Title: DNA barcoding for identification of agarwood source species using *trn*L-*trn*F and *mat*K DNA sequences

Authors: Seiji Tanaka[1], Michiho Ito[1]

[1] Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida-Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

Corresponding author

Michiho Ito

e-mail: michihoi@pharm.kyoto-u.ac.jp

Abstract

Agarwood is a type of resinous wood found in the trunks of *Aquilaria*, *Gonystylus*, and *Gyrinops* species [1]. High quality agarwood is extraordinarily expensive and therefore its source plant species have become depleted due to exploitation. In 2005, these species were added to Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora [1]. Because these wild agarwood resources have become depleted, commercial production of agarwood has long been a desirable goal. In addition, inauthentic agarwood is sometimes produced from non-agarwood species. Few reports have attempted to identify source species in order to distinguish genuine from false agarwood. In this study, DNA was extracted from putative agarwood samples collected from Japanese, Indonesian, Thai, and Vietnamese markets. The *trn*L-*trn*F region and *mat*K gene were amplified from each sample by PCR to serve as DNA barcodes for identifying the plant species to which each sample belonged. One of the wood samples did not originate from a genuine agarwood species. Although some species were identified, sequence data for agarwood source species currently available in GenBank is insufficient to identify the species to which all of these putative agarwood samples belonged. Thus, positive identification of remaining samples will require further exploration.

Keyword

agarwood, DNA barcoding, *trn*L-*trn*F, *mat*K,

Introduction

Agarwood is a type of resinous wood obtained from the trunks of *Aquilaria*, *Gonystylus*, and *Gyrinops* species [1] and has been used as an herbal medicine for sedation, detoxification, treatment for stomachaches or as incense. According to the *Japanese Standards for Non-Pharmacopoeial Crude Drugs 2018*, 5 species of *Aquilaria* have been designated as source species for agarwood [2]. Because high quality agarwood is extraordinarily expensive, exploitation of these source species has decreased their abundance in their native environments and in 2005 they were added to Appendix II of CITES [1]. Commercial production of agarwood has long been a desirable goal. However, because agarwood is sometimes produced from wood of non-agarwood source species, the identification of source species has become important. To date, identification of the *Aquilaria* and *Gyrinops* species has been carried out by comparing fruit morphology. This method is not particularly suitable for confirming the constituents of agarwood herbal medicines, which normally consist of resin-containing wood pieces. Barcoding techniques for species identification of plants using DNA sequences has become popular, and the availability of DNA sequence data for medicinal plant species has expanded dramatically. Barcoding is useful for identifying the source species of natural medicines and has also been used to detect foreign matter contamination in natural medicines, especially those in small pieces or powder forms [3]. Whereas, there have been some reports of plant taxonomy using leaves of agarwood source species [4, 5, 6, 7, 8], few reports described identification of source species using wood pieces of

agarwood [9, 10]. In this study, barcoding of DNA extracted from fresh and dried leaves as well as dried fruits of *Aquilaria* and *Gyrinops* plants was used to identify source species. The *trn*L(UAA)-*trn*F(GAA) region, the intergenic spacer region between coding regions of the leucine and phenylalanine tRNA genes on the chloroplast genome, and the *mat*K gene, which encodes the enzyme maturase K that excises introns from the chloroplast genome, were analyzed to identify agarwood source species [11, 12]. This method was employed to identify agarwood samples collected in Japanese, Indonesian, Thai, and Vietnamese markets using the same regions.

Materials and methods

Plant materials and DNA extraction

Fourteen samples of leaves and dried fruits from *Aquilaria* plants and 3 samples of leaves from *Gyrinops* plants were used as samples, and whose morphological characteristics are shown in Fig. 1. Tentative species names of these samples were given by agarwood suppliers (Table 1). And also 1 sample of leaves from an unknown agarwood source species (Table 1), and 15 samples of resin-containing portions of agarwood with no species identification were used (Table 2) (Fig. 1). Voucher specimens have been deposited at the Experimental Station for Medicinal Plants at the Graduate School of Pharmaceutical Science, Kyoto University, Japan. All DNA extractions were performed using the DNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer’s protocol.

Primer design

The sequence of primers used for PCR amplifications are shown in Table 3. The *trn*L-*trn*F region was amplified using the forward primer A1 and reverse primer A2, and the *mat*K gene was amplified using the forward primer C1 and reverse primer C2. For nested-PCR, the second PCR amplifications were performed using the forward primer B1 and reverse primer B2 within the *trn*L-*trn*F region and using the forward primer D1 and reverse primer D2 within the *mat*K gene (Fig. 2).

Amplification by PCR

PCR amplifications were performed on a PC320 (Astec, Japan) or GeneAtlasG02 (Astec) thermocycler. Amplification of PCR products from DNA derived from leaves and dried fruits was achieved using a final reaction volume of 20 μL containing 0.5 μL of genomic DNA template, 2.0 μL of 10× PCR Buffer for KOD Dash, 0.2 mM of dNTPs, 0.2 μM each of forward and reverse primers (A1/A2, C1/C2), and 0.5 U of KOD Dash® (Toyobo, Japan). The temperature cycling program for PCR consisted of 1 min at 94°C, followed by 30 cycles of 30 s at 94°C, 2 s at annealing temperature and 30 s at 74°C, with a final extension for 1 min at 72°C. To amplify PCR products from DNA derived from resin-containing portions of agarwood, the final reaction volume of 20 μL contained 2 μL of genomic DNA template, 2.0 μL of 10× PCR Buffer for KOD -Plus-, 0.2 mM of dNTPs, 1.2 μL of $MgSO_4$, 0.3 μM each of forward and reverse primers (A1/A2, C1/C2), and 0.4 U of KOD -Plus- (Toyobo). The temperature cycling program for PCR consisted of 2 min at 94°C, followed by 35 cycles

of 10 s at 98°C, 30 s at annealing temperature and 30 s at 68°C. The annealing temperatures for PCR depended on the combinations of primers (Table 4).

Amplification by nested PCR

DNA samples from 6-1 through 6-6 and 7-1 through 7-3 were further amplified after the first amplification. PCR products from the first amplification were applied to PCR clean-up Gel extraction (Macherey-Nagel, Germany) and the DNA concentrations of the purified products were measured using a Qubit 4 Fluorometer (Invitrogen, USA) and adjusted for a second round of PCR. The final reaction volume of 20 μL contained 10 ng of genomic DNA template, 2.0 μL of 10× PCR Buffer for KOD -Plus-, 0.2 mM of dNTPs, 1.2 μL of $MgSO_4$, 0.3 μM each of forward and reverse primers (B1/B2 D1/D2), and 0.4 U of KOD -Plus-. The temperature cycling program for PCR consisted of 2 min at 94°C, followed by 35 cycles of 10 s at 98°C, 30 s at annealing temperature, and 30 s at 68°C. The annealing temperatures for PCR depended on the combinations of primers (Table 4).

DNA sequencing and data analysis

PCR amplification products were separated on 0.5% agarose/TAE gels, purified using PCR clean-up gel extraction (Macherey-Nagel, Germany) and sequenced on a 3730xl DNA Analyzer using BigDye Terminator V3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Sequences were compared and aligned using DNASIS Pro version 2.09 (Hitachi Solutions Ltd., Japan). The sequences obtained from

leaves and fruits were deposited in GenBank. (Tables 5 and 6)

Results

Sequence analyses using DNA extracted from leaves and dried fruits

1. *trn*L-*trn*F region

Samples of leaves and fruits from 18 putative agarwood species were classified into 6 groups according to the DNA sequences of their *trn*L-*trn*F regions, and sequence patterns of source species were determined by comparing sequences isolated in the present study with those deposited in GenBank (Table 5). Samples 1-1, 1-2, and 1-3 were morphologically identified as *Aquilaria malaccensis* but the sequence data for these samples showed that they were actually *A. crassna*. Samples 1-4, 1-5, and 1-6 were morphologically identified as *A. crassna* and this was confirmed by their sequence data. Samples 1-7, 1-8, and 1-9 were morphologically identified as *A. sinensis* and this was confirmed by their sequence data. Samples 2-1, 2-2, and 2-3 were morphologically identified as *A. malaccensis* and this was confirmed by their sequence data. 0Sample 2-4 was morphologically identified as *A. malaccensis*, but its sequence data showed that it was actually *A. microcarpa*. Sample 2-5 was morphologically identified as *A. microcarpa* and its sequence data verified that it belongs to *A. microcarpa*. Samples 3-1, 3-2, and 3-3 were morphologically identified as *Gyrinops* sp. but their

sequence data revealed that they were actually *Gyrinops versteegii*. Sequence data from sample 4-1 differed from the other samples and did not match any GenBank sequences from agarwood source species.

2. *mat*K gene

Eighteen putative agarwood samples were classified into 3 groups according to the DNA sequences of their *mat*K gene (Table 6); results from *trn*L-*trn*F regions (Table 5) were combined with those from matK genes and sequences of matK gene deposited in GenBank were further added to make conclusion as Table 6. The DNA sequences of the *mat*K genes in *A. malaccensis* and *A. microcarpa* were identical, and those from *A. crassna* and *G. versteegii* were also identical.

Identification of source species of agarwood samples

1. Extraction of DNA

Extraction of sufficient DNA from agarwood pieces for amplification by PCR proved difficult (Fig. 3). Adequate PCR products were obtained using DNA extracted from samples 5-1 through 5-6 by changing both the reaction enzyme and program (Fig. 4). However, adequate results were not achieved for samples of 6-1 through 6-6 and 7-1 through 7-3 with this modified protocol so nested PCR was used to amplify PCR products.

2. *trn*L-*trn*F region

The *trn*L-*trn*F sequences from all 8 samples were identical among leaves and dried fruits and

sequences identical to those from 4 samples were found in GenBank. However, the sequences of the remaining 3 samples included one or more mutations compared with known sequences, and could not be identified (Table 7). In particular, the sequence of sample 5-5 differed substantially from those of other samples in regions other than those shown in the table (Online Resource 1).

3. *mat*K gene

The sequences from 12 out of 15 samples were identical between leaves and dried fruits. For the remaining 3 samples, the sequences of samples 5-3 and 7-3 carried the same mutation and were new to GenBank (Table 8), and the sequence of sample 5-5 again showed very different sequence from those of other samples (Online Resource 1).

Discussion

Development of genetic identification method using leaves or fruits of agarwood

DNA sequence data for the *trn*L-*trn*F region stored in GenBank (as of March 2019) includes 115 sequences from *Aquilaria* species and 19 sequences from *Gyrinops* species. The sequence information obtained from the specimens in this study were consistent with that in GenBank for *A. malaccensis*, *A. crassna*, *A. sinensis*, and *G. versteegii*. Sequence information of the *trn*L-*trn*F region of *A. microcarpa* matched 5 similar sequences in GenBank (KT364474.1, KU244042.1, KU244041.1, KU244040.1, KT726322.1), differing by only a single base pair. Further studies with a larger number

of samples will be needed to clarify whether our samples belong to the species *A. microcarpa*.

The DNA sequence information found for the *mat*K gene in GenBank (as of March 2019) includes 104 sequences from *Aquilaria* species and 2 sequences for *Gyrinops* species. There is currently very little DNA sequence information for *Gyrinops* species in GenBank, and the accumulation of more sequence data for this species will be necessary to more definitively identify source species for some samples. The sequence of the *mat*K genes from our study and from GenBank were identical between *A. malaccensis*, *A. microcarpa* and *A. beccariana* and was also identical between *A. crassna* and *G. versteegii*, which makes identification of the source species of agarwood using only *mat*K difficult or impossible. Species could not be determined using morphological characteristics or DNA sequences in samples 1-1, 1-2, 1-3, and 2-4. Identifications of samples by only morphological characteristics were provided by agarwood suppliers. This indicates that classification of agarwood source species using morphological characteristics is unreliable, and thus that identification using DNA barcoding will be of significant benefit to the agarwood trade.

Application of DNA barcoding for identification of agarwood source species using resin-containing wood

Sample 5-2 was predicted to belong to *A. beccariana* according to the match to its *trn*L-*trn*F region in GenBank, but the *mat*K gene of this sample matched *A. sinensis*. Sample 5-6 was predicted to belong

to *A. microcarpa* based on the sequence match to its *trn*L-*trn*F region in GenBank, but its *mat*K sequence matched the *A. sinensis mat*K gene. Similarly, samples 6-2 and 6-6 were predicted to belong to *G. ledermannii* based on the sequence matches of their *trn*L-*trn*F regions in GenBank. However, the *mat*K gene of 6-2 matched *A. crassna* and that of 6-6 matched *A. sinensis* (Table 9). Those results suggested that samples of 5-2, 5-6, 6-2, and 6-6 might be derived from hybrid plants or that the samples were mixtures of 2 or more species. However, the *trn*L-*trn*F region and *mat*K gene are both located on chloroplast DNA and are very likely to be conserved by maternal inheritance. Thus, hybridization is not a likely explanation for the differing species identifications derived from these 2 sequences. An alternative possibility that these samples were mixtures of tissues from multiple plants was ruled out for samples 5-6 and 6-6, which consisted of chunks of wood, but not for samples 5-2 and 6-2, which consisted of small pieces of wood (Table 2). The *trn*L-*trn*F and *mat*K sequences of samples 5-3 and 7-3 were identical and were thus considered to belong to the same species. Their sequence patterns were similar to those of other agarwood samples (Tables 7 and 8), although they included deletions of adjacent bases at identical sites within the *trn*L-*trn*F region, and no other sequences in GenBank harbor this deletion. These results demonstrate the difficulty of identifying agarwood-producing species, and that increases in the amount of available sequence data for agarwood source species will be useful and may even lead to the identification of additional source species. Interestingly, the sequences of both *trn*L-*trn*F and *mat*K from sample 5-5 differed significantly from those of the *Aquilaria* species.

Although its sequences turned out to be similar to those belonging to the order Thymelaeaceae, sample 5-5 could not be classified at the species level. This result indicates that wood from plants that are not agarwood source species is likely introduced into the agarwood market as a false substitute for agarwood. Although fake agarwood has been a common problem, especially when intended for medicinal purposes, identification of source plant species by DNA barcoding could help validate whether material is genuine for purposes of quality and safety assurance. Seven specimens other than the above-mentioned 8 samples of original species identifications based on comparisons of *trn*L-*trn*F sequence and those from the *mat*K gene all matched (Table 9). Samples 7-1 and 7-2 were randomly selected specimens from one lot of agarwood pieces purchased at a market in Vietnam and were shown here to belong to different species (Table 9). This probably occurred because agarwood is collected by brokers before distribution to markets and is classified according to morphological characteristics, resin composition, and other characteristics as the brokers have no other means of confirming the source species of agarwood.

Conclusions

The results of this study showed a sequence comparison of *trn*L-*trn*F region and *mat*K gene can be used to identify agarwood source species, but indicated that the sequences of either one of these regions alone is insufficient for complete identification, due to the inadequacy of sequence data in existing sequence databases. Species identification of agarwood source plants has long been

accomplished by comparing morphological characteristics of fresh fruits. Thus, after harvest, there was no way to confirm the source species of agarwood without information about the fruit morphology of their source plants. Thus, the information about agarwood source plants often seemed incorrect and species names provided by suppliers often contradicted those from DNA barcoding analyses. The DNA barcoding analyses in this study also revealed that the plant source species of a sample represented as agarwood was not a genuine agarwood species, demonstrating the utility of this method. However, this method requires sequence analysis, which is not always convenient, and insufficient sequence data for agarwood source plant species can make it difficult to positively identify source species. Therefore, increasing the amount of sequence data available for agarwood source species will address this issue. Several attempts have been made to identify source species of agarwood. Eurlings et al. (2005) sought to distinguish agarwood source plant species using *trn*L-*trn*F sequence data from old wood specimens but did not describe extraction of DNA from tree trunk or agarwood [4]. Lee et al. (2016) performed DNA analyses of resinous portions of 8 agarwood samples and found that DNA extraction from this tissue took a long time [9]. Lichao et al. (2014) examined the efficacy of DNA extraction according to changes in drying conditions of *A. sinensis* wood, and they also found that DNA extraction took a long time and that the DNA of some samples could not be extracted [10]. DNA from agarwood samples harvested many hundreds of years ago is fragmented, which makes amplification of PCR products difficult. Improving DNA extraction methodology, while necessary, is

often not sufficient for improving amplification. In this study, we obtained PCR products of sufficient quality for sequencing by modifying both the reaction enzyme and temperature cycling program. These modifications were useful for obtaining PCR products from fragmented DNA isolated from tissues exposed to long-term storage, drying, and heat. Our method could also be used to identify other plant species used in natural medicines and powdered herbal medicines that have not previously been analyzed using DNA. For appropriate use of natural resources and also for regulatory purposes, the species identification methods developed in this study could be useful for ensuring the quality and safety of natural medicines, in addition to validating the authenticity of natural materials.

References

1. ANNUAL REPORT OF THE SECRETARIAT 2004, Convention on International Trade in Endangered Species of Wild Fauna and Flora

2. The Ministry of Health, Labour and Welfare (2018) The Japanese standards for non-Pharmacopoeial crude drugs 2018:51.

3. The Ministry of Health, Labour and Welfare (2016) The Japanese pharmacopoeia, 17th edn. The MHLW Ministerial Notification No. 64, Tokyo, pp2434-2437

4. Eurlings MCM, Gravendeel B. (2005) *Trn*L-*trn*F sequence data imply paraphyly of *Aquilaria* and *Gyrinops* (Thymelaeaceae) and provide new perspectives for agarwood identification. Plant Syst Evol.; 254: 1–12

5. Shiou Yih Lee,1 Maman Turjaman,2 and Rozi Mohamed (2018) Phylogenetic Relatedness of Several Agarwood-Producing Taxa (Thymelaeaceae) from Indonesia. Trop Life Sci Res. 2018 Jul; 29(2): 13–28.

6. Tingting Feng et al. (2018) Phylogenetic analysis of *Aquilaria* Lam. (Thymelaeaceae) based on DNA barcoding. Holzforschung 73(6):517-523

7. Woratouch Thitikornpong, Chanida Palanuvej, Nijsiri Ruangrungsi (2018) DNA barcoding for authentication of the endangered plants in genus *Aquilaria.* Thai Journal of Pharmaceutical Sciences, 42, 214-220

8. Qiwei Li et al. (2018) Molecular Identification of Three Aquilaria (Thymelaeaceae) Species through DNA Barcoding Biol. Pharm. Bull., 41, 967-971

9. Lee SY, Ng WL, Mahat MN et al. (2016) DNA Barcoding of the Endangered *Aquilaria* (Thymelaeaceae) and Its Application in Species Authentication of Agarwood Products Traded in the Market. PLoS One;11:e0154631.

10. L Jiao, Y Yin, Y Cheng, X Jiang (2014) DNA barcoding for identification of the endangered species Aquilaria sinensis: comparison of data from heated or aged wood samples. Holzforschung 68(4):487-494

11. Taberlet P., Gielly L., Patou G., Bouvet J. (1991) Universal primers for amplification of three noncoding regions of chloroplast. DNA. Plant Mol. Biol. 17: 1105–1109

12. CBOL Plant Working Group et al. (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106:12794 - 12797

Tables

Table 1 Details of leaf and fruit samples from agarwood source species used in this study

Sample No.	Species according to morphological characteristics	Part used	Collection site	Collection date
1-1	*A. malaccensis*	**Leaf**	**Japan**	**June, 2014**
1-2	*A. malaccensis*	**Leaf**	**Japan**	**June, 2014**
1-3	*A. malaccensis*	**Leaf**	**Japan**	**June, 2014**
1-4	*A. crassna*	**Leaf**	**Japan**	**June, 2014**
1-5	*A. crassna*	**Leaf**	**Japan**	**June, 2014**
1-6	*A. crassna*	**Leaf**	**Japan**	**June, 2014**
1-7	*A. sinensis*	**Leaf**	**Japan**	**June, 2014**
1-8	*A. sinensis*	**Leaf**	**Japan**	**June, 2014**
1-9	*A. sinensis*	**Leaf**	**Japan**	**June, 2014**
2-1	*A. malaccensis*	**Leaf**	**Indonesia**	**February, 2015**
2-2	*A. malaccensis*	**Leaf**	**Indonesia**	**February, 2015**
2-3	*A. malaccensis*	**Leaf**	**Indonesia**	**February, 2015**
2-4	*A. malaccensis*	**Fruit**	**Indonesia**	**February, 2015**
2-5	*A. microcarpa*	**Fruit**	**Indonesia**	**February, 2015**
3-1	*Gyrinops.* sp.	**Leaf**	**Indonesia**	**January, 2017**
3-2	*Gyrinops.* sp	**Leaf**	**Indonesia**	**January, 2017**
3-3	*Gyrinops.* sp.	**Leaf**	**Indonesia**	**January, 2017**
4-1	Unknown	**Leaf**	**Indonesia**	**February, 2017**

Samples 1-1 through 1-9 and 2-1 through 2-5 were collected from a greenhouse at the Experimental Station for Medicinal Plants at the Graduate School of Pharmaceutical Science, Kyoto University, Japan

Table 2 Details of forms resin-containing agarwood samples used in this study

Sample No	Sample type	Collection site
5-1	**small pieces**	**Japan**
5-2	**small pieces**	**Japan**
5-3	**chunk**	**Thailand**
5-4	**chunk**	**Indonesia**
5-5	**chunk**	**Indonesia**
5-6	**chunk**	**Indonesia**
6-1	**small pieces**	**Japan**
6-2	**small pieces**	**Japan**
6-3	**small pieces**	**Japan**
6-4	**small pieces**	**Japan**
6-5	**small pieces**	**Japan**
6-6	**chunk**	**Japan**
7-1	**chunk**	**Vietnam**
7-2	**chunk**	**Vietnam**
7-3	**chunk**	**Indonesia**

Table 3 List of primers used in this study

Target region	Code	Name	Sequence 5′-3′	References
*trn*L-*trn*F	A1	B49873	5′-GGTTCAAGTCCCTCTATCCC-3′	[11]
	A2	A50272	5′-ATTTGAACTGGTGACACGAG-3′	[11]
	B1	trnLaq	5′-ACAGGCGTATCCGAGCATCA-3′	This study
	B2	trnFaq	5′-CCGACCATTACCAAGACATCATCC-3′	This study
matK	C1	3FKIM_f	5′-CGTACAGTACTTTTGTGTTTACGAG-3′	[12]
	C2	1RKIM_r	5′-CCCAGTCCATCTGGAAATCTTGGTTC-3′	[12]

D1	matkaq_f	5′-GCAATCTTTCTTGAACGGATCT-3′	This study
D2	matkaq_r	5′-AATCGACCCAAGTTGGCTTA-3′	This study

Table 4 List of primer combinations used in this study

Combinations	Length of target region	Annealing temperature for PCR
A1/A2	500 bp	56 ℃
B1/B2	300 bp	56 ℃
C1/C2	1000 bp	52 ℃
D1/D2	500 bp	56 ℃

Table 5 SNPs in the *trn*L-*trn*F IGS region amplified from DNA extracted from leaf and fruit samples

Sample	SNPs								Species identified	Genbank
No.	159	291	301	327	328	339	368	381	by DNA sequence	Accession No.
1-1	A	T	G	T	-	T	C	C	*A. crassna*	LC467499
1-2	A	T	G	T	-	T	C	C	*A. crassna*	LC467500
1-3	A	T	G	T	-	T	C	C	*A. crassna*	LC467501
1-4	A	T	G	T	-	T	C	C	*A. crassna*	LC467502
1-5	A	T	G	T	-	T	C	C	*A. crassna*	LC467503
1-6	A	T	G	T	-	T	C	C	*A. crassna*	LC467504
1-7	A	G	T	T	T	T	A	A	*A. sinensis*	LC467505
1-8	A	G	T	T	T	T	A	A	*A. sinensis*	LC467506
1-9	A	G	T	T	T	T	A	A	*A. sinensis*	LC467507
2-1	C	T	T	T	-	G	C	A	*A. malaccensis*	LC467508
2-2	C	T	T	T	-	G	C	A	*A. malaccensis*	LC467509
2-3	C	T	T	T	-	G	C	A	*A. malaccensis*	LC467510
2-4	A	T	T	-	-	G	C	A	*A. microcarpa*	LC467511
2-5	A	T	T	-	-	G	C	A	*A. microcarpa*	LC467512
3-1	A	T	T	T	-	T	C	A	*G. versteegii*	LC467513
3-2	A	T	T	T	-	T	C	A	*G. versteegii*	LC467514
3-3	A	T	T	T	-	T	C	A	*G. versteegii*	LC467515

4-1	A	G	T	-	-	G	C	A	Unknown	-

All samples in this experiment were amplified using the A1/A2 primer combination

Table 6 SNPs in the *mat*K region amplified from DNA extracted from leaf and fruit samples

Sample No.	SNPs						Species identified by DNA sequence	GenBank Accession No.
	156	265	294	358	371	406		
1-1	C	C	G	C	C	A	*A. crassna*	LC467516
1-2	C	C	G	C	C	A	*A. crassna*	LC467517
1-3	C	C	G	C	C	A	*A. crassna*	LC467518
1-4	C	C	G	C	C	A	*A. crassna*	LC467519
1-5	C	C	G	C	C	A	*A. crassna*	LC467520
1-6	C	C	G	C	C	A	*A. crassna*	LC467521
1-7	A	T	C	C	A	C	*A. sinensis*	LC467522
1-8	A	T	C	C	A	C	*A. sinensis*	LC467523
1-9	A	T	C	C	A	C	*A. sinensis*	LC467524
2-1	C	C	G	T	C	A	*A. malaccensis*	LC467525
2-2	C	C	G	T	C	A	*A. malaccensis*	LC467526
2-3	C	C	G	T	C	A	*A. malaccensis*	LC467527
2-4	C	C	G	T	C	A	*A. microcarpa*	LC467528
2-5	C	C	G	T	C	A	*A. microcarpa*	LC467529
3-1	C	C	G	C	C	A	*G. versteegii*	LC467530
3-2	C	C	G	C	C	A	*G. versteegii*	LC467531
3-3	C	C	G	C	C	A	*G. versteegii*	LC467532
4-1	C	C	G	T	C	A	Unknown	-

All samples in this experiment were amplified using the C1/C2 primer combination

Table 7 SNPs in the *trn*L-*trn*F IGS region amplified from DNA extracted from extracted from wood samples

Sample	Primer	SNPs	Species identified

No.	combinations	159	291	301	327	328	339	368	381	by DNA sequence
5-1	A1/A2	C	T	T	T	-	G	C	A	*A. malaccensis*
5-2	A1/A2	A	T	T	T	-	G	C	A	*A. beccariana*
5-3	A1/A2	A	G	G	T	-	T	-	A	Unknown
5-4	A1/A2	C	T	T	T	-	G	C	A	*A. malaccensis*
5-5	A1/A2	A	-	T	T	-	-	C	T	(Thymelaeaceae)
5-6	A1/A2	A	T	T	-	-	G	C	A	*A. microcarpa*
6-1	A1/A2, B1/B2	A	T	T	T	-	T	C	A	*G. versteegii*
6-2	A1/A2, B1/B2	A	G	T	T	T	T	C	A	*G. ledermannii*
6-3	A1/A2, B1/B2	A	T	T	T	-	T	C	A	*G. versteegii*
6-4	A1/A2, B1/B2	A	G	T	T	T	T	A	A	*A. sinensis*
6-5	A1/A2, B1/B2	A	T	T	T	-	T	C	A	*G. versteegii*
6-6	A1/A2, B1/B2	A	G	T	T	T	T	C	A	*G. ledermannii*
7-1	A1/A2, B1/B2	A	T	T	T	-	G	C	A	*A. beccariana*
7-2	A1/A2, B1/B2	A	T	G	T	-	T	C	C	*A. crassna*
7-3	A1/A2, B1/B2	A	G	G	T	-	T	-	A	Unknown

The sequence data is shown in Online Resource 1

Identification of *A. beccariana*, and *G. ledermannii* was based on GenBank sequence data (KT726319) [4] and (AY216755) [4]

Table 8 SNPs in the *mat*K region extracted from wood samples

Sample No.	Primer combinations	SNPs						Species identified by DNA sequence
		156	265	294	358	371	406	
5-1	C1/C2	C	C	G	T	C	A	*A. malaccensis*
5-2	C1/C2	A	T	C	C	A	C	*A. sinensis*
5-3	C1/C2	A	T	C	C	A	A	Unknown
5-4	C1/C2	C	C	G	T	C	A	*A. malaccensis*
5-5	C1/C2	T	C	C	A	C	A	(Thymelaeaceae)
5-6	C1/C2	A	T	C	C	A	C	*A. sinensis*
6-1	C1/C2, D1/D2	/	C	G	C	C	A	*G. versteegii*

6-2	C1/C2, D1/D2	/	C	G	C	C	A	*G. versteegii*
6-3	C1/C2, D1/D2	/	C	G	C	C	A	*G. versteegii*
6-4	C1/C2, D1/D2	/	T	C	C	A	C	*A. sinensis*
6-5	C1/C2, D1/D2	/	C	G	C	C	A	*G. versteegii*
6-6	C1/C2, D1/D2	/	T	C	C	A	C	*A. sinensis*
7-1	C1/C2, D1/D2	/	C	G	T	C	A	*A. beccariana*[*]
7-2	C1/C2, D1/D2	/	C	G	C	C	A	*A. crassna*
7-3	C1/C2, D1/D2	/	T	C	C	A	A	Unknown

1 The sequence data is shown in Online Resource 1

2 We were unable to analyze 156 base pairs when samples were amplified with the D1/D2 primer

3 combination

4 Identification of *A. beccariana*, was based on GenBank sequence data (FJ572802)

5 *The DNA sequences of the matK genes in *A. malaccensis* and *A. beccariana* were identical

6

7

8 Table 9 Comparison of species identifications based on *trn*L-*trn*F or *mat*K regions

Sample No.	Species identification based on sequence of ***trn*L-*trn*F** region	Species identification based on sequence of ***mat*K** region
5-1	*A. malaccensis*	*A. malaccensis*
5-2	*A. beccariana*	*A. sinensis*
5-3	Unknown	Unknown
5-4	*A. malaccensis*	*A. malaccensis*
5-5	(Thymelaeaceae)	(Thymelaeaceae)
5-6	*A. microcarpa*	*A. sinensis*
6-1	*G. versteegii*	*G. versteegii*
6-2	*G. ledermannii*	*G. versteegii*
6-3	*G. versteegii*	*G. versteegii*
6-4	*A. sinensis*	*A. sinensis*
6-5	*G. versteegii*	*G. versteegii*
6-6	*G. ledermannii*	*A. sinensis*

7-1	*A. beccariana*	*A. beccariana*
7-2	*A. crassna*	*A. crassna*
7-3	Unknown	Unknown

Figure legends

Fig. 1 Examples of morphology of agarwood samples used in this study

Fig. 2 Positions and directions of primers used for amplifying the *trn*L-*trn*F IGS region and the *mat*K gene

Fig. 3 DNA samples extracted from plant materials

Fig. 4 PCR products amplified from extracted DNA separated by agarose-gel electrophoresis

Fig. 1

Leaf sample

Fruit sample

Wood samples

Fig. 2

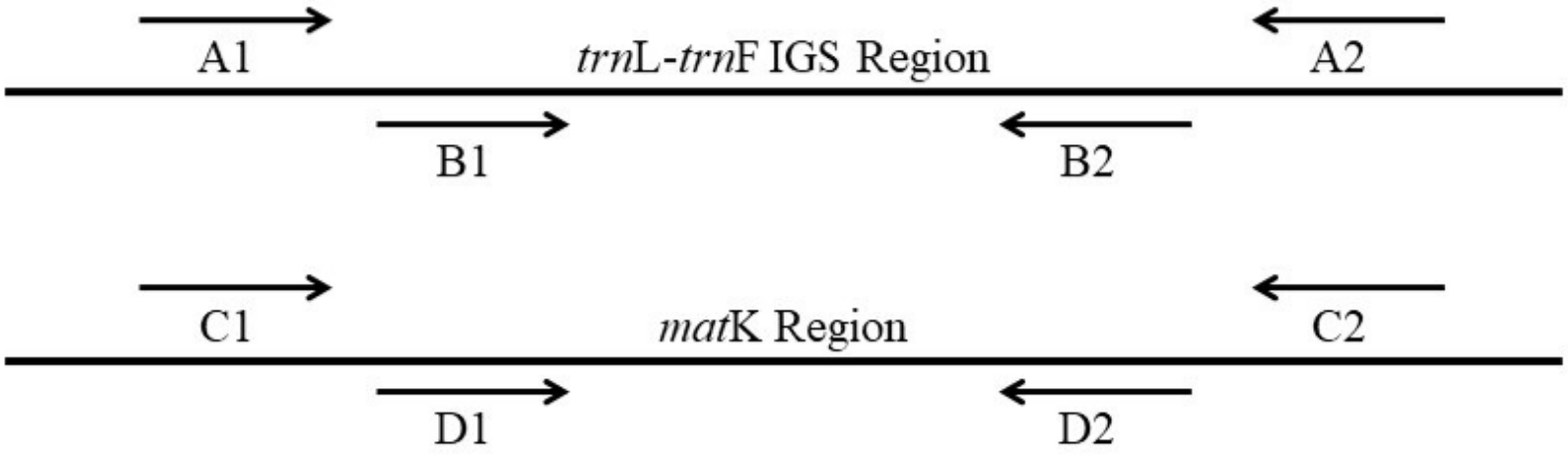
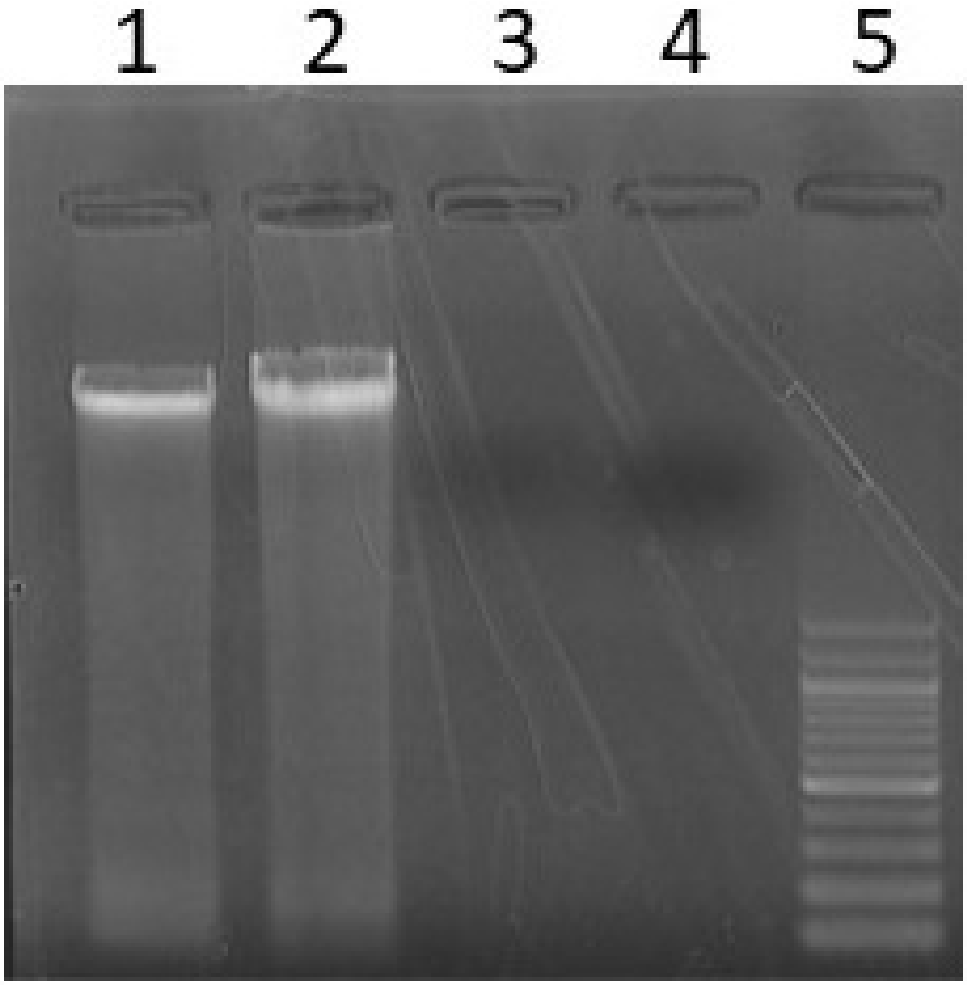


Fig. 3



Lane 1: leaf or fruit

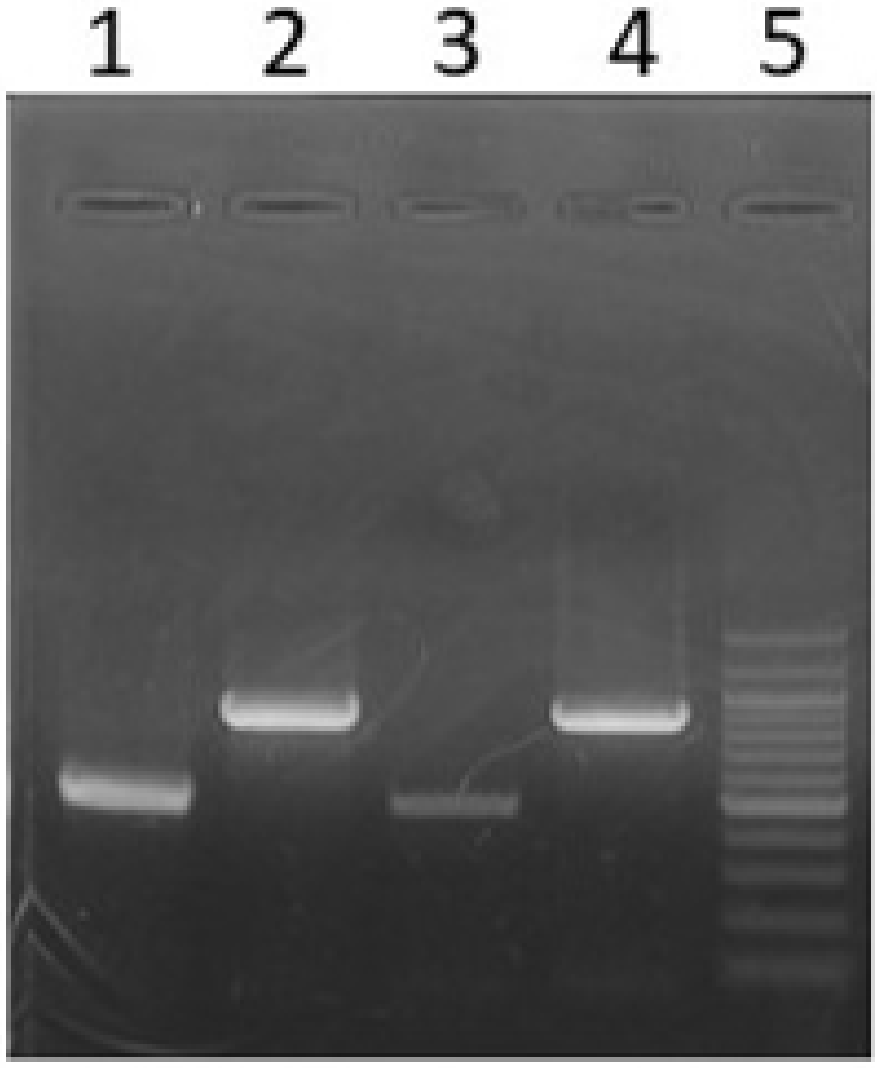
Lane 2: leaf or fruit

Lane 3: wood

Lane 4: wood

Lane 5: 100 bp DNA ladder

Fig. 4



Lane 1: leaf or fruit by A1/A2

Lane 2: leaf or fruit by C1/C2

Lane 3: wood by A1/A2

Lane 4: wood by C1/C2

Lane 5: 100 bp DNA ladder