Effect of discrete (individual) and mixed (bulk) genomic DNA on genetic diversity estimates and population structure in Teak (*Tectona grandis* L. f.)

Naseer Mohammad¹, Shashank Mahesh¹, Yogesh Kumar Jain¹, & SA Ansari¹,²,*

¹Genetics and Plant Propagation Division, Tropical Forest Research Institute, Jabalpur-482 021, Madhya Pradesh, India
²Institute of Forest Productivity, Lalgutwa, Ranchi-835 303, Jharkhand, India

Received 15 July 2014; revised 14 September 2015

Teak (*Tectona grandis* L.f.), a paragon timber tree of tropical deciduous forests of Central and Peninsular India, is highly prized for its wood colour, decorative grains, durability and lightness. An experiment was carried out to compare the genetic variation detected and genetic relationships inferred in five teak populations via 10 genomic DNA samples per population each of either single seed or bulk of 3- or 5- seeds with the help of ISSR markers. The genomic DNA of single seed exhibited higher number of polymorphic loci, per cent polymorphism, nei’s genetic diversity and shannon Information Index than the bulk genomic DNA of 3- or 5- seeds. The bulking of genomic DNA of 3- and 5- seeds using Nei’s genetic distance coefficient revealed similar genetic relationships, which were at variance with those in single seed treatment. Mantel’s correlation test among the genetic distance matrices of single seed sampling, 3-seed bulk and 5-seed bulk sampling also confirmed the trend. Since the bulking of genomic DNA did not generate compatible estimates of diversity parameters and genetic relationship of five populations from its single seed sampling, we recommend strict guarding of identities of genotypes within the collected samples for obtaining precise estimates and drawing accurate conclusions about the genetic diversity and clustering of populations.

**Keywords**: AFLP, Bulking, Genetic diversity, Genetic relationship, ISSR, Molecular marker, RAPD

Tropical forests comprise dense communities and populations, and molecular evaluation of genetic diversity of such natural forest tree populations involves expensive and time consuming experimental efforts. Bulking of genomes (genotypes) has been widely used in molecular characterization of plant germplasm to reduce the cost, experimental efforts and increase the characterization scope. There are many reports advocating application of genome bulking for molecular characterization of germplasm of agricultural crops¹. However, such investigations are lacking in forest tree species, which are known for long gestation period, high heterozygosity and mostly of out crossing nature. Therefore, it is important to investigate changes in genetic diversity estimates as influenced by mixing (bulking) of genomic DNA of individuals of populations. In this study, we investigated five diverse populations of teak, a premiere tropical timber species..

Teak is widely distributed in Central and Peninsular India, which are considered to be centre of its origin and diversity². Teak wood is globally famous for its golden colour, decorative grains, durability and lightness³. The species has been extensively explored for genetic diversity estimates in India and abroad using isozymes⁴,⁵, RAPDs⁶, ISSRs⁶,⁷, AFLPs⁸,⁹,¹⁰ and nuclear SSRs¹¹. Further, the teak is highly cross pollinated species, possessing genetically non-identical seeds within individual fruits²². Considering above points in view, the present study was undertaken for detection of change in ISSR marker based genetic diversity estimates and genetic relationships among five teak populations as influenced by genomic DNA of single seed and bulk (3- and 5-) seeds.

**Materials and Methods**

**Plant materials**

Teak fruits were procured from five populations belonging to Rajasthan (Antri), Maharashtra (Central Chanda), Andhra Pradesh (ShivRam), Kerala (Walyer) and Karnataka (Tuppur). Ninety seeds were extracted mechanically from the fruits of each teak population and made into three groups (single seed, bulk of 3 seeds and bulk of 5 seeds) each of 10 samples for DNA extraction. Genomic DNA from 90 seeds (genotypes) per population or 450 seeds (genotypes) from five populations was extracted and employed for the present investigation.
DNA Isolation and ISSR analysis

Genomic DNA was extracted from single seed and bulk of 3- and 5- seeds using SDS based method. Extracted DNA was quantified using spectrophotometer (Cintra 404, GBC) and finally diluted to 30 ng genomic DNA for ISSR analysis. Prescreened five polymorphic ISSR primers, viz., UBC-801, UBC-834, UBC-880, UBC-899 and UBC-900 from university of British Columbia, Canada were used for amplification of genomic DNAs. The reaction mixture for ISSR amplification assay had a total volume 10 µL, which contained 30 ng genomic DNA, 1X Taq polymerase buffer, 0.1 mM of each dNTPs, 2.5 mM MgCl₂, 1 U Taq polymerase and 0.8 µL primers. PCR conditions were programmed for initial denaturation at 94°C for 3 min, and 30 cycles of 30 s denaturation at 94°C, 30 s annealing at 47-50°C, 1 min extension at 72°C, and final extension for 10 min at 72°C. PCR products were electrophoresed on 3% SFR agarose gel containing 0.5 µL/mL ethidium bromide in horizontal gel electrophoresis apparatus, using 0.5 X TBE as running buffer (pH 8.0) and constant power of 100 V for 2 h. After electrophoresis, the amplified DNA bands were visualized on Gel documentation system (Alpha Imager 1200TM, Alpha InfoTech Corporation, USA).

Data analysis

The ISSR bands were scored as present (1) or absent (0), each of which was treated as an independent character regardless of its intensity. Software NTSYS-pc (Numerical Taxonomy System, version 2.1) was used for estimation of genetic distances among the populations in all three sampling treatments. Genetic relationships among the populations were investigated by cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA) based on genetic distance matrices. Correlation coefficients between the three genetic distance matrices were calculated and their significance was tested by Mantel matrix correspondence test using 1000 random permutations. Analysis of molecular variance (AMOVA) was performed using Arlequin (ver. 2.001) to assess ISSR variations among and within the populations.

Results

Five ISSR primers detected significantly highest values for number of polymorphic loci, per cent polymorphism, Nei’s genetic diversity and Shannon Information Index in single seed DNA and lower values for these estimates in the bulked 3- and 5- seed DNA. However, the bulk of 5-seed DNA obtained on par values for Nei’s genetic diversity and Shannon Information Index with single seed DNA. The single seed DNA enhanced the number of polymorphic loci by 26%, per cent polymorphism by 30%, Nei’s genetic diversity by about 19% and Shannon Information Index by 24% in comparison to the bulk of 3-seed DNA (Table 1).

AMOVA analysis revealed that the variation among populations was non-significant in single seed DNA but was significant in a bulk of both, 3- and 5- seed DNA. Nevertheless, the single seed DNA as well as the bulk of 3- and 5- seed DNA allocated significantly highest proportion of variations to within populations. As for gene flow, single seed DNA demonstrated the highest gene flow and the bulk of 3- and 5- seed, the lowest gene flow (Table 2). Critical perusal of CD (critical difference) values revealed that the estimate of Nei’s genetic diversity.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Variance components</th>
<th>% of variation</th>
<th>P value</th>
<th>Fst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single seed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>0.00779</td>
<td>0.09</td>
<td>0.24145</td>
<td>0.00088</td>
</tr>
<tr>
<td>Within populations</td>
<td>8.85659</td>
<td>99.91</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8.86437</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three seeds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>0.03140</td>
<td>0.49</td>
<td>0.00587</td>
<td>0.00488</td>
</tr>
<tr>
<td>Within populations</td>
<td>6.40395</td>
<td>99.51</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6.43535</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five seeds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>0.02131</td>
<td>0.39</td>
<td>0.01173</td>
<td>0.00394</td>
</tr>
<tr>
<td>Within populations</td>
<td>5.38736</td>
<td>99.61</td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5.40868</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1—Mean, Standard error and critical differences for the variability indices observed in the teak populations using single seed and bulk seed samples

<table>
<thead>
<tr>
<th>No. of polymorphic loci</th>
<th>% polymorphism</th>
<th>Nei’s genetic diversity</th>
<th>Shannon Information Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-seed</td>
<td>2-seed bulk</td>
<td>3-seed bulk</td>
<td>1-seed bulk</td>
</tr>
<tr>
<td>Mean</td>
<td>10.16</td>
<td>8.04</td>
<td>8.68</td>
</tr>
<tr>
<td>SE</td>
<td>0.28</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td>CD</td>
<td>1.06</td>
<td>8.83</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 2—Analysis of molecular variance (AMOVA) for single seed and bulked-samples-based approaches using
number of polymorphic loci and percent polymorphism was at par in bulk of 3-seed and 5-seed sampling and differed significantly in single seed DNA. Unlike, polymorphic loci, percent polymorphism and Nei’s genetic diversity, estimate of Shannon index was at par in single seed sampling & 5-seed sampling compared to estimate obtained in 3-seed bulk treatment (Table 2).

The UPGMA dendrogram constructed on the basis of Nei’s genetic distance obtained in single seed DNA and the bulk of 3- and 5- seed DNA made two groups of five teak populations. However, these groups included different teak populations as per the single seed DNA and the bulk DNA. For example, single seed DNA categorized Rajasthan, Maharashtra, Kerala and Andhra Pradesh in one group and Karnataka in the other group. On the other hand, the bulk of 3- and 5- seed DNA made first group of four populations, i.e. Maharashtra, Andhra Pradesh, Kerala and Karnataka and second group of single population, i.e. Rajasthan. However, the categorization in the second group of four populations by single seed analysis was poorly supported by bootstrap values, which were <50%. On the other hand, the categorization by bulk of 3- and 5- seeds was not only similar but also supported by robust bootstrap values, exceeding >50% (Fig. 1). Besides, the distance matrix of single seed did not significantly correlate with distance matrices obtained in bulk of 3- and 5- seeds. But distance matrices of bulk of 3- and 5- seeds were significantly correlated with each other (Table 3).

**Discussion**

On several instances in natural forests, grasslands as well as in agricultural crops, it becomes impossible to identity individual germplasm at the time of collections due to intermingling of genotypes produced by root suckers, shoot runners or rhizomes. On the other hand, the genetic conservation programs heavily depend on the minimal sample size required for the trustworthy assessment of intra- and inter-population genetic variation\textsuperscript{19}. Narayanan et al.\textsuperscript{6} have proposed 10 individuals per population and five polymorphic primers being adequate for molecular diversity analysis of Indian teak. In bulk method, using small sample size produces distortion in the detection of rare alleles\textsuperscript{20,21}. Contrary, if samples size of bulk becomes large, a large difference in banding pattern will no longer be detectable\textsuperscript{20}. This indicates the importance of the number of individuals which represent the variance of the species used for analysis\textsuperscript{22}. The same sample size per teak population and polymorphic primers as proposed by Narayanan et al.\textsuperscript{6} has been used in the present investigation to compare polymorphic loci, per cent polymorphism, genetic diversity and Shannon Information Index in single seed versus bulk of 3- and 5- seeds. In fact, the individual teak seed is equivalent to a genotype. The bulking of seeds are akin to mixing of indistinguishable genotypes in a natural sample.

Comparison of the three sampling treatments to detect the polymorphic loci in five teak populations reveals decrease in polymorphism in bulk of three seeds compared to that in single seed sample. In
contrast, the single seed sample and the bulk of five seeds detect ‘on par’ polymorphism. This creates a paradoxical situation, for bulk samples mask the effect of rare alleles, thereby detecting lower polymorphism\textsuperscript{23,24}. However, the observed situation in single seed versus bulk of three seeds gets well explained only. It may be possible that the ‘on par’ polymorphism between single seed and bulk of five seeds reflects the rare alleles gaining detectable effect in the bulk of five seeds due to the large sample size. However, the significant decrease in the estimates of gene diversity and Shannon Information Index in bulk of 3- and 5- seeds compared to the single seed broadly confirms the contention of Fu \textit{et al.}\textsuperscript{23} and Wakui \textit{et al.}\textsuperscript{24}, which indicates less detection probability for rare alleles with increase in bulk sample size.

Range of Nei’s genetic distance values varied considerably in individual (single seed DNA) versus bulk (3- and 5- seed DNA) sampling treatments. This is also corroborated by the estimates of the correlation between three matrices analyzed using Mantel test, indicating that the estimates of genetic distances might be biased in bulked samples. Wakui \textit{et al.}\textsuperscript{20} reported highly significant correlation coefficient between the similarity matrices based on individuals and bulked population samples using 20 random primers (10 each of RAPD and ISSR). Therefore, it is also necessary to investigate the number as well as sensitivity of marker techniques along with bulking size.

Estimates of the variation within population are greater than the estimates of the variation among population irrespective of the bulking treatments. Similar trend has been observed in almost all the earlier published reports on teak diversity\textsuperscript{4,11}. The larger portion of variation within population is due to the high heterozygosity and out crossing nature of the species. This suggests that the bulking did not disturb the partition of variation into its components. In the present investigation, inference of gene flow has indirectly been drawn through fixation index (Fst). High gene flow in single seed compared to the bulk of 3- and 5- seeds also pointed towards the earlier contention\textsuperscript{23,24} of decreasing detection probability for rare alleles with increase in bulk sample size.

Dendrogram constructed based on the genetic distance values using UPGMA has exhibited similar trend in clustering pattern of teak populations in individual (single seed) and bulk (3-seed and 5-seed) sampling treatment. In bulk samples (3-seed and 5-seed), Maharashtra, Andhra Pradesh, Kerala and Karnataka have formed a single cluster leaving second group of one population, i.e. Rajasthan. Conversely in single seed sample, Rajasthan, Maharashtra, Kerala and Andhra Pradesh are in one group and Karnataka in the other. But the grouping in single seed dendrogram is weak as revealed by the smaller bootstrap values. In a RAPD assessment of the congruency of grouping five \textit{Brassica oleracea} accessions between bulk and individual plant samples, Divaret \textit{et al.}\textsuperscript{20} found identical patterns of genetic relationship for the five accessions. The same pattern was found by Fu \textit{et al.}\textsuperscript{25} in their RAPD assessment on the grouping of five flax landrace accessions. An identical AFLP-based grouping of six \textit{Agropyron} accessions between bulked and individual plant samples was also reported by Mellish \textit{et al.}\textsuperscript{26}. One reason for deviation of clustering in bulk sampling treatment in the present experiment may be that the biased estimates of genetic variation with increasing bulk size reduced the accuracy of genetic distance estimates among the germplasm accessions, and thus affected inferences of genetic relationships among the populations/germplasm\textsuperscript{27}. Alternatively, the multi-parental mating in teak, even up to the extent of having genetically non-identical seeds within the same individual fruits\textsuperscript{22} is also possible for such deviation in single versus bulk samplings. This suggests the bulking in teak is more complicated as compared to agricultural crops and underlines the need for more critical investigations on optimization of bulking procedures in forest tree species.

Use of bulking is increasing in molecular characterization of germplasm banks holding a large number of accessions, especially of agricultural crops. However, this strategy has not been assessed for conservation and utilization of forest genetic resources. From the present investigation, it is clear that the bulking DNA of 3 seeds/genotypes is preferable as there were no changes in diversity indices and structure on further increase in bulking size. With reference to teak bulking and similar highly cross pollinated tree species, we strongly advocate that the identities of genotypes within the collected populations need to be strictly guarded for obtaining precise estimates of their genetic diversity, clustering and structure.

References
2. Vaishnav V, Mohammad N, Wali SA, Kumar R, Tripathi SB, Negi MS & Ansari SA, AFLP markers for analysis of genetic
9. Minn Y, Genetic diversity within and among populations of Teak (*Tectona grandis* Linn. f.) and Dahat (*Tectona hamiltoniana* Wall.) in Myanmar detected by AFLP markers and cpSSRs: From Forest Genetics and Forest Tree Breeding, (Goettingen University, Goettingen), 2007, 70.
18. Schneider S, Roessli D & Excoffier L, Arlequin ver 2.001; A Software for Population Genetics Data Analysis, (Genetics and Biometry Laboratory, University of Geneva, Geneva), 2002.