



Assessment of morphological and genetic variability through genic microsatellite markers for essential oil in Sandalwood (*Santalum album* L.)

Tanzeem Fatima¹ · Ashutosh Srivastava¹ · P. V. Somashekar¹ · Vageeshbabu S. Hanur² · M. Srinivasa Rao³ · Surendra Singh Bisht⁴

Received: 22 January 2019 / Accepted: 11 May 2019 / Published online: 7 June 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

Sandalwood (*Santalum album* L; family Santalaceae) is a highly significant aromatic oil yielding tree. It is valued for two important traits, heartwood and essential oil obtained from the heartwood. This study was proposed to assess the morphological and genetic variability of sandalwood accessions. For this, genotypes were randomly selected ($n = 177$) from the 14 populations from three states in southern India. The total heartwood oil content and quality was estimated by UV method and GC–MS. Total 14 oil-specific genic SSR markers were procured to evaluate the genetic diversity among the sandalwood accessions. Total core size, heartwood content, and oil of *S. album* ranged from 4.4 to 19.1 cm; 0.0 to 17.3 cm; and 0.0 to 5.96% with covariance 27.61, 85.25, and 73.12% followed by mean 9.74, 3.77, and 2.71, respectively. Genetic diversity estimates were highly polymorphic in terms of Na 7.28, Ne 5.89, He 8.0 PIC 0.891, with little Ho, and F-0.922. AMOVA revealed that minimal genetic variation among populations and highest variation was found among individuals with Nm (58.4). The UPGMA reveals the cluster favored the grouping pattern by the PCA analysis. Structure and PCA analysis clustered the entire populations into two major groups with F_{ST} 0.046 in which population of Kerala and Karnataka were pure and Telangana accessions were found admixtures. No significant correlation ($r^2 = 0.23$, $P = 0.00$) was observed between heartwood oil and genetic structures. A high degree of transferability of genic markers would facilitate the assessment of novel genotypes for future tree improvement and conservation of Sandalwood populations.

Keywords GC–MS · Genic SSRs · Genetic variability · Heartwood oil · PIC · *Santalum album*

Abbreviations

F Inbreeding coefficient
GC–MS Gas Chromatography–Mass spectrometry
He Expected heterozygosity

Ho Observed heterozygosity
Na Number of alleles
Ne Effective number of alleles
Nm Gene flow
PIC Polymorphic information content
PCA Principal component analysis
UPGMA Unweighted Pair Group Method with Arithmetic mean

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1758-9>) contains supplementary material, which is available to authorized users.

✉ Tanzeem Fatima
tanzeem.fatima@gmail.com

- ¹ Genetics and Tree Improvement Division Institute of Wood Science and Technology, Bangalore, 03, India
- ² Division of Biotechnology, Indian Institute of Horticultural Research, Hessaraghatta Lake Post, Bangalore, 89, India
- ³ Woodworking Division, Institute of Wood Science and Technology, Bangalore, 03, India
- ⁴ Chemistry and Bio Prospecting Division, Institute of Wood Science and Technology, Bangalore, 03, India

Introduction

Santalum album is one of the evergreen hemi-parasitic and economically important aromatic tree species. The Sandalwood species identified as a xylem tapping root hemiparasites belongs to the family Santalaceae, which consist of 18 species (Subasinghe 2013). Sandalwood has been categorized as ‘vulnerable’ by International Union for Conservation of natural and natural resources (Arunkumar et al.

2019). It is culturally and economically important species for more than 15 countries (Rao et al. 2007). The sandalwood tree is indigenous to peninsular India. In India, its natural distribution is estimated all over the country with over 90% of the area in Karnataka (5245 km²) (Sreenivasan et al. 1992) and Tamil Nadu state covering around 3600 km² (Jain et al. 2003). Apart from India, it is also found in Australia, Indonesia, Japan, Belgium, China, Cambodia, Fiji, and Madagascar (Purohit 2018). Sandalwood heartwood and oil considered being highly prized products, which are used in incense, perfumes, cosmetics, toiletries, and pharmaceuticals. The genetic resources in the country are threatened by various biotic and abiotic causes including illegal logging, habitat fragmentation, and poor natural regeneration (Sreenivasan et al. 1992). Despite the protection status, the natural resources of sandalwood are being erratically exploited in India (Rao et al. 2007). These activities have selectively eroded the best trees of sandalwood from habitat considering large genetic distance between provenances. *S. album* trees differed significantly in their percentage of heartwood and this varied independently of stem diameter. The biosynthesis of Sandalwood oil is tissue and developmental stage-specific, and occurs in the transition zone between sapwood and heartwood, especially in ray parenchyma after a certain age (Jones et al. 2006). To authenticate the oil constituent's variability, the GC–MS analysis process was carried out to facilitate characterization of oil constituents with oil concentration variants in different climatic zones. The oil composition was α -santalol%, *trans*- β -santalol%, santalol E-*cis*-*epi*- β %, and bergamotol *trans* Z- α santalol%, and various trace element concentrations were recognized to differ across its geographical locations. α -santalols (49–59%) and *trans*- β -santalols (29–35%) are the major contributors of Sandalwood oil fragrance. This study highlights the knowledge of morphological and genetic variation in sandalwood across the herd regions of Karnataka, Kerala, and Telangana states. Various genes play a significant role in oil biosynthesis, but there are no available data on the genetic characterization of sandalwood oil biosynthesizing genes and their polymorphic information. Precisely, four parameters were tested in this study: (1) estimation of heartwood and heartwood oil in collected core samples, (2) estimation of α and β santalol concentration by GC–MS analysis of selected four high and low oil yielding clones, (3) estimation of genetic diversity and genetic differentiation of the selected *S. album* accessions, and (4) association of genetic structure with morphological (heartwood and heartwood oil variability) characters of *S. album*. The estimates of genetic diversity by genic SSR markers would be more useful than the previously available genomic SSR marker in *S. album* and would facilitate the assessment of novel genotypes for future tree improvement and conservation of Sandalwood populations.

Materials and methods

Sample collection

Variation in the species was assessed by undertaking survey and sampling along the area of natural populations and plantations during the period 2014–2018. The morphological and genetic study was conducted in major sandalwood places in Karnataka, Kerala, and Telangana states in southern India (Fig. 1).

Core and leaf sampling

For estimation of morphological and genetic diversity, natural and plantations grown trees were randomly ($N=177$) *S. album* accessions which were collected from three states Southern India including Karnataka ($N=80$), Telangana state ($N=67$), and Kerala ($N=30$) (Table S1). The cores were collected through the conventional increment borer (Mora coretax original Sweden 100 mm) (core diameter 5.15 mm) up to 1.3 m from the ground level. Each core was passed through the trunk of the tree in such a way that it contains intact bark sapwood and heartwood. The hole in the tree was properly sealed with a paraffin wax to prevent infection. The core samples were stored into a butter paper bags and kept it for drying at room temperature before analysis. For molecular work, mature leaves of *S. album* were collected and immediately placed in dried silica gel-containing zip lock covers.

Heartwood percentage and oil yield

A total of 177 core samples were used to estimate the heartwood radius (Fig S2) (ArunKumar et al. 2011). The heartwood content of each tree at GBH (Girth at breast height) (1.3 m) was calculated by converting the tree girth into radius using the formula $2\pi r$ ($\pi=3.14$, r =radial length) from the collected core samples.

Total heartwood oil estimation was done as per UV–spectrophotometer based on the method developed by Shankararayna et al. (1997).

Quantitative analysis of oil

Based on the oil concentration, qualitative analysis was carried out for four high oil yielding *S. album* using GC–MS based on the method developed by Bisht and Hemanthraj (2014).

Gas chromatography–mass spectrometry (GC–MS) analysis

The resulting essential oil masses were evaluated by gas chromatography–mass spectrometry (GC–MS-QP-2010

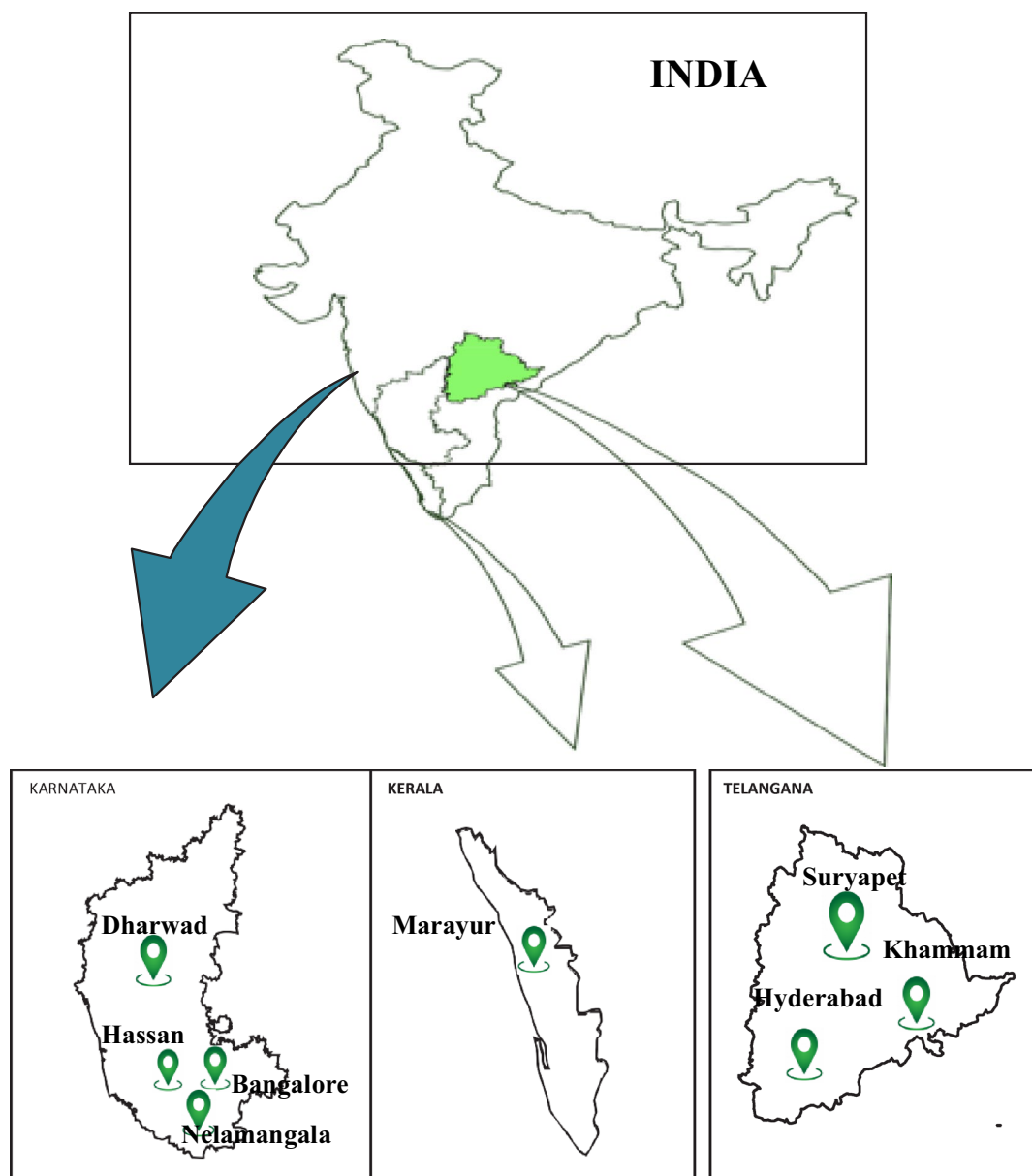


Fig. 1 Site map for location of *S. album* populations in Karnataka, Kerala, and Telangana states in southern India

ultra Auto sampler), equipped with a 30 m (I) 0.25 mm (i.d.), 0.25 mm film thickness, nonpolar, RTX-WAX column, connected to anion trap quadruple (ITQ) mass selective detector, with a unit mass resolution. The split was 1:25.0, with helium as the carrier gas at a flow rate of 1.40 mL/min, while the damping gas flow was 0.3 mL/min. The initial oven temperature was set to 80 °C for 2 min. The GC oven temperature program was as follows: 80–240 °C, by ramping at 3 °C, and held at 240 °C for 6 min. Injector temperature was maintained at 240 °C and the transfer line was held at 220 °C. Detection was performed by a GC–MS–QP2010 Ultrain EI mode (ionization

energy of 70 eV, ion source temperature of 200 °C, and emission current of 220 mA). Acquisition was made in full-scanning mode (mass range 50–900 m/z ; 3 scans/s). Maximum ionization time was 25 ms. A solvent delay time of 3 min was used to avoid overloading the mass spectrometer with chloroform. Resulting GC–MS profile was analyzed using National Institute of Standards and Technology (NIST, Washington DC, USA) and Dr. Duke's Phytochemical and Ethnobotanical Database (<http://www.ars-grin.gov/duke/>). Estimation of metabolite was considered by percentage peak area at the total ion chromatogram in GC–MS analysis.

Genomic DNA isolation and quantitative analysis of DNA

Total genomic DNA was extracted from stored dried mature leaves of *S. album* by using modified CTAB method by Fatima et al. (2018). The purity and intactness of DNA was assessed by nanodrop and running the samples on 0.8% (w/v) agarose gel stained with ethidium bromide.

Sequence data and primer designing

Gene-specific sequence data for eight fully characterized genes and full-length cDNA FASTA sequences (CytochromeP450; 1931 bp, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA); 2616 bp, Geranyl geranyl pyrophosphate synthase 1 (GGPS); 1261 bp and Geranyl pyrophosphate synthase (GPS); 1320 bp) were downloaded from *Arabidopsis* gene data bank TAIR (<http://www.arabidopsis.org>) to develop functional genic SSR markers. The genic microsatellite markers were designed by Primer3 version 0.4.0 software online (<http://bioinfo.ut.ee/primer3-0.4.0>). The primer parameters included the following elementary criteria: (1) primer length should range from 17 to 25 bp. (2) G + C content should not be larger than 45–55%. (3) Melting temperature should range from 55 to 61 °C. Total eight genic primers were synthesized and carried out for further analysis (Table 1).

Selected oil-specific genic microsatellite markers of Sandalwood

Genetic analysis of *S. album* was performed by screening more than 25 primers and selected 14 oil producing genic SSRs which were showing high polymorphism: SaFDSF, SaFDSE, SaSSF, and SaSSE (Rani et al. 2013), DXS, FPPS, STPS and MTPS (Misra and Dey 2013), SESQUITPS1, SESQUITPS2 (Jones et al. 2008), CytP450, GPS1, GGPS1, and HMG-CoA (Developed at IWST laboratory Bangalore). The detailed information of selected genic primers is listed below in Table 2.

PCR standardization of PCR (polymerase chain reaction) conditions for genic primers

Polymerase chain reaction (PCR) was performed according to the protocol developed by Williams et al. (1990). The PCR amplifications were carried out in 0.2 mL tube in Mastercycler gradient (Thermo scientific, Germany) in standardized 14 µL reaction volume. The reaction mixture containing 60 ng of genomic DNA as template DNA, 10× PCR buffer, 1.5 mM MgCl₂ (10 pm/µL) 2 µL (1 µL for each forward and reverse) primers (Eurofins Pvt. Ltd.), 0.2 mM of each of dNTPs, and 0.2 µL (3U/µL) Taq polymerase. Amplification

reactions were carried out with cycle profiles, viz., initial denaturation at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 30 s, primer annealing at 55–65 °C for 1 min, extension at 72 °C for 1 min, final extension of 3 min at 72 °C temperature, and hold ∞ at 4 °C.

Separation and detection of amplified PCR products

The amplified product size was separated by gel electrophoresis in 4% high-resolution agarose (Guimaraes et al. 2018) gel stained in ethidium bromide solution in 1× TAE buffer at a constant 80 V current for 4 h. To visualize the bands, gel was observed in UV light provided by a *trans*-illuminator in (Syngene G: Box) gel documentation system. The size and molecular weight of the amplification products was estimated by a standard 50 bp ladder (Fermentas ThermoFisher Pvt. Ltd).

Genetic diversity analysis

Banding profiles generated by the genic SSRs compiled into a data matrix based on the molecular weight of amplicons using Syngene G: box compatible software Gensys tools.

Genetic diversity parameters were characterized for each marker, and each population includes Na, Ne, Ho, He, I, and Nei's genetic distance by Gen Alex 6.0 (Peakall and Smouse 2012). To compare the efficiency of markers, PIC as a marker differentiation tendency was calculated by a software Cervus v 3.0.3 using formula:

$$PIC = 1 - \sum (P_i)^2,$$

where P_i is the frequency of the i th allele at a given locus (Botstein et al. 1980).

Markers were considered as informative when PIC was ≥ 0.5 .

Genetic differentiation

To determine the percentage of genetic distribution of genetic variation, analysis of molecular variance (AMOVA) and Wright's F statistics were performed in GenAlex v.6.5 (Peakall and Smouse 2012). A phylogenetic tree based on UPGMA was constructed using Jaccard's similarity coefficient in poptree2 (Takezaki et al. 2010).

Structure analysis

An unbiased Bayesian approach using Markov chain Monte Carlo (MCMC) clustering of samples was conducted via the Structure v.2.3.4 software (<http://www.stats.ox.ac.uk/>) (Pritchard et al. 2000). Parameters were set as diploid data for each individual and assessed for values of K ranging

Table 1 List of selected oil biosynthesizing gene markers of *S. album*

| Sl. no. | Primer name | Accession number | Primer sequence 5'–3' | No. of bases | Tm °C | Expected amplification range (bp) | Observed amplification range (bp) |
|---------|-----------------------|------------------|--|--------------|-------|-----------------------------------|-----------------------------------|
| 1 | <i>Sa</i> FDSF | KF011939.1 | F: 5'-ATGGGCGATCGGAAAACC AAATTCTCGAG-3' R: 5'-GGCGAAGATATACAAGCG GCAGAAGTAG-3' | 30/28 | 60.5 | 151–161 | 160–180 |
| 2 | <i>Sa</i> FDSE | HQ343283.1 | F: 5'-TCCTCCGGGACCCCGCTT TC-3' R: 5'-GCACCAACCGAGTGCAGA TGC-3' | 20/21 | 64.3 | 156–166 | 150–200 |
| 3 | <i>Sa</i> SSF | KM108327.1 | F: 5'-ATGGTTCTTCCACCGCCA CCGCCATGAC-3' R: 5'-CTACTCCTCGCCGAGAGG AATAGGGTCG-3' | 28/28 | 65.8 | 159–169 | 200–220 |
| 4 | <i>Sa</i> SSE | KM108326.1 | F: 5'-AGGCATATGAGCAAGAAG CGAATATG-3' R: 5'-GTCCATTCAAGCTATATT CTCC-3' | 26/23 | 57.3 | 165–175 | 169–180 |
| 5 | <i>Sa</i> DXS | HM232852 | F: 5'-GACTCGGACATGATTGTC ATTCT-3' R: 5'-AATAGAGCCCAAGCTTTT CG-3' | 23/20 | 54.7 | 277–287 | 240–250 |
| 6 | <i>Sa</i> FPPS | HM232854 | F: 5'-TGAGAATCTGGACAGCCA CA-3' R: 5'-CAGGATGACCAAAGCAGT CC-3' | 20/20 | 56.0 | 101–111 | 110–123 |
| 7 | <i>Sa</i> MTPS | HM232855 | F: 5'-TCGACAATTTGCATGCTA CG-3' R: 5'-AAATCCTCACCTTCCCAA CC-3' | 20/20 | 54.9 | 187–197 | 200–210 |
| 8 | <i>Sa</i> STPS | HM232858 | F: 5'-AGATGACATTTATGATGC TTATGG-3' R: 5'-TCATTGTAGAAATCCAAA ACTCCTT-3' | 24/25 | 52.8 | 144–154 | 150–170 |
| 9 | <i>Sa</i> monoTPS | JF746807.1 | F: 5'-TATGGAAAATCAAAAAGT GCCTATTTCTTCTGT-3' R: 5'TTCAAACAATCATGGGAT CTATAAGTAGGGATGCAATG-3' | 34/38 | 60.8 | 178–188 | 190–210 |
| 10 | <i>Ssp</i> isesquiTPS | HQ343282.1 | F: 5'-TGGAAAATCAAAAAGTGC CTATTTCTTCTGTCC-3' R: 5'-GGGATCTATAAGTAGGGA TGC-3' | 33/21 | 59.3 | 183–193 | 185–195 |
| 11 | <i>Sa</i> CytP450 | NM101178.4 | F: 15'-TAGGTTACACCCGGCACT TC-3' R: 5'-GGACAAATCCTTCGACCA GA-3' | 20/20 | 59.4 | 240–250 | 220–240 |
| 12 | <i>Sa</i> GPS1 | AT2G34630 | F: 15'-CAGCAGAAGTTGCCGTGT TA-3' R: 15'-TCTGACAACGATCCCTTT CC-3' | 20/20 | 61.0 | 205–215 | 210–220 |
| 13 | <i>Sa</i> GGPS1 | AT4G36810 | F: 5'-TCGGAGGTGAAGAATCAA CC-3' R: 5'-GCTCGAAAGCGAAAGAGA GA-3' | 20/20 | 61.0 | 280–300 | 260–280 |
| 14 | <i>Sa</i> HMG-CoA | AT1G76490 | F: 5'-ACAAGAAGGCTTCGGCTG TA-3' R: 5'-GCGCTGACAATATTGCTT GA-3' | 20/20 | 61.0 | 150–180 | 170–200 |

Table 1 (continued)

| Sl. no. | Primer name | Accession number | Primer sequence 5'–3' | No. of bases | Tm °C | Expected amplification range (bp) | Observed amplification range (bp) |
|---------|-------------|------------------|---|--------------|-------|-----------------------------------|-----------------------------------|
| 15 | ACT | HM232850 | F: 15'-GTATTGTGTTGGACTCTG GTGATGGTGT-3' R: 15'-GATGGATCCTCCAATCCA GACACTGTA-3' | 28/27 | 60.3 | 460–470 | 450–480 |

SaFDSF farnesyl diphosphate synthase, *SaFDSE* E, E, Farnesyl diphosphate synthase, *SaSSF* mutant *Santalum album* santalene synthase, *SaSSE* santalene synthase, *SaDSX* 1-Deoxy-D-xylulose5-phosphate synthase, *SaFPPS* farnesyl pyrophosphate synthase, *SaMTPS* monoterpene synthase, *SaSTPS* sesquiterpene synthase *Santalum album*, *SasesquiTPS* sesquiterpene synthase *Santalum album*, *SspisesquiTPS* sesquiterpene synthase *Santalum spicatum*, *SaCYTP450* cytochrome P450 synthase, *SaGPSI* geranyl pyrophosphate synthase, *SaGGPSI* geranyl geranyl pyrophosphate synthase, *SaHMG-CoA* reductase; 3-Hydroxy-3-methylglutaryl-CoA reductase, *Act* actin (+regulator of PCR reactions)

Table 2 Major volatile metabolites variations in selected *S. album* samples

| SI no. | Sample no. | α -Santalol | <i>Trans</i> - β -Santalol | Santalol E- <i>cis</i> -epi- β | Bargamotol, <i>trans</i> Z- α |
|--------|------------|--------------------|----------------------------------|--------------------------------------|--------------------------------------|
| 1. | S1 | 59.30 | 29.21 | 4.46 | 3.45 |
| 2. | S2 | 49.52 | 32.76 | 3.94 | 6.66 |
| 3. | S3 | 57.01 | 26.60 | 3.89 | 10.41 |
| 4. | S4 | 55.59 | 31.94 | 4.24 | 5.81 |

from 1 to 11 Burn-in MCMC iteration setting were 50,000 and 100,000, respectively. Allele frequencies were correlated for each value of K , 10 replicate simulations using a model allowing for admixture used correlated allele frequencies. The appropriate value for ΔK was estimated from $\ln P(K)$ value described by Rosenberg et al. (2001) and based on the second-rate order rate of change of the likelihood (ΔK) described by Evanno et al. (2005). Data from each sample analysis were combined and ΔK statistics calculated using online software Structure harvester (<http://taylor0.biology.ucla.edu/structureHarvester/>) (Earl and Vonholt 2012).

Principal component analysis (PCA) and multiple regression analysis

PCA is a prominent development in genotypes collections from different geographical locations that allows a better understanding of the entirely different populations. PCA was executed for all the 177 accessions with selected genic SSR markers using Minitab v.18. It used to represent genetic relations among the accessions from different geographical locations. Statistical analysis investigating the relationship between oil yielding and genetic structure data was conducted using the statistical tools SPSS (version 2015). Linear models (multiple linear regression for dependent variables) were used to test whether oil biosynthesizing gene variables were related to high and low oil yielding parameters. Stepwise multiple regression analysis was attempted to

delineate the correlation of genic marker data (independent variables) with oil concentration (dependent variable) for all the *S. album* accessions.

Results

Estimation of core size, heartwood percentage, and oil content

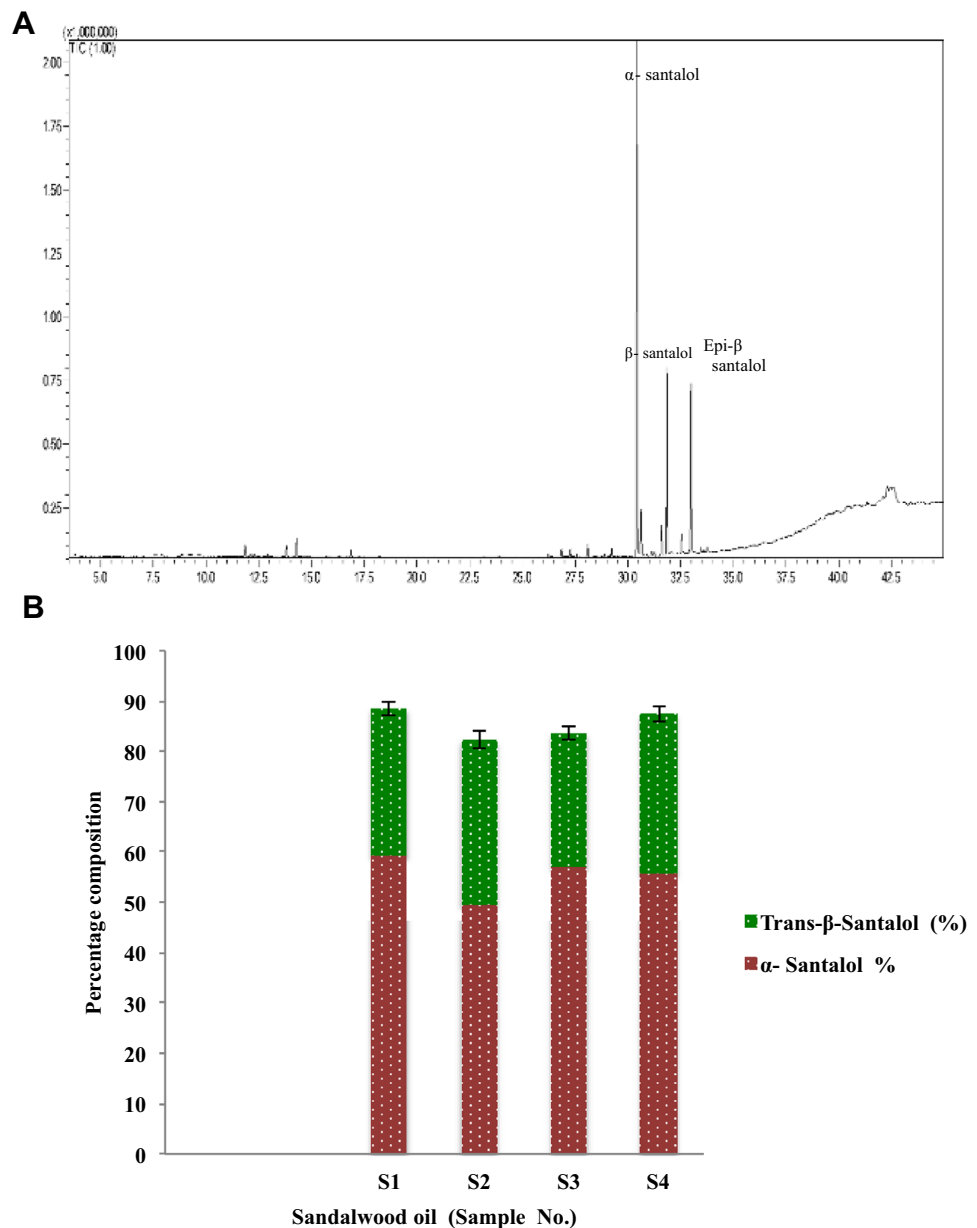
Overall core size of all 177 accessions of *S. album* from three states ranged from tree no 9 (4.4 cm) in Kodada Telangana to Marayur location 1 Kerala tree no 1 (19.1 cm) with variance 27.61% and mean of 9.74 cm (Table S3). Overall Heartwood size from 14 populations of 177 *S. album* accessions ranged from IFB Hyderabad (0.0) to Kerala Marayur location 1 Tree no 1 (17.3 cm) with variance 85.25% and average 3.77 cm (Table S3).

The oil concentration estimated based on optical density through UV spectrophotometer showed R^2 value of 0.90 (Fig S4). Overall total heartwood oil % in 14 populations of three states ranged from IFB Hyderabad (0.0) to IWST Bangalore tree no 5 (5.96%) followed by variance 73.12% and mean of 2.27. The lowest oil concentration was found in (0.0) and highest oil concentration was obtained in (Table S3).

Qualitative GC–MS analysis

Among the 177 accessions, four samples were selected based on the higher oil yield, and for those samples, qualitative analysis was carried out through GC–MS analysis. Variability in the essential oil constituents was estimated by percentage peak area (% A) in GC–MS chromatogram. Highest peak shows α -Santalol, followed by *Trans*- β -Santalol, E-*cis*-epi- β Santalol, and *trans* Z- α Bargamotol (Fig. 2a). The above four selected samples, high oil yielding samples, were showing variation in α -Santalol 49.52–59.30%, *Trans*- β -Santalol, 29.21–32.76%, Santalol E-*cis*-epi- β 3.89–4.46% and Bargamotol, and *trans* Z- α ranges between 1.41 and 3.45%,

Fig. 2 a GC–MS total ion chromatograms of a trade *Santalum album* essential oil using a 30 m × 0.25 mm i.d. × 0.25 (i. d) m RTX-WAX column. Peak identification: (1) α -Santalol, (2) β -Santalol, (3) E-*cis*-epi- β Santalol, and (4) *Trans Z*- α -Bargamol. **b** Percentage composition of α -Santalol and Z- β -Santalol in trade sandalwood essential oils from *Santalum album*; results determined using GC–MS



respectively (Table 2). The percentage composition of Santalols in the essential oils from the above samples, sample no 1 and sample no 4 met the ISO (2002) specification for α -Santalol and β -Santalol (up to 90%) (Fig. 2b).

Genetic diversity analysis

The analysis of total 177 *S. album* accessions using 14 genic loci revealed a moderately high polymorphism. In DXS markers, gradient polymorphic bands were obtained and ranged between 250 and 280 bp in which high oil yielding clones were 270–280 bp ranges, whereas, in low oil yielding clones, it was low base pair (250–220). SaFDSF and Sesquiterpene synthase was conserved and monomorphic in

nature. MTPS and STPS were showing bi-allelic polymorphism in high oil yielding clones, while in low oil yielding genotypes were monomorphic. FFPS primers were showing highly polymorphic and heterozygous in the selected genotypes. In SaSSF, i.e., Santalene synthase produced gradient polymorphic bands, whereas SaFDSE showed clear polymorphic bands with the range of 150–200 bp. The genes, which involve in high oil biosynthesis, were heterozygous allelic form (Fig. 3). The number of alleles (N_a) was varying from (6.0–9.0) SaSSE and SaFDSF with an average 7.28. The selected 14 genic SSR markers of *S. album* among populations of three states, locus wise N_a and N_e ranged from 6.0–9.0 and 4.82–6.88 with an average value of 7.28 and 5.89 respectively. The H_o and H_e ranged from 0.01–0.27 and

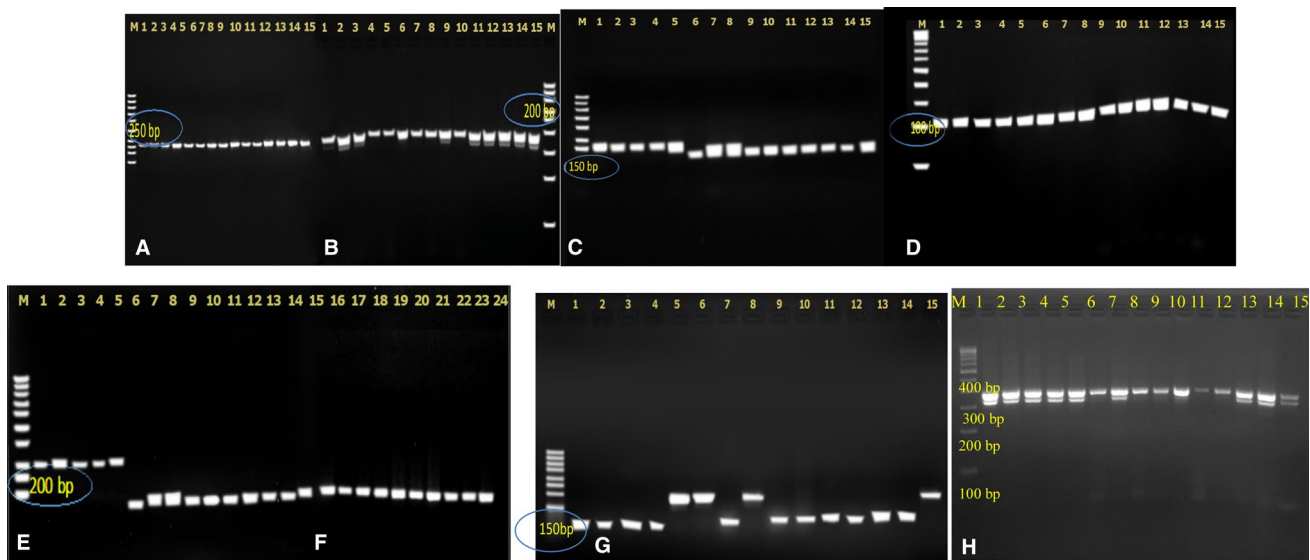


Fig. 3 Amplification profile of *S. album* L. accessions using oil biosynthesis genic SSR markers. **m** Ladder (50 bp. Fermentas Thermofisher pvt. ltd..) **a** *SaDSX* 1-Deoxy-D-xylulose5-phosphate synthase. **b** *SaCytochrome 450* cytochrome P450 synthase *S. album*. **c** *SaMonoSTPS* monoterpene synthase *Santalum album* L. **d** *SaFPPS*

farnesyl pyrophosphate synthase. **e** *SaSSF* mutant *S. album* santalene synthase. **f** *SaGPS1* geranyl pyrophosphate synthase *S. album*. **g** *SaFDSE* E, E, Farnesyl diphosphate synthase. **h** *SaMTPS* monoterpene synthase

0.80–0.88 with the mean value of 0.06 and 0.85 respectively (Table 3). The F ranged per loci 0.56–1.0 with an average 0.922. The PIC value was observed per loci ranged between 0.77–0.89 with the mean value of 0.92 and the Shannon information index ranged between 1.63–1.92 with the mean value of 1.83 (Table 3). High expected heterozygosity was observed with the selected genic SSR markers among fourteen populations of Southern India (Fig. S5). Overall these selected genic SSR markers were considered to be highly informative due to high PIC 0.5 value (Table 3).

Among fourteen populations of *S. album* from three states in southern India, populations wise N_a and N_e per population ranged between 5.83–10.41 and 4.94–7.2 with an average value of 7.28 and 5.89 respectively. The H_o and H_e per population ranged between 0.0–0.15 and 0.83–0.88 with the mean value of 0.06 and 0.85 respectively (Table 4). The F per population ranged between 0.76–1.00 with the mean value of 0.93. The Shannon information index per population ranged between 1.68–2.14 with the mean value of 1.83 and the PIC ranged from 0.77–0.93 with an average value of 0.88 (Table 4).

F varied Suryapet (0.84) to Marayur L2 (1.0) with the mean value of 0.93. I ranged from Dharwad (1.69) to Kodada (2.14) with the mean value of 1.83. PIC value ranged from Marayur Kerala (0.77) to (IFBHyd Telangana)-0.93 (Kodada Telangana-IWST Bangalore) with an average of 0.88 (Table 4).

Pairwise population matrix by Nei's unbiased genetic identity revealed that, at the population level, genetic

similarity between selected populations ranged from 0.0393 (Marayur location 2 Kerala-Dharwad Karnataka) to 0.787 (Khammam Telangana-IWST Bangalore, Hassan Karnataka-Kodada Telangana state) using 14 genic markers (Table 5). The above result revealed that there was high exchange of gene pool between Karnataka and Telangana populations.

Genetic differentiation and F statistics

The AMOVA method was carried out to partition the genetic differentiation among all 14 populations and the state-wise (Karnataka, Kerala, and Telangana). Analysis revealed that major portion of genetic variation resided among the individuals (90, 83, 88, and 84%) are much higher than the variation among the populations (3, 4, 5, and 1%) and within the individuals (7, 13, 7, and 15%). All variance components were found highly significant ($P=0.001$). When populations were analyzed state-wise-based, among the population, variation increased up to 5% which were obtained in Kerala populations. Distribution of genetic variation for the 14 genic markers of interest revealed that most of the total variance is attribute to genetic differentiation among individuals of *S. album*. A degree of genetic differentiation F_{ST} value among all the populations was 0.004 and state-wise (0.08, -0.03, and 0.06) which indicated that the lowest genetic differentiation was found in Kerala populations and the highest genetic differentiation was obtained in Karnataka followed by Telangana. The gene flow was 58.40 individual per generation among all the populations. N_m

Table 3 Diversity parameters of the entire collection for the genic SSR markers used: number of alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), fixation index (F), Shannon information index (I), and polymorphic information content (PIC)

| SI no. | Locus | Na | Ne | Ho | He | F | I | PIC |
|--------|--------------|------|------|------|------|------|------|------|
| 1. | SaFDSF | 8.69 | 6.88 | 0.10 | 0.88 | 1.00 | 1.92 | 0.89 |
| 2. | SaFDSE | 8.29 | 6.05 | 0.27 | 0.85 | 0.56 | 1.84 | 0.89 |
| 3. | SaSSF | 6.79 | 5.35 | 0.01 | 0.84 | 0.99 | 1.85 | 0.89 |
| 4. | SaSSE | 6.43 | 4.82 | 0.12 | 0.80 | 0.92 | 1.99 | 0.89 |
| 5. | SaDXS | 7.21 | 6.13 | 0.04 | 0.86 | 0.95 | 1.63 | 0.79 |
| 6. | SaFPPS | 6.36 | 5.24 | 0.05 | 0.84 | 0.99 | 1.72 | 0.89 |
| 7. | SaMTPS | 6.07 | 4.98 | 0.02 | 0.83 | 0.98 | 1.89 | 0.89 |
| 8. | SaSTPS | 7.43 | 6.20 | 0.02 | 0.86 | 1.00 | 1.99 | 0.89 |
| 9. | SamonoTPS1 | 7.43 | 5.93 | 0.04 | 0.86 | 0.94 | 1.68 | 0.89 |
| 10. | SasesquiTPS2 | 7.93 | 6.84 | 0.12 | 0.88 | 0.76 | 1.85 | 0.89 |
| 11. | SaCytP450 | 7.50 | 6.32 | 0.08 | 0.87 | 0.94 | 1.90 | 0.89 |
| 12. | SaGPS1 | 7.29 | 5.96 | 0.01 | 0.85 | 0.99 | 1.73 | 0.89 |
| 13. | SaGGPS1 | 6.59 | 5.57 | 0.13 | 0.78 | 0.85 | 1.64 | 0.77 |
| 14. | SaHMGC0-A | 7.85 | 6.85 | 0.15 | 0.85 | 0.75 | 1.73 | 0.86 |
| Total | | 7.28 | 5.89 | 0.06 | 0.85 | 0.92 | 1.83 | 0.89 |

Table 4 Mean and standard error of Na, Ne, Ho, He, F, and I for overall populations using genic SSR markers

| Population | N | Na | | Ne | | Ho | | He | | F | | I | | PIC |
|-------------|-----|-------|------|------|------|-------|------|------|------|------|------|------|------|------|
| | | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE | Mean | SE | |
| KodaL1 TS | 48 | 10.41 | 0.06 | 7.92 | 0.98 | 0.135 | 0.29 | 0.88 | 0.03 | 0.86 | 0.05 | 2.14 | 0.29 | 0.93 |
| KodaL2 TS | 46 | 9.66 | 0.05 | 6.99 | 0.18 | 0.15 | 0.23 | 0.86 | 0.03 | 0.84 | 0.05 | 2.05 | 0.23 | 0.89 |
| Khamm TS | 20 | 6.75 | 0.96 | 5.40 | 0.94 | 0.20 | 0.15 | 0.85 | 0.03 | 0.76 | 0.04 | 1.77 | 0.15 | 0.90 |
| IFB Hyd TS | 20 | 6.58 | 1.44 | 5.48 | 1.50 | 0.11 | 0.26 | 0.84 | 0.06 | 0.90 | 0.00 | 1.75 | 0.26 | 0.83 |
| MaraL1 KL | 24 | 7.08 | 1.08 | 5.63 | 1.12 | 0.02 | 0.17 | 0.85 | 0.03 | 0.98 | 0.05 | 1.82 | 0.17 | 0.89 |
| MaraL2 KL | 16 | 5.83 | 0.93 | 5.16 | 0.97 | 0.00 | 0.18 | 0.85 | 0.04 | 1.00 | 0.04 | 1.68 | 0.18 | 0.83 |
| MaraL3 KL | 20 | 6.50 | 1.24 | 5.45 | 1.36 | 0.00 | 0.21 | 0.84 | 0.04 | 1.00 | 0.04 | 1.75 | 0.21 | 0.77 |
| IWST KA | 40 | 8.50 | 1.97 | 6.77 | 1.63 | 0.00 | 0.28 | 0.86 | 0.05 | 0.99 | 0.06 | 1.98 | 0.28 | 0.89 |
| Nela KA | 20 | 6.83 | 0.93 | 5.84 | 1.05 | 0.02 | 0.17 | 0.86 | 0.03 | 0.95 | 0.04 | 1.82 | 0.17 | 0.89 |
| Hassan KA | 20 | 6.83 | 1.58 | 5.69 | 1.31 | 0.07 | 0.22 | 0.85 | 0.04 | 0.94 | 0.04 | 1.80 | 0.22 | 0.89 |
| Chenna KA | 20 | 7.25 | 1.42 | 6.08 | 1.39 | 0.06 | 0.20 | 0.87 | 0.03 | 0.95 | 0.03 | 1.87 | 0.20 | 0.88 |
| Dharwad KA | 20 | 6.16 | 0.93 | 4.94 | 0.86 | 0.00 | 0.16 | 0.83 | 0.03 | 1.00 | 0.05 | 1.69 | 0.16 | 0.89 |
| Shiva KA | 20 | 6.83 | 1.19 | 5.73 | 1.34 | 0.00 | 0.20 | 0.85 | 0.04 | 1.00 | 0.04 | 1.81 | 0.20 | 0.87 |
| Gotti KA | 20 | 6.75 | 1.54 | 5.40 | 1.76 | 0.06 | 0.28 | 0.83 | 0.08 | 0.87 | 0.04 | 1.75 | 0.28 | 0.88 |
| Overall pop | 354 | 7.28 | 1.00 | 5.89 | 1.17 | 0.06 | 0.21 | 0.85 | 0.04 | 0.93 | 0.07 | 1.83 | 0.21 | 0.88 |

varied state-wise Karnataka—42.86; Kerala—13.50; and Telangana—20.98, respectively. This result indicated that Kerala has a lower level of gene flow among the populations than the other provenances (Table S6). F_{IS} of 14 populations with selected *S. album* accessions ranged from (0.76 to 1.00) with average range 0.923 (Table S7) indicated the substantial heterozygosity by selected genic primers. F_{ST} , which determines the population differentiation coefficient, measures the degree of genetic differentiation among the populations. The F_{ST} ($P=0.001$) of the selected primers for the overall populations were varying from (0.0053 to 0.1243) and in natural populations F_{ST} was 0.026 which indicated 2.6% genetic diversity in natural populations, whereas, in

plantation areas, F_{ST} was 0.006 which indicated 6% diversity among the natural populations. The average F_{ST} was 0.008, which signified 8% genetic variation among the populations. Gene flow obtained due to migration of pollens or genetic drift, habitat fragmentation in different populations. The estimates of gene flow among the populations of Telangana ($Nm=20.98$); in which Kodada individuals $Nm=9.81$ were highest. In Karnataka populations, $Nm=42.866$ in which highest gene flow was found in IWST Bangalore $Nm=7.39$ followed by Chennarayapatna $Nm=5.50$. This result indicated that, in natural population, gene flow was higher than the plantation populations in Karnataka. In Kerala, $Nm=13.50$ in which highest Nm was obtained in Marayur

Table 5 Pairwise Nei's unbiased genetic distance table between *S. album* populations using genic microsatellite markers

| | Kod1TS | Kod2TS | KhmTS | IFBTS | MKL1 | MKL2 | MKL3 | IWKA | NelKA | DharKA | HasKA | ChenaKA | GoKA | SHIKA |
|--------|--------|--------|-------|-------|------|------|------|------|-------|--------|-------|---------|------|-------|
| Kod1TS | *** | | | | | | | | | | | | | |
| Kod2TS | 0.68 | *** | | | | | | | | | | | | |
| KhmTS | 0.63 | 0.64 | *** | | | | | | | | | | | |
| IFBTS | 0.69 | 0.69 | 0.57 | *** | | | | | | | | | | |
| MKL1 | 0.64 | 0.65 | 0.47 | 0.58 | *** | | | | | | | | | |
| MKL2 | 0.57 | 0.60 | 0.44 | 0.50 | 0.55 | *** | | | | | | | | |
| MKL3 | 0.55 | 0.53 | 0.42 | 0.41 | 0.49 | 0.52 | *** | | | | | | | |
| IWSKA | 0.65 | 0.75 | 0.78 | 0.59 | 0.66 | 0.53 | 0.65 | *** | | | | | | |
| NelKA | 0.54 | 0.53 | 0.43 | 0.49 | 0.58 | 0.56 | 0.61 | 0.65 | *** | | | | | |
| DharKA | 0.47 | 0.51 | 0.47 | 0.42 | 0.74 | 0.39 | 0.58 | 0.55 | 0.51 | *** | | | | |
| HasKA | 0.48 | 0.49 | 0.78 | 0.54 | 0.44 | 0.47 | 0.44 | 0.52 | 0.53 | 0.62 | *** | | | |
| ChenaK | 0.54 | 0.58 | 0.51 | 0.56 | 0.54 | 0.54 | 0.53 | 0.62 | 0.57 | 0.51 | 0.51 | *** | | |
| GoKA | 0.56 | 0.61 | 0.46 | 0.51 | 0.52 | 0.48 | 0.55 | 0.59 | 0.53 | 0.49 | 0.48 | 0.63 | *** | |
| SHIKA | 0.60 | 0.58 | 0.57 | 0.55 | 0.54 | 0.53 | 0.48 | 0.56 | 0.55 | 0.44 | 0.47 | 0.54 | 0.48 | *** |

location 3 $N_m = 6.28$. Gene flow N_m of the selected markers was ranged from (1.76 to 2.67); the gene flow was ranged from 3.76 in natural populations and 2.24 in plantations. N_m was higher in stretch of Karnataka populations 5.025 followed by Telangana 4.27; on the other hand, it was found low in a Kerala state 1.89 (Table S8).

Cluster analysis, Structure, and Principal component analysis (PCA)

On the 0.05 level of genic marker Jaccard's similarity coefficient and UPGMA of *S. album* accessions were clustered into two major clusters. Cluster I subdivided into two groups in which one consists Khammam, IFB Hyderabad

Telangana, and Marayur 1 Kerala. In group two: Marayur 2, Marayur 3, Dharwad, IWST, Nelamangala Gottipura and Shimogga were included. In cluster II: Marayur location 3, Hassan and Chennarayapatna Karnataka were included. The dendrogram supported the grouping and showed two major groups of *S. album* accessions in which cluster representing the Telangana state genotypes was having intermixing of accessions (Fig. 4). The UPGMA dendrogram revealed clusters which were favored with the grouping by the structure and principal component analysis.

Population structure analysis was performed on the A set and two possible sub populations were identified. To confirm this result, the number of suggested ΔK was plotted against the calculated obtained from structure harvester software.

Fig. 4 UPGMA dendrogram for 14 populations of *S. album* based on Nei's genetic distance method showing clustering of all accessions into two major groups respective to their geographical locations

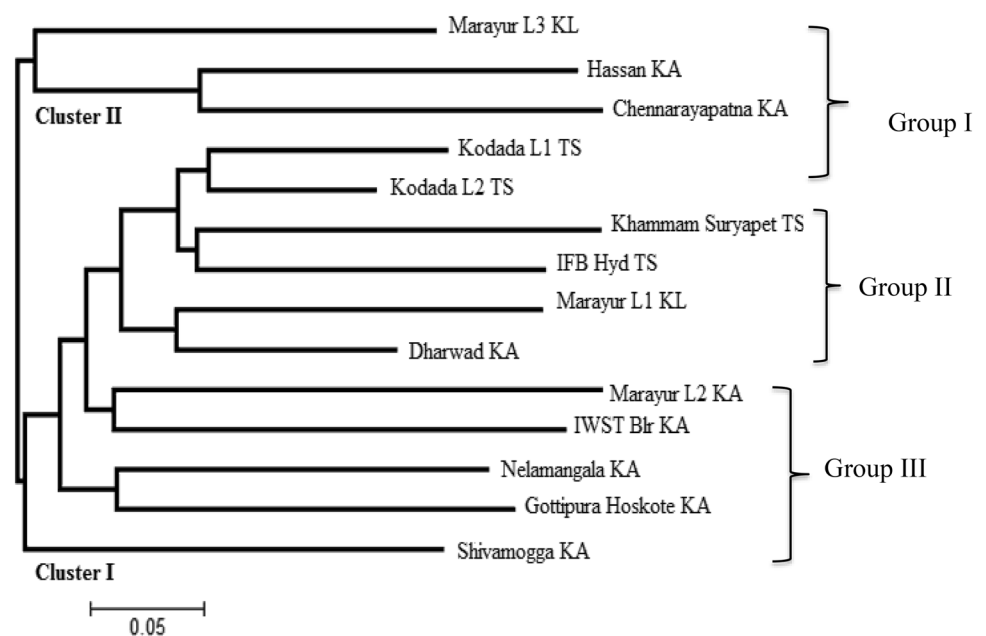
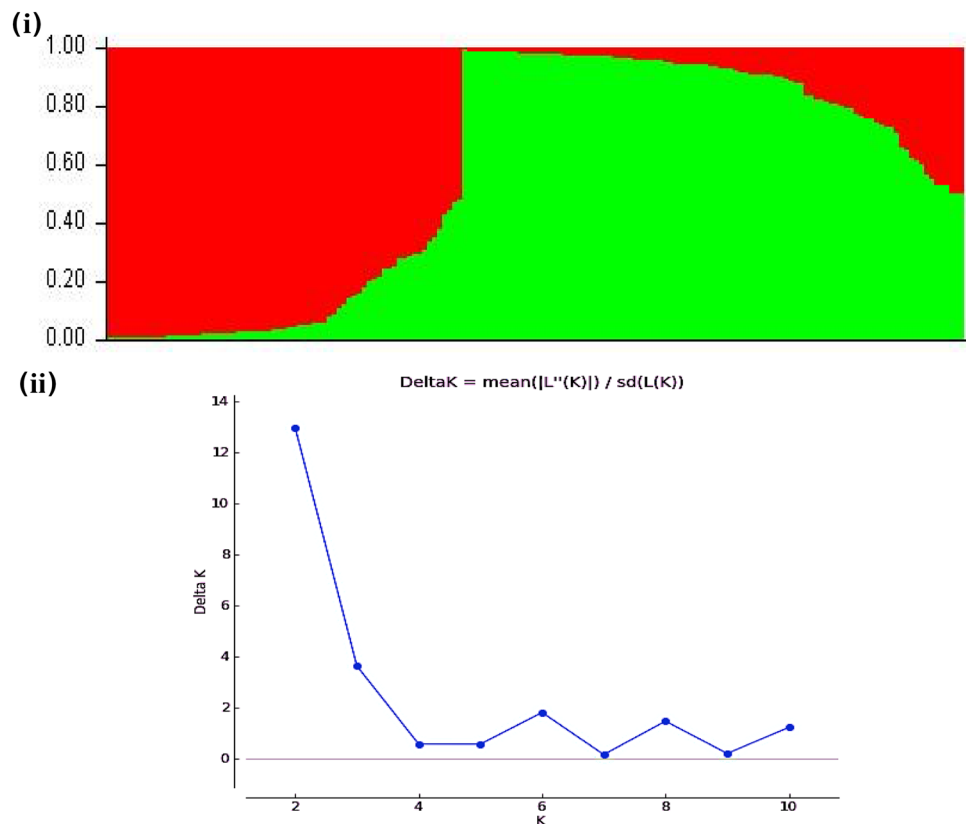


Fig. 5 Analysis of population structure using 14 genic SSR markers (i) Bayesian model-based estimation of overall population structure of 177 genotypes ($K=2$). The Y-axis is the subpopulation membership and the X-axis is the genotypes (ii). Estimated LnP (D) of possible clusters from (2–11) ΔK

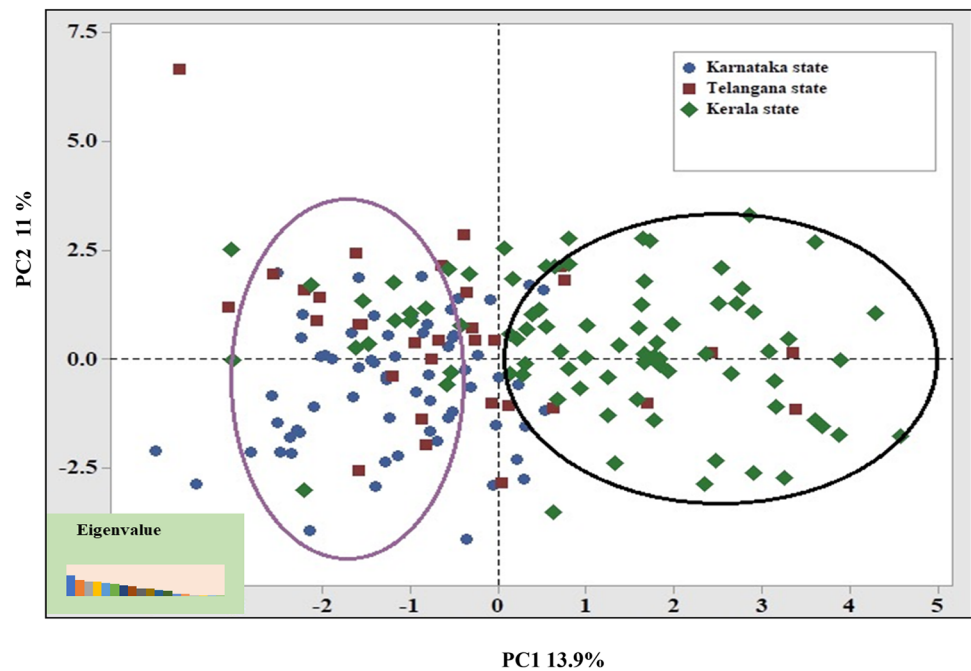


The highest value was shown at $\Delta K = 2$. Therefore, structure clustered entire populations into two distinct genetic structure groups with inferred clusters: (1) 0.439 (red lines) 37.4% and (2) 0.561 (green lines) 53.80%. The estimated Ln probability of all data was < 0.9 and predominantly admixture was found in Telangana state samples (Fig. 5). When we analyzed the state-wise structure data, the clustering largely supported the grouping of dendrogram. In Karnataka at $\Delta K = 2$, mean probability < 0.8 with inferred clusters (1) 0.871 (43.42%) (red lines) and (2) 0.876 (51.31%) (green lines) (Fig. S9a). Kerala and Telangana at $\Delta K = 3$ in populations inferred cluster value (1) 0.337 (33.33%) (red lines), (2) 0.349 (33.33%) (green lines), and (3) 0.313 (26.66%) (blue lines) revealed that genetic structure of Telangana and Kerala populations was admixtures (Fig. S9b). In Telangana at $\Delta K = 3$, mean probability > 0.7 in populations inferred cluster value (1) 0.298 (27.69%) (red lines), (2) 0.341 (30.76%) (green lines), and (3) 0.361 (38.0%) (blue lines) (Fig. S9c). Structure and cluster analysis indicated the entire germplasm selected in this study, basically belongs to two genetic stocks for the (Wright 1978). The essential oil concentration profile across the three regions in southern India was reflected the genetic profiles with the provenances. The oil concentration was more similar in Karnataka and Kerala than that of the Telangana populations, but the genetic structure was more polymorphic in Karnataka regions followed by Telangana and Kerala.

The PCA exhibits similarity to clustering and population structure analysis which represented the scatter plot based on 14 selected genic markers showed much more vague borders among the three states of India. PCA with 177 individuals of sandalwood accessions clearly segregated into two groups in which Karnataka (blue dots) and Kerala (green dots) individuals and Telangana (red dots) populations were being mixed represented the overlapping in Karnataka group with few out layered individuals (Fig. 6). Telangana populations were overlapped and indicated that the genotypes of Telangana might be originated from Kerala or Karnataka populations. Eigenvalues of each principal component were higher than 1 excluding (PC11, PC12, PC13, and PC14) showed high variation and executed the importance for the selection of the individuals. The PC1 (Component1) 13.9% and PC2 (Component2) 11.0% of the combined variation 89.6% with the cumulative variation 66.27% found in selected populations, respectively (Table S10).

Furthermore, to understand the origin of Telangana mixture, a secondary level of PCA analysis was carried out between the selected states, i.e., (1) Kerala–Telangana; (2) Karnataka–Kerala; and (3) Karnataka–Telangana, to divide the group within genetic structure similarity. Two distinct clustering was observed in Kerala (red dots) and Telangana (blue dots) with the PC1 (component1) 17.8% and PC2 (component2) 14.3% (Fig. S11, i). Within Kerala and Karnataka, populations also show clustered in two discrete

Fig. 6 Discriminant analysis of Principal Component analysis for selected genotypes collected from the 14 populations of three states in India. Colored vertical graph represented Eigenvalues



groups with PC1 (component 1) 15.8%, PC2 (component2) 10.85% (Fig. S11, ii). However, there was no distinct group observed between Karnataka and Telangana individuals with PC1 (component1) 15.80% and PC2 (component2) 11.84% (Fig S11, iii). These results signified the major variation for all the accessions to identify the populations found across the study area belong mainly falls into the clusters belonging to Kerala and Karnataka meaning most of the planting materials originated from these sources. Telangana *S. album* accessions overlapped both of these states accessions; the reason that the samples from Telangana originally belong to Karnataka and few accessions belongs to Kerala state.

Multiple regression analysis further revealed that the concentration of Sandalwood oil and heartwood content is correlated ($r^2=0.426$, $P=0.001$). The oil concentration and optical density showed significant correlation ($r^2=0.90$, $P=0.001$), but the GC–MS results exhibited variation in the concentration of α -santalol, β -santalol, and some traces of other components (correlation $r^2=0.45$, $P=0.001$). The genic SSR marker data and oil concentration results were not significantly correlated ($r^2=0.23$, $P=0.001$) (Table 6).

Discussion

The present study describes the first determined attempt to establish the morphological (quantitative and qualitative) and genetic diversity of *S. album* using genic SSR markers. A high morphological variation was observed in *S. album* accessions. Heartwood variation ranged from 0.0 to 17.9 with an average 3.77 cm (Table S3). The result shows a significant correlation between heartwood oil and heartwood content (Table 6). Likewise in the previous studies, a considerable variation was found in *S. album* tree diameter 8.12–15.44 cm and heartwood diameter 1.34–9.55 cm and heartwood oil % varied from 0.39–2.64. The result shows a positive correlation between tree diameter and heartwood diameter ($r^2=0.63$, $P<0.01$). However, tree diameter and heartwood oil percentage was not correlated ($r^2=0.095$, $P>0.05$) (ArunKumar et al. 2011). In the present study high oil % was found in hexane solvent method and ranged was observed from 0.0 to 5.96% with average 2.27% and CV 73.12%, respectively (Table S3). A significantly positive correlation generated using linear regression data ($R^2=0.90$)

Table 6 Multiple regression results for oil concentration and heartwood %, optical density, α - and β -santalol, and genic SSR markers of *S. album* with all selected accessions from the 14 populations of three states in India

| Correlation | Oil concentration | | Heartwood % | | Optical density | | α -Santalol, β -santalol | | Genic SSRs | |
|-------------|-------------------|-----------|-------------|-----------|-----------------|-----------|---------------------------------------|-----------|------------|-----------|
| | R^2 | P value | R^2 | P value | R^2 | P value | R^2 | P value | R^2 | P value |
| Oil conc. | 0.00 | 0.00 | 0.726 | 0.001* | 0.90 | 0.001* | 0.451 | 0.001 | 0.23 | ns |

Significant differences are indicated * $P<0.05$, ** $P<0.01$, *** $P<0.001$; ns not significant

(Fig. S4). Similarly, Hettiarachchi et al. (2010) evaluated the *S. album* oil extraction by four different solvent methods in which hexane solvent extraction method showed high oil % content ranged from 0.46 to 2.16% with an average 1.21%. Result of the % oil content was showing considerable variation (CV = 74.16%) and a significant positive correlation ($R^2 = 0.99$). GC–MS reveals that the concentration of the heartwood oil was not significantly correlated to components (α - and β -santalol) of the oil content ($r^2 = 0.45$, $P = 0.001$) (Table 6). Among 177, four high and low oil yielding samples were showing variation in α -Santalol 49.52–59.30%, *Trans*- β -Santalol, 29.21–32.76%, *E-cis-epi*- β Santalol 3.89–4.46%, and *trans Z*- α bargamolol varied between 1.41 and 3.45%, respectively (Table 2). Likewise, (Doran et al. 2005) qualitatively analyzed *S. album* and *S. yasi* solvent extracted oil by gas chromatography to estimate the % of α and β santalol of total oil. In *S. yasi*, α santalol 30.5% and β santalol 15.5%, whereas, in *S. album*, it was found 55.5% and 22.21%, respectively. In our study, α - and β -Santalol content was high and met the ISO (2002) specification for α -Santalol and β -Santalol (up to 90%) (Fig. 2b) which revealed the high quality of Sandalwood oil present in the respective samples.

Genic markers are derived from transcripts (reside in the functional genes) and they are useful for determining the phenotypic trait, i.e., heartwood oil in sandalwood and functional diversity in populations (Varshney et al. 2005). These markers are valuable because of their higher level of transferability to related species, and they can be used as anchor markers for comparative mapping and evolutionary studies. The genetic diversity coefficient measured by genic markers was higher than SSR markers and cluster results generated using these markers showed some differences, but the results are more accurate, because they are produced by functional genes (Lin et al. 2012). The average Na, Ho, He, and I determined by genomic SSRs is higher than genic SSRs, but cluster analysis indicated that the EST-SSR capacity of genotypic identification is more accurate than that of genomic SSRs and suitable for the study of genetic diversity and genetic differences between the species (Song et al. 2012).

In this present study, high variation was observed in product size range of each primers and showing high polymorphism (Table 1). However, the genic SSRs show low base pair differences and low polymorphism than genomic SSR markers in *Abies alba* Mill (Postolache et al. 2014).

The number of alleles and effective number of alleles was high in the selected populations per loci (Table 3). Likewise, 12 genic SSR markers showed polymorphism in castor genome. The number of alleles was ranged two to three with an average of 2.33 (Thatikunta et al. 2016). Genic SSR markers, Farnesyl diphosphate synthase (SaFDS), and Santalene synthase (SaSS) gene expressions were high and characterized as chief genes in regulating heartwood oil

biosynthesis in *S. album* (Rani et al. 2013). In our study, these selected genes were showing high number of alleles and higher level of expected heterozygosity. This present study reveals the polymorphism two (*S. album*) santalene synthases from *S. album* (SaSSY), Monoterpene synthase (TPS-a: 1731 bp), and Sesquiterpene synthase (TPS-b: 1680 bp). These genes encode enzymes which catalyze the formation of Farnesyl diphosphate synthase. However, these two gene products were not found in considerable quantity in sandalwood oil (Jones et al. 2011). In this study, expected heterozygosity (He) was ranged 0.78–0.88 with an average 0.85 per loci (Tables 3, 4), which indicated their high level of polymorphic nature. Similarly, the expected heterozygosity ranged from 0.374 to 0.876 in Oil tea tree (*Camellia oleifera*) by 15 genic SSR markers (Jia et al. 2014). In this study, diversity in *S. album* was higher, although measures of heterozygosity and rare alleles were higher in Karnataka and Telangana trees compared with those in Kerala. The oil concentration of the trees in Karnataka and Kerala was more similar despite the being more genetically similar to Telangana populations. Likewise, the average number of effective alleles per locus was ranged between 3.13 in 24 varieties of olives by 46 genic SSR markers (Arbeiter et al. 2017).

In the present study, PIC value ranged 0.74–0.91 with the average 0.88 and represented high level of polymorphism in *S. album* accessions. Similarly, EST-SSR markers evaluated the genetic diversity in *Salix psammophila*, a sandy shrub. The PIC value was varied between 0.386 and 0.88 with an average of 7.4 and shared high genetic diversity (Jia et al. 2016). PIC value ranged between 0.27 and 0.49 in Castor (*Ricinus communis* L), which indicated moderate polymorphism level or genetic diversity in castor genome (Thatikunta et al. 2016). The above results indicated that the *S. album* or other forestry species shows high polymorphism than the agricultural or other crop species. This might be due to cross-pollination and self-incompatibility behavior of *S. album*. In this study, the Jaccard's similarity coefficient ranged from 0.39 to 0.69 (Table 5) and UPGMA revealed that *S. album* accessions were supported the grouping, and showed Karnataka and Kerala major groups in which Telangana genotypes were having intermixing accessions (Fig. 4) Similarly, the similarity coefficient values of the 45 accessions of *Jatropha curcas* with 36 genic SSR markers ranged from 0.55 to 0.92 with the mean of 0.76. UPGMA dendrogram revealed the accessions correlated with their geographic origin (Wen et al. 2010).

Population differentiation statistics also showed a clear separation of individual populations based on their geographical origin in India, indicating high level of gene flow among individuals within a geographical region than across the geographical regions. AMOVA revealed that the major variation was found among *S. album* individuals (90%). Only

7.0% of the total variability was found in within individuals and 3.0% variability was detected in among the populations with $F_{ST}=0.023$, $P=0.001$ (Table S6), which indicated that the population is sharing their genetic material through high level of breeding in among individuals. Populations in the Eastern dry zone of Karnataka showed greater level of diversity and less population differentiation than the north west dry zone, i.e., Telangana state and Semi-dry forest loam that belongs to Kerala populations. In contrast, Telangana populations showed a relationship between drift and gene flow indicated that the seedlings of Telangana populations belong to Karnataka and Kerala populations. Likewise, population differentiation in 77 individuals of *Eucalyptus tereticornis* by 64 genic SSR markers indicated 1.2% variation among populations and within populations accounting for the majority of variations showing low $F_{ST}=0.012$ (Zhijiao et al. 2016). Analysis of structure infers accessions ancestry from genotypic information. The most likely member of clusters was evaluated considering the plateau criterion and using the ΔK method. The highest value was for $K=2$, indicating two clusters, first containing 70 accessions of Karnataka. The second comprises 20 accessions of Kerala populations. Additional accessions clearly showed mixed ancestry of 67 accessions of Telangana population and 20 accessions of Karnataka and Kerala (membership values lower than 80% in either of these clusters (Fig. 5). In the previous studies, the two genetic clusters of the *Juglans regia* collection fit well with the geographical origin of the accession using genic markers and showed that the genetic structure of *J. regia* populations separated into two main clusters. Each cluster was analyzed independently to explore if a lower structure could be found within a group (Bernard et al. 2018). The PCA results revealed that *S. album* accessions from Karnataka ($n=80$) and Kerala ($n=30$) clustered separately. However, Telangana ($n=67$) accessions scattered among Kerala and Karnataka, and were closer to Karnataka than Kerala (Fig. 6). Similarly (Fu et al. 2014) developed 63 EST-SSR markers for genotyping of 30 accessions of Celery (*Apium graveolens* L.). PCA results revealed that accessions of the same horticultural types clustered together. Accessions of local celery were well separated from those of celery. The middle-type celery accessions were scattered among celery accessions and were closer to celery cluster than local celery cluster. In our study, we found that *S. album* genotypes from three states were clustered together which belong to the same geographical regions except Telangana genotypes was found to be intermixture of Karnataka and Kerala (Fig. 6). In the present study, we found that the oil composition and genetic structure by the genic markers were not significantly correlated in *S. album* (Table 6). Likewise, the association of genetic structure as well as oil composition in *S. spicatum* was independent variables and determined by environmental factors mainly on geographical regions (soil-type factor) of

the populations (Moniodis et al. 2018). In this study, we found that the pattern of sandalwood oil component variation (α -santalol, *trans*- β -santalol, E-*cis*-epi- β bergamotol, and *trans* Z- α santalol) was significantly associated with environmental factors, which might affect the genetic structure variability. Similarly (Shams et al. 2016) showed that the essential oil content of coriander (*Coriandrum sativum* L.) fruits influenced by genetic and environmental factors.

The pattern of genetic profile across the three states was broadly reflected the sandalwood oil concentration. In this study, we observed that the highest concentration of oil was obtained in natural populations of Kerala, but this might be due to old and unknown age of the trees. The genetic structure of Kerala was less differentiated than the Karnataka and Telangana populations. It also revealed that the superior genotypes of *S. album* can be selected at individual level among the populations mostly belongs to Karnataka populations. This study revealed the conservation of genic SSR markers and considerable degree of transferability between the species. Our work, therefore, provides another practical demonstration of the advantage of genic marker in the study of genetic diversity association mapping that can be used in comparing the other population of sandalwood as well as related species.

Conclusion

From the study of *S. album* genotypes of three states (Karnataka, Kerala, and Telangana) from south India, it was concluded that there is a large variability in the oil content and high genetic differentiation in natural and plantation populations of *S. album* accessions. This research also reveals that the Karnataka populations are having high genic polymorphism than Telangana and Kerala state. The above results suggested high degree of genic SSR marker transferability among sandalwood populations and that would be more useful to estimate the genetic diversity studies than the available genomic SSR markers. Thus, the study provides genic SSR markers not only for *S. album*, but also for genetic studies in related species that comprise the important trait, i.e., sandalwood oil gene pool for improvement of sandalwood. The existing variability in genic SSR markers gives way for genetic diversity analysis, marker-assisted selection, functional map construction, QTL mapping, and selection of elite genotypes for future genetic improvement and tree breeding program.

Acknowledgements Authors are thankful to the Director, IWST, Group co-ordinator Research, Head-Genetics and Tree Improvement Division, Dr. D. Annapurna GTI Division, Head-Chemistry of Forest Product Division, Institute of Wood Science and Technology. The authors also wish to thank Hassan DFO, Dharwad DFO, Director, Institute of Forest Biodiversity Hyderabad, State farmers of Suryapet Khammam

Telangana, and their assistance on field trips, tree identification, and sample collection.

Author contributions All authors contributed equally. All authors read, reviewed, agreed, and approved the final manuscript.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

References

- Arbeiter AB, Hladnik M, Jakse J, Bandelj D (2017) Identification and validation of novel EST-SSR markers in olives. *Sci Agric* 74:215–225
- ArunKumar AN, Srinivasa YB, Joshi G, Seetharam A (2011) Variability in and relation between tree growth, heartwood and oil content in sandalwood (*Santalum album*). *Curr Sci* 6:827–830
- Arunkumar AN, Dhyani A, Joshi G (2019) *Santalum album*. The IUCN red list of threatened species 2019:e.T31852A2807668
- Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLoS One* 13:e0208021
- Bisht SS, Hemanthraj KPM (2014) Gas Chromatography–Mass spectrometry (GC–MS) profiling of heartwood oil composition from 15 years old Sandalwood. *Int J Pharm Phytol Res* 6:387–392
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Doran J, Thomson L, Brophy J, Goldsack B, Bulai P, Fakaosi T, Mokoia T (2005) Variation in heartwood oil composition of young sandalwood trees in the South Pacific (*S. yasi*, *S. album* and F1 hybrids in Fiji, and *S. yasi* in Tonga and Niue). *Sandal Res Newsl* 3–4
- Earl DA, Vonholt BM (2012) STRUCTURE HARVESTER: a website and program A for visualizing STRUCTURE output and implementing the Evanno method. *Cons Gen Resour* 4:359–361
- Evanno G, Regnaut S, Goudet J (2005) Detectin the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Fatima T, Srivastava A, Hanur VS, Rao SM (2018) An effective wood DNA extraction protocol for three economic important timber species of India. *Am J Plant Sci* 9:139–149
- Fu N, Wang PY, Liu XD, Shen HL (2014) Use of EST-SSR markers for evaluating genetic diversity and fingerprinting celery (*Apium graveolens* L.) cultivars. *Molecules* 19:1939–1955
- Guimaraes AG, Junior ATDA, Filho EDA, Pena GF, Vittorazzi C, Pereira MG (2018) Population structure and impact of recurrent selection on popcorn EST-SSR markers. *Gen Plant Breed* 40:E35218
- Hettiarachchi DS, Gamage M, Subasinghe U (2010) Oil content analysis of sandalwood: a novel approach for core sample analysis. *Sandal Res Newsl* 1–4
- International Organisation for Standardisation (2002) second ed., ISO 3518
- Jain SH, Angadi VG, Shankaranarayana KH (2003) Edaphic environmental and genetic factors associated with growth and adaptability of sandal (*Santalum album* L.) in provenances. *Sandal Res Newsl* 17:6–7
- Jia B, Lin Q, Zhang L, Tan X, Lei X, Hu X, Shao F (2014) Development of 15 Genic-SSR markers in oil-tea tree (*Camellia oleifera*) based on transcriptome sequencing. *Genet* 46(3):789–797
- Jia H, Yang H, Sun P, Li J, Zhang J, Guo Y, Han X, Zhang G, Lu M, Hu J (2016) *De novo* transcriptome assembly, development of EST-SSR markers and population genetic analyses for the desert biomass willow (*Salix psammophila*). *Sci Rep* 6:39591
- Jones CG, Ghisalberti EL, Plummer JA, Barbour EL (2006) Quantitative co-occurrence of sesquiterpenes; a tool for elucidating their biosynthesis in Indian sandalwood, *Santalum album*. *J Phytochem* 67:2463–2468
- Jones CG, Keeling CI, Ghisalberti EL, Barbour EL, Plummer JA, Bohlmann J (2008) Isolation of cDNA and functional characterization of two multi-product terpene synthase enzymes from sandalwood from sandalwood, Sandalwood, *Santalum album* L. *Arch Biochem Biophys* 477:121–130
- Jones CG, Moniodis J, Zulak GK, Scaffidi A, Plummer JA, Ghisalberti EL, Barbour EL, Bohlmann J (2011) Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase TPS-a and TPS-b subfamilies, including santalene synthases. *J Biol Chem* 286:17445–17454
- Lin Z, Wang Y, Zhang X, Zhang J (2012) Functional Markers for cellulose synthase and their comparison to SSRs in cotton. *Plant Mole Biol Rep* 30:1270–1275
- Misra BB, Dey S (2013) Developmental variations in sesquiterpenoid biosynthesis in East Indian sandalwood (*Santalum album* L.). *Trees* 27:1071–1086
- Moniodis J, Renton M, Jones CG, Barbour EL, Byrne M (2018) Genetic environment parameters show associations with essential oil composition in Western sandalwood (*Santalum spicatum*). *Aust J Bot* 66:48–58
- Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in excel. Genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinfo* 28:2537–2539
- Postolache D, Leonarduzzi C, Piotti A, Spanu I, Roig A, Fady B, Roschanski A, Liepelt S, Vendramin GG (2014) Transcriptome versus genomic microsatellite markers: highly informative multiplexes for genotyping *Abies alba* mill and congeneric species. *Plant Mol Biol Rep* 32:750–760
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genet* 155:945–959
- Purohit PM (2018) Reviving the royal tree *Santalum album* Linn.: Santalaceae. *J Med Plants Stu* 6(2):273–276
- Rani A, Ravikumar P, Reddy MD, Kush A (2013) Molecular regulation of santalol Biosynthesis in *Santalum album* L. *Gene* 527:642–648
- Rao NM, Ravikanth G, Ganeshiah KN, Rathore TS, Shaanker RU (2007) assessing threats and identifying the ecological niche of sandal resources to identify ‘hot-spots’ for in situ conservation in Southern India. *Sandal Proc Natl Semin* 23–31
- Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, Maki-Tanila A, Tixier-Biochard M, Vignal A, Wimmers K, Weigend S (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *G E N* 159:699–713
- Shams M, Ramezani M, Esfahan SZ, Esfahan EZ, Dursun A, Yildirim A (2016) Effects of climatic factors on the quantity of essential oil and dry matter yield of coriander (*Coriandrum sativum* L.). *J Sci Technol* 9:6
- Shankaranarayana KH, Angadi VG, Rajeevalochan KS, Sharma CR, Rangaswamy CR (1997) A rapid method of estimating essential oil content in heartwood of *Santalum album* Linn. *Curr Sci* 72:41–242
- Song YP, Jinag XB, Zhang M, Wang ZL, Bo WH, An XM, Zhang DQ, Zhang ZY (2012) Differences of EST-SSR and genomic markers in assessing genetic diversity in poplar. *For Stud China* 14(1):1–7

- Sreenivasan VV, Shivaramkrishnana VR, Rangaswamy CR, Ananthapadmanabha HS, Shankaranarayana KH (1992) Sandal. ICFRE, Dehradun, p 233
- Subasinghe SMCUP (2013) Sandalwood research: a global perspective. *J Trop For Environ* 3:1–8
- Takezaki N, Nei M, Tamura K (2010) POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Mol Biol Evol* 27(4):747–752
- Thatikunta R, Sankar AS, Sreelakshmi J, Palle G, Leela C, Rani CVD, Shankar VG, Lavanya B, Reddy PN, Dudhe MY (2016) Utilization of in silico EST–SSR markers for diversity studies in castor (*Ricinus communis* L.). *Phys Mol Biol Plants* 4:535–545
- Varshney RK, Graner A, Sorrells ME (2005) Genetic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23:48–55
- Wen M, Wang H, Xia Z, Zou M, Lu C, Wang W (2010) Development of EST-SSR and genomic SSR markers to assess genetic diversity in *Jatropha curcas* L. *BMC Res Notes* 3:42
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA Polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Res* 18:2213–2222
- Wright S (1978) Variability within and among natural populations. *Evol and the Gen of Pop*, 4, University of Chicago Press, Chicago
- Zhijiao S, Heyu Y, Qijie W, Changpin Z, Fagen L, Mei L, Wanhong L, Jianzhong L, Siming G (2016) Genetic diversity and selective loci in *Eucalyptus tereticornis* populations. *Sci Sil Sini* 52(9):40–47