still serve these functions, but they also have potential to be enormous resources for DNA studies, providing abundant and reliable resource materials for establishing DNA barcoding reference databases, though few researchers to date have taken advantage of xylaria in this way<sup>11</sup>.

In this study, we selected xylarium wood specimens of six important commercial *Pterocarpus* species and evaluated four candidate DNA barcodes (ITS2, *matK*, *ndhF-rpl32*, and *rbcL*) for their efficacy at species-level separation. The specific objectives were to (i) test the discrimination ability of the barcodes using both TaxonDNA as well as a tree-based method, (ii) determine the efficacy of the mini-barcode *ndhF-rpl32* to separate *P. santalinus* from *P. tinctorius*, (iii) provide the essential data for establishing a reference library for *Pterocarpus* using the chosen barcodes individually and in combination, and (iv) verify the feasibility of building wood DNA barcode reference libraries using wood specimens from a xylarium.

## Results

**Wood anatomical separation of *P. santalinus* and *P. tinctorius*.** The wood anatomical features of *P. santalinus* and *P. tinctorius* are almost identical (Fig. 2). Wood diffuse-porous; vessels exclusively solitary, occasionally with radial multiples of 2 to 3, and often filled with dark gums; intervessel pits alternate; vessel-ray pits



**Figure 1.** A schematic representation of the potential strengths and weaknesses of source tissue (fresh, herbarium, xylarium) for developing DNA barcoding reference libraries.

similar to intervessel pits; perforation plates simple; axial parenchyma, aliform, confluent and narrow bands of 1–4 cells wide; prismatic crystals in chambered axial parenchyma cells; axial parenchyma cells storied; fibres thick-walled, storied; rays exclusively uniseriate, occasionally 2 cells wide, 2 to 10 cells high, homocellular, consisting of procumbent cells; All rays storied. In addition to these anatomical similarities, we report that ethanol extract color, heartwood surface fluorescence, heartwood water extract fluorescence, and heartwood ethanol extract fluorescence are all also indistinguishable. It is thus impossible to make a forensically valid separation of *P. santalinus* from *P. tinctorius* based on wood anatomical features.

**Barcode Recovery and Sequence Characteristics.** The recovery success rate was the highest for *ndhF-rpl32* (90%), followed by *matK* (82%) and *rbcl* (70%), while ITS2 exhibited the lowest rate (67%). In total, 123 sequences generated in this work were deposited to GenBank (Accessions, ITS2: KY829137–KY829162; *matK*: KY829163–KY829195; *ndhF-rpl32*: KY829196–KY829232; *rbcl*: KY829233–KY829259) (Supplementary Table S1).

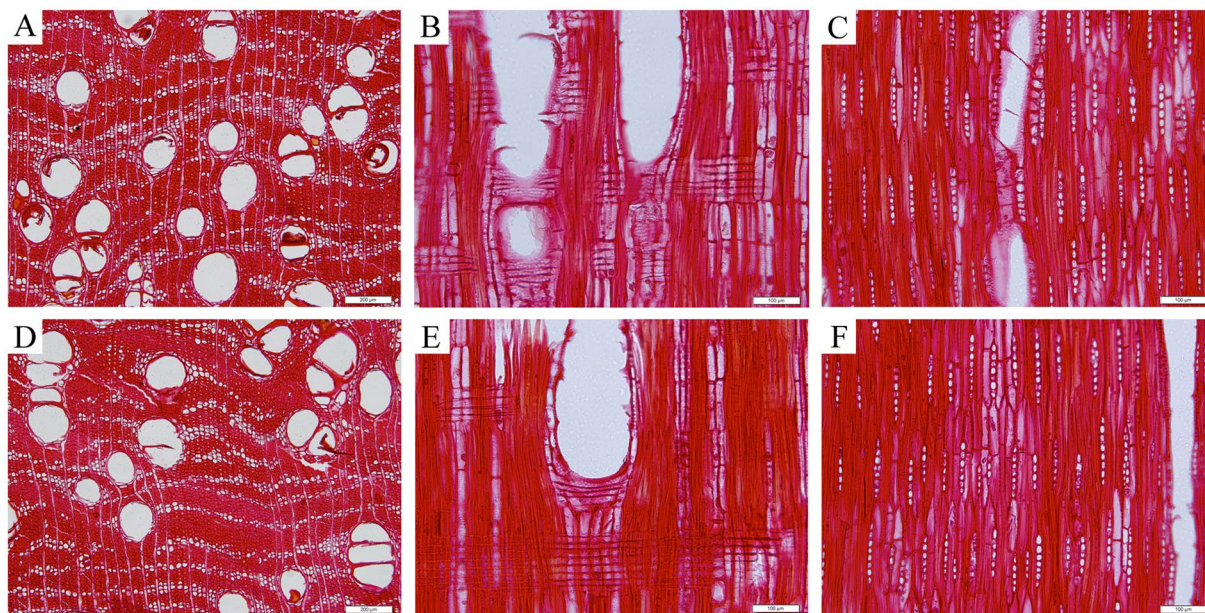
The features of the four DNA barcodes were shown in Table 1. The length of the aligned *rbcl* sequences was 350 bp with 47 variable sites and 15 informative sites. The aligned *matK* sequence was 239 bp long, with 10 variable sites and 10 informative sites. In the ITS2, the sequence was 234 bp in length, with 75 variable sites, 69 informative sites and 16 indels. For the sequence of *ndhF-rpl32*, the aligned length was 173 bp, with 14 variable sites, 12 informative sites and six indels. Among the four DNA barcodes, ITS2 had the highest proportion of variable (32.05%) and informative (29.49%) sites, followed by *rbcl* (13.43% and 4.29%) and *ndhF-rpl32* (8.09% and 6.94%), with *matK* showing the lowest values (4.18% and 4.18%).

The pairwise intraspecific distances for the barcodes ranged from a minimum value of 0.0000 for all four barcodes to a maximum value of 0.0962 (ITS2), and the mean intraspecific distances ranged from 0.0026 (*matK*) to 0.0200 (ITS2). The pairwise interspecific distances for the barcodes ranged from 0.0000 for all four barcodes to 0.1681 (ITS2), and the mean interspecific distances ranged from 0.0073 (*rbcl*) to 0.0800 (ITS2) (Table 2). ITS2 shows the highest mean intra- and inter-specific distances. The pairwise intraspecific distances for combined barcodes ranged from 0.0000 for all combinations to 0.0638 (*ndhF-rpl32* + *rbcl*), and the mean intraspecific distances ranged from 0.0027 (*matK* + *ndhF-rpl32* + ITS2) to 0.0091 (*ndhF-rpl32* + *rbcl*). The pairwise interspecific distances for combined barcodes ranged from 0.0000 (*matK* + *ndhF-rpl32*, *matK* + *rbcl* and *ndhF-rpl32* + *rbcl*) to 0.0954 (*rbcl* + ITS2) and the mean interspecific distances ranged from 0.0133 (*matK* + *rbcl*) to 0.0524 (*ndhF-rpl32* + ITS2).

**DNA Barcoding Gap Assessment.** Barcoding gaps, the absence of overlapping regions between intra- and interspecific distances, were evaluated by the results of the distribution graph obtained in the “pairwise summary” function in TaxonDNA (Supplementary Figure S1). In the study, no single- or multi-barcodes exhibited clear barcoding gaps; all barcodes overlapped between the intra- and interspecific distances. However, the mean interspecific divergence was higher than that of the corresponding intraspecific variation for each of the barcodes (Table 2). Among the single barcodes, ITS2 had the highest variation in interspecific divergence compared to the range of intraspecific distances (Table 2). When barcodes were individually analyzed, ITS2 presented the best barcode gap performance, with 69.6% of pairwise interspecific distances greater than 0.05 and 95.9% of pairwise intraspecific distances lower than 0.05. Conversely, unsatisfactory results were observed for *matK*, *ndhF-rpl32* and *rbcl* separately, with almost total overlap of intra- and interspecific variation (Supplementary Figure S1) for each.

As for the barcode combinations, the best results were found for *matK* + ITS2 and *matK* + *ndhF-rpl32* + ITS2, with 98.9% and 95.1% of pairwise interspecific distances greater than 0.05, respectively, and 91.8% of pairwise intraspecific distances lower than 0.05, both of which also outperformed any single barcode. All other barcode combinations showed clear overlap (Supplementary Figure S1).





**Figure 2.** Wood anatomical features of *P. santalinus* and *P. tinctorius*. (A,B and C) Transverse, radial, and tangential sections of *P. santalinus* wood, respectively. (D,E and F) Transverse, radial, and tangential sections of *P. tinctorius* wood, respectively. Scale bars, 200  $\mu\text{m}$  (A and D) and 100  $\mu\text{m}$  (B,C,E and F).

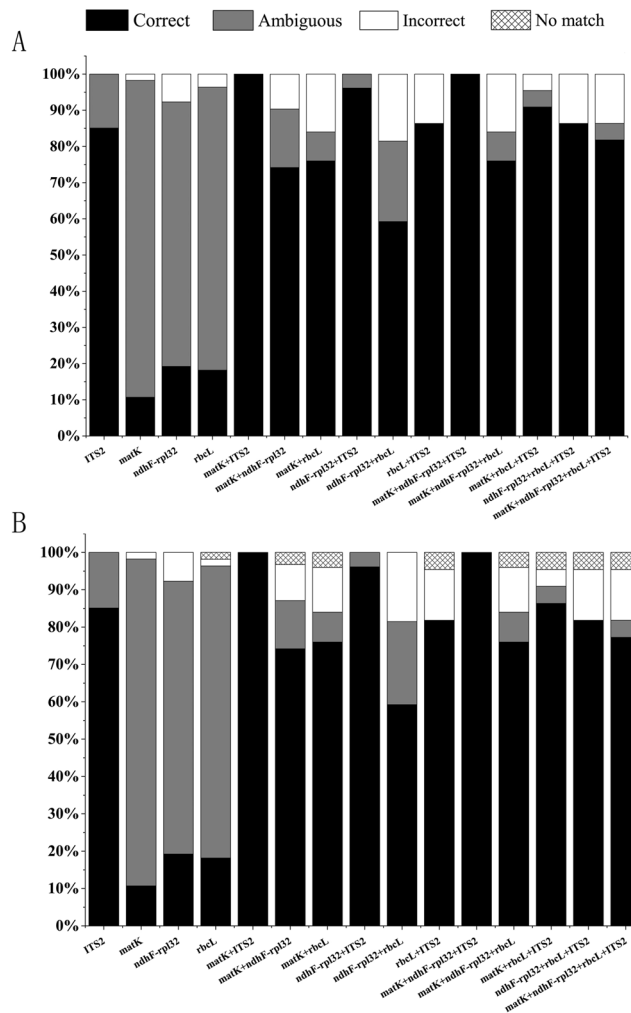
DNA marker	Recovery rate (%)	Sequence length (bp)	Aligned sequence length (bp)	G + C ratio (%)	No. variable sites (%)	No. parsimony informative sites (%)	Indel length (bp)
ITS2	66.67	219–224	234	66.1	75 (32.05)	69 (29.49)	16
<i>matK</i>	82.05	273	239	36.0	10 (4.18)	10 (4.18)	0
<i>ndhF-rpl32</i>	89.74	167–173	173	26.6	14 (8.09)	12 (6.94)	6
<i>rbcL</i>	69.67	485	350	40.6	47 (13.43)	15 (4.29)	0

**Table 1.** The characteristics of the four DNA barcode loci.

Barcode loci and combinations	Intraspecific distance			Interspecific distance		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
a) ITS2	0.0000	0.0962	0.0200	0.0000	0.1681	0.0800
b) <i>matK</i>	0.0000	0.0293	0.0026	0.0000	0.0335	0.0099
c) <i>ndhF-rpl32</i>	0.0000	0.0226	0.0045	0.0000	0.0292	0.0091
d) <i>rbcL</i>	0.0000	0.0914	0.0063	0.0000	0.0943	0.0073
e) <i>matK</i> + ITS2	0.0000	0.0101	0.0030	0.0084	0.0671	0.0433
f) <i>matK</i> + <i>ndhF-rpl32</i>	0.0000	0.0122	0.0015	0.0000	0.0146	0.0083
g) <i>matK</i> + <i>rbcL</i>	0.0000	0.0578	0.0060	0.0000	0.0608	0.0133
h) <i>ndhF-rpl32</i> + ITS2	0.0000	0.0131	0.0035	0.0019	0.0824	0.0524
i) <i>ndhF-rpl32</i> + <i>rbcL</i>	0.0000	0.0638	0.0091	0.0000	0.0755	0.0173
j) <i>rbcL</i> + ITS2	0.0000	0.0464	0.0067	0.0013	0.0954	0.0410
k) <i>matK</i> + <i>ndhF-rpl32</i> + ITS2	0.0000	0.0091	0.0027	0.0065	0.0582	0.0392
l) <i>matK</i> + <i>ndhF-rpl32</i> + <i>rbcL</i>	0.0000	0.0467	0.0058	0.0012	0.0551	0.0159
m) <i>matK</i> + <i>rbcL</i> + ITS2	0.0000	0.0374	0.0053	0.0049	0.0739	0.0335
n) <i>ndhF-rpl32</i> + <i>rbcL</i> + ITS2	0.0000	0.0388	0.0059	0.0010	0.0787	0.0383
o) <i>matK</i> + <i>ndhF-rpl32</i> + <i>rbcL</i> + ITS2	0.0000	0.0327	0.0064	0.0050	0.0638	0.0340

**Table 2.** Genetic distance generated using Kimura 2-parameter model analysis for the candidate barcode loci and their combinations.

**Species Discrimination based on TaxonDNA and Tree-based Analysis.** The parameters “best match” and “best close match” from Taxon DNA were used to analyze all sequences generated in this study as well as those downloaded from the GenBank database (Fig. 3). For single-locus barcodes, both the “best match”



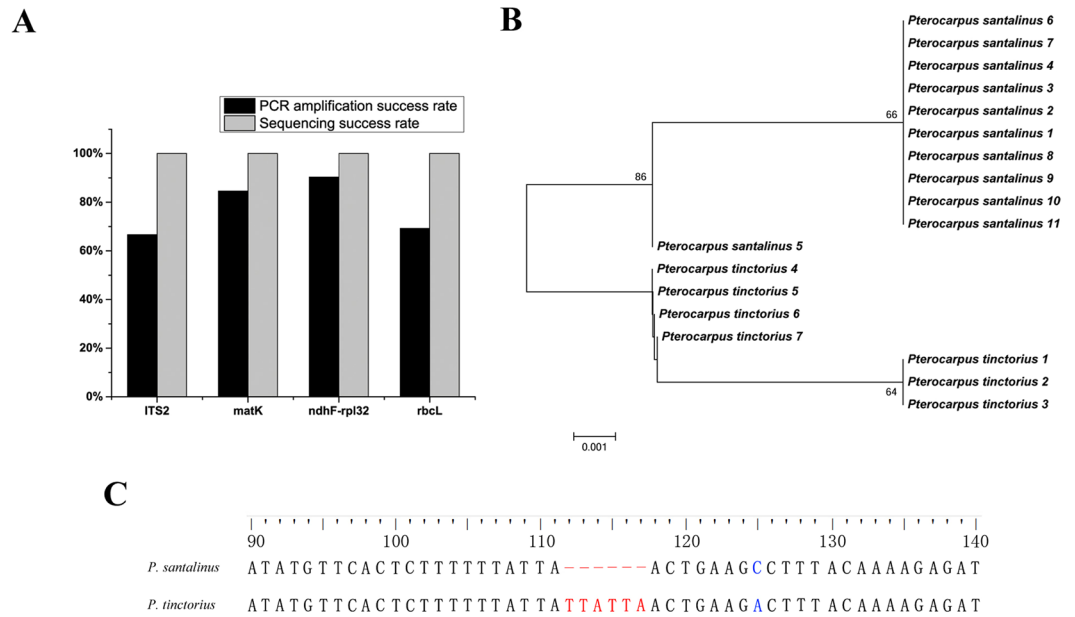
**Figure 3.** Success of species identification based on analysis of the “best match” (A) and “best close match” (B) functions of TaxonDNA program for the four DNA barcodes and their combinations.

and “best close match” methods provided the similar species discrimination success rate. ITS2 showed the highest success rate (85.1%), followed by *ndhF-rpl32* (20.0%), *rbcl* (18.2%), while *matK* exhibited the lowest rate (10.7%). The identification success rates for all barcode combinations were generally higher than those of the single barcodes. The highest success rate (100%) of barcode combinations based on the “best match” and “best close match” analysis was obtained by the two-barcode combination of *matK* + ITS2 and three-barcode combination of *matK* + *ndhF-rpl32* + ITS2. The *ndhF-rpl32* + *rbcl* combination exhibited the lowest performance for correct identification. All barcode combinations that included ITS2, i.e. *matK* + ITS2, *ndhF-rpl32* + ITS2, *rbcl* + ITS2, *matK* + *ndhF-rpl32* + ITS2, *matK* + *rbcl* + ITS2, *ndhF-rpl32* + *rbcl* + ITS2 and *matK* + *ndhF-rpl32* + *rbcl* + ITS2, provided higher identification success rates than other chloroplast DNA barcode combinations (Fig. 3).

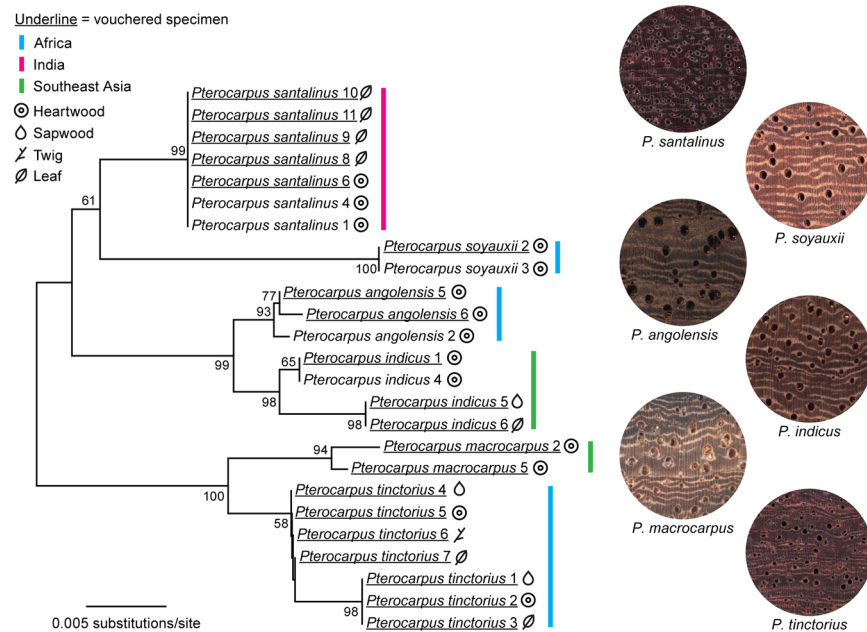
Bootstrap support for species-specific clusters based on unrooted neighbor-joining (NJ) trees for the four barcodes and their combinations were calculated (Supplementary Figure S2). When barcodes were individually analyzed, the highest species discrimination successes were obtained by ITS2 and *rbcl* (16.7%), whereas the barcodes *matK* and *ndhF-rpl32* could not distinguish any *Pterocarpus* species (Supplementary Table S4). The mini-barcode *ndhF-rpl32*, 167–173 bps in length, can separate the two anatomically similar species, *P. santalinus* and *P. tinctorius* using neighbor-joining tree analysis (Fig. 4B). Furthermore, six continuous diagnostic characters (insertion/deletion) at nucleotide positions from 112 to 117 (TTATTA) were found within the *ndhF-rpl32* region (Fig. 4C), which was a distinguishing feature based on the character-based approach. Discrimination of all six species using only one barcode was insufficient to provide an accurate resolution among the *Pterocarpus* species studied here. When combining two to four barcodes, the highest discrimination rate (100%) was obtained by *matK* + *ndhF-rpl32* + ITS2 and *matK* + *rbcl* + ITS2 (Fig. 5). Moreover, the barcode combinations that included ITS2 yielded higher success rates than other chloroplast DNA barcode combinations (Supplementary Table S4).

## Discussion

**Assessment of DNA Barcodes for *Pterocarpus*.** An ideal DNA barcode should be short making it easy for recovery, and have sufficient information to provide maximal species discrimination<sup>30–32</sup>. While this is true for any barcode as a general principle, it is a key concern for barcodes for wood identification, because wood



**Figure 4.** Analysis of discrimination ability of *P. santalinus* and *P. tinctorius* based on the specific mini-barcode *ndhF-rpl32*. **(A)** PCR amplification and sequencing success rate of the four DNA barcodes, **(B)** neighbor-joining tree constructed based on the barcode *ndhF-rpl32*, **(C)** variable sites of the barcode *ndhF-rpl32* between the two species.



**Figure 5.** Taxon identification tree constructed using neighbor-joining analysis of P-distance showing the best-performing barcode combination *matK* + *ndhF-rpl32* + ITS2. Bootstrap values (>50%) are shown above the relevant branches. Photomicrographs ( $\times 16$ ) of *Pterocarpus* xylarium specimens.

is a DNA-poor botanical material in the living tree, and the quality and quantity of DNA in wood degrades with industrial processing, necessitating barcodes known to be recoverable from dry wood. In these *Pterocarpus*, shorter amplicons showed a generally higher recovery rate than longer ones, with the shortest fragment *ndhF-rpl32* having the highest success rate, which is in line with several previous studies<sup>22,28,33</sup>. We expect that the DNA in xylarium wood specimens is typically highly fragmented<sup>33,34</sup>. Additionally, the nuclear ribosomal DNA region ITS2 yielded lower recovery success rate (67%) compared to the chloroplast DNA regions although it is present in multiple copies in the genome. In spite of some amplification disadvantages, ITS2 provided the best



discrimination performance among the four barcodes. The superior identification power of nuclear DNA region ITS2 over plastid barcodes is also consistent with the results of other previous studies<sup>15,19,35–37</sup>.

Although the chloroplast DNA regions *rbcL* and *matK* were proposed as core barcodes for seed plants<sup>31</sup>, the two regions gave low species resolution for *Pterocarpus* in this study. Both *rbcL* and *matK* are widely used in phylogenetic analyses with over 130,000 sequences available in Genbank. Kress *et al.*<sup>30</sup> showed that the *rbcL* sequence evolves slowly and this barcode has been recognized as the lowest divergence of studied plastid genes in flowering plants<sup>30</sup>. Consequently, on average it is not likely to be useful for identification at the species level<sup>15,31,38–40</sup>. It is reported that *matK* shows different discrimination success rates when it comes to different taxonomic groups (e.g. discriminating more than 90% of species in the Orchidaceae<sup>41</sup>) but less than 49% of species in the nutmeg family<sup>40,42</sup>. Meanwhile, despite its power in phylogenetic studies of other species<sup>17,18</sup>, *ndhF-rpl32* showed low resolution for distinguishing all six *Pterocarpus* species in this study.

No single barcode was found to be able to distinguish all six *Pterocarpus* species in this study. Overall, combined barcodes provided higher species resolution than any single barcode, which was consistent with previous studies<sup>12,43,44</sup>. The CBOL Plant Working Group recommended the combination barcode of *rbcL* + *matK* as the core barcode for land plants. Yan *et al.*<sup>32</sup> also demonstrated that the three barcode combination of ITS + *psbA-trnH* + *matK* could give better discrimination performance than single barcodes, and was the best choice for the genus of *Rhododendron*<sup>32</sup>. In this study of *Pterocarpus*, the highest success rate of barcode combinations based on the “best match” and “best close match” analysis of TaxonDNA was obtained by *matK* + ITS2 and *matK* + *ndhF-rpl32* + ITS2. When the tree-based analyses (NJ) were conducted, the combination *matK* + *ndhF-rpl32* + ITS2 and *matK* + *rbcL* + ITS2 gave the best results. We conclude that the combination *matK* + *ndhF-rpl32* + ITS using these two methods is the best combination DNA barcode to resolve six of *Pterocarpus* species (Figs 3 and 5).

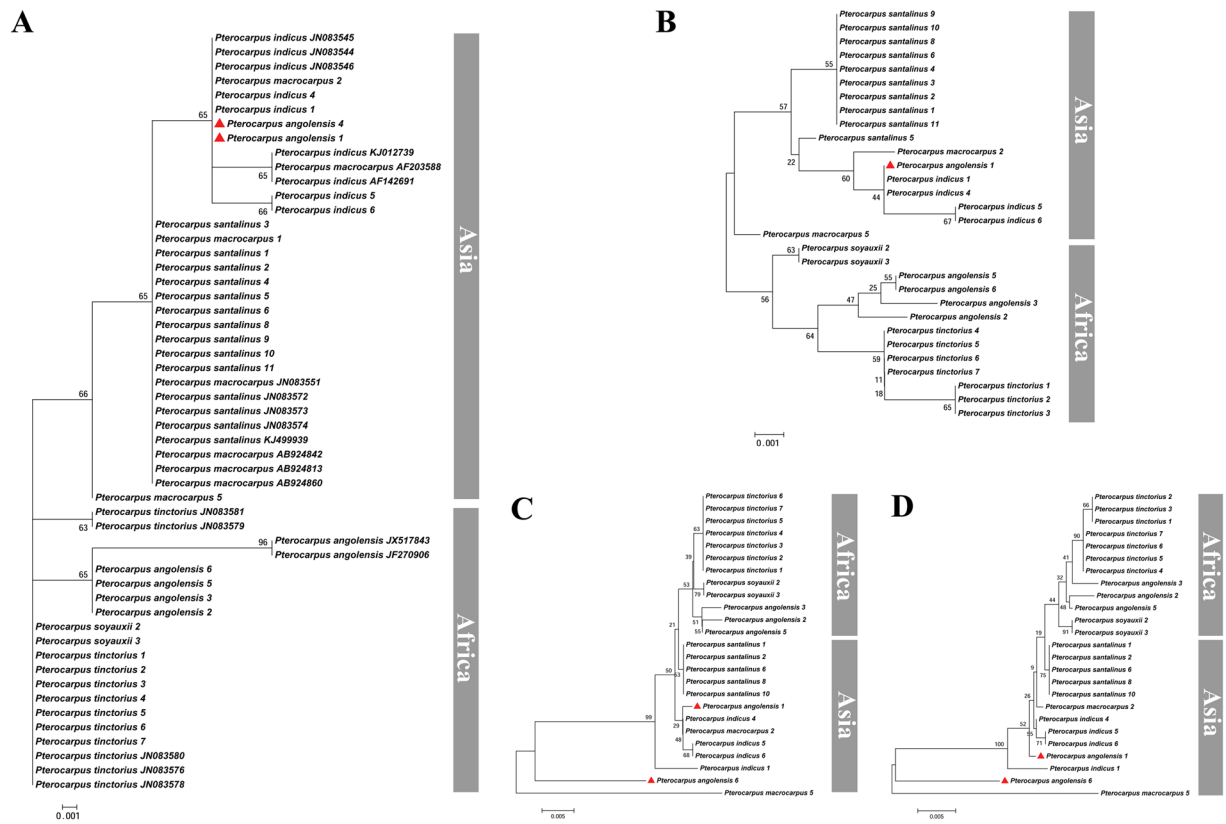
Although the barcode *matK* individually or in combination with other chloroplast DNA barcodes yielded a low success rate for species discrimination, interestingly it has the ability to cluster studied *Pterocarpus* species according to their broad geographic origins (Fig. 6). We found that Asian and African species clustered together except for 1 or 2 samples of *P. angolensis* (Fig. 6). Here we suggest the two chloroplast locus combination of *matK* + *ndhF-rpl32* as a potential barcode for geographic origin tracking of *Pterocarpus* species when the recovery success rate is considered. It has been reported that the chloroplast DNA barcodes that are variable enough to reveal geographic structure could be used to differentiate the origin of timber<sup>45–47</sup>. Additionally, Lee *et al.*<sup>12</sup> also showed that the DNA barcode combination *matK* + *trnL-trnF* + ITS2 had the ability of geographic clustering for *Aquilaria* species<sup>12</sup>.

**Species Discrimination between *P. santalinus* and *P. tinctorius* based on the Special Mini-barcode.** Inasmuch as *P. santalinus* and *P. tinctorius* cannot be separated by wood anatomy but are mixed in trade, an effective method to separate these woods is critically needed. A single DNA barcode targeted to this question alone would be an effective tool, especially if the barcode were easily recovered from both species. The DNA mini-barcode *ndhF-rpl32* could give good performance for distinguishing the two closely related *Pterocarpus* species.

DNA mini-barcodes, short DNA sequences of 100–250 bp, are suitable for species identification within a given taxonomic group of old herbarium/museum specimens when high-quality DNA is not available and seriously degraded DNA is retrieved<sup>9,48,49</sup>. We suggest that the DNA mini-barcoding approach is suitable for species identification of woody tissues, especially in narrow cases to separate a small number of anatomically indistinguishable woods. In this study, the recovery success rate of *ndhF-rpl32* was highest among the four DNA barcodes in the study (Fig. 4) and this parameter has been used as an important criterion to determine whether DNA could be effectively isolated from wood tissues<sup>11,22,23,34</sup>. The reduced taxonomic discriminatory power of a mini-barcode compared to that of a full-length barcode and the taxon-specific nature of which mini-barcodes are most effective are the primary detriments of this approach. If for every group of taxa a new mini-barcode is needed, the basic principle of standardization is violated. Therefore, the choice of position of mini-barcodes from DNA genome is significant in their ability of discriminating species<sup>48–50</sup>. A good DNA mini-barcode candidate should be of high PCR and sequencing success without much loss of species discrimination power, and as broadly applicable as possible.

**Regarding the Utility of DNA Barcoding in the Conservation of and Controlled Trade in *Pterocarpus* Wood.** Biodiversity conservation has rapidly become a focus of attention due to the sharp increase of global forest resources trade, over-exploitation and illegal logging activities. For forest protection and global trade monitoring, developing accurate species-level identification and geographic traceability for wood is a crucial and significant technical prerequisite<sup>2</sup>. The application of DNA barcoding to identify the species and track the geographic origin of internationally traded timber has attracted increasing interest as a potential part of global systems to support sustainable forestry and especially to reduce the behaviors of illegal logging<sup>51</sup>. In addition to this work, previous studies have reported the potential of DNA barcoding to support conservation efforts of wood species, e.g. *Aquilaria*<sup>11,12</sup>, *Dalbergia*<sup>23,51,52</sup> and *Populus*<sup>22</sup>.

DNA barcoding can play an increased role in identification and conservation of *Pterocarpus* species, and of wood species worldwide. Availability of a reliable reference DNA barcode library remains the main obstacle of application of DNA barcoding for the next few years. Our study confirms that xylarium wood specimens are rich sources for reliable DNA sequence data. Xylarium wood specimens could certainly enhance the construction of global DNA barcode reference libraries to support species conservation worldwide, and thus continue to play a critical role as repositories of wood anatomical, chemical, and molecular information for the future.



**Figure 6.** Neighbor-joining tree constructed using the barcode *matK* individual and combined showing geographic clustering pattern in *Pterocarpus* species. (A) *matK*, (B) *matK* + *ndhF-rpl32*, (C) *matK* + *rbcL*, (D) *matK* + *ndhF-rpl32* + *rbcL*.

## Materials and Methods

**Plant Materials.** All wood specimens were taken from the xylarium (wood collections) of the Chinese Academy of Forestry (WOODPEDIA), the largest wood collection in China. A total of 39 specimens of 6 species of *Pterocarpus* were sampled. Four types of specimens, i.e., heartwood, sapwood, twig, and silica gel-dried leaf were collected in this study. 4–11 individuals per species were sampled. Details of the collected reference samples, including the location of vouchers, are listed in Supplementary Table S1.

**Molecular Methods.** Exposed surfaces of xylarium wood specimens were removed with a sterile scalpel to avoid external contamination. Each wood sample of 500 mg was frozen in liquid nitrogen and then ground into a fine powder in a 6770 Freezer/Mill (SpexSamplePrep, Metuchen, NJ, USA).

All DNA isolations were carried out under sterile conditions. DNA from the wood specimens was extracted following the DNeasy Plant Maxi Kit (Qiagen, Hilden, Germany) protocol, modified<sup>11</sup> according to Jiao *et al.*<sup>11</sup>. For silica gel-dried leaves, DNA was isolated using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations.

PCR amplification was performed in a 30  $\mu$ l reaction with 15  $\mu$ l of TaKaRa Premix Ex Taq (containing 0.75 units of Ex Taq DNA polymerase, 2 mM of MgCl<sub>2</sub>, and 200 Mm of each dNTP), 0.2  $\mu$ M of each primer and approx. 10 ng of template DNA. The amplification was conducted in a Veriti 96-Well Thermal Cycler (Applied Biosystems, Foster City, CA, USA). PCR Primers and PCR cycling conditions used in this study are listed in Supplementary Table S2. The PCR products were purified using a UNIQ-10 Spin Column DNA Gel Extraction Kit (Sangon, Shanghai, China) and sequenced in both directions with the same primers used for PCR on an ABI PRISM 3730xl (Applied Biosystems, Foster City, CA, USA).

In addition to the sequences generated in this work, we downloaded sequences (from loci ITS2, *matK*, *ndhF-rpl32* and *rbcL*) (Supplementary Table S3) for specimens of *Pterocarpus* from GenBank for analysis.

**Light Microscopy.** Sectioning blocks [10 mm (L)  $\times$  10 mm (R)  $\times$  10 mm (T)] were cut with razor blades and then softened in 2% ethylenediamine at 60  $^{\circ}$ C for 48 hours. Thereafter, 15  $\mu$ m thick transverse, radial and tangential sections were cut on a sliding microtome. Sections were stained with a 1% aqueous safranin solution, rinsed, then mounted on glass slides and then observed under a light microscope (Olympus BX61, Japan).

**Data Analysis.** Raw sequences for each region were assembled and edited using ContigExpress in Vector NTI Advance v. 10.1 (Invitrogen InforMax, Frederick, MD, USA), saved in FASTA format and deposited to GenBank (Supplementary Table S1). The edited sequences were then aligned with Clustal X 1.81<sup>53</sup> followed by a manual



adjustment with BioEdit software<sup>54</sup>. To assess the barcoding gap, the relative distribution of pairwise genetic distances was calculated using TaxonDNA<sup>55</sup> under the K2P-corrected pairwise distance model<sup>32</sup>.

To evaluate species discrimination success, two widely used methods, TaxonDNA and a neighbor-joining tree-based approach, were applied to the four single barcode and all their possible combinations. For the TaxonDNA analysis, we used the “best match” and the “best close match” functions in the software to test the species-level discrimination rates under the K2P-corrected distance model for each barcode singly and all possible combinations of barcodes<sup>52,56</sup>. The “best close match” method required a threshold value which was calculated for each barcode from pairwise summary. All the results above the threshold were treated as “no match”. For the tree-based method, unrooted neighbour-joining (NJ) trees were constructed in MEGA 5<sup>57</sup> with pairwise deletion and the P-distance model<sup>32,51,58–60</sup>. Only when all the conspecific individuals were clustered a single clade and at least one specimen in each clade was derived from a botanically vouchered collection was it considered a successful species discrimination.

## References

- Tittensor, D. P. *et al.* A mid-term analysis of progress toward international biodiversity targets. *Science* **346**, 241 (2014).
- Dormontt, E. E. *et al.* Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biol. Conserv.* **191**, 790–798 (2015).
- Sun, L. & Bogdanski, B. E. C. Trade incentives for importers to adopt policies to address illegally logged timber: The case of non-tropical hardwood plywood. *J. Forest Econ.* **27**, 18–27 (2017).
- The Plant List Version 1. 1. <http://www.theplantlist.org> (2013).
- Maclachlan, I. R. & Gasson, P. PCA of cites listed *Pterocarpus santalinus* (Leguminosae) wood. *IAWA J.* **31**, 121–138 (2010).
- The IUCN Red List of Threatened Species Version. <https://www.iucnredlist.org> (2016).
- The State Council of the People's Republic of China. The first part directory of national protected flora on emphasis (1999).
- Hebert, P. D. N., Cywinska, A., Ball, S. L. & Dewaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **270**, 313–321 (2003).
- Hajibabaei, M., Singer, G. A., Hebert, P. D. & Hickey, D. A. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **23**, 167–172 (2007).
- Li, D. *et al.* Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *PNAS* **108**, 19641–19646. <https://doi.org/10.1073/pnas.1104551108> (2011).
- Jiao, L., Yin, Y., Cheng, Y. & Jiang, X. DNA barcoding for identification of the endangered species *Aquilaria sinensis*: comparison of data from heated or aged wood samples. *Holzforchung* **68**, 487–494 (2014).
- Lee, S. Y., Ng, W. L., Mahat, M. N., Nazre, M. & Mohamed, R. DNA Barcoding of the Endangered *Aquilaria* (Thymelaeaceae) and Its Application in Species Authentication of Agarwood Products Traded in the Market. *PLoS ONE* **11**, e0154631. <https://doi.org/10.1371/journal.pone.0154631> (2016).
- Erickson, D. L., Spouge, J., Resch, A., Weigt, L. A. & Kress, W. J. DNA barcoding in land plants: developing standards to quantify and maximize success. *Taxon* **57**, 1304 (2008).
- Kress, W. J., GarcíaRobledo, C., Uriarte, M. & Erickson, D. L. DNA barcodes for ecology, evolution, and conservation. *Trends Ecol. Evol.* **30**, 25–35 (2015).
- Chen, S. *et al.* Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **5**, e8613. <https://doi.org/10.1371/journal.pone.0008613> (2010).
- Timme, R. E., Kuehl, J. V., Boore, J. L. & Jansen, R. K. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* **94**, 302–312 (2007).
- Shaw, J., Lickey, E. B., Schilling, E. E. & Small, R. L. Comparison of Whole Chloroplast Genome Sequences to Choose Noncoding Regions for Phylogenetic Studies in Angiosperms: The Tortoise and the Hare III. *Am. J. Bot.* **94**, 275–288 (2007).
- Song, Y. *et al.* Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Front. Plant Sci.* **6**, 662. <https://doi.org/10.3389/fpls.2015.00662> (2015).
- Ly, T. *et al.* DNA barcodes for the identification of *Anoectochilus roxburghii* and its adulterants. *Planta* **242**, 1167 (2015).
- Pei, N., Chen, B. & Kress, W. J. Advances of Community-Level Plant DNA Barcoding in China. *Front. Plant Sci.* **8**, 225. <https://doi.org/10.3389/fpls.2017.00225> (2017).
- Tang, X., Zhao, G. & Ping, L. Wood identification with PCR targeting noncoding chloroplast DNA. *Plant Mol. Biol.* **77**, 609–617 (2011).
- Jiao, L., Liu, X., Jiang, X. & Yin, Y. Extraction and amplification of DNA from aged and archaeological *Populus euphratica* wood for species identification. *Holzforchung* **69** (2015).
- Yu, M., Liu, K., Zhou, L., Zhao, L. & Liu, S. Testing three proposed DNA barcodes for the wood identification of *Dalbergia odorifera* T. Chen and *Dalbergia tonkinensis* Prain. *Holzforchung* **70**, 127–136 (2016).
- Borisenko, A. V., Sones, J. E. & Hebert, P. D. The front-end logistics of DNA barcoding: challenges and prospects. *Mol. Ecol. Resour.* **9**(Suppl s1), 27 (2009).
- Ekrem, T., Willassen, E. & Stur, E. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Mol. Phylogenet. Evol.* **43**, 530 (2007).
- Taylor, H. R. & Harris, W. E. An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Mol. Ecol. Resour.* **12**, 377–388 (2012).
- Cho, S. Y., Suh, K. I. & Bae, Y. J. DNA barcode library and its efficacy for identifying food-associated insect pests in Korea. *Entomol. Res.* **43**, 253–261 (2013).
- Xu, C. *et al.* Accelerating plant DNA barcode reference library construction using herbarium specimens: improved experimental techniques. *Mol. Ecol. Resour.* **15**, 1366 (2015).
- Index Xylariorum 4. 1. [http://www.iawa-website.org/downloads/IndexXylariorum\\_4-1\\_updated.pdf](http://www.iawa-website.org/downloads/IndexXylariorum_4-1_updated.pdf) (2016).
- Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. & Janzen, D. H. Use of DNA Barcodes to Identify Flowering Plants. *PNAS* **102**, 8369. <https://doi.org/10.1073/pnas.0503123102> (2005).
- Group, C. P. W. *et al.* A DNA barcode for land plants. *PNAS* **106**, 12794–12797. [pnas.0905845106](https://doi.org/10.1073/pnas.0905845106) (2009).
- Yan, L. J. *et al.* DNA barcoding of *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya–Hengduan Mountains. *Mol. Ecol. Resour.* **15** (2015).
- Deguilloux, M. F., Pemonge, M. H. & Petit, R. J. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **269**, 1039–1046 (2002).
- Jiao, L. *et al.* Comparative Analysis of two DNA Extraction Protocols from Fresh and Dried wood of *Cunninghamia Lanceolata* (Taxodiaceae). *IAWA J.* **33**, 441–456 (2012).
- Yao, H. *et al.* Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* **5**, e13102. <https://doi.org/10.1371/journal.pone.0013102> (2010).
- Pang, X., Shi, L., Song, J., Chen, X. & Chen, S. Use of the potential DNA barcode ITS2 to identify herbal materials. *J. Nat. Med.* **67**, 571–575 (2013).

37. Han, J. *et al.* An authenticity survey of herbal medicines from markets in China using DNA barcoding. *Sci. Rep.* **6**, 18723, <https://doi.org/10.1038/srep18723> (2016).
38. Fazekas, A. J. *et al.* Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **3**, e2802, <https://doi.org/10.1371/journal.pone.0002802> (2008).
39. Lahaye, R. *et al.* From the Cover: DNA barcoding the floras of biodiversity hotspots. *PNAS* **105**, 2923, <https://doi.org/10.1073/pnas.0709936105> (2008).
40. Li, X. *et al.* Plant DNA barcoding: from gene to genome. *Biol. Rev. Camb. Philos. Soc.* **90**, 157 (2015).
41. Kress, W. J. & Erickson, D. L. 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, <https://doi.org/10.1371/journal.pone.0000508> (2007).
42. Newmaster, S. G., Fazekas, A. J., Steeves, R. A. D. & Janovec, J. Testing candidate plant barcode regions in the Myristicaceae. *Mol. Ecol. Resour.* **8**, 480 (2008).
43. Burgess, K. S. *et al.* Discriminating plant species in a local temperate flora using the *rbcl* + *matK* DNA barcode. *Methods Ecol. Evol.* **2**, 333–340 (2011).
44. Chen, S. *et al.* A renaissance in herbal medicine identification: from morphology to DNA. *Biotechnol. Adv.* **32**, 1237 (2014).
45. Deguilloux, M., Pemonge, M., Bertel, L., Kremer, A. & Petit, R. Checking the geographical origin of oak wood: molecular and statistical tools. *Mol. Ecol.* **12**, 1629–1636 (2003).
46. Thah, L. H., Lee, S. L., Ng, K. K. S., Faridah, Q. Z. & Faridah-Hanum, I. Forensic DNA profiling of tropical timber species in Peninsular Malaysia. *Forest Ecol. Manag.* **259**, 1436–1446 (2010).
47. Ng, K. K. *et al.* Forensic timber identification: a case study of a CITES listed species, *Gonystylus bancanus* (Thymelaeaceae). *Forensic Sci. Int. Genet.* **23**, 197–209 (2016).
48. Meusnier, I. *et al.* A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**, 214 (2008).
49. Dong, W. *et al.* A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genet.* **15**, 138 (2014).
50. Little, D. P. A. DNA mini-barcode for land plants. *Mol. Ecol. Resour.* **14**, 437 (2014).
51. Hartvig, I., Czako, M., Kjær, E. D., Nielsen, L. R. & Theilade, I. The Use of DNA Barcoding in Identification and Conservation of Rosewood (*Dalbergia* spp.). *PLoS ONE* **10**, e0138231, <https://doi.org/10.1371/journal.pone.0138231> (2015).
52. Hassold, S., Bauert, M. R., Razafintsalama, A., Ramamonjisoa, L. & Widmer, A. DNA Barcoding of Malagasy Rosewoods: Towards a Molecular Identification of CITES-Listed *Dalbergia* Species. *PLoS ONE* **11**, e0157881, <https://doi.org/10.1371/journal.pone.0157881> (2016).
53. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
54. Hall, T. A. “BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT”, in: *Nucleic Acids Symp. Ser.*: [London]: Information Retrieval Ltd., c1979–c2000, 95–98 (1999).
55. Meier, R., Shiyang, K., Vaidya, G. & Ng, P. K. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst. Biol.* **55**, 715–728 (2006).
56. Meyer, C. P. & Paulay, G. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol.* **3**, e422, <https://doi.org/10.1371/journal.pbio.0030422> (2005).
57. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
58. Liu, J., Provan, J., Gao, L. & Li, D. Sampling Strategy and Potential Utility of Indels for DNA Barcoding of Closely Related Plant Species: A Case Study in *Taxus*. *Int. J. Mol. Sci.* **13**, 8740–8751 (2012).
59. Srivathsan, A. & Meier, R. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics* **28**, 190–194 (2012).
60. Collins, R. A. & Cruickshank, R. H. The seven deadly sins of DNA barcoding. *Mol. Ecol. Resour.* **13**, 969–975 (2013).

## Acknowledgements

This work was supported financially by Fundamental Research Funds of CAF (Grant No. CAFYBB2017MA013), National Natural Science Foundation of China (Grant No.31600451), Fundamental Research Funds of CAF (Grant No. CAFYBB2017ZE003), China Postdoctoral Science Foundation (Grant No. 2016M590152), and the China Scholarship Council (Grant No. 2016–3035). We would like to express our gratitude to Mr. Di Xiao and Mr. Changyu Xu for the help on collecting samples, and to Sarah Friedrich, Department of Botany, University of Wisconsin-Madison for her technical assistance with the specimen origin and phylogram figures.

## Author Contributions

Y.Y., L.J., M.Y., X.J. conceived and designed the study. L.J., M.Y., T.H. performed the experiments. L.J., M.Y., Y.Y., A.C.W. analyzed the data. L.J., M.Y., T.H., J.L., B.L., Y.Y. contributed reagents/materials/analysis tools. L.J., M.Y., A.C.W., Y.Y. wrote the paper. All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-20381-6>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018