5. DISCUSSION

_Eucalyptus_ constitute approximately 15% of global plantations, and are most commonly exploited for its characteristic superior fiber and pulping properties. _Eucalyptus_ is the dominant and most productive planted forest in Brazil, covering around 6.3 million ha (Abraf, 2011). The ability of epicormic buds to re-sprout effectively after fire possibly has allowed the species to dominate multiple niches and speciate widely (Crisp _et al._, 2011). Plantations are grown on short- and medium-rotations for the production of pulp, charcoal, fuel wood, reconstituted and solid wood. Brazilian _Eucalyptus_ plantations are sustaining the mean growth rates of about 50 m³/ha/year over the 4.7 million hectares planted across the country, which could be achieved because of the investment and developments in research and technology (Binkley and Stape, 2004; Campoe _et al._, 2013). India is the second major planter of _Eucalyptus_ with an area of 3.943 M ha (http://git-forestry.com/Global_Eucalyptus_Map.htm). The recent review on the family Myrtaceae validates the importance of genomic studies in the most vital genus _Eucalyptus_ (Grattapaglia _et al._, 2012). The genus has probably developed quite unique adaptive strategies unlike other tree genera such as _Populus_ (Myburg _et al._, 2006). Genetic improvement in _Eucalyptus_ principally focuses on traits related to productivity and wood quality. Although, breeding and propagation technologies play a significant part in tree improvement programs, only limited studies target on adventitious rooting competency of elite clones, the trait which demonstrates great variation within and across the species.

Tree improvement programs have exploited clonal propagation to deploy superior performers selected from natural and hybrid populations. Poor rooting of the propagated clones results in high economic losses (Sorin _et al._, 2006) and hence formations of adventitious roots play a vital role in any successful propagation systems (Davies _et al._, 1994). Various molecular and gene based studies on the trait corroborates the importance for the trait (Butler and Gallagher, 1999; 2000, Sedira _et al._, 2005, Smolka, 2009). In _Eucalyptus_, propagation through stem cuttings is commonly employed, due to its ease for mass propagation and cost-effective production. In addition, vegetative propagation can capture both additive and non-additive genetic variation in tree breeding (Zobel and Talbert, 1984). In Brazil, _Eucalyptus_ have been
mainly established through vegetative propagation based on rooted cuttings and grown in commercial forest plantations thereby significant enhancement in productivity was achieved and tolerance to various diseases was improved (Wingfield, 2003; Wingfield et al., 2008).

Several studies have explicitly revealed the interest of tree breeders on the use of DNA markers for precise breeding in eucalypts for commercial trait improvement (Grattapaglia and Kirst, 2008). Particularly, in the countries where the species is introduced, limited seed sources form the basis for breeding population and therefore integration of DNA markers in the genetic improvement program of these species would have a major impact on productivity.

5.1 ADVENTITIOUS ROOTING TRAITS

The significance of cloning serves as a means for the establishment of productive clonal forestry in eucalypts and has its achievable outcome on end product quality. Hence studies regarding the genetic control on the vegetative propagation traits were initiated (Borralho and Wilson, 1994; Grattapaglia et al., 1995; Tibbits et al., 1997; Marques et al., 1999). Genetic control and architecture of adventitious rooting in forest trees was recently reviewed by Shepherd et al., (2009). Heritability for adventitious rooting of stem cuttings in *E. globulus* was estimated and found a high narrow sense heritability (*h*² = 0.54) suggesting that large gains can be achieved by direct selection for rooting ability (Lemos et al., 1997). Genetic control of vegetative propagation traits estimated from related species *Corymbia* revealed large phenotypic variances for rooting and most other propagation traits, with significant proportions attributable to differences between clones (*H*²= 0.2–0.5) (Shepherd et al., 2007). Mankessi et al., (2011) assessed the prospect of clonal field selected individuals by rooted cuttings on *E. urophylla × E. grandis*. The study explained the influence of donor plant age, within-shoot position and seasonal effects on the adventitious rooting ability of the cuttings. Very recently the detailed attributes of the clonal propagation of *E. globulus* have been reported by Iwasaki et al., (2012 and the references there in). The study highlighted the significance of the rooting in the tree species and its future implications in the improvement of the plant material.
In our study, higher variation in rooting percentage was observed among the individuals of *E. tereticornis* accessions (0 to 100%), followed by *E. camaldulensis* accessions (7 to 98%) (Table 4.1 and 4.2). Awang et al., (2009) detected similar variations in rooting among five indigenous and one exotic landscape tree species. The rooting process may be highly affected by many possible factors including concentration of inhibitors in the stem tissues (Kibbler et al., 2002) and food reserves (Druege et al., 2004), mineral nutrients (Schwambach et al., 2005) and genetics (Titon et al., 2006). Rooting studies in *Eucalyptus* also shows a greater variation for rooting (Dasgupta et al., 2010), which concur with this study. Moreover, *E. camaldulensis* revealed higher estimates of rooting parameters than *E. tereticornis*. A major source of phenotypic variation for rooting and root quality traits due to the species effects/differences were observed in *Pinus* and *Corymbia* (Shepherd et al., 2006; 2008). In eucalypts, easy to root and rooting recalcitrant species were identified and also high within species variability was reported (Marques et al., 1999). Variable rooting potential (0 to 100%) among the selections are considered a hindrance to the clonal propagation and moreover a bottleneck in the production of elite genotypes (England and Borralho, 1995).

The operationally accepted limit for rooting success in eucalypt vegetative propagation programs in Aracruz (Brazil) and Smurfit Carton de Colombia (Colombia) is at least 70% (Zobel, 1993). Hence the breeding program for commercially exploited eucalypt species could focus on selection for growth, form, adventitious rooting competence and resistance traits for successful clonal deployment. The average number of roots for *E. camaldulensis* and *E. tereticornis* were recorded as 6.9 ± 3.3 and 2.8 ± 2.2 respectively. The maximum number of roots observed for *E. camaldulensis* and *E. tereticornis* were 17 and 9 respectively (Table 4.1 and 4.2). Large phenotypic variances were reported for rooting and most other propagation traits, with significant proportions attributable to differences between clones (H² = 0.2-0.5) (Shepherd et al., 2007). Shepherd et al., (2008) identified one of the QTL explaining ~66% variation for rooting percent in *Corymbia torelliana × Corymbia citriodora* subspecies variegata hybrid family (n=186) using the SSR markers. Thumma et al., (2010) reported higher percentage of genetic variation associated with vegetative propagation traits compared to growth traits. The percentage of variation explained by the tissue culture and stem cutting methods were 52.3% and 69.0% respectively.
The results of ANOVA explained high significant (p< 0.01) correlation between the rooting parameters in both *E. tereticornis* and *E. camaldulensis*. Extreme level of correlation was observed between the root length and the length of the longest main root (r =0.972, P< 0.01) in *E. tereticornis* and for the root length and the length of the longest main root in *E. camaldulensis* (r =0.955, P< 0.01) (Table 4.4 and 4.5). Comparable results were observed by Shepherd *et al.*, (2007) in hybrids of *Corymbia torelliana* and *Corymbia variegate*, where the rooting traits were highly and favorably inter correlated with the root quality traits. The study moreover showed strong favorable genetic correlations between rooting percentage and the root quality traits (root biomass, volume, and length). Awang *et al.*, (2009) observed comparable level of correlation between rooting percentage and the shoot length during the examination of adventitious rooting of in five indigenous forest species and one exotic landscape tree species. The study demonstrated high level of correlation between the rooting ability and the number of leaves (r = 0.869-0.981) and shoot length of the mother plants (r = 0.690-0.954).

In the case of hybrid population used in this study, observed mean rooting percentage was 81.5 ± 14 .8 (SD) (Table 4.3). The explanation for high rooting potential of the hybrids may be due to their parents. *E. tereticornis* (pollen parent) and *E. camaldulensis* (seed parent) were recorded with moderate and high rooting ability respectively (Table 4.3). Shepherd *et al.*, (2005) reported higher rooting percentage in F₁ hybrids of *Pinus elliottii*× *P. caribaea* families during the study of genetic control of adventitious rooting. Substantiations were demonstrated by the utilization of *E. tereticornis* in hybridization programs to improve adventitious rooting of other eucalypt species (Eldridge *et al.*, 1994). Less variation in rooting percentage was observed among the *Eucalyptus* hybrids. Genetic control of vegetative propagation traits from interspecific hybrids of *Corymbia torelliana* and *Corymbia variegate* revealed large phenotypic variances for rooting and most other propagation traits (Shepherd *et al.*, 2007). However establishing that *Eucalyptus* interspecific hybrids are being most successfully employed in plantation forestry (Dungey and Nikles, 2000; Potts and Dungey, 2004). Quantitative trait loci (QTL) detection was carried out for adventitious rooting and associated propagation traits in a mapping pedigree of cross *E. tereticornis* (seed parent) and an *E. globulus* (pollen parent) wherein putative QTLs accounted for 2.6–17.0% of the phenotypic variation. The study reported *E. tereticornis* as a potential rooter with a rooting percentage of 60–90% (Marques *et al.*, 1999). Hence many
breeding programs target intra- and inter-specific hybridization for the transfer of vegetative propagation traits and it is an important objective in many breeding programs. In Brazil and South Africa, interspecific hybrids of *E. grandis* X *E. urophylla* are quite common with the involvement of *E. grandis* to form shoots, while *E. urophylla* contributes its ability in rooting. However, in India the potential of hybrid eucalypt forestry is yet to take off with the commonly grown drought tolerant species like *E. tereticornis* and *E. camaldulensis*.

**5.2 MICROSATELLITE AMPLIFICATION**

**5.2.1 ASSOCIATION PANEL**

Microsatellite markers used in this study were transferred from related species of *Eucalyptus* and *Corymbia*. They have been proven to be an extremely valuable tool for paternity analysis, construction of high density genome maps, mapping of useful genes, marker assisted selection, etc. (Parida *et al*., 2009). PCR amplification involved standardization, where non-specific amplification was greatly reduced by the addition of Bovine serum albumin (BSA). Recently, Nevill *et al*., (2010) have reported the amplification of *Eucalyptus* SSRs employing BSA. In the present study, 56.9% transferability was observed for the first set of natural population, by providing 62 informative loci out of 109 loci (Table 3.2). The transferability of microsatellite across species of the subgenus Symphyomyrtus varied between 78 and 100%. The cross amplification rate around 50 to 60% for species of different subgenera such as Idiogenes and Monocalyptus and excessively low for the related genus *Corymbia* (25%) (Kirst *et al*., 1997).

The transferability of SSR loci among the species of eucalypts is amply represented (Agrama *et al*., 2002; Da Silva *et al*., 2009) and it was highly successful in the present study (56.9%) for *E. camaldulensis* and *E. tereticornis*. Dos Santos *et al*., (2007) reported 44.5% transferability of microsatellite markers from *Eucalyptus* spp to *Acca sellowiana* and its successive utilization in genetic characterization of the species. Ekué *et al*., (2009) reported the transferability of 12 SSR markers developed in *Litchi chinensis* to *Blighia sapida* (Sapindaceae). 58.3% transferability was stated for the species with seven primers showing amplification and four loci revealing polymorphism for the species in 16 trees. Faria *et al*., (2010) reported the
development of 20 EST-SSRs and its subsequent transferability across the 6 major *Eucalyptus* species, providing excellent resolution for population genetic studies within the subgenus Symphyomyrtus. Recently Nagabhushana *et al.* (2011) developed a novel set of genic SSR markers in *E. camaldulensis* and effectively established the power of orthologous SSR in the species. Recently, Acuña *et al.* (2012) developed a set of novel functional markers from *E. globulus* and reported 25% transferability among seven *Eucalyptus* species, including *E. camaldulensis* and *E. tereticornis*.

### 5.2.1.1 Microsatellite Allele diversity

Number of alleles generated primarily contributes to the diversity existing among the samples. In this study, number of alleles ranged from 5 to 30 (total of 1067), with a mean value of 17.2 ± 5.2 in 93 accession of *Eucalyptus* (Table 4.6). Similar study in *E. grandis* (n=192) have detected loci with highly variable number of alleles 6 to 33, with an average of 19.8 ± 9.2 (Kirst *et al.*., 2005). Brondani *et al.*., (2002) reported mean allele number of 14.34 and heterozygosity 0.87 in the species belonging to subgenus *Symphyomyrtus*. Analogous studies involving analysis and characterization of microsatellites in *Eucalyptus* species reported loci with highly variable numbers of alleles (Marcucci-Poltri *et al.*, 2003; Brondani *et al.*, 2006). Allelic variability observed in the two species, *E. tereticornis* and *E. camaldulensis* was very similar to other species of eucalypts. The existing diversity in eucalypts may be due to the natural hybridization that had contributed greatly to the improvement of the species (Barbour *et al.*, 2008). However, in *E. camaldulensis*, Butcher *et al.*., (2009) reported 40 alleles for the locus Embra11 with allele size range varying from 72 – 172 bp and concurrently in the present study 22 alleles in the range of 124 - 150 bp were identified (Table 4.6 and Table 4.7).

The number of alleles reported by Dasgupta *et al.*, (2010) for the primers Embra6 and Embra18 were comparable to our results. The polymorphic information content (PIC) values provide an estimate of the marker informativeness (Cordeiro *et al.* 2000). PIC is commonly used to estimate polymorphism status of a marker locus and values higher than 0.8 indicated their informativeness for all kinds of genetic studies. PIC values were higher than those observed by Da Silva *et al.*, (2009), where the PIC ranged from 0.37 to 0.88 (average of 0.72) in *E. camaldulensis*. The PIC value of the loci Embra10 in *E. tereticornis* (0.89) and *E. camaldulensis*
(0.90) (Table 4.7) is comparable to that reported by Hong-xin et al., (2008) in *E. globulus* (0.88). The results demonstrate the relevance of the subset of microsatellites loci (n=62) employed in the study which may be satisfactory for any mapping studies.

Roger's genetic distance based dendrogram generated using Powermarker (Liu and Muse 2005) revealed the genetic diversity among the accessions utilized in the association panel (Figure 4.8). Marker-based genetic distances have been used to improve the structure of breeding populations (Marcucci-Poltri et al., 2003) and seed orchards (Zelener et al., 2005). Given the wide genetic diversity and multiple sources of germplasm for eucalypt breeding, choices have to be made as to which elite parents should be mated. Any means of predicting tree performance would be valuable for the breeder. Vaillancourt et al., (1995) reported the ability of RAPD-based genetic distance to predict heterosis in *E. globulus* progenies. De Aguiar et al., (2007) explained the significance of microsatellite-based genetic divergence in *Eucalyptus*. The preferences of natural hybridization in eucalypts have contributed a major role in the evolution of the present diversity of species (Barbour et al., 2008).

Assessment of the diversity in the forest species *Eugenia uniflora* L. using seven microsatellites (n=84) revealed a mean of 14.4 for the total number of alleles and a low genetic differentiation (mean $F_{ST}$= 0.031) was observed which indicated higher genetic diversity within the populations (Ferreira-Ramos et al., 2008). These results concur with the present study, where an $F_{ST}$ estimate of 0.04 was observed between the populations in *E. tereticornis* and *E. camaldulensis*. However, genetic diversity and population structure assessment in *Populus balsamifera* employing 11 nuclear loci revealed significant population structure and low nucleotide diversity, due to a recent population expansion (Keller et al., 2010). Genetic diversity among 43 grapevine accessions analyzed exploiting nine microsatellite markers revealed a mean of 8.3 alleles and high genetic differentiation among the populations. Most of the genetic diversity (78%) occurred within populations suggesting considerable isolation and gene flow among the populations (Ramezani et al., 2009). Recently Acuña et al., (2012) demonstrated the significance of microsatellite markers in diversity assessment of candidate genes of *Eucalyptus globulus* in a study for wood properties.
5.2.1.2 Species discrimination in *Eucalyptus*

In eucalypts, the predominance of natural hybrids is well documented and landraces are considered as putative hybrids (Doran and Burgess, 1993; Varghese *et al.*, 2009). Further, few of the provenances of *E. tereticornis* were reclassified under *E. camaldulensis* and many provenances show mixed morphology between the two species (Doran and Burgess, 1993). It could be due to the natural hybridization occurring at the seed source locations in Australia and unclear delineation of the distribution boundaries between the species (Doran and Burgess, 1993, Brondani *et al.*, 2006). Moreover, the complicated taxonomy of the family Myrtaceae to which the eucalypts belong makes the distinctiveness of many widely recognized genera in problem (Biffin *et al.*, 2010; Edwards *et al.*, 2010). In this study, SSR analysis of *E. camaldulensis* and *E. tereticornis* with 38 loci could differentiate both the species due to the presence of species-specific most common alleles (Table 4.9). Although limited number of landrace accessions were used, the presence of SSR alleles belonging to both the species was obvious in the present study, suggesting that these SSR markers could be used to identify pure species and their hybrids. The effectiveness of SSR markers along with AFLPs was demonstrated in species and hybrid discrimination employing a set of samples of *Populus alba*, *P. tremula* and *P. canescens*.

Species-specific AFLPs and SSR alleles were recorded in *P. alba* and *P. tremula* samples. The outcome explained a clear distinction between the two poplar species and their hybrids. These diagnostic markers were able to detect the hybrids of *P. canescens* intermixed with *P. alba* trees (Fossati *et al.*, 2004). Butcher *et al.*, (2009) reported marked separation of the two species of *Eucalyptus*, viz *E. camaldulensis* and *E. tereticornis* using 15 SSRs. The study revealed 100% consent for separation of *E. tereticornis* subsp. *tereticornis* from *E. camaldulensis*. However, other subspecies *mediana* of *E. tereticornis* showed mixed pattern with *E. camaldulensis*. A different study in Cattails was in agreement of species-specific SSR alleles employed for identification of its hybrids (*Typha latifolia* X *T. angustifolia*). The report demonstrated the documentation of backcrossed plants in hybridizing cattail populations using the species-specific SSR alleles from seven loci (Snow *et al.*, 2010). Examination of spatial genetic structure and clustering of individuals in *Corymbia* using five microsatellite loci could not reveal any delineation between the populations, where the spotted gums were found to be
molecularly homogeneous (Ochieng et al., 2010). Recently, Vargas et al., (2011) exemplified few species specific alleles in sea turtle. Analysis of 258 individuals comprising hybrids amplified 5 microsatellite markers revealed a total of 37, 40 and 40 alleles for Eretmochelys imbricate, Caretta caretta and the hybrid population respectively. Out of these, 18 E. imbricata-specific alleles and 21 C. caretta-specific alleles were reported with the hybrids revealing 13 E. imbricata-specific alleles, 12 C. caretta-specific alleles and 15 shared alleles.

5.2.1.3 Clonal discrimination in Eucalyptus

Countries such as Brazil and Uruguay have included the use of microsatellite markers in addition to morphological descriptors, in their legislation for clonal discrimination of eucalypts. Individual identification of clones finds several applications in tree breeding and clonal forestry (Torres-Dini et al., 2011). Clonal forestry use specific clones with specific wood properties which results in uniform raw material for the industry and hence material traceability becomes possible. Seed orchards established with the clones are tested for their breeding value and specific combing ability thus genetic structure and clone combination can be designed in an improved seed orchard in which clones have the most effective genetic contribution to the seed lots (Li et al., 2011a). The 28 samples of E. camaldulensis selected for the study of clonal discrimination were identified as the best clones from a total of 101 individuals from CSO, established by IFGTB (Table 3.1). Hence suitable tools such as DNA markers are required for simple and exact identification of individuals in the nursery and field plantations. DNA markers particularly SSR markers are one of the best tools for the assessment of genetic diversity among the clones in seed orchards and clonal plantations, wherein the maintenance of suitable genetic diversity is highly advocated to minimize the losses of clonal plantations (Bishir and Roberds, 1999). UPIC values indicate the maximum number of accessions could be differentiated with a particular SSR loci and it is more informative for selecting subsets of SSRs than the use of its PIC alone (Arias et al., 2009). In the present study, UPIC values were ranging from 1 (for the EST-SSR loci CD668704) and 26 (for the loci Embra5) indicating that the loci CD668704 could differentiate only one clone whereas Embra5 could differentiate 26 clonal accessions (Table 4.10).
5.2.1.4 Discriminatory power of microsatellites

Probability of identity (PI) estimate provides the power of discrimination of a SSR locus with the given genotypes. This estimator can be used as a conservative upper bound for the probability of observing identical multilocus genotypes sampled from a population (Waits et al., 2001). Two types of PI estimates were calculated (Table 4.11) which showed that the sib-based PI values for individual markers were around $10^{-1}$ however the unbiased PI estimates ranged from $10^{-1}$ to $10^{-3}$ thus indicating a low probability that any two individuals drawn from these accessions will share the same multi-locus genotype. The cumulative PI indicating discriminatory power of the markers were found to be higher wherein the sib-based cumulative PI calculated over 62 SSR markers was 2.95$x$ 10$^{-01}$ to 2.20$x$ 10$^{-31}$ and unbiased cumulative PI estimates were ranging from 9.50$x$ 10$^{-03}$ to 9.24$x$ 10$^{-109}$. The estimated values of the study were higher than those reported by Acuña et al., (2012) in diversity assessment of six candidate genes of *Eucalyptus globulus* (n=54) employing eight SSRs. However, the results were analogous to that reported by Kirst et al., (2005) and Cupertino et al., (2011) authenticating the potential of the 62 SSR markers for the identification of related individuals, parentage testing and genetic diversity studies on *Eucalyptus* species. The use of cumulative PI values is useful in absolute identification of clonal accessions compared to the use of single marker. Similar observations were in Coffee germplasm and it was suggested that the use of sib-based PI for the clonal discrimination would be more authentic because the sib-based PI accounts the possible relatedness in the target germplasm arising due to common parentage (Hendre et al., 2008).

5.2.1.5 Analysis of Population structure

The presence of population structure, which may lead to spurious associations, is one of the potential limitations to the wide usage of linkage disequilibrium mapping in plants (Buckler and Thornsberry, 2002). Though population structure in outbreeding species is low, to avoid uncertainty in the association tests (Thornsberry et al., 2001), the genetic structure existing among the total accessions of *Eucalyptus* and within species were carried out using STRUCTURE v. 2.1 (Pritchard et al., 2000) employing the whole set of SSRs (n=62). The concern about sensitiveness of LD to distance and consequently increasing the probability of detecting spurious clustering is eliminated, since similar results were experienced with both
observations (41 and 62 SSRs) (Kaeuffer et al., 2007). In this study, results at higher K values (K = 7) (Fig. 4.9B), indicated the species distinctness however mixing of genome between the species was obvious, which may be due to the natural hybridization occurring at intergrade zones or taxonomic uncertainty of the species. The major subpopulations (n=3) in E. tereticornis included 34 pure species of E. tereticornis (see species discrimination), the landraces and few putative hybrids. Other divisions in the species included only very few individuals forming a mixed group (Fig. 4.9B). The significance of SSRs in the inference of structure in two maize (Zea mays L. ssp. mays) inbred panels has been explained by (Camus-Kulandaivelu et al., 2007).

The present study showed 1.92% variation (AMOVA) among E. camaldulensis and E. tereticornis accessions. However, Butcher et al., (2009) studied the genetic structure of E. camaldulensis and the results indicated 9% variation among the populations which was essentially due to the sampling of huge number of populations distributed throughout Australia. Analysis of molecular variance revealed low genetic differentiation (F_ST) between the populations (0.04; P < 0.000). Tripiana et al., (2007) showed low differentiation among populations (F_ST = 0.04) using ten microsatellite markers in E. urophylla natural populations, which concur with our study. Related results were observed in other Eucalyptus species, where the level of genetic differentiation among populations was low (F_ST = 0.031) in E. loxophleba (RFLP; Hines and Byrne, 2001). Shepherd and Raymond, (2010) demonstrated the significance of the microsatellite markers (n = 13) in evaluating the genetic pattern of two species of Blackbutt (genus Eucalyptus subgenus Eucalyptus section Pseudophloius). Analysis of genetic structure illustrated the major division within Eucalyptus pilularis. Similarly, analysis of the populations of E. archeri, E. gunnii and E. urnigera with eight microsatellite loci showed extremely low levels of genetic differentiation (Mc Kinnon et al., 2008). Traditional F-Statistics, or fixation indices, which describe the level of heterozygosity revealed very low differentiation between the individuals of putative hybrids of association panel (F_IS = -0.1528; P = 0.9071) (Table 4.15).

The genetic structure of the eucalypts collection within the species, supported K = 3 and K = 5 for E. camaldulensis and E. tereticornis respectively (Table 4.13). The result does not reveal any delineation between the provenances used in E. camaldulensis, except for few
individuals of Northern territory provenance. But in case of *E. tereticornis*, 5 subgroups (3 apparent and 2 mixed) were identified (Figure 4.11). The first group includes individuals of Orobay provenance (n = 8), the second group with a mixture of landraces and some F$_1$ hybrids of *E. tereticornis* X *E. urophylla* (n = 12) and the third comprising of F$_1$ hybrids of *E. tereticornis* X *E. grandis* and *E. tereticornis* X *E. pellita* (n = 6). The other two groups are of mixed type consisting of other provenances of Kupiano and Australia (Fig. 4.11 A and B). The subpopulations of *E. tereticornis* are similar to those results obtained from total 93 accessions of eucalypts. The only differentiation observed is the subpopulation of Orobay provenance and was distinctly different due to its geographical isolation. However the individuals belonging to the PNG populations like Kupiano and Sogeri plateau did not show any similarity with Orobay. The presence of physical barrier, Owen Stanley Ranges in between Orobay and other PNG populations could be attributed for the isolation of Orobay population. The study revealed a high genetic differentiation with in *E. tereticornis* provenances.

### 5.2.1.6 Linkage disequilibrium estimation

Estimation of LD in eucalypt species was carried out using multiallelic SSR markers from related species. Different measures of LD are available, but $r^2$ estimates are predominantly considered. The reason that $r^2$ value has an increased gain over other estimates is due to the fact that, it is directly linked to the proportion of variance of the QTL (whose position is usually unknown), which will be captured by the genotyped marker (Zhu *et al.*, 2008). Threshold 0.1 is the minimum $r^2$ value to detect associations for rather large quantitative trait loci (QTLs explaining 10% of the phenotypic variation) with a reasonable population size (Ersoz and Buckler, 2007). In addition, diverse measures of LD are available for bi-allelic markers; however use of such measures with SSR markers tends to decrease the LD by averaging the allelic effect as a single estimate (Zapata *et al.*, 2001). In heterozygous species, the interpretation of LD becomes complicated because of the non availability of haplotypic phases of the alleles (Zapata *et al.*, 2001).
In this study, interallelic LD was estimated using composite linkage disequilibrium (GDA) (Lewis and Zaykin, 2001) and gametic linkage disequilibrium (MIDAS) (Gaunt et al., 2006). The pattern of linkage disequilibrium (LD) decay determines the marker density required and the level of resolution that may be achieved in an association study (Flint Garcia et al., 2003). Patterns of linkage disequilibrium and SNP variation in spring and winter wheat (Triticum aestivum L.) showed that in a 3,500 cM hexaploid wheat map, placing markers at 0.2 cM will require at least 17,500 markers (Chao et al., 2010). To describe the relationship between LD decay and genetic distance, two methods of establishing baseline $r^2$ values were established. Critical values of $r^2$ were based on a fixed value of 0.1 (Nordborg et al., 2002; Palaisa et al., 2003; Remington et al., 2001) and from the parametric 95th percentile of the distribution of the unlinked markers (Breseghello and Sorrells, 2006).

Based on the NLR curve it is clear that the LD in eucalypts decays faster as in other tree species such as Populus nigra (Marroni et al., 2011). Generally, LD decays more rapidly in outcrossing species as compared to selfing species (Nordborg, 2000; Flint Garcia et al., 2003), particularly in case of forestry species (De-qiang and Zhi-yi, 2005). The low LD detected in this study is expected because compared to crop species, perennial out-crossing tree species has a higher effective recombination rate, which leads to a rapid decay of LD (Krutovsky and Neale, 2005). Low levels of LD in the genome may require an exponentially increasing population size for detection of marker-trait associations (Inghelandt et al., 2011). Rapid LD decay was reported in many other out-crossing tree species such as Pseudotsuga menziesii var. menziesii, where the SNP marker based LD was very low ($r^2 < 0.05$) (Eckert et al., 2009). In Populus nigra decay of $r^2$ with distance in CAD4 gene was observed at about 16 bp (Marroni et al., 2011), while the previous studies reported the decay between 50-500 bp (Ingvarsson, 2005; 2010). In Eucalyptus globulus 20 wood quality candidate genes was analysed using SNP markers and LD was estimated to decay rapidly except in few genes where LD extended beyond 500 bp (Thavamanikumar et al., 2011). Recently, it was observed that Populus nigra and P. balsamifera genome had rapid LD decay across the gene sequences (Kelleher et al., 2012; Marroni et al., 2011; Olson et al., 2010).
In some of the out-crossing domesticated crops such as ryegrass (Li et al., 2011) and maize (Inghelandt et al., 2011) also the mean $r^2$ was well below 0.05 when estimated with SSR markers. Recent study in Perennial ryegrass (Lolium perenne L.) demonstrated a rapid LD ($r^2$) decay that occurred within 0.4 cM, when analysed with 40 SSRs and 2 STS (Brazauskas et al., 2011). Most recently Truntzler et al., (2012) performed LD analysis in maize inbred lines from public institutes (n=113) and a private company (n=201) and explained higher extent of LD ($r^2$~0.2) between polymorphisms (200 kb). Comparatively, public material established lower level of LD and higher genetic diversity. In perennial fruit species like Prunus persica and Vitis vinifera, having a history of domestication and breeding showed a long range LD among the cultivated varieties (Aranzana et al., 2010; Barnaud et al., 2006) while in wild grapevine, Vitis vinifera L. subsp. silvestris LD decayed rapidly, with $r^2$ values decreasing to 0.1 within 2.7/1.4 cM for genotypic/haplotypic data (Barnaud et al., 2010).

In the present study, pairwise genotypic LD measured by $r^2$ was very low (mean $r^2 = 0.019$ and 0.012 for E. camaldulensis and E. tereticornis respectively) (Table 4.19), since the 62 SSRs had an average marker interval of approximately 24.8 cM according to the integrated consensus map of Brondani et al., (2006). Further, the higher number of allele pairs in LD observed in E. tereticornis accessions could be due to the existence of structure in the analyzed population.

The haplotype based interallelic LD, employed in human and animal populations for LD and association analysis (Kirsten et al., 2009; Li and Merila, 2010; Li et al., 2010) was calculated in the two species of eucalypts. The minimum and maximum interallelic $r^2$ value for E. camaldulensis was 0.11 and 0.51 respectively with the mean of 0.19. Similarly, in E. tereticornis, the minimum and maximum interallelic $r^2$ was 0.16 and 0.41 with the mean of 0.25. Results demonstrated more significant locus pairs in haplotypic LD than genotypic LD (Table 4.19) (Figure 4.12 and 4.13). The observed differences may be due to the larger power of independence test with haplotypes than with genotypes. Most of the linked locus pairs with significant genotypic LD showed significant haplotypic LD.
Although limited number of allele pairs showed significant $r^2$, the interallelic LD estimated in *E. camaldulensis* and *E. tereticornis* was high (mean $r^2 = 0.19$ and 0.25), indicating the possibilities of using SSRs for association analysis. The study revealed more significant pairs above the critical values of $r^2 (> 0.1)$, for which strong evidences are assisted (Nordborg *et al.*, 2002; Palaisa *et al.*, 2003; Remington *et al.*, 2001). Similarly, Inghelandt *et al.*, (2011) observed that multiallelic SSRs have four to five times higher mutation rate than SNPs suggesting higher power of SSRs to detect LD than biallelic SNPs, if marker density is ignored. Further, Stich *et al.*, (2006) showed a clear advantage of SSRs over AFLPs to detect LD in a population with short history of recombination. Verhaegen *et al.*, (1998) estimated LD of eucalypt hybrids using RAPD markers and reported significant relationship between the cumulative number of marker alleles in the parents with the full-sib family performance for various traits.

Most of the LD estimation studies in tree species are conducted on candidate genes (Krutovsky and Neale, 2005; Chu *et al.*, 2009; Marroni *et al.*, 2011) and LD estimates in nongenic regions are still unavailable (Ingvarsson, 2010). Further, the development of locus-specific assays for single nucleotide polymorphisms (SNPs) is difficult unless the specific gene sequence is isolated from the particular species. In maize, an outcrossing crop, genome-wide sample of 47 SSRs demonstrated higher levels of LD than SNPs in candidate genes (Remington *et al.*, 2001). In humans, LD detected using microsatellite markers were significantly wider, about 3 Mb apart than those detected using SNPs with only about 0.5 Mb (Varilo *et al.*, 2003). Unlike humans, the linkage disequilibrium (LD) rapidly decays within candidate genes in forest trees (De-qiang and Zhi-yi, 2005). In *Pinus radiata*, LD was tested in full sib families created with limited number of pollen and seed parents (45 individuals), using 34 SSR markers and significant correlation was observed between trait and marker loci (Kumar *et al.*, 2004). Recently, Brazauskas *et al.*, (2011) demonstrated rapid LD decay in Perennial ryegrass (*Lolium perenne* L.) with 40 SSRs and 2 STS.

LD determinations using the HaploView program explained only very few significant LD pairs in both *E. tereticornis* (7 pairs) and *E. camaldulensis* species (10 pairs) (Table 4.19). Cochery-Nouvellon *et al.*, (2009) investigated the four IL10 promoter polymorphisms associated with maternal immune tolerance (early pregnancy loss) employing 2 SNPs and 2 SSRs. Linkage
disequilibrium pattern of the genomic region located between SNPs and the microsatellites were displayed as haplotypes using Haploview program and demonstrated that the polymorphisms of the IL10 gene promoter is related to constitutional risk factors for early (embryonic) pregnancy failure. However, the present study did not display any significant haplotype blocks with the microsatellite data scored as bi-allelic data for the natural population of *E. tereticornis* and *E. camaldulensis*. Though strong evidences of recombination has been demonstrated (multiallelic D' displayed in squares) no significant haplotype blocks were detected in both the species, which may be due to weak LD existing among the genotypes.

5.2.2 HYBRID POPULATION

Hybridity conformation is one of the major issues in the clonal deployment programs. SSR markers are commonly used to address these issues (Bohra *et al*., 2011). Genetic diversity among the selected high yielding F1 hybrids is important for clonal forestry. Moreover, genetic diversity is desirable for any long-term culture improvement, reducing the susceptibility to inbreeding depression (Kantartzi *et al*., 2009). Highly heterozygous loci for the parents were selected for the amplification of hybrid population. Cupertino *et al*., (2011) observed 100% transferability of SSRs developed from *E. grandis* and *E. urophylla* in 112 hybrids of *Eucalyptus* spp. The progenies included individuals from 14 hybrid full sib families of *E. grandis, E. urophylla, E. dunnii, E. globulus*, and *E. camaldulensis*. However, our study revealed about 23.4% transferability to the F1 individuals.

5.2.2.1 Microsatellite Allele diversity

The mean observed and expected heterozygosity values were 0.64 and 0.87 respectively (Table 4.18). A comparison between the observed and expected heterozygosity under the expectation of Hardy-Weinberg equilibrium (HWE) explain that if the observed heterozygosity is lower than expected, we seek to attribute the discrepancy to forces such as inbreeding, which is demonstrated in the F1 population. Hybrids, besides having higher genetic diversity, provide some advantages in breeding program as they combine the best traits of interest of both parental species.
5.2.2.2 Linkage disequilibrium estimation

The rate of LD decay in the population determines the type of association approach to be conducted, whether a genome-wide approach or candidate gene based approach (Flint Garcia et al., 2003). Both the approaches exploit thousands and more markers ending up with utilization of large number of resources for phenotyping, genotyping and also with the statistical issues. One of the possible way of circumvent this problem is the use of F1 derived mapping populations. Efficient genome scan with the exploitation of only a few hundred markers are possible in case of the mapping populations. The advantage of more statistical power to evaluate epistasis is possible in case of these populations, since only two alleles are being evaluated (Flint Garcia et al., 2003; Sorkeh et al., 2008). Myles et al., (2009) examined the pros and cons of the application of association mapping and bi-parental mapping for the dissection of QTLs. The added advantage of F1 mapping population is inflating the frequency of rare alleles (that really contribute to the variation in the trait) and encountering the confounding effects of relatedness (Balasubramanian et al., 2006; Manenti et al., 2009) should be taken into consideration. Moreover, the generation of controlled crosses can break up the covariance between genotypes and phenotypes and enhance power to detect QTL. Though the explanation concluded in recommendation of the use of Nested association mapping (NAM), we have just exploited a related cross of the species for the examination of LD.

The present study exploited F1 hybrids derived from a cross between *E. tereticornis* and *E. camaldulensis* for estimating LD. The population was selected based on the anticipation that LD may extend in the hybrid population compared to the diverse set of association panel. However the estimated genotypic LD measure ($r^2$) was very low although it extended to 11.5 cM. The NLR curve obviously explains the decay of LD in the hybrid population of *Eucalyptus* (Fig 4.14 A). This is similar to that observed in grapevine where genotypic LD extends to about 16.8 cM (Barnaud et al., 2006). Tommasini et al., (2007) reported that LD on chromosome 3B extended up to 0.5 cM in 44 varieties and 30 cM in 240 RIL populations of winter wheat, surveyed with 91 SSR and STS markers. The effect of population on LD is clearly observed in many species including barley, where LD decays within 0.4 kb in wild material and extends up to 212 kb in elite lines (Caldwell et al., 2006). In outcrossing species maize (*Zea mays*), LD
decayed within 1 kb in land races (Tenaillon et al., 2001), within 2 kb in diverse inbred lines (Remington et al., 2001) but extended up to 500 kb in commercial elite inbred lines (Rafalski, 2002; Jung et al., 2004). In outcrossing plant species such as maize where LD typically declines very rapidly, extensive LD can persist when strong artificial selection is maintained (Palaisa et al., 2003). Somers et al., (2007) observed only a small fraction of locus pairs with $r^2$ values > 0.2, but extended for longer distances with higher $r^2$ values in a subpopulation. In the present study similar results were observed in the hybrid population, where the LD decayed rapidly in diverse accessions of association panel, but extended up to 11.5 cM in case if hybrids.

LD in other cross-pollinated species, like douglas fir, maize and ryegrass (Krutovsky and Neale, 2005; Tenaillon et al., 2001; Xing et al., 2007) were compared to our study. Estimated LD values were high, however decayed rapidly in the hybrid population (Figure 4.14). In addition to the factor that eucalypts, is an outcrossing species, LD decay can also vary considerably when divergent materials are crossed. Extensive “admixture LD” can be generated that does not decay as a function of genetic distance (Hamblin et al., 2010). LD decay can also vary considerably from locus to locus due to different recombination rates and selection pressures at different regions of the genome. LD generated in the F1 population might be useful for association mapping in specific/related (preferably E. terticornis X E. camaldulensis) population that ensure the possibility of reduction in number of markers needed for association mapping (Stich et al., 2005; 2006).

The maximum value of $r^2$ (0.86) was observed for the locus pair Embra186 and Embra179 of linkage group 4. Recently, Salazar et al., (2010) observed higher inter-allelic $r^2$ values ($p<0.001$) across the HmB region in Heliconius heurippa and provide molecular evidence for adaptive introgression of red band using 670 SNPs distributed among 29 unlinked coding genes. The proportion of critical $r^2$ values (>0.1) observed were higher (86%). Threshold 0.1 is the minimum $r^2$ value to detect associations for rather large quantitative trait loci (QTLs explaining 10% of the phenotypic variation) with a reasonable population size (Ersoz and Buckler, 2007). About 19 significant pairs with $r^2 > 0.5$ were observed in the present study (Table 4.19) and hence these loci can detect QTL explaining even moderate proportion of phenotypic variation (adventitious rooting in this study) with a reasonable population size.
In the present study significant haplotype block is observed for the hybrid population at Linkage group 4 (Figure 4.15), showing strong evidences of LD between the locus pairs Embra186 – Embra66 and Embra66 – Embra36. About 56 locus pairs revealed significant LD with $r^2 > 0.1$ (Table 4.19). Evidences for strong LD may be due to the recent recombination events occurred in the population. LD decay distances differed among chromosomes in 632 maize inbred lines from temperate, tropical, and subtropical public breeding programs for a whole genome LD scan (Yan et al., 2009) used. A decrease of LD over time in the most important maize hybrids was investigated by Reif et al., (2005) with SSR markers.

5.3 ASSOCIATION ANALYSIS FOR ADVENTITIOUS ROOTING TRAITS

Adventitious rooting traits are appropriate for QTL dissection in Eucalyptus as there is abundant intra- and inter-specific variation (Reuveni et al., 1990), genetic control is reasonable ($h^2 \sim 0.4$) (Borralho and Wilson, 1994) and clonal propagation allows an increased accuracy in trait measurement (Bradshaw and Foster, 1992). The significance of genetic control of rooting parameters for the success of vegetative propagation has been reported. Grattapaglia et al., (1995) positioned 4 QTLs explaining 8.5-26.3 % variation for rooting traits in $E. \text{grandis} \times E. \text{urophylla}$ using RAPD markers. The results evidently explained the effect of small number of major-effect QTLs for a large proportion of variation in the trait. Subsequently, Marques et al., (1999) reported QTLs explaining 2.6–17.0% variation for of the vegetative propagation traits in $Eucalyptus \text{tereticornis}$ and $Eucalyptus \text{globules}$ using AFLP markers. The results remarkably indicated that the phenotypic variation in the trait has a meaningful genetic component and stable QTLs can be establish in a family of reasonable size.

In Corymbia, 12 QTL distributed on four linkage groups were identified, for the ten adventitious rooting and associated propagation traits. Most traits QTL clustered to one LG (CM12) and many of these QTL, were major effects QTL explaining more than 60% of the phenotypic variation (Shepherd et al., 2008). In $E. \text{nitens}$, QTL analysis was performed and found four QTLs for percentage of roots and the percentage of genotypic variation explained by the QTL for percentage of roots produced ranged from 4.9% to 15.4%. The QTL on LG7 explained the highest percentage of genetic variation (11.1%) in roots produced by the stem cutting method. Recently, Thumma et al., (2010) observed four QTLs for percentage of roots
produced by both stem cutting and tissue culture methods. Highest percentage of genetic variation (11.1%) in roots was observed for the QTL on LG7 by the stem cutting method. The percentage of genotypic variation explained by vegetative propagation traits was generally higher than growth traits. Single-marker analysis of vegetative propagation traits revealed four markers that were significantly associated with percentage of roots produced by stem cuttings.

In forest trees, tremendous efforts have been made on QTL mapping using the LD generated through the interspecific hybrids (F₁ and F₂) for several economically important traits (Grattapaglia and Kirst, 2008). Such QTL identification process is a time intensive where the researcher has to wait for many years to assess the phenotypic variations expressed at intermittent stages of growth period. In association mapping approach, LD present in the extant population of interest is exploited and hence it is highly attractive for tree species. Using the approach it may be possible to achieve very high resolution due to the very low extent of linkage disequilibrium. The basis for linkage disequilibrium based association analysis in plant breeding is to ultimately develop markers tightly linked to trait loci or to identify exact causal loci for marker assisted selection (Rafalski, 2010). This can be achieved by whole genome scans using a large number of neutral markers either associated or not associated with a phenotypic trait or by selecting particular genes as candidates for testing more specific associations with putatively correlated phenotypic traits.

One of the goals in the present study was to investigate the potential use of STS markers linked to vegetative propagation QTLs through bi-parental mapping of E. grandis X E. urophylla, E. grandis X E. tereticornis and Corymbia torelliana × Corymbia citiodora subspecies variegata marker identification in E. tereticornis a species belong to Subgenus Symphyomyrtus. The repeatable detection and collocation of QTL for propagation traits in interspecific F₁ of Eucalyptus spp. (Grattapaglia et al., 1995; Marques et al., 1999; 2002; 2005; Shepherdet al., 2008) showed a common genetic basis for vegetative propagation traits in Eucalyptus species. Shepherd et al., (2008) has indicated that species retains the allelic variation at loci affecting propagation-related traits. The present study reports the significant association of four (two STS and two SSR) markers associated with the adventitious rooting trait located on three linkage groups in E. tereticornis (Table 4.16). Embra7 and Embra40 which have been
collocated across 4 species of *Eucalyptus* Marques *et al.*, (2002) were found to be associated with shoot length and rooting parameter respectively. Across the species the linkage group 5 and 10 shows the presence of rooting QTLs. Dasgupta *et al.*, (2010) identified four putative SSR markers linked to the adventitious rooting trait in *Eucalyptus tereticornis*. Non rooting specific alleles (EMBRA 10110, EMBRA 13139, EMBRA 13138 and EMBRA 13135) were significantly correlated with rooting per cent (-0.777).

In crops, the DNA markers linked to QTLs for various traits developed through biparental mapping were utilized in other breeding populations (association populations) for detection of marker-trait linkages using association analysis (Wang *et al.*, 2008; Shi *et al.*, 2010, Zhou *et al.*, 2012). SSR markers linked with important QTLs through conventional QTL mapping strategy has increased the power of association mapping in cotton (Abdurakhmonov *et al.*, 2009) and *Brassica* (Hasan *et al.*, 2008), rice (Zhou *et al.*, 2012), sorghum (Upadhyaya *et al.*, 2012) and many other crops. In *Eucalyptus*, QTLs linked with vegetative propagation traits were identified through SSRs/STS markers and putative QTLs influencing vegetative propagation traits were located on homeologous linkage groups of few species in *Symphyomyrtus* subgenus (Marques *et al.*, 2002).

The repeatable detection and collocation of QTL for propagation traits in interspecific F$_1$ of *Eucalyptus* spp. (Grattapaglia *et al.*, 1995; Marques *et al.*, 1999, 2002, 2005) and closely related *Corymbia torelliana* × *Corymbia citriodora* subspecies variegata (Shepherd *et al.*, 2008) supported a common genetic basis for propagation traits in *Eucalyptus* spp. and some species retain considerable standing allelic variation at loci affecting propagation-related traits. Comparative mapping could be executed and linkage map synteny has been established across multiple pedigrees (Freeman *et al.*, 2006; Hudson *et al.*, 2011; Kullan *et al.*, 2012).

A number of association studies in forestry have been successful. Thumma *et al.*, (2005) demonstrated the association of polymorphisms in Cinnamoyl CoA Reductase (CCR) gene with the variation in microfibril angle (MFA). The study in *Eucalyptus* spp. lead the way for the association studies in forestry species by establishing the feasibility of LD mapping to identify alleles associated with a trait in natural populations. Association genetic study in *Pinus taeda* for the wood property traits were regarded the first multigene association genetic study in forest
trees that has shown the possibility of candidate gene strategy for dissecting complex adaptive traits (Gonzalez-Martinez et al., 2007). Significant associations of SNPs with cellulose content and pulp yield were validated in an E. nitens full-sib pedigree (Thumma et al., 2009). Recently, Southerton et al., (2010) explained the association of allelic variation in xylem genes with wood properties in Eucalyptus nitens. A number of 37 significant associations across 11 candidate genes and 19 traits explaining 1.8 -6.1% phenotypic variation by each polymorphism have been reported in E. globulus (Kulheim et al., 2011).

Most of the studies described display the significance of SNPs in association genetics. On using neutral markers like SSRs for LD estimation it was reported that high ratio of LD between unlinked and adjacent loci for SSR markers is a major disturbing force in gene mapping, which indicates association between markers and genes located on different chromosomes resulting in a high rate of false positive detections of marker trait association (Inghelandt et al., 2011). Moreover SSRs were widely used in various plant species for the genome wide LD estimation and association analysis The use of SSR markers for the significant allelic associations were recommended for early selection of individuals for mass propagation or for clonal testing in Pinus radiata have been reported (Kumar et al., 2004).

In Vitis vinifera an outcrossing perennial species with high diversity, the LD was estimated with SSR markers and demonstrated its importance in genome wide analysis (Barnaud et al., 2006; 2010). Interallelic LD estimation of multiallelic markers like SSRs has high significance in association mapping (Gaunt et al., 2006; Kirsten et al., 2009). In Gossypium hirsutum, SSR markers linked with important fiber QTLs has increased the power of association mapping (Abdurakhmonov et al., 2009). Similarly, Hasan et al., (2008) investigated the potential use of Brassica-Arabidopsis comparative genomics data and found gene-linked SSR markers associated with seed glucosinolate content in oilseed rape. Fifty-nine SSR markers were significantly associated with six fiber traits in an exotic germplasm of Gossypium species (Zeng et al., 2009). Recently Liu et al., (2010) reported 10 SSR markers that were significantly associated with six agronomic traits in wheat and moreover some of the associated markers were in agreement with previous QTL analysis.
Further, SSRs could be more powerful for association mapping if they were available in the genome with the same density as SNP markers. However, the SSR markers employed in the present study would not be adequate for association analysis, because of insufficient marker density for the germplasm evaluated. Although the use of 62 SSR markers is limited for association mapping of QTL analysis in this study, the QTL for adventitious rooting postulated from this research confirmed the QTL region previously identified by linkage analysis. These results indicated that the association analysis is a complementary approach for detection of QTL associated with adventitious rooting. The representation of microsatellite markers screened in the present study is based on *Eucalyptus* consensus map by Brondani *et al.*, (2006) and Thamarus *et al.*, (2002). Several SSR markers have been linked with QTLs for important traits in various eucalypt species using different species combinations. Nevertheless, conserved QTLs have been located on homeologous linkage groups of the taxonomically related species (Marques *et al.*, 2002; Thumma *et al.*, 2010) and several candidate genes co-located to QTL positions controlling different traits (Grattapaglia and Kirst, 2008). Collocation of QTL for different rooting traits was identified in a single region on linkage group (Shepherd *et al.*, 2008).

Low LD in eucalypts promises a higher resolution in genome-wide association mapping, however, many more markers are required to span the whole genome. In *Eucalyptus*, the estimated genetic distance of 1.0 cM was about 385 kb (Thamarus *et al.*, 2002) and hence the SSRs surrounding the genes/QTLs would be a perfect target for LD estimation and association studies. Given the moderate genome size (~ 650 Mb) and the availability of whole genome sequence of eucalypt species, it should be possible to develop high density SSR markers for characterizing the genome. The information generated in the past research on QTL mapping could be used in eucalypts by understanding the pattern and extent of LD in the QTL hot spots. Genome wide association mapping in combination with eQTL data and whole genome marker data will yield significant insight into the genetic architecture of complex traits and help to elucidate the contribution of gene expression to natural trait variation (Ingvarsson and Street, 2011).
5.4 COMPARISON OF LD IN ASSOCIATION PANEL AND HYBRID POPULATION

Pairwise LD measured by $r^2$ based on SSRs was very low because the SSRs used have an average marker interval of 24.8 cM and 30.8 cM (Brondani et al., 2006) for natural and hybrid population respectively. Similar observations were recorded in case of rye grass, which encompassed a very low LD measure (mean $r^2 = 0.01$), when evaluate with 37 SSRs at an average marker interval of 21 cM (Li et al., 2011b). The genotypic LD decayed rapidly in association panel, however extended to 11.5 cM in hybrid population. The low LD detected in the species could be expected, because compared to crop species, perennial outcrossing tree species has a higher effective recombination rate, which leads to a rapid decay of LD (Thumma et al., 2005). Rapid LD decay was reported in many other out-crossing tree species such as Pseudotsuga menziesii, Populus nigra and P. balsamifera (Eckert et al., 2009; Chu et al., 2009; Olson et al., 2010).

Estimation of LD in a panel of 20 diverse genotypes of the Arabidopsis thaliana (selfing species), LD was estimated to decay to $r^2=0.10$ within 250 kb (~1 cM) in the region of the FRI gene (Nordborg et al., 2002) is comparable to our results. Similarly, Bresghello and Sorrells, (2006) observed a consistent LD on chromosome 2D (<1 cM) and concluded the same. The reason for rapid LD decay in the Eucalyptus population depict that the causative polymorphism may very close to the QTL loci and the recombinations are ineffective in breaking down LD between them. Another likely cause is the broad study material selected for our study, which include individuals from various provenances from Australia. The LD decay on the eucalypts representing the whole genome, demonstrates that more than one marker per centi Morgan would be needed to achieve a reasonable power of detection. Consequently it would persuade the fine mapping of QTL.

More number of significant pairs with $r^2 > 0.1$ was observed in case of hybrid population compared to the two species, E. tereticornis and E. camaldulensis in the association population (Table 4.19). This may be due to the relatedness/structure within the hybrid population. To determine the relevance of these alleles ($r^2 > 0.1$) for association analysis, identification of more number of SSR alleles with higher LD may be required. Between the species, accessions from
E. tereticornis showed more number of allele pairs in LD than E. camaldulensis. Recently, Bouchet et al., (2012) exploited the critical value of \( r^2 > 0.3 \) for examining the extent of LD in highly diverse individuals of sorghum using DArT markers. In the present study, \( r^2 > 0.5 \) (95\textsuperscript{th} percentile value for hybrid population) were observed with 19 significant pairs, which enable the possibility of dissecting QTL that explain smaller variation in the phenotype (Table 4.19). The percentage of significant pairs above the threshold value (\( r^2 > 0.5 \)) is higher in case of hybrids compared to those observed for the species in the natural population and demonstrates the fact that the alleles showing significant LD can be possibly exploited for a related population of Eucalyptus, which may include either half-sib or full sib families of E. camaldulensis and E. tereticornis. Threshold 0.1 is the minimum \( r^2 \) value to detect associations for rather large quantitative trait loci (QTLs explaining 10\% of the phenotypic variation) with a reasonable population size (Ersoz and Buckler, 2007). However, threshold 0.3 was considered as the minimum value to enable detection of a QTL explaining around 5 to 10\% of the phenotypic variation (Bouchet et al., 2012). Hence increasing threshold values may contribute to the detection of QTLs explaining smaller phenotypic variations that would results in still finer mapping of the trait.

The number of markers and their density needed for much finer resolution are defined by genome size and LD decay. Hence this will vary considerably among species. For example, while 140,000 markers provide reasonable coverage of the 125 Mb Arabidopsis genome (Kim et al., 2007), a rough estimate suggests that over two million markers will be required to cover the 475 Mb genome of the grapevine, and 10 to 15 million may be necessary for diverse maize varieties. More than 100,000 markers may be required for whole genome scans within sorghum if an \( r^2 \) threshold of 0.1 is considered and 350,000 for a threshold of 0.3 (Bouchet et al., 2012). In the current study, although it is clear that more markers will be required to obtain accurate estimates of the marker density required for associating adventitious rooting trait with accounting the genome wide LD, we roughly estimated that 17 million markers may be required for whole genome scans within Eucalyptus if an \( r^2 \) threshold of 0.1 is considered and 85 for a threshold of 0.5 (590 Mb).
Brazauskas et al., (2011) recommended two strategies for the high resolution association studies with the species encountering a rapid decay of LD extending to few cM or bp. A genome-wide association study (GWAS) with several hundred markers may be relevant for synthetic varieties which direct the identification of large genomic regions linked with the trait (up to 10 cM) or addition of new markers to the current set are needed to accomplish resolution in elite breeding materials. However, higher marker density of several million SNPs might be required for impending outcome in QTL associated marker within a set of highly diverse individuals.

LD determinations using the HaploView program explained only very few significant LD pairs in both *E. tereticornis* and *E. camaldulensis* species (Table 4.19). No significant haplotype blocks were observed in both the species, which may be due to weak LD existing among the genotypes. But in case hybrid population, significant haplotype block is observed at Linkage group 4 (Fig. 4.15), showing strong evidences of LD between the locus pairs. Coupled with the LD blocks found, the higher level and decay of LD observed on the LG4 suggests that LD mapping may be possible on the LG 4 in a related population using low-density marker maps.