

RESEARCH

Open Access



# Development of DNA-based species identification and barcoding of three important timbers

Tanzeem Fatima<sup>1\*</sup> , Ashutosh Srivastava<sup>1</sup>, P. V. Somashekar<sup>1</sup>, Vageeshbabu S. Hanur<sup>2</sup> and M. Srinivasa Rao<sup>3</sup>

## Abstract

**Background:** Wood identification is extremely important in the modern forest industry. Teak (*Tectona grandis* L. family verbanaceae), Black Rosewood (*Dalbergia latifolia* L. family Fabaceae), and Ben Teak (*Lagerstroemia lanceolata* L. family Lytheraceae) are significant economic timber species in tropical countries. *T. grandis* is one of the most expensive hardwoods. It occurs naturally in India, Myanmar, Thailand, Java, and Laos in South Asia. *L. lanceolata* a strong associate of *T. grandis* occurs in deciduous forest-adjointing states of Karnataka such as Andhra Pradesh and Telangana state in India. *D. latifolia* distributed in India, Indonesia, Myanmar, Nepal, Philippines, Sri Lanka, and Vietnam. The selected valuable timbers threatened illegal logging and deforestation. It is challenging for non-taxonomist to accurately recognize and identify the timber species based on the conventional methods. DNA barcoding technologies and large-scale screening of DNA variation in trees can be done routinely and much faster with higher taxonomic resolution than morphological determination methods. In this study, we selected two plastid-specific ribulose biphosphate carboxylase (rbcL), maturase K (matK), and one nuclear-specific trnH-ITS (intergenic spacer region)-conserved barcoding marker to compare species inter- and intradetailed discrimination. The aim of this study was to develop an accurate species and their provenance identification to support conservation efforts of three selected species viz *T. grandis*, *D. latifolia*, and *L. lanceolata* in Karnataka, India.

**Results:** Among the three selected markers, only two markers (rbcL and trnH-ITS) were having potential to amplify and discriminate the intra- and interspecific regions of selected timbers and would have been used to get barcodes for each species. Additionally, rbcL (90–96%) has a higher discrimination dominance over trnH-ITS (80–86%). From the barcoding results, it was found that the *T. grandis* and *D. latifolia* samples that belong to the Institute of Wood Science and Technology Bangalore are genetically more related to Dharwad than Hassan region in Karnataka.

**Conclusion:** The present study provided a database in which DNA sequences from unknown samples could be matched as well as required sequences from which other timbers or plant species could be identified. The use of inbuilt unique properties of DNA database within a timber would serve as an important molecular element in forensic forestry to support the determination of identifying the species and their adulterants.

**Keywords:** Plastid markers, DNA barcoding, rbcL, matK, Timbers

\* Correspondence: [tanzeem.fatima@gmail.com](mailto:tanzeem.fatima@gmail.com)

<sup>1</sup>Genetics and Tree Improvement Division, Institute of Wood Science and Technology, Bangalore 03, India

Full list of author information is available at the end of the article

## Introduction

Black Rosewood (*Dalbergia latifolia* f: Fabaceae), Ben Teak (*Lagerstroemia lanceolata* f: Lythraceae), and Teak (*Tectona grandis* f: Verbenaceae) are incredibly significant economic timber species. *T. grandis* is one of the most expensive hardwoods in the world (Kollert and Walotek 2015). It is indigenous to peninsular and central India, Myanmar, Indonesia, Thailand, and Laos in South Asia and has also emerged as a major hardwood plantation species in a number of other countries in Africa, West Indies, and Sri Lanka (Mohanani et al. 1997) (Hansen et al. 2017). *L. lanceolata* is found in the deciduous forest-adjointing state of Karnataka such as Andhra Pradesh and Telangana states, and it is a strong associate of *T. grandis* in the forest of Karnataka, India. *D. latifolia* mostly distributed in India, Indonesia, Myanmar, Nepal, Philippines, Sri Lanka, and Vietnam (Orwa et al. 2009). These three valuable timber species are threatened by illegal logging and deforestation. In an effort to reduce and ultimately halt illegal logging and trade in, a challenge for the molecular biologists is the identification of plants at an intraspecific and interspecific level using barcode sequences that differentiate among closely related species or intervarietal is the most significant (Han et al. 2016). DNA barcoding is a widely used molecular marker technology, which uses a standardized and conserved but diversified DNA sequence to identify species and the origins of timber and to track logs along supply chains (Kang et al. 2017). This technique could serve as another tool to help identify and trace valuable timbers (Finkeldey et al. 2009). A significant assessment of DNA variation can be done faster and with higher taxonomic resolution than morphological determination methods (Lowe 2007). The development of DNA marker for species identification method would greatly contribute to identify wood species and wood products through region and source verification to trace legally from illegally harvested logs or timber products (Lowe and Cross 2011). The major influence of DNA barcoding is that once a solid reference database has been established, the method does not require expert taxonomic knowledge in order to identify specific samples (Hartvig et al. 2015). This technique has facilitated marketing and sale of poor-quality wood of other species in place of the novel timber species. Although DNA barcoding has been successfully used to discriminating plants or crop species, applying this technique for discriminating forest tree species is difficult. Plant plastid and mitochondrial genomes exhibit a low rate of mutation and nucleotide substitution (Palmer 1985). DNA barcoding studies in the plants have been proposed a few plastid-coding regions (rbcL and matK) and non-coding region (ITS-trnH) as potential barcodes (Pettengill and Neel 2010). The study was

mainly focused on the sequence analysis, calculation of genetic distance, variable sites, molecular weight of bases, amplified polymorphism, barcode development of multiple sequence alignment, use of a neighbor-joining algorithm, and tree construction analysis. In the present study, the main focus was to develop an accurate species identification method for the three selected *T. grandis*, *D. latifolia*, and *L. lanceolata* species and this was employed by developing potential DNA barcodes to support conservation efforts of the genus.

## Materials and methods

### Sampling sites

The samples of selected species were collected from three broad natural-growing population regions along Haliyal, Barchi range Dharwad under Western Ghats, Hassan (Alur taluka) and Bangalore, Karnataka, India (Fig. 1). Details of the sample location were described in (Table 1). Superior phenotypes were chosen for this research study based on height, girth, and disease-free tree.

### Sample collection

For the development of barcodes from the selected three species, three samples (three leaves and three cores) of each species were selected and core and leaf sampling was carried out. Core sample was collected at 1.3 m from the ground level through using the conventional increment borer (Mora coretax original, Sweden; 100 mm) (core diameter 5.15 mm). The hole in the tree was properly sealed with paraffin wax to prevent infection. Core and leaf samples were put into a plastic cover and immediately placed into dried silica gel. After the arrival to the laboratory, the samples that had been placed in the silica gel were cleaned and kept for drying at least 3 to 4 days with the change of silica gel which absorb all moisture content of the samples and then transferred the dried samples to  $-20^{\circ}\text{C}$  cryogenic freezer to maintain the quality for further analysis.

### Genomic DNA isolation

Total genomic DNA was isolated from silica-dried leaves and wood samples by using a modified CTAB protocol (Fatima et al. 2018a, 2018b).

### DNA quantification

The yield of extracted DNA was quantified by nanodrop at 260 and 280 nm wavelength. The purity of DNA shows a ration between 1.80 to 2.0 (Sambrook et al. 1989). DNA suitability for sequencing and barcoding was checked by running the samples on 1% (w/v) agarose gel (Jiao et al. 2013) stained with 4  $\mu\text{L}$  of 0.3% ethidium bromide.



**Fig. 1** Map of Karnataka illustrating the geographic positions of sampling sites. Western Karnataka (IWST, Alur Hassan) and North Karnataka (Halyal Dandeli Dharwad)

**PCR amplification through universal barcode primers (RbcL, matK, and trnH-ITS)**

Three candidate barcoding loci *rbcl*, *matK*, and ITS-*trnH* (CBOL Plant Working Group 2009) were selected and synthesized by Integrated DNA Technologies Pvt. Ltd. (Table 2). DNA amplification was carried in a 13- $\mu$ L reaction volume containing 1.5  $\mu$ L genomic DNA (30 ng/ $\mu$ L), 2  $\mu$ L primers (10 mM both forward and reverse), 1.5  $\mu$ L PCR buffer, 1.5  $\mu$ L dNTPs (10 mM), 1.5  $\mu$ L MgCl<sub>2</sub>, 0.3  $\mu$ L Taq polymerase (3 U/ $\mu$ L) (Bangalore genie), and 4.2  $\mu$ L double distilled water. Amplification cycle consists of an initial 3-min denaturation at 94 °C, 30 cycles for 1 min at

60–61 °C, 1 min at 72 °C, and final extension step for 10 min at 72 °C.

**Separation of amplified PCR products and purification**

The amplified product loaded with 5  $\mu$ L loading buffer was size fractionated by electrophoresis on a 2% agarose gel (Saiki et al. 2017) with 0.3% ethidium bromide. The bands are visualized on UV transilluminator (Syngen G:Box) to test out the bands. PCR amplified products were purified using EZ-10 spin column PCR purification kit (Life Technologies, California USA) following the manufacturer’s standard instructions.

**Table 1** Details of *D. latifolia*, *T. grandis*, and *L. lanceolata* populations identified across the major regions in Karnataka

Sl. no.	Location	Taluka	Latitude (north)	Longitude (east)	Annual rainfall (mm)	Agroclimatic zone
1	IWST Bangalore	Bangalore	13° 00' 41.3"	077° 34' 17.6"	698	Eastern dry zone
2	Hassan KFD	Hassan	13° 00' 29.9"	076° 06' 10.8"	1013	Northwest transition zone
3	Dharwad	Dandeli Haliyal (Barchi range)	13° 00' 39.6"	077° 34' 21.7"	899	Northwest dry zone

**Table 2** Universal primers used for barcoding and their information

Barcode loci	Primer information	Size (bp)	Primer temperature (°C)
rbcL	rbcL 1 F ATGTCACCACAAACAGAAAC R724 R TCGCATGTACTCTGAGTAGC	679	60 °C
matK	F' matK 472 F CCRTCATCTGGAATCTTGGTT R'1248R GCTRTRATAATGAGAAAGATT TCTGC	710	60 °C
trnH-ITS	F' trnH-psbAFCGCGCATGGTGGATTCACAATCC R' psbA RGTWATGCAYGAAACGTAATGCTC	443	60.8 °C

### DNA sequencing and primer development

Sanger sequencing of the purified PCR products was carried out in Eurofins Genomics India Pvt. Ltd. Database search was done in the nucleotide database of the US National Center for the Biotechnology Information (NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) or the Barcode of Life Database (BOLD, <http://www.boldsystems.org/>) to identify homologs of the related genes. Gene sequences from ITS-trnH and rbcL were submitted to NCBI sequence data submission bank (<http://www.ncbi.nlm.nih.gov/genbank/>) to obtain accession numbers. The provenance-specific barcode primers from the sequenced data of rbcL and ITS-trnH were designed by Primer3 version 0.4.0 software online (<http://bioinfo.ut.ee/primer3-0.4.0>) (Untergasser et al. 2012).

### Multiple sequence alignment and barcode generation

Multiple sequence alignment comparison brought out the genus identification in between the provenances. Nucleotide base FASTA files of all the species aligned through online European Bioinformatics Institute (EMBL-EMBI) (Clustal Omega; <http://www.ebi.ac.uk/Tools/msa/clustalo/>) program. The nucleotide sequences of ITS-trnH and rbcL were aligned separately and analyzed. Barcode construction for rbcL and ITS-trnH region of each selected species was done by Eurofins Genomics India Pvt. Ltd.

### Tree construction and phylogenetic analysis

Tree construction is a method that uses the MEGA software to construct a neighbor-joining (NJ) tree for each

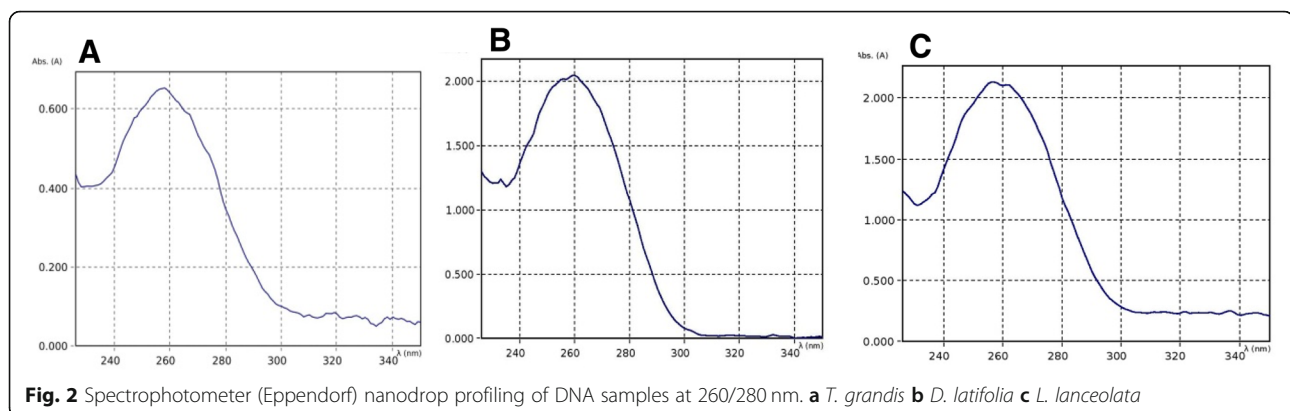
marker and with the different combinations of barcode markers under the Kimura-2 parameter substitution model (Hall 2013). The phylogenetic trees were constructed by using algorithm programs in molecular evolutionary genetic analysis (MEGA7) software (Kumar et al. 2007). The bootstrap support of the NJ tree was assessed using 1500 replicates. The selected species were effectively differentiated if all the individuals of the corresponding species formed a group in phylogenetic tree (Hollingsworth et al. 2009).

### Results and discussion

The total genomic DNA that has been extracted from the selected species concentration was estimated by UV spectrophotometer using the standard formula  $50 \times OD_{260} \times \text{dilution factor}/100$ , and the graphs were recorded (Fig. 2).

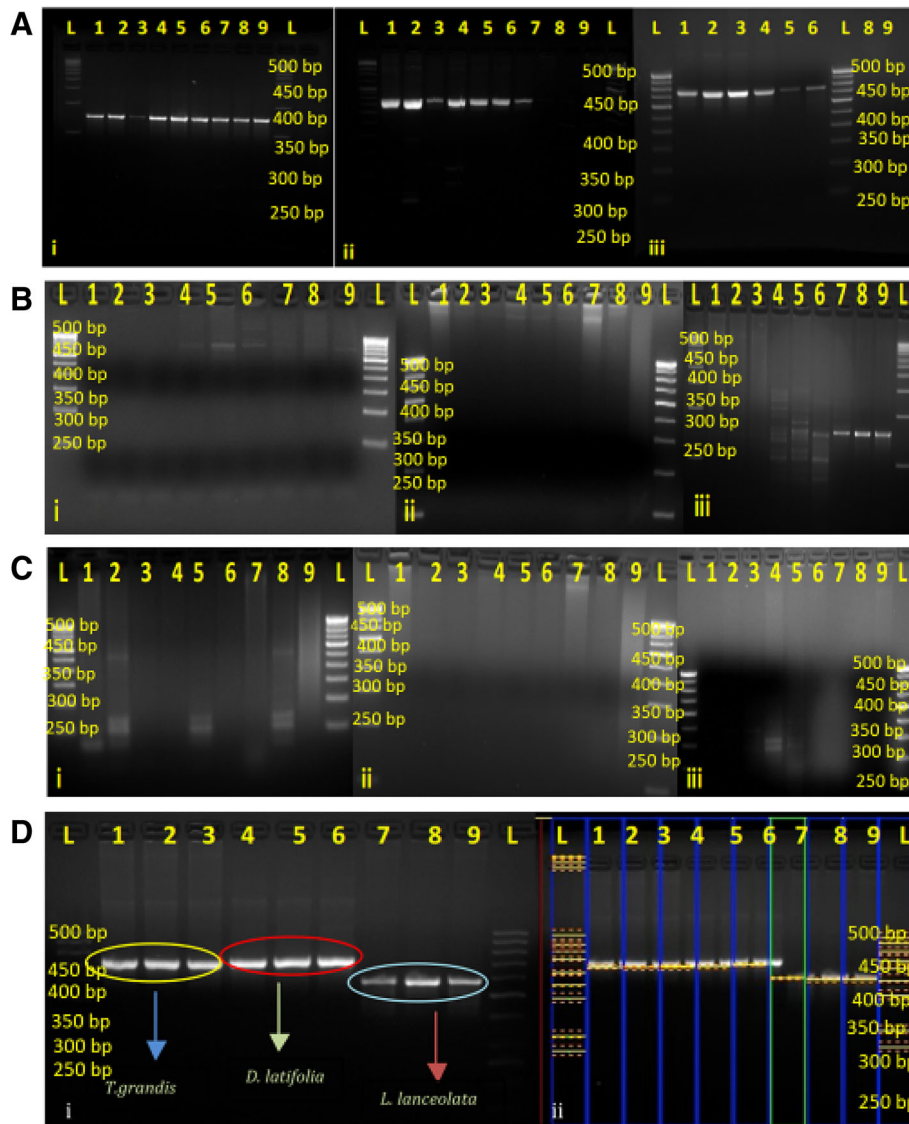
### Barcode primer amplification

Three barcode primer PCR products were randomly chosen from the wood samples of selected species to verify the sequencing and barcode development of the selected species (Kang et al. 2017). A total of seven individuals representing all the three species were successfully amplified and sequenced using two DNA barcodes (rbcL and ITS-trnH). The selected barcode regions were amplified with the barcode primers, and the eluted product gave a good read frame length in the forward direction (Table 3). However, there was a difficulty in the amplification and alignment sequencing with matK primer (Fig. 3a–c). From the above gel pictures, it was



**Table 3** The success rates of PCR amplification and sequencing of three barcode markers in the sample collected from three different locations

Location	Markers	PCR amplification rate (%)	Sequencing rate (%)
Bangalore, Karnataka	rbcl	96	98
	ITS	80	91
	matK	30	51
Hassan, Karnataka	rbcl	90	97
	ITS	86	84
	matK	40	48
Dharwad, Karnataka	rbcl	96	94
	ITS	81	80
	matK	59	43



**Fig. 3** PCR amplification product generated by using plastid DNA. **a** *rbcl* for (i) *T. grandis*, (ii) *D. latifolia*, and (iii) *L. lanceolata*. **b** *matK* for (i) *T. grandis*, (ii) *D. latifolia*, and (iii) *L. lanceolata*. **c** *ITS-trnH* for (i) *T. grandis*, (ii) *D. latifolia*, and (iii) *L. lanceolata*. **d** Selected primer *rbcl* amplified and showing clear bands in all three timbers: (i) *T. grandis*, (ii) *D. latifolia*, and (iii) *L. lanceolata*

clearly depicting that only *rbcL* and ITS-trnH were amplified in all the tree species. Further analysis revealed the species-wise difference in *D. latifolia*, *T. grandis*, and *L. lanceolata*, but it did not discriminate the provenances of the samples (Fig. 3d). The success rate of PCR amplification sequencing was the higher for *rbcL* compared to ITS-trnH in the selected tree species (Fig. 4).

**Nucleotide variation and sequencing success rate**

The average sequence characteristics of the studied barcodes have been tabulated (Table 4). The total sequences obtained from the *rbcL* and ITS-trnH region of all the selected species were aligned based on their provenances by using (MSA Clustal W1.2).

The final aligned partial sequences of *rbcL* in *T. grandis* had a length of 619 bp for Dharwad, 528 bp for Hassan, and 653 bp for IWST Bangalore with the variable sites, and for *D. latifolia*, 502 bp for Dharwad, 514 bp for Hassan, and 608 bp for IWST Bangalore which revealed the location specific base differences in the selected species. In *L. lanceolata*, it was 620 bp length sequence. For *T. grandis*, the average bp sequences obtained by using *rbcL* marker was 600 bp with the 75 bp variable sites. In case of *D. latifolia*, the average bp sequence length by using *rbcL* primer is 541 bp with the variable sites of 85 bp. In *L. lanceolata* sequences, the bp average was found to be 620 bp with the 56 variable sites. The total number of SNPs (single nucleotide polymorphisms) was found more in *rbcL* than in ITS-trnH marker sequences. The total number of SNPs produced by using *rbcL* in *T. grandis* was 97 followed by 68 in *D. latifolia* and 91 in *L. lanceolata*. CpG% was more in *rbcL* sequences compared to ITS-trnH sequences. In *T. grandis*, CpG% was found 97% by using *rbcL* marker

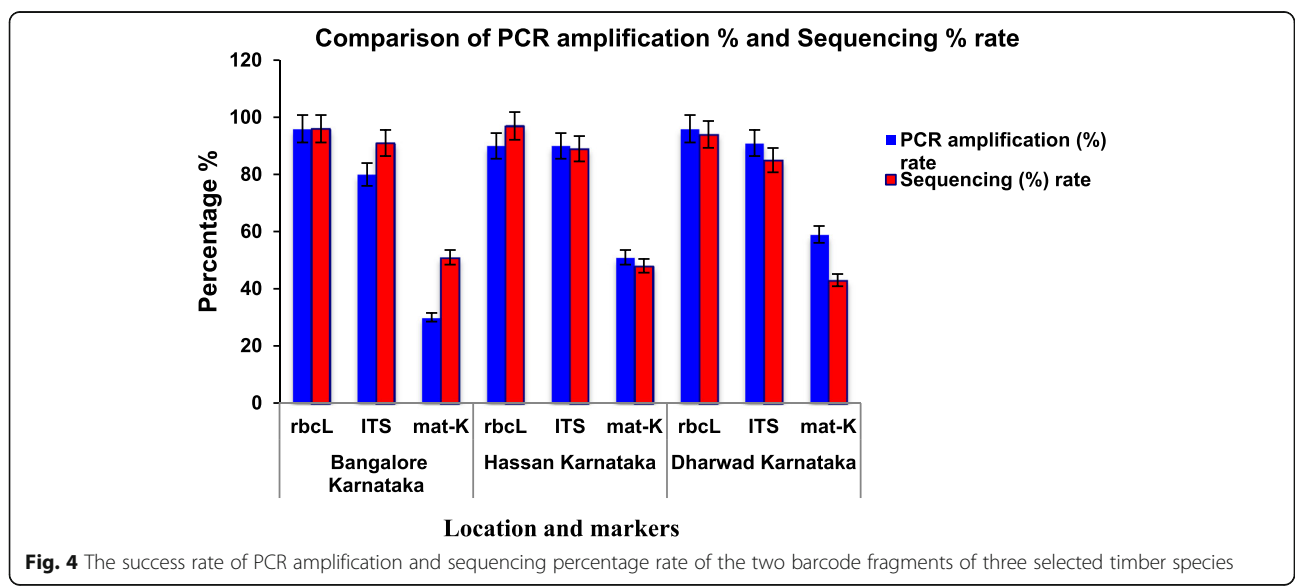
followed by 85% in *D. latifolia* and 91% in *L. lanceolata* (Table 4).

For ITS-trnH region, the final aligned partial sequences of *T. grandis* had a length of 477 bp for Dharwad, 455 bp for Hassan, and 314 bp for IWST Bangalore, and for *D. latifolia*, 741 bp for Dharwad, 761 bp for Hassan, and 775 bp for IWST Bangalore which revealed the location specific base differences in the selected species. For *L. lanceolata*, it was 305 bp length sequence. For *T. grandis*, the average of bp sequences obtained by using ITS-trnH marker was 415 bp with 93 bp variable sites. In case of *D. latifolia*, the average bp sequence length by using ITS-trnH primer was 759 bp with the variable sites of 115 bp. In *L. lanceolata* sequences, the bp average by using ITS-trnH primer was found to be 305 bp with the 68 variable sites. The total number of SNPs produced by using ITS-trnH in *T. grandis* was 16 followed by 13 in *D. latifolia* and 11 in *L. lanceolata*. In *T. grandis*, CpG% was found 85% by using ITS-trnH marker followed by 59% in *D. latifolia* and 73% in *L. lanceolata* (Table 4).

The numbers of polymorphic sites among the species with different region samples were ranged from 78 for *D. latifolia*, 85 for *L. lanceolata*, and 267 for *T. grandis*, and the number of variable sites of the *rbcL* regions increased when the sequence length varies (Table 4).

**Nucleotide discrimination-based approach for species identification**

ITS-trnH and *rbcL* could assign the characteristic nucleotide positions, and differences also showed for all the species with 100% correct division. Nucleotide sequence analysis was done using *rbcL* and ITS-trnH loci for all the species with 100% correct classification. In case of ITS-trnH in *D. latifolia*, A and T nucleotides



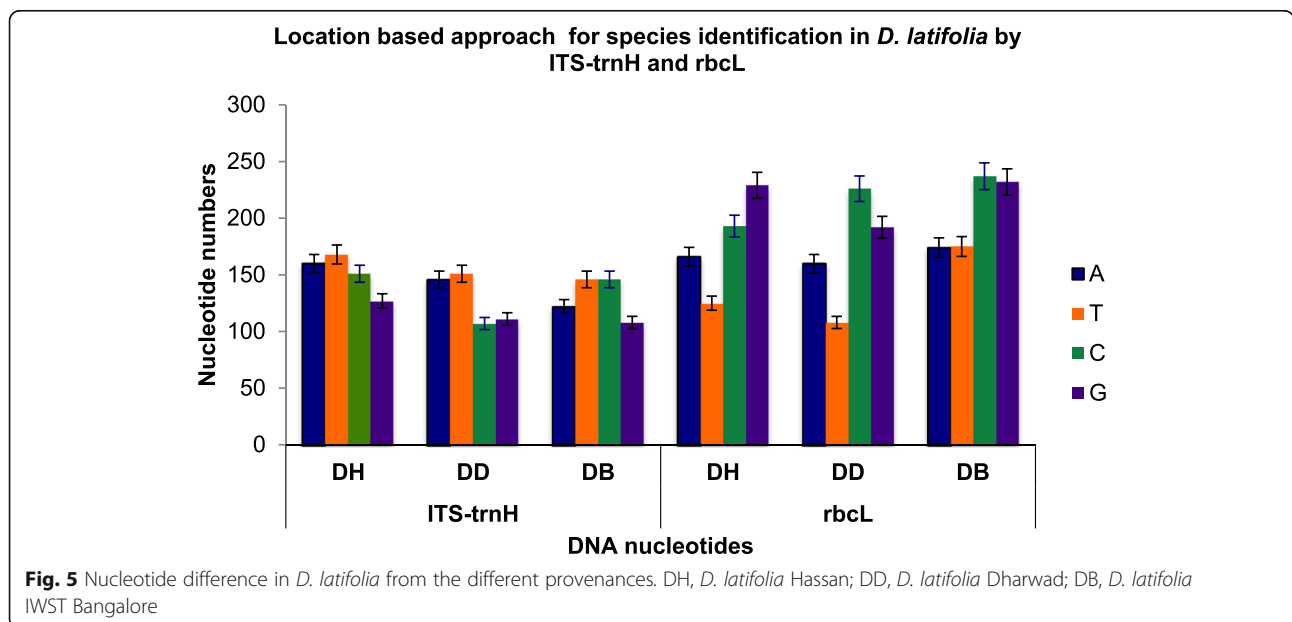
**Fig. 4** The success rate of PCR amplification and sequencing percentage rate of the two barcode fragments of three selected timber species

**Table 4** Basic barcode analysis of the two DNA markers and the combination of the two plastid DNA regions

Loci	Aligned average sequence length (bp)	Homology (%)	Conserved	Variable	No. of SNPs	Percentage of SNP	CpG %
<i>T. grandis</i>							
rbcL	600	98	653	75	73	43.8	97
ITS-trnH	415	97	314	93	115	15.58	85
<i>D. latifolia</i>							
rbcL	541	99	608	85	105	37.52	68
ITS-trnH	759	96	745	115	139	12.58	59
<i>L. lanceolata</i>							
rbcL	620	100	60	56	43	30.25	91
ITS-trnH	305	95	305	68	96	10.58	73

showed low variable whereas C and G were highly variable. Adenine ranged from 160 (Dharwad) to 174 (IWSTB) with the average of 66. Thymine base was varied in between 108 (Dharwad) and 125 (Hassan) with the mean value 114 followed by cytosine which ranged from 193 (Hassan) to 237 (IWSTB) with the average 219 base and by guanine which ranges between 192 (Dharwad) and 232 (IWSTB) with the mean value of 218 base. Similarly, in case of rbcL, G nucleotide was a low variation whereas A, T, and C were showing high variation in *D. latifolia*. Adenine variation ranged from 122 (IWSTB) to 160 (Hassan) with the average value of 143. Thymine base was ranged from 146 (IWSTB) to 168 (Hassan) with the mean value 155 followed by cytosine which ranged from 91 (IWSTB) to 151 (Hassan) with the average of 117 and by guanine nucleotide variation which ranged from 108 (IWSTB) to 127 (Hassan) with the average value of 116, respectively (Fig. 5; Table 5).

In *T. grandis*, adenine nucleotide was showing low variation while thymine, cytosine, and guanine were found to be extremely variable which corresponds to the provenances. In case of ITS-trnH sequences of *T. grandis*, A and T nucleotides showed low variable whereas C and G were highly variable. Adenine ranged from 40 (IWSTB) to 97 (Dharwad) with the average of 78 nucleotides. Thymine base was diverse in between 49 (IWSTB) and 85 (Dharwad) with the mean value of 114 followed by cytosine which ranged from 114 (IWSTB) to 160 (Dharwad) with the average of 135 nucleotides and by guanine which ranges between 99 (IWSTB) and 144 (Dharwad and Hassan) with the mean value of 129 base. Similarly, in case of rbcL, adenine nucleotide was showing low variation whereas cytosine, guanine, and thymine were showing high variation in *T. grandis*. Adenine variation ranged from 147 (Hassan) to 176 (Dharwad) with the average value of 166 nucleotides. Thymine base was ranged from



**Table 5** Nucleotide sequence based difference in *D. latifolia*, *T. grandis* and *L. lanceolata*

Locus	Location						
	DH	DD	DB	TH	TD	TB	LH
ITS-trnH	A = 166	A = 160	A = 174	A = 97	A = 97	A = 40	A = 123
	T = 125	T = 108	T = 125	T = 86	T = 85	T = 49	T = 95
	C = 193	C = 226	C = 237	C = 131	C = 160	C = 114	C = 74
	G = 229	G = 192	G = 232	G = 144	G = 144	G = 99	G = 184
rbcL	A = 160	A = 146	A = 122	A = 147	A = 176	A = 175	A = 186
	T = 168	T = 151	T = 146	T = 151	T = 156	T = 170	T = 166
	C = 151	C = 107	C = 91	C = 113	C = 97	C = 129	C = 139
	G = 127	G = 111	G = 108	G = 118	G = 123	G = 146	G = 124

DH *D. latifolia* Hassan, DD *D. latifolia* Dharwad, DB *D. latifolia* IWST Bangalore, TH *T. grandis* Hassan, TD *T. grandis* Dharwad, TB *T. grandis* IWST Bangalore, LH *L. lanceolata* Hassan

151 (Hassan) to 170 (IWSTB) with the mean value of 159 followed by cytosine which ranged from 97 (Dharwad) to 129 (IWSTB) with the average of 113 and by guanine nucleotide variation which ranged from 118 (Hassan) to 146 (IWSTB) with the mean of 129 nucleotides (Fig. 6; Table 5).

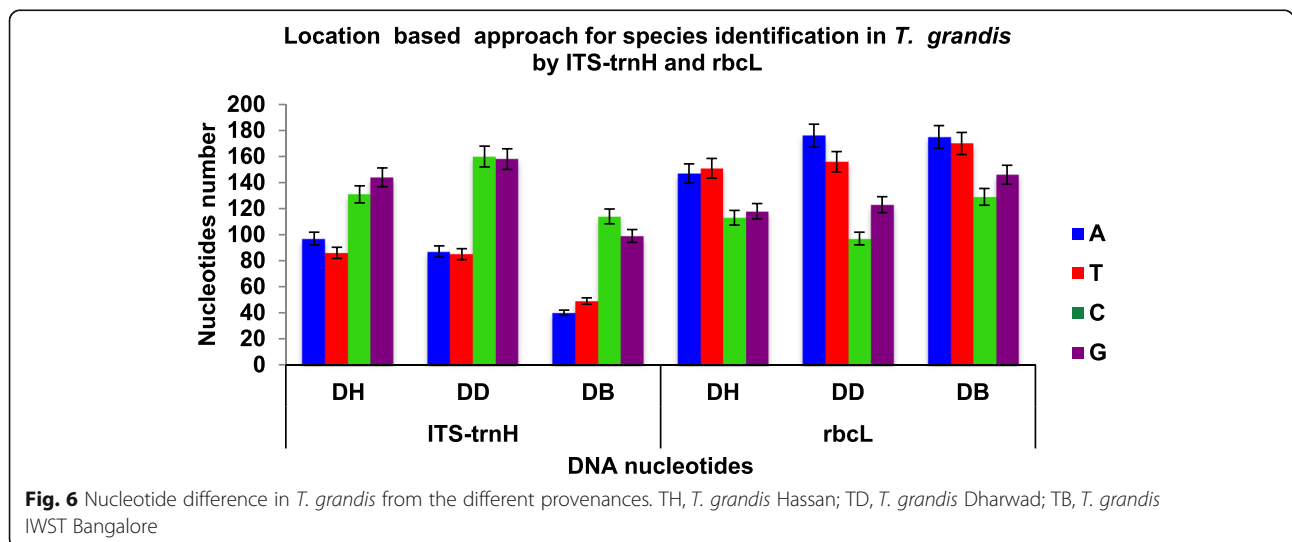
**Multiple sequence alignment and discrimination of species**

We analyzed three individuals of the selected species from different regions of Karnataka, and the sequences of the two examined regions of plastid DNA (rbcL) and nuclear DNA (ITS-trnH) were examined. Multiple sequence alignment of the selected three species sequences was carried out by using the online software ClustalW (v 1.2.4). We manually calculated the sequence alignments for SNPs and found unique SNPs in all the selected species which were observed with rbcL and ITS-trnH barcoding markers.

The total number of sequences of *D. latifolia* by using ITS-trnH markers was found to be 742 bp in IWST, 742 bp in Dharwad, and 762 bp in Hassan, Karnataka. Total rbcL sequences found in *D. latifolia* were 609 bp in

IWSTB, 503 bp in Dharwad, and 515 bp in Hassan, Karnataka. We found 28 unique nucleotide sites in *D. latifolia* by using ITS-trnH. These were the 116th and 762th sites of the ITS-trnH sequences at which A and T respectively were present in *D. latifolia* of IWSTB and in Dharwad while G was present in *D. latifolia* Hassan. In case of *D. latifolia*, a total of 38 unique sites were found by using rbcL markers. These were the rbcL sequences at which A and T were present in Dharwad and Hassan while C and G were found in Hassan, Karnataka (Fig. 7).

The total number of sequences of *T. grandis* which was obtained by using ITS-trnH markers was 315 bp in IWSTB, 478 bp in Dharwad, and 456 bp in Hassan. The sequence number of rbcL markers found in IWSTB was 654 bp, in Dharwad 620 bp, and in Hassan region 529 bp. In *T. grandis*, 25 unique sites were found by ITS-trnH sequences. These were the 133th to 478th sites of the ITS-trnH at which C and G were present in *T. grandis* of Dharwad and IWSTB, respectively, while A and T were present in *T. grandis* of Hassan, Karnataka. A total of 53 unique sites were found in *T. grandis*

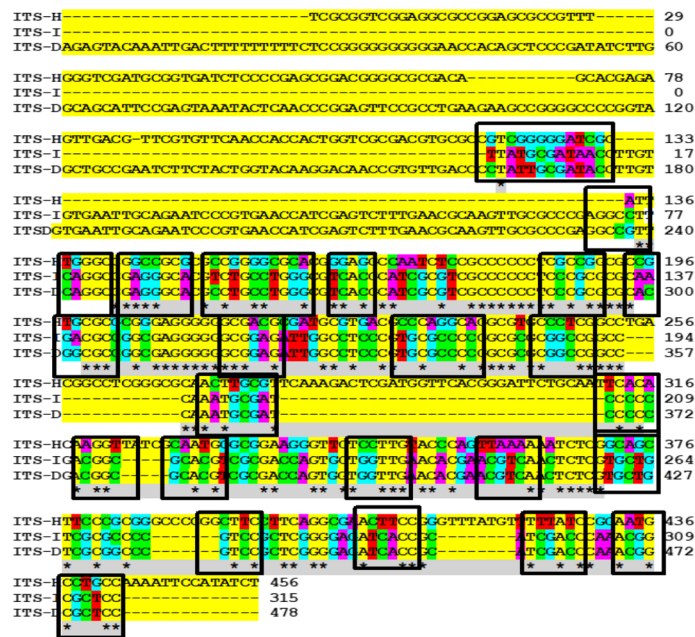


**Fig. 6** Nucleotide difference in *T. grandis* from the different provenances. TH, *T. grandis* Hassan; TD, *T. grandis* Dharwad; TB, *T. grandis* IWST Bangalore

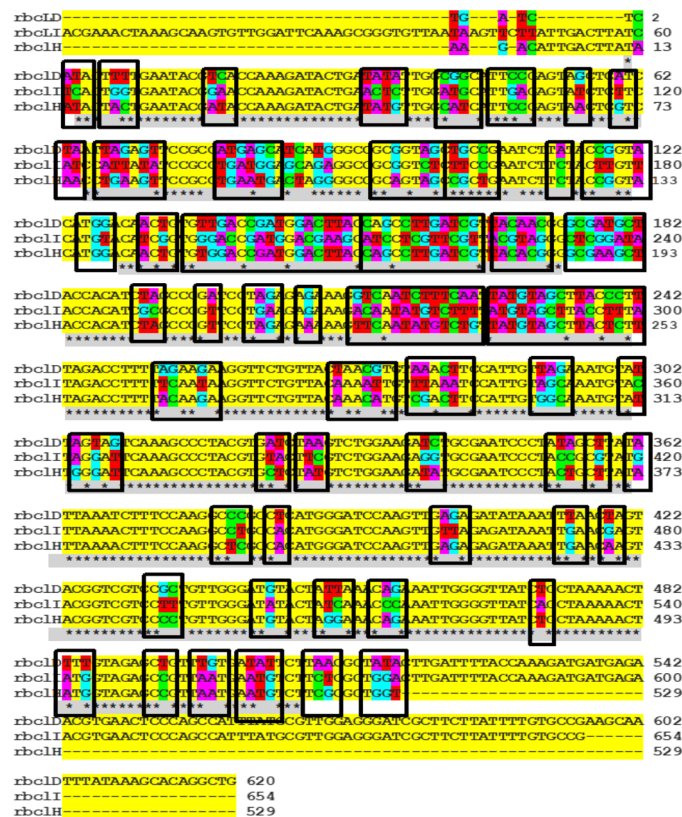




**(A)** ITS marker (*T. grandis*)

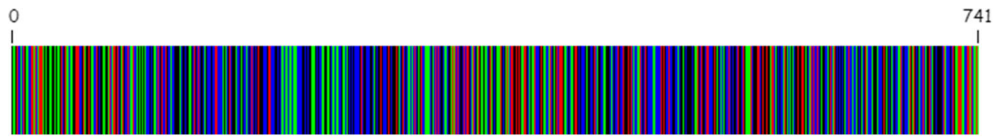


**(B)** rbcL (*Tectona grandis*).

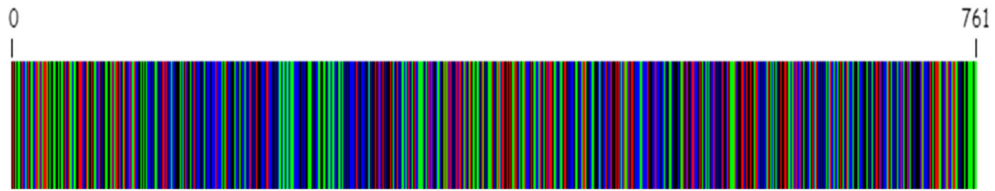


**Fig. 8** Multiple sequence alignment screening single nucleotide polymorphism (SNP) highlighted in squares in *T. grandis*. **a** ITS-trnH sequences. **b** rbcL sequences

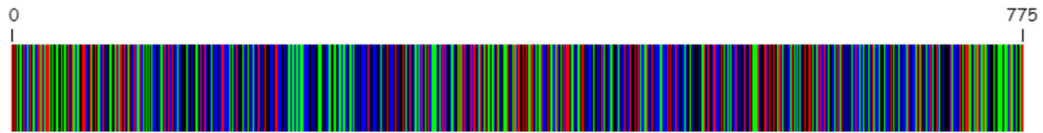
**(A) ITS region DNA Barcodes for *D. latifolia***



(i) (*D. latifolia* Haliyal Barchi range Dharwad)

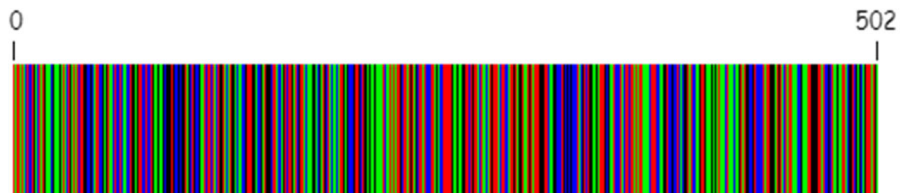


(ii) (*D. latifolia* Alur Taluka Hassan)

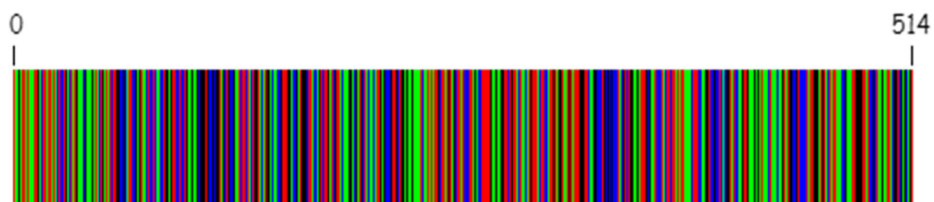


(iii) (*D. latifolia* IWST Bangalore)

**(B) rbcL DNA barcodes for *D. latifolia***

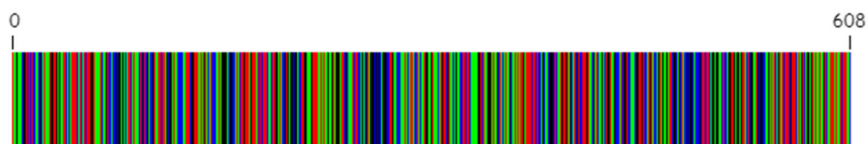


(i) (*D. latifolia* Haliyal Barchi range Dharwad)



(ii) (*D. latifolia* Alur Taluka Hassan)

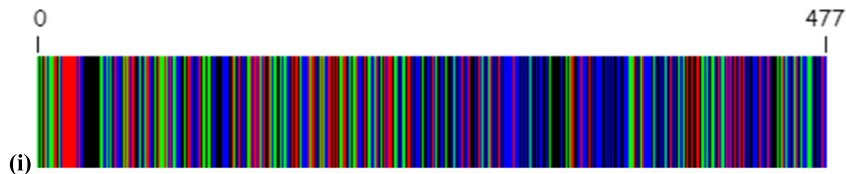
(iii)



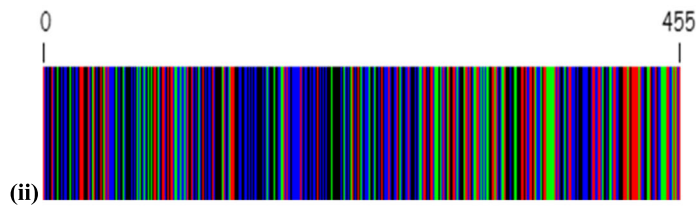
(iii) (*D. latifolia* IWST Bangalore)

**Fig. 9 a** DNA barcodes of ITS-trnH *D. latifolia*. (i) Dharwad. (ii) Hassan. (iii) Bangalore. **b** DNA barcodes of rbcL *D. latifolia*. (i) Dharwad. (ii) Hassan. (iii) Bangalore Karnataka

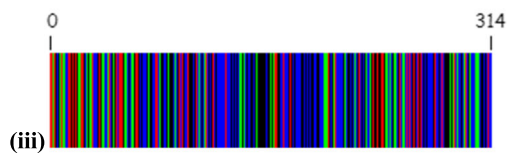
**(A) ITS-trnH region barcodes for *T. grandis***



*(T. grandis Haliyal Barchi range Dharwad)*

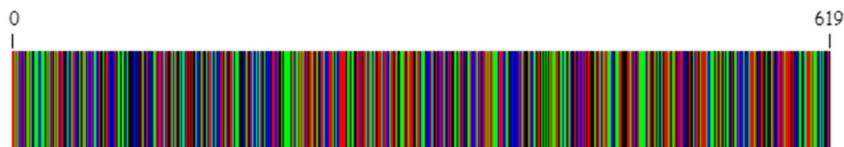


*(T. grandis Alur Taluka Hassan)*

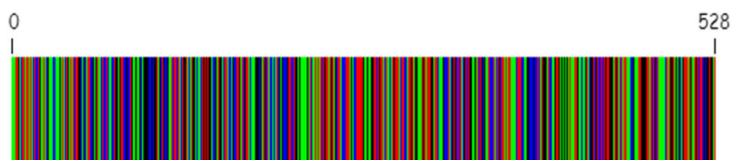


*(T. grandis IWST Bangalore)*

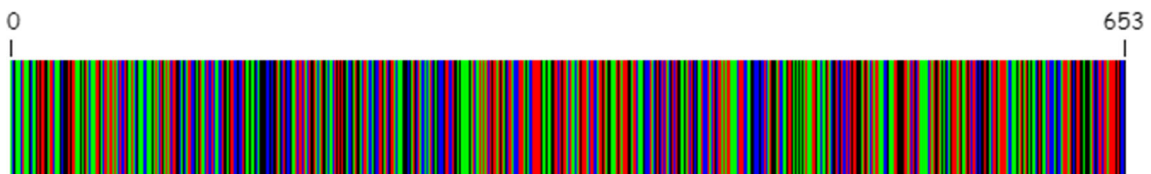
**(B) rbcL DNA barcodes for Teak (*T. grandis*)**



*(T. grandis Haliyal Barchi range Dharwad)*

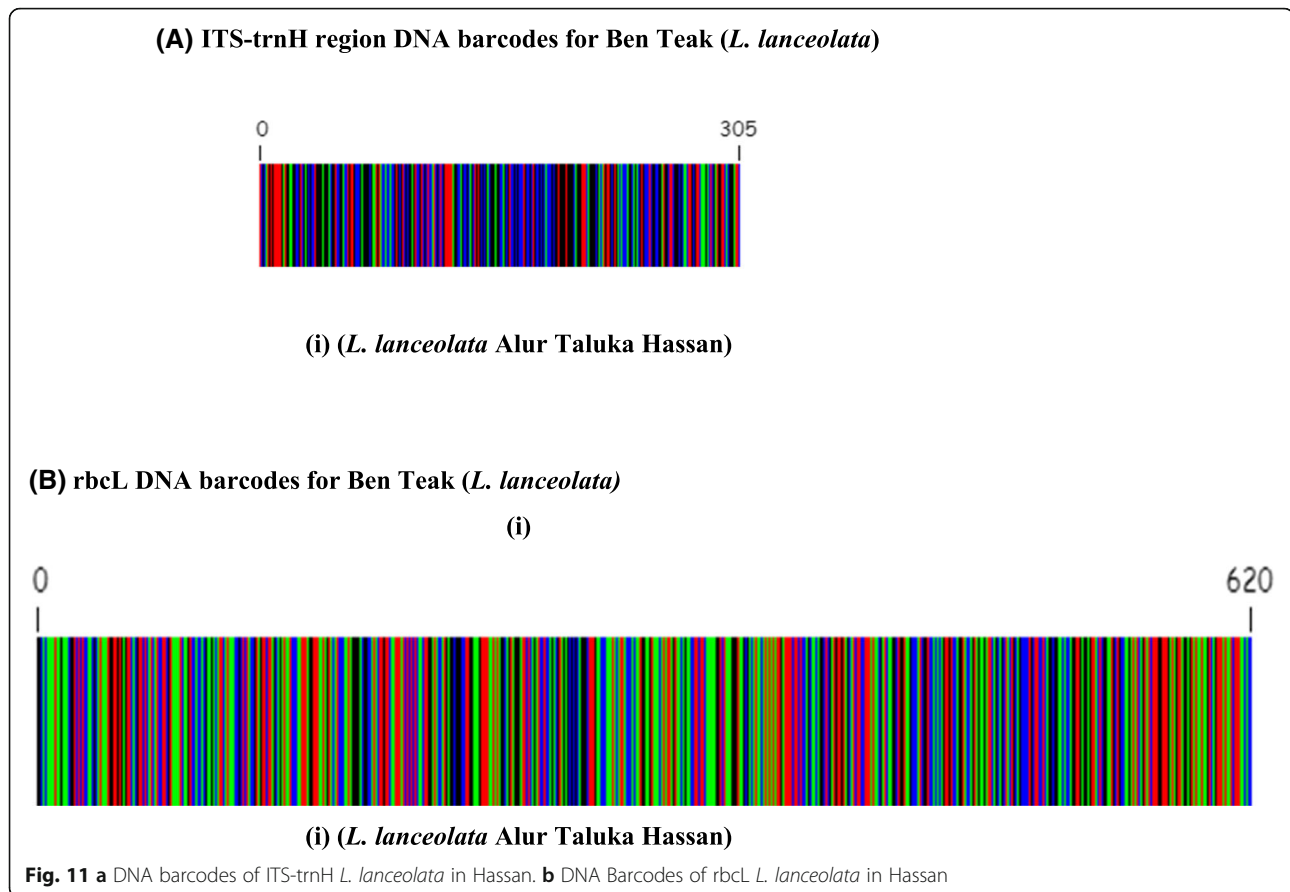


*(T. grandis Alur Taluka Hassan)*



*(T. grandis IWST Bangalore)*

**Fig. 10 a** DNA barcodes of ITS-trnH *T. grandis*. (i) Dharwad. (ii) Hassan. (iii) Bangalore. **b** DNA barcodes of rbcL *T. grandis*. (i) Dharwad. (ii) Hassan. (iii) Bangalore, Karnataka



*T. grandis* belongs to Dharwad and IWST Bangalore region which were more related to each other (0.004–0.048) than Hassan samples. Hassan and IWST Bangalore were found to be quite distantly related (0.045–0.008), whereas Hassan and Dharwad were more related than IWST Bangalore (0.002–0.008) (Fig. 12b).

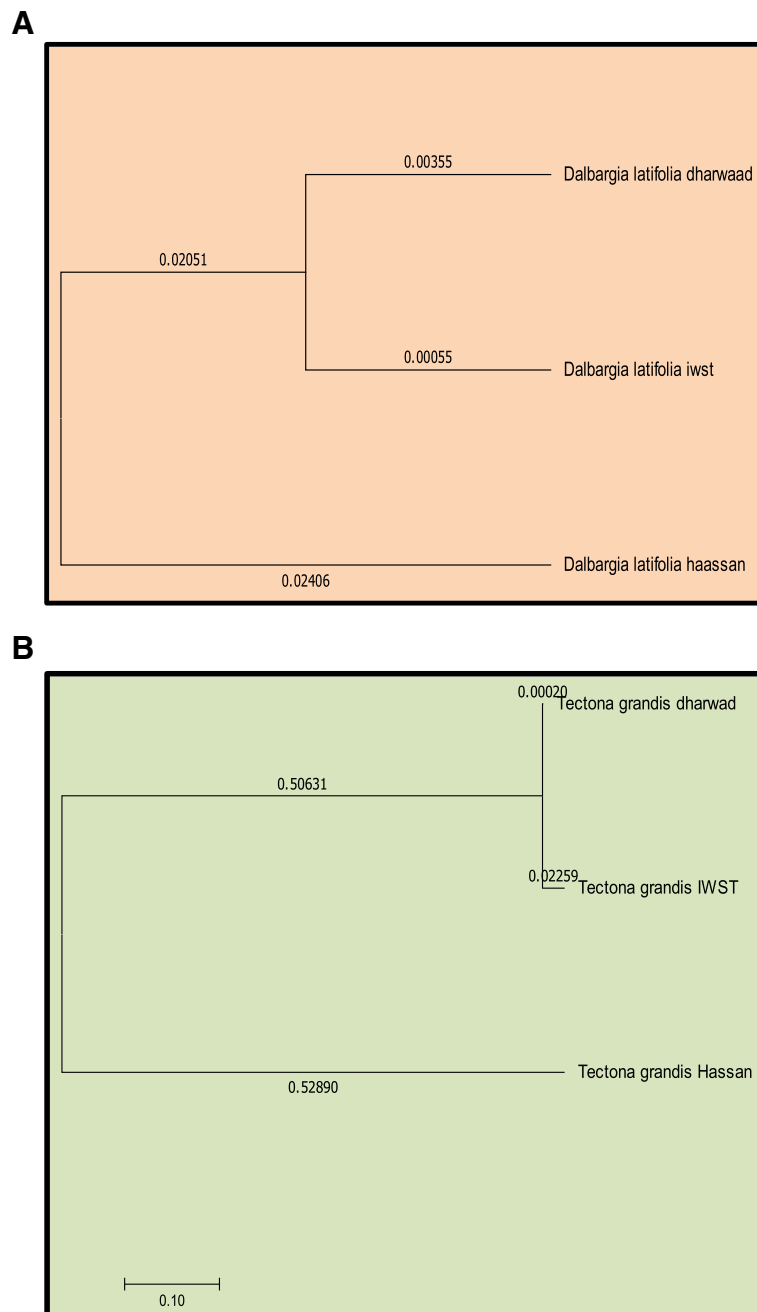
On the other hand, *L. lanceolata* samples from Hassan and Dharwad were diverged at the rate of 0.04634–0.3290 (Table 6). From the above results, it was concluded that the geographical effect and genetic effect were both high in Bangalore and Hassan populations rather than in Dharwad populations. Both factors were acted upon the selected tree species and gene low variability based on all the climatic as well as genetic conditions. Conserved gene variability was slow in Dharwad, IWST Bangalore, and Hassan populations and resulted less variability.

#### Primer development

In this study, we developed region-specific (IWSTB, Dharwad, and Hassan, Karnataka) barcode primers for each selected species (*D. latifolia*, *T. grandis*, and *L. lanceolata*) by using rbcL- and ITS-trnH-generated marker sequences Table 7.

#### Discussion

This study aims to develop a DNA marker to identify the species of *T. grandis*, *D. latifolia* and *L. lanceolata* by a multilocus DNA barcoding approach. The chloroplast DNA region rbcL and nuclear region ITS could be amplified successfully through the fresh wood samples and confirmed that fresh wood is a good source of DNA for wood tissue barcoding in *Aquilaria sinensis* (Jiao et al. 2013). In this study, the rbcL and ITS-trnH regions of chloroplast and nuclear DNA were amplified by using the wood tissues of respective timber samples (Fig. 3a–c). Lee et al. (2016) developed a DNA barcode to identify seven *Aquilaria* species using eight barcode candidate loci (matK, rbcL, rpoB, rpoC1, psbA-trnH, trnL-trnE, ITS, and ITS2) and the results revealed that all single barcodes amplified and sequenced with the selected primers. The results revealed that all single barcodes could be easily amplified and sequenced with the selected primers. In the present study, we found that rbcL marker is individually able to discriminate species of the selected timbers (Fig. 3d). Previous studies revealed that *matK* barcode individually or in combination with other chloroplast DNA barcodes yielded a low success rate for species discrimination, but it has the ability to cluster studied *Pterocarpus* species according to their broad geographic origins (Jiao et al.



**Fig. 12** Dendrogram depicted by the maximum likelihood method genetic relationship of *D. latifolia* (a) and *T. grandis* (b) (values indicate genetic distances)

2018). Amplification and sequencing of the matK barcoding region is problematic due to high sequence variability in the primer binding sites (Hollingsworth et al. 2011). In the present study, the overall lowest rate of amplification and sequencing was reported with matK barcoding marker (Fig. 1b). Hassold et al. (2016) characterized diagnostic polymorphisms in the three DNA barcoding markers (matK, rbcL, and trnL) that allowed rapid discrimination between the *Dalbergia* species of Madagascar

and its distribution range. Li et al. (2011) suggested that trn-F would be a good alternative locus due to high variation within trn-F marker and recommended the adoption of rbcL + matK for fern studies. The ITS barcoding markers which perform as an influential marker at the species level show high levels of interspecific divergence (Alvarez and Wendel 2003). In this study, we used three barcode markers in which rbcL and ITS-trnH performed 97% of the amplification and

**Table 6** Estimates of evolutionary divergence between sequences

<i>D. latifolia</i> Dharwad		0.008	0.002
<i>D. latifolia</i> Hassan	0.048		0.008
<i>D. latifolia</i> IWST Bangalore	0.004	0.045	
<i>T. grandis</i> Dharwad		0.115	0.008
<i>T. grandis</i> Hassan	1.040		0.122
<i>T. grandis</i> IWST Bangalore	0.023	1.071	

sequencing Table 3. Kang et al. (2017) reported a success rate of amplification and sequencing of matK fragment (50–57%). In the present study, the success rate of amplification and sequencing of the selected markers is 96 to 94% for rbcL, 91 to 80% for ITS-trnH, and 43 to 30% for matK with the selected species. Nuclear ITS-trnH fragment represents a more promising DNA barcode marker than the maternally inherited chloroplast barcodes in Aceraceae (Han et al. 2016), while in this study, both rbcL and ITS barcoding markers were amplified and generate good results with all the selected species Figs. 9, 10, and 11. DNA barcoding with ITS marker alone revealed unknown species and proved useful in identifying species listed in endangered species appendixes (Meullner et al. 2011).

Nithaniyal et al. (2014) studied that rbcL and matK markers correctly identified 136 species out of 143 tropical dry evergreen species. The success rate of amplification and sequencing of ITS marker could be variable for tropical forest plant in different regions (Kang et al. 2017). In this study, the selected species belongs to tropical forest, and amplification followed by sequencing was found highly successful by using ITS markers. *D. oliveri* species were accurately identified by rbcL, matK, and ITS barcode markers (Hartvig et al. 2015). Bhagwat et al. (2015) reported matK and matK + rbcL as the most suitable barcodes to differentiate *Dalbergia* species. In this present study, we found that not only combinations of the barcodes but also alone maker could identify

the species. ITS sequences ranged from 663 to 666 bp with 94 variable sites and 80 informative sites. The alignment length was 69 bp with 19 insertion and deletion of 1–3 bp within the aligned regions in the genus *Decalepis* (Apocynaceae) (Mishra et al. 2017a, 2017b). In our study, ITS-trnH ranged from with variable sites 68 to 115 (Table 4). Tsumura et al. (2011) developed a dataset and found nucleotide substitutions between the four wood plant groups, which provide convenient sites for developing markers. The highest nucleotide diversity was found in White Meranti species out of four species (Yellow, Valau, and Red Meranti). The two unique nucleotides found in *S. albida* were not present in any other *Shorea* species which were the 11th and 1225th sites of the trnH-psbA-trnK sequence, at which G and C respectively were present in *S. albida*, while A and G were present in all other *Shorea* species. In this study, we found 28 unique nucleotide sites in *D. latifolia* by using ITS-trnH and 38 unique sites was found by using rbcL markers (Fig. 7; Table 5). In *T. grandis*, 25 unique sites were found by using ITS-trnH sequences and 53 unique sites were found by using rbcL markers which indicated that rbcL sequences having more bp variation in their unique sequences than ITS-trnH unique sequences. The above rbcL and ITS-trnH sequencing and barcoding results of *T. grandis*, *D. latifolia* and *L. lanceolata* trees samples, revealed a considerable discrimination in intra-species and inter-species level in the selected regions of Karnataka (Table 1). (Hassan, IWST Bangalore, and Dharwad. Genetic distance base neighbor-joining analysis revealed the character-based identification of the traded species in *Terminalia*. Hassold et al. (2016) tested matK, rbcL, and trnL barcoding markers to differentiate *Dalbergia* species between Madagascar and from other areas of its distributional range. Phylogenetic analysis revealed that the *Dalbergia* studies from monophyletic groups contain two subgroups, one of which corresponds to a single species. Phylogenetic study of *Terminalia* species revealed the average intraspecific variations with the combination of 0.0188 to 0.0019

**Table 7** Primer developed for the selected species

Species	Barcode loci	Primer information	Size (bp)	Product size (bp)	Primer temperature (°C)
<i>T. grandis</i>	ITS	F' CGATATCTTGCCAGCATTCC R' TTGCGTTCAAAGACTCGATG	20/20	232	61.0
	rbcL	F' GACAACTGTGTGGACCGATG R' AGAGCACGTAGGGCTTTGAA	20/20	201	60.0
<i>D. latifolia</i>	ITS	F' AGGAAGCAACAACCGAACAG R' TTGCGTTCAAAGACTCGATG	20/20	192	60.0
	rbcL	F'GCTGCCGAATCTTCTACTGG R' GGATTCGCAAAATCTTCCAGA	20/20	250	60.0
<i>L. lanceolata</i>	ITS	F' GCTCCATCGCTCGTTTTTAG R' CCGTGAACCATCGAGTCTTT	20/20	210	60.0
	rbcL	F' CAGGGCTTTGAACCCAAATA R' GACAACTGTGTGGACCGATG	20/20	193	60.0

were less than the distance to the nearest neighbor (0.106 ± 0.009) with the matK and ITS (Mishra et al. 2017a, 2017b). In the present study, we calculated divergence among the individuals of the species (intraspecific divergence) with the combination of rbcL and ITS-trnH phylogenetic analysis revealed in *T. grandis* (0.004–0.048), *D. latifolia* (0.008–0.122), and *L. lanceolata* (0.046–0.329) (Table 6). The genetic distance between sequence pair after multiple sequence alignment was helpful in determining the genetic relatedness between the species of selected timbers at the molecular level. Li et al. 2011 performed the DNA barcoding technique to generate barcode sequences for *Aquilaria* species identification and applied the barcodes to identify the source species of agarwood found in the market. DNA barcoding of eight *Dalbergia* species from Vietnam recommended ITS locus as a potential barcode based on UPGMA analysis and nucleotide diversity (Phong et al. 2014), but in this study, we found that rbcL marker was a strong barcode marker for *D. latifolia* and *T. grandis* and based on genetic distance UPGMA study. Phylogenetic analysis of *Santalum album* with its adulterants *Osyris wightiana* and *Erythroxylem monogynum* formed the most genetically similar cluster by using rbcL, matK, and trnH-psbA barcoding markers (Dev et al. 2014). In the present study, genetic similarity with low intraspecific variation was reported in *T. grandis*, *D. latifolia* and *L. lanceolata* trees samples by using the rbcL and ITS-trnH markers. Through the barcoding sequence results and phylogenetic analysis, *D. latifolia* samples of IWST Bangalore are genetically more related to Dharwad than Hassan, Karnataka (Fig. 12a). Phylogenetic analysis of *T. grandis* IWST Bangalore samples was genetically more related to Dharwad than Hassan, Karnataka (Fig. 12b). Out of three universal primers, rbcL and ITS-trnH detected suitable and distinguishable gene sequences for the selected timber species, and that could have been used to get barcodes for forest timber species.

## Conclusions

DNA-based barcoding is an invaluable tool for detecting errors in identifications of the tree species. The ratio of successfully identified species to all three sampled species was calculated as the proportion of species that were discriminated. From the above sequences, barcodes and pairwise distance matrix indicated their closeness during evolution and help to draw their phylogenetic tree. The present study provided a database to which the selected DNA sequences from unknown samples could be matched as well as necessary sequences from which other timbers or plant species could be identified and used for species identification. This study was able to identify wood samples, which represent that DNA barcoding has the potential to be an identification tool in the selected timbers.

## Abbreviations

A: Adenine; C: Cytosine; CTAB: Cetyl trimethyl ammonium bromide; G: Guanine; PCR: Polymerase chain reaction; SSR: Simple sequence repeats; T: Thymine; TAE: Tris-acetate-EDTA

## Acknowledgements

The authors are thankful to the Director, IWST, Group Coordinator Research, Head Genetics and Tree Improvement Division, molecular biology lab members in the Institute of Wood Science and Technology, the Additional Principal Chief Conservator of Forest (research), and the Karnataka Forest Department for financial support. The authors also wish to thank Hassan DFO, Dharwad DFO, and Barchi Haliyal forest range officers for their assistance on field trips, tree identification, and sample collection.

## Funding

The authors declare that this work was funded by the Karnataka Forest Department in Karnataka, India (Project ID-KFD/DNA 2014-2016).

## Availability of data and materials

We declare that all data generated or analyzed during this study are included in this manuscript.

## Authors' contributions

TF completed the sample collection in the field and laboratory experiment work, the data analysis, and the writing of the draft of the manuscript. AS was involved in designing the experiments and selecting the species and the study location. SPV carried out the sample identification and collection in the field. VSH was involved in the result interpretations and reviewed the manuscript. MSR was involved in reviewing the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## 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.

## Author details

<sup>1</sup>Genetics and Tree Improvement Division, Institute of Wood Science and Technology, Bangalore 03, India. <sup>2</sup>Division of Biotechnology, Indian Institute of Horticultural Research, Hessaraghatta Lake Post, Bangalore 89, India. <sup>3</sup>Woodworking Division, Institute of Wood Science and Technology, Bangalore 03, India.

Received: 18 January 2019 Accepted: 11 April 2019

Published online: 10 May 2019

## References

- Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* 29:417–434
- Bhagwat RM, Dholakia BB, Kadoo NY, Balasundaran M, Gupta VS (2015) Two new potential barcodes to discriminate *Dalbergia* species. *PLoS One* 10(11): e0142965
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proc Natl Acad Sci U S A* 106(31):12794–12797
- Dev SA, Muralidharan EM, Sujjanpal P, Balasundaran M (2014) Identification of market adulterants in east Indian sandalwood using DNA barcoding. *Ann For Sci* 71:517–522
- Fatima T, Srivastava A, Hanur VS, Rao SM (2018a) An effective wood DNA extraction protocol for three economic important timber species of India. *Am J Plant Sci* 9:139–149
- Fatima T, Srivastava A, Somashekar PV, Hanur VS (2018b) Optimization of genomic DNA extraction routine from ROSEWOOD (*Dalbergia latifolia*) and BEN TEAK (*Lagerstroemia lanceolata*). *Int J Gen* 6:446–451



- Finkeldey R, Leinemann L, Gailing O (2009) Molecular genetic tools to infer the origin of forest plants and wood. *Appl Microbiol Biotechnol* 5:1251–1258
- Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol* 30:1229–1235
- Han WY, Duan D, Ma XF, Jia Y, Liu ZL, Zhao GF, Li ZH (2016) Efficient identification of the forest tree species in Aceraceae using DNA barcodes. *Front Plant Sci* (7):1707
- Hansen OK, Changtragoon S, Ponoy B, Lopez J, Richard J, Kjaer ED (2017) Worldwide translocation of Teak- origin of landraces and present genetic base. *Tree Genet Genomes* 13:87
- Hartvig I, Czako M, Kjaer ED, Nielson LR, Theilade I (2015) The use of DNA barcoding in identification and conservation of Rosewood (*Dalbergia spp.*). *PLoS One* 9(1371):1–24
- Hassold S, Lowry PP, Bauert MR, Razafintsalama A, Ramamonjisoa L, Widmer A (2016) DNA barcoding of Malagasy rosewoods: towards a molecular identification of CITES-listed *Dalbergia* species. *PLoS One* 10(1371):1–17
- Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol Ecol Resour* 9:439–457
- Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. *PLoS One* 6:e19254
- Jiao L, Yin Y, Cheng Y, Jiang X (2013) DNA barcoding for identification of the endangered species *Aquilaria sinensis*: comparison of data from heated or aged wood samples. *J Holzforschung* 4:487–494
- Jiao L, Yu M, Wiedenhoef AC, He T, Li J, Liu B, Jiang X, Yin Y (2018) DNA barcode authentication and library development for the wood of six commercial *Pterocarpus* species: the critical role of *Xylarium* specimens. *Sci Rep* 28:1945
- Kang Y, Deng Z, Zang R, Long W (2017) DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forest. *Sci Rep* 7: 12564. 1–12564. 9
- Kimura MA (1980) Simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kollert W, Walotek J (2015) Global Teak trade in the aftermath of Myanmar's log export ban. Food and agriculture organization of the united nations. <http://www.fao.org/3/a-i5023e.pdf>.
- Kumar S, Stecher G, Tamura K (2007) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874
- Lee SY, Ng WL, Mahat MN, Nazre M, Mohamed R (2016) DNA barcoding of the endangered *Aquilaria* (Thymelaeaceae) and its application in species authentication of agarwood products traded in the market. *PLoS One* 11(4): e0154631. <https://doi.org/10.1371/journal.pone.0154631>
- Li F-W, Kuo L-Y, Rothfels CJ, Ebihara A, Chiou W-L, Windham MD, Pryer KM (2011) rbcL and matK earn two thumbs up as the core DNA barcode for ferns. *PLoS One* 6(10):e26597
- Lowe AJ (2007) Can we use DNA to identify the geographic origin of tropical timber. Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins", Bonn
- Lowe AJ, Cross HB (2011) The application of DNA methods to tree tracking and origin verification. *J IAWA* 2:251–262
- Meullner AN, Schaefer H, Lahaye R (2011) Evaluation of candidate DNA barcoding loci for economically important tree species of the mahogany family (Meliaceae). *Mol Ecol Resour* 11:450–460
- Mishra P, Kumar A, Nagireddy A, Shukla AK, Sundaresan V (2017a) Evaluation of single and multilocus DNA barcodes towards species delineation in complex tree genus *Terminalia*. *PLoS One* 12(8):e0182836
- Mishra P, Kumar A, Sivaraman G, Shukla AK, Kaliyamoorthy RK, Slater A, Velusamy S (2017b) Character-based DNA barcoding for authentication and conservation of IUCN red listed threatened species of genus *Decalepis* (Apocynaceae). *Sci Rep* (7):14910
- Mohanan C, Sharma JK, Florence EJM (1997) Nursery diseases of teak in India. Proceeding of the International Teak Symposium, Thiruvananthapuram, pp 107–112
- Nithaniyal S, Newmaster SG, Ragupathy S, Krishnamoorthy D, Vassou SL, Parani M (2014) DNA barcode authentication of wood samples of threatened and commercial timber trees within the tropical dry Evergreen Forest of India. *PLoS One* 9(9):e107669
- Orwa CA, Mutua R, Jamnadass RS, Anthony S (2009) Agroforestry tree database: a tree reference and selection guide version 4.0. World Agroforestry Centre, Kenya
- Palmer JD (1985) Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MACINTY RJ, ILI MI (eds) Monographs in evolutionary biology: mole evol genet, pp 131–240
- Pettengill JB, Neel MC (2010) An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae). *Am J Bot* 97(8):1391–406
- Phong DT, Tang DV, Hien VTT, Ton ND, Van HN (2014) Nucleotide diversity of a nuclear and four chloroplast DNA regions in rare tropical wood species of *Dalbergia* in Vietnam: a DNA barcode identifying utility. *Asian J Appl Sci* 2:116–125
- Saiki FA, Bernardi AP, Reis MS, Faoro H, Souza EM, Pedrosa FO, Mantovani A, Guidolin AF (2017) Development and validation of the forest SSR markers for *Mimosa scabrella* Benth. *Genet Mol Res* 16;16(1). <https://doi.org/10.4238/gmr16019571>
- Sambrook J, Fritesh EDF, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd edn. Cold spring Harbor Laboratory Press, NY, p 2344
- Tsumura Y, Kado Y, Yoshida K, Abe H, Ohtani M, Taguchi Y, Fukue Y, Tani N, Ueno S, Yoshimura K, Kamiya K, Harada K, Takeuchi Y, Diway B, Finkeldey R, Naiem M, Indrioko S, Ng KKS, Muhammad N, Lee SL (2011) Molecular database for classifying *Shorea* species (Dipterocarpaceae) and technique for checking the legitimacy of timber and wood products. *J Plant Res* 124:35–48
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40:15

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---