6. SUMMARY

Expansion of highly managed industrial tree plantations is benefited by augmentation of modern biotechnological methods and products from genomic strategies. Although traditional silvicultural methods have contributed for increasing the productivity of plantations, biotechnological approaches like controlled breeding and clonal forestry has assured increase productivity and uniform products for industry. *Eucalyptus* is one of the genuses benefitted from such biotechnological applications. Molecular markers, the breeder’s tool, are continuously exploited in genetic diversity estimation, propagation systems, seed orchard management and marker-trait associations to accelerate the genetic gains. The availability of complete genome sequence increases the possibility for marker assisted studies in *Eucalyptus*.

*Eucalyptus* plantations are mainly established as clonal forests and adventitious root formation on vegetative cuttings is a key parameter influencing the viability of a clonal forestry program for eucalypt species. The variable rooting potential among the collections reflect on the execution of the clonal plantation. Understanding the genetic architecture of the trait helps in for further improvement, which involves the dissection of the QTLs (Quantitative Trait Loci) using molecular markers. Linkage disequilibrium (LD) based association mapping (AM) approach, which does not require specific pedigreed populations as in conventional QTL approach, gains significance recently. This involves the examination of populations of unrelated individuals, wherein the phenotypic variation among individuals is correlated with DNA polymorphism. Limited studies have been carried out for dissection of QTLs in forestry species using AM approach and at present the approach is in progress fo several forest tree species including *Eucalyptus*.

The present study aimed at estimating the LD in the eucalypts species using SSR primers developed from others *Eucalyptus* species and associating SSR/STS markers with the adventitious rooting traits. Vegetative propagation traits were measured for the selected association panel for rooting and the related traits. The hybrid population was considered for the rooting percentage only. Higher variation for the rooting percentage was observed in *E.*
tereticornis compared to *E. camaldulensis* and the hybrid population. Analysis of Variance (ANOVA) revealed significant correlation between the rooting related traits (r = 0.972, P < 0.01) in *E. tereticornis* and *E. camaldulensis* (r = 0.955, P < 0.01). The rooting percentage measured for the hybrid population revealed less contrasting rooting characteristics between the parents and the hybrids. For this reason they were considered for the analysis of LD only and excluded for the association analysis for the rooting trait.

Linkage maps for *Eucalyptus* display high synteny across species and have been extensively used to map QTLs for a number of traits including growth, wood quality, disease and insect resistance. The present study employed SSR markers developed and mapped from other *Eucalyptus* and *Corymbia* species including few EST SSR loci. The selected set of primers included 26 STS markers (Sequence Tagged Site) that were linked to the vegetative propagation traits in *Eucalyptus* and *Corymbia* species. *Eucalyptus* accessions of association panel revealed 56.9 % transferability. However only 23.4% transferability was observed for hybrid population since the polymorphism between the parents was low.

The existence of genetic diversity is highly imperative for every tree improvement programs. The present study demonstrated high genetic diversity among the total eucalypt populations and as well as within the populations of *E. tereticornis* and *E. camaldulensis*. The largest fraction of the diversity was observed within the populations. For the total eucalypts accessions highest number of alleles, 30 and 27 were found at locus Embra35 and Embra207 respectively. The highest observed heterozygosity was observed for the locus Embra35 of LG 1. The hybrid population revealed very low number of alleles when compared to the pure species but had higher variability in heterozygosity estimate, indicating the recent recombination of two different genomes. The loci with higher diversity estimates are relevant for mapping studies in the species and a subset of the loci (n=10) with these estimates are sufficient to discriminate the total set of 93 eucalypt accessions.

Investigation of structure revealed a K value of 7, where *E. tereticornis* showed subdivision of the population and *E. camaldulensis* did not separate into subpopulations. The major divisions of the subpopulation in *E. tereticornis* included the landraces and few putative hybrids. Other divisions in the species included only very few individuals forming a mixed
group. Similar observations were recorded when *E. tereticornis* was analyzed for the existence of structure, except for the additional separation of Orobay provenance. Population structure was incorporated as a covariate into the model (GLM and MLM) to perform association analysis of the provided SSR loci for the trait. Excluding the input of structure estimates for the detected k values may end up in spurious marker-trait correlations. Analysis of molecular variance exhibited higher partitioning of the variation within individuals (74%) and low $F_{ST}$ values (0.019) between the two populations.

Examination for the presence of most common alleles in the two species and the landraces identified 23, 14 and 38 SSR alleles that were specific to *E. tereticornis*, *E. camaldulensis* and the landraces respectively. These most common alleles could differentiate the three groups of population.

The PI estimates indicated that the microsatellite markers used in the study could distinguish two closely related genotypes by the order of $9.50 \times 10^{-03}$ to $9.24 \times 10^{-109}$. The UPIC values have provided the actual number of SSR loci required to discriminate the clones analyzed. Embra5 could differentiate 26 clonal accessions. The primers showing high UPIC values can be utilized in cases of alleged infringement and DNA fingerprinting. The provided unique pattern information also exploited for determining clonal distinctness and hybridity. All the clonal accessions are genetically distinct having a unique pattern information combination. These clones were already proven to be a best for growth performance and they could be selected as base population for clonal seed orchard establishment.

Estimation of LD using GDA revealed 135 significant pairs with 17 pairs showing $r^2 \geq 0.1$ in *E. camaldulensis* and in *E. tereticornis* out of 136 significant pairs, 18 pairs showed $r^2 \geq 0.1$. The hybrid population explained very low $r^2$ values with 61 significant pairs. However the interallelic LD computed using MIDAS revealed higher number of significant allelic pairs for hybrid population compared to the *E. tereticornis* and *E. camaldulensis* accessions. The decay of LD ($r^2$) with genetic distance (cM) evaluated by nonlinear regression (NLR) set with the population specific threshold value ($P_{95}$) for genotypic and haplotypic LD in *E. camaldulensis* and *E. tereticornis* indicated the rapid decay of LD. However, in case of hybrid population, LD decay was observed at 11.5 cM with very low $r^2$ values (0.04) for the genotypic LD.
Nevertheless for haplotypic LD, NLR curve fitted on the $r^2$ estimates was below the 95th percentile (0.51) baseline. LD determinations using the HaploView program explained only very few significant LD pairs in both *E. tereticornis* and *E. camaldulensis* species. However in hybrid population, significant haplotype block is observed at linkage group 4, showing strong evidences of LD between two loci pairs (Embra186 – Embra66 and Embra66 – Embra36).

Low LD in *Eucalyptus* promises a higher resolution in genome-wide association mapping. A challenge, however, is that many more markers are required for covering the whole genome. The small genome size of *Eucalyptus*, 590 Mbp/C and 580 Mbp/C for *E. camaldulensis* and *E. tereticornis* respectively and the availability of whole genome sequence of *Eucalyptus* make it possible to identify few more SSRs with less than <1 cM to corroborate the alleles showing high LD ($r^2$$>$0.1) identified in the present study. While SSRs are currently used for the LD based association analysis in *Eucalyptus*, a high number of SNP markers are being developed, providing additional opportunities for high-resolution LD mapping. These studies require understanding of LD extension and intensity. Candidate gene-based association genetics have successfully found marker–trait associations for wood and fiber traits. The recently released *E. grandis* genome will open up excellent opportunity to advance *Eucalyptus* genetics and genomics research.

Even though a higher marker density will be required to accurately describe LD variability at the whole genome scale, this study contributed to reach a better image of LD decay in *Eucalyptus* and provided the guidelines for the marker densities required for association mapping. In addition, the number of individuals analyzed in each subpopulation was small (40 to 53 individuals) and it has been stated that populations with a broad genetic basis have lower levels of LD than narrow-based populations, so the use of more diverse material had exploited the recombinational history. Also it is likely a population with reduced LD extension is ideal for association mapping purposes.

The outcome revealed three SSR loci Embra40, Embra54 and Embra101 associated with rooting percent and mortality. Embra7 was associated with shoot length. When the significance level was decreased three loci (Embra5, Embra6 and Embra39) were observed with the association for number of roots. Locus Embra72 was found to correlate with the rooting
percentage. One of the STS loci (Embra40) linked with vegetative propagation traits in eucalypt species and Corymbia species (Marques et al., 2002; Shepherd et al., 2008) showed association with mortality trait in E. tereticornis. In E. camaldulensis two SSR loci Embra167 and Embra39 were associated with shoot length and root length. Decreasing the significance level resulted in detection of one STS locus Embra24 associated with shoot length. Furthermore, two SSR loci Embra19 and Embra167 were associated with the length of the longest main root.

The present study has successfully identified a number of SSR markers linked to previously reported QTLs associated with vegetative propagation traits in addition to few newly identified markers. The results suggest that association analysis could be a useful alternative to linkage mapping for the identification of putative gene-linked SSR markers which can be potentially used in marker-assisted selection in Eucalyptus. The identified marker–trait associations can be useful in breeding and helpful in determination of genetic mechanisms underlying adventitious rooting.