Forest tree breeding is basically aimed at producing quality products to solve specific problems or to produce specially desired product. The major limitation to tree breeding is the time required for one cycle of breeding and selection. Molecular markers could be quite effective in reducing the time of selection in breeding programs in two ways: either by early identification of superior genotypes or by identification of parents that will yield superior progeny. Resin is a valuable non wood forest product obtained from Chir pine. The early identification of high resin yielding genotypes could be a valuable contribution towards proper utilization of this natural resource.

2.1 Resin tapping from pines

Most *Pinus* species bleed when the stem wood (xylem) is cut or otherwise injured, but probably only a few dozen out of approximately hundred species have ever been tapped commercially as a source of resin for rosin and turpentine production. In the other species, poor yields and/or quality of the resin makes the exploitation uneconomic. Resin tapping system was started in fifteenth century as a naval stores industry in America and in India it started in 1896 (Chaudhari, 1995).

2.1.1 *Pinus* species tapped for resin

*Pinus* species are tapped for resin in several countries across the world. Generally the species which are tapped at a commercial scale are invariably chosen from those existing within the country either as natural stands or in the form of plantations (Coppen and Hone, 1995). The principal species of *Pinus* which are tapped in different countries include *P. elliottii* (Brazil, Argentina, South Africa); *P. massoniana* and *P. kesiya* (People’s republic of China); *P. pinaster* (Portugal); *P. merkusii* (Indonesia);
*P. oocarpa* (Mexico, Honduras); *P. caribaea* (Venezuela); *P. sylvestris* (Russia); *P. halepensis* (Greece); and *P. radiata* (Kenya).

### 2.1.2 Pinus species tapped in India

India has been producing naval stores for a long time. Both natural forests as well as plantations of *P. roxburghii* have been used for resin tapping in the northern states of Jammu and Kashmir, Uttarakhand and Himachal Pradesh. *P. wallichiana* grows at higher elevations along the same Himalayan belt but as it gives lower yields of resin as compared to *P. roxburghii*, it is rarely tapped for commercial purposes (Coppen and Hone, 1995).

### 2.1.3 Methods of resin tapping

Different methods are practiced for resin tapping across the world such as cup and lip method, Silva hill basula method, bark chipped method, rill method and bore hole method. The rill method of tapping is considered as economically more viable and biologically acceptable. The modern rill method, which was introduced in 1983 to replace the traditional “cup-and-lip” method, was not better as far as damage to trees was concerned. Therefore, bore-hole method was introduced as a new technique by Dr YS Parmar University of Horticulture and Forestry, Nauni, to improve the quality of resin and to minimize the damage to pine trees (reported by Tribune News service, Feb, 2011).

Method of tapping plays a significant role in determining the resin yield. The turpentine content and its relative flow rate were found to be significantly higher in borehole method of tapping as compared to the rill method. It was reported that the use of bore-hole method (15.24cm depth and 2.54cm diameter) followed by 10% ethephone with sulphuric acid (upto 25%) treatment resulted in higher yield and minimum damage to the *P. roxburghii* trees (Rawat, 2000). Later it was reported that freshening up of bore hole improves the resin yield in *P. wallichiana*. Freshening at forty-five days interval significantly affected the oleoresin yield (Sharma *et al.*, 2005).
India, although tapping entails removal of wood from the tree in the form of ‘blazes’, the ‘rill’ method, in which an acid-based stimulant is applied to small channels cut in the xylem in a ‘herring bone’ fashion, is advocated (Coppen and Hone, 1995).

2.1.4 Factors affecting resin yield

There are several factors which determine the resin yield such as species used for tapping (yield varies from species to species), tree morphology (trees with bigger crown size/crown height and twisted trees yield more resin), method of tapping (cup and lip method yields less than rill method whereas in bore hole method yield is much higher), aspect (more resin is secreted on a warmer or sun facing aspect), nature of forest (more yield in open forest as compared to crowded forest), slope of the area (higher yield in slopes as compared to plains), tree diameter, environmental factors and genetic factors.

In a study, it was found that the size and frequency of vertical resin canals does not have any bearing on the variation in resin yield (Anon, 1972). The tapping of resin for *P. roxburghii* trees of four diameter groups (20-30cm) and four blaze width groups (8-20cm) was studied in Himachal Pradesh using rill method (Brahmi *et al.*, 2000). It was found that resin yield increased with increase in diameter and blaze width and also as blaze width increased within the blaze group and as diameter increased within the diameter class. The yield was found to be maximum in the month of May and minimum in October. In another study, it was observed that more than 50% of the clonal variation in resin yield was due to genetic factors (Tadesse *et al.*, 2001 a). In twenty-six high resin yielders and ten check trees of *P. roxburghii*, the genotypic, phenotypic and environmental correlations were studied for wood and oleoresin characters in Himachal Pradesh (Nimkar *et al.*, 2003). The genotypic correlation coefficients were found to be higher in magnitude for all the combinations as compared to those at phenotypic and environmental levels. It was found that constitutive and induced resin yield in loblolly pine (*P. taeda*) were affected by fungal inoculation (with the southern pine beetle associated fungus *Ophiostoma minus*). Mass wounding (400 wounds m$^{-2}$) and
inoculated trees produced higher induced resin yields than control or wounded-only trees (Klepzig et al., 2005). The number of resin ducts, diameter of resin ducts, needle length, bark percentage of wood, height and diameter were found to contribute directly to the oleoresin yield in *P. roxburghii*, indicating their direct involvement and relative importance for the increase in the ultimate oleoresin yield (Nimkar and Sharma, 2005).

2.2 Tree improvement in *P. roxburghii*

Tree improvement work in India was started by Prof. H.G. Champion who realized the importance of geographical variations and application of the knowledge of forest genetics during 1930. He established a provenance trial of chir pine at New Forest, Dehradun and found that spiral gains in chir pine are inherited. In a study on half sib progeny of chir pine, located at Solan comprising of 58 families from the state of Himachal Pradesh (Western Himalaya), some growth and wood characters were analyzed (Sehgal et al., 1995). Significant family and within family differences were observed for all the traits studied. Heritability and genetic gains were also determined. The monoterpenes variation in stem oleoresin of half-sib progenies of high resin yielders were studied by Sharma et al. (1998) in chir pine. Sharma, (2001) reviewed the oleoresin tapping of pines in India. An improved method for the development of seed production areas of chir pine was proposed by Gera, (2002). Roy et al. (2004) studied the seed source variation in cone, seed and seedling characteristic across the natural distribution of Himalayan low level pine *P. roxburghii*.

Sharma et al. (2006) reviewed the attempts made on various aspects of genetic improvement of chir pine in the country including the studies on phenotypic variation, plus tree selection, progeny tests, variation in cone and seed characters and gene marker studies specifically highlighting the work on isozyme analysis of the provenances and plus trees. The high resin yielding trees and check trees of chir pine were evaluated for anatomical characters (Nimkar and Sharma, 2006). An assessment of heterozygosity and fitness in chir pine was carried out using isozymes by Sharma et
al. (2007). Studies are also available on genetic variation among
provenances/genotypes in *Pinus* and other conifer species. Variation studies were
carried out in provenances and plus trees of *P. roxburghii* by Sharma, (2007). The
additive genetic variation in seedling growth and biomass of fourteen *P. roxburghii*
provenances from Garhwal Himalaya was evaluated by Ghildiyal et al. (2009a). The
 genetic variation in wood specific gravity among half-sib progenies of chir pine was
studied (Raj et al., 2010). Recently, the seasonal influence on the rooting response of
chir pine was assessed (Sharma and Verma, 2011)

2.2.1 Selection of plus trees for high resin yield

Under various tree improvement programs in the country, attempts have been
made to select plus trees for growth and high resin yield. A total of one hundred and
fifteen plus trees of *P. roxburghii* have been reported for high resin yield by Gupta et al.
(1970). Fifty-eight plus trees of *P. roxburghii* were selected for growth and ten for resin
yield (Dogra, 1985). Forty-two higher growth genotypes were identified from seven
provenances from Himachal Pradesh (Bhalaik et al., 2000). Forty nine plus trees were
selected for growth form and high resin yield from the distributional range of chir pine
falling in different forest divisions of Himachal Pradesh (Chauhan and Kanwar, 2001).
Different traits including oleoresin yield, total height and diameter were measured in *P.
taeda* (Roberts et al., 2003). All traits studied were found to be highly genetically
variable and to have much greater additive than dominant variance. It was observed
that the additive genetic correlations between oleoresin yield and growth traits were
positive and moderately high in *P. taeda* suggesting that directional selection to
improve growth in loblolly pine will also result in increased production of oleoresin.

2.3 DNA markers in tree improvement

DNA based markers are popular means for identification and authentication of
plant species. The power of discrimination of DNA based markers is comparatively high
and hence very closely related varieties can be differentiated. The most commonly used
DNA based markers are Random Amplified Polymorphic DNA (RAPD), Inter Simple Sequence Repeats (ISSRs), Amplified Fragment Length Polymorphisms (AFLPs), Restriction Fragment Length Polymorphisms (RFLPs) and Microsatellites or Simple Sequence Repeats (SSRs). Most of these techniques rely on the advantages of PCR detection, including speed, sensitivity and selectivity except RFLP. The major limitation with RAPD markers is their low reproducibility.

2.3.1 Use of DNA markers to assess genetic diversity

Research on genetic diversity is important for developing strategies for breeding and provides an understanding for the scientific and sustainable management of genetic resources in forest tree species (Poltri et al., 2003). Generally the aim of long term breeding is to achieve a right balance between continuous genetic gains and maintenance of substantial genetic variation. Recurrent selection, including choosing the best trees in the current population as parents for the next generation is most commonly adopted method for long-term forest tree breeding. This cycle is repeated every generation to increase the frequency of favourable alleles at loci that control economically important traits (Namkoong et al., 1988). In hybridization programs also best parents are identified for crossing to optimize the expression of genes of interest (Ruas et al., 2003).

Traditionally, morphological observations and progeny tests were used as major tools for the analysis of genetic diversity. The limitation of this traditional approach was that it failed to reveal the exact inherent similarity among the individuals because most of the morphological characters are plastic and influenced by environmental factors. For the past two decades, DNA markers have become a key strategy for the study of genetic diversity in tree species and they provide an increasingly accurate assessment of taxonomic relationships and history of gene flow. The variation within the populations of tree species has been exploited for the selection of superior provenances and for evolving strategies for conservation of genetic diversity. Unlike morphological traits which are subjected to environmental effects, data from DNA
fingerprinting provide direct measure of genetic variation (Brown and Moran, 1981). The characterization is possible in terms of alleles, gametes or genotypes.

2.3.2. Genetic diversity studies in pines

Various markers have been used for estimating genetic diversity levels in different Pinus species. Reports are available for DNA based markers such as isozyme, RAPD, ISSR, AFLP etc.

2.3.2.1 Isozymes

Isozymes are the multiple forms of enzymes (proteins) produced by living tissues. Since enzymes are primary gene products, variation in their structure gives reliable information about the variability in the genome. Isozymes have been used as genetic markers in many facets of plant breeding and genetics including studies of provenance identification (Muhs, 1974) and can be helpful in selecting breeding populations, seed stands and for conservation of biodiversity. Isozymes can serve as convenient and reliable gene markers allowing for observation of allelic variation at individual gene loci and thus, facilitating a reliable estimation of genetic variation in populations and species.

Unlike other pine species not much work on isozyme analysis in chir pine has been done. In allozyme variation studies carried out among twelve populations of P. roxburghii from Pakistan, 49.6 per cent loci were reported to be polymorphic (Hussain, 1995). Eight natural populations of P. roxburghii covering Shiwalik and Himalayan ranges of species distribution in Himachal Pradesh were analyzed isozymatically at eleven enzyme systems (Sharma, 1999). For the enzyme systems studied, twenty-five gene loci were identified out of which eighteen were polymorphic. In another study, variation in allozymes was studied in thirty-three plus trees in seven provenances of P. roxburghii from the state of Himachal Pradesh (Sharma, 2003). Isozyme variation studies carried out in the species suggested that the differences are attributed to the
adaptations of the populations/provenances to the local environment over a period of time.

2.3.2.2 Random Amplified Polymorphic DNA (RAPD)

Among different types of DNA based markers used for genetic diversity analysis, RAPD analysis is fast and easily generated by PCR (Williams et al., 1990). Advantages of RAPDs include suitability of work on anonymous genomes, high efficiency and low expense. This technique overcomes many of the limitations of RFLP and has shown promise in diversity estimations in many plant species (Gupta et al., 1999) including forest trees (Skabo et al., 1998).

RAPD markers have proven to be very effective in assessing the genetic diversity in East Asian Pinus species (Kim and Lee, 1998; Nkongolo et al, 2002); P. halepensis (Gomez et al., 2001); P. massoniana (Peng et al., 2003); P. brutia (Kandedmir et al., 2004); P. koraiensis (Kim et al., 2005); P. squamata (Zhang et al., 2005); P. sylvestris (Naugzemys et al., 2006), P. pinea (Cuesta et al., 2010); P. strobus and P. monticola (Mehes et al., 2007).

In P. roxburghii, the use of RAPD markers for genetic variability analysis is limited to only two reports. Malabadi and Nataraja (2007) highlighted the key role of peroxidases influencing somatic cells towards embryogenic pathway and assessed the clonal identity of somatic seedlings using RAPD and ISSR markers. The lowest peroxidase activity was observed in embryogenic cultures on maintenance medium showing elongated cells with cleavage polyembryony as compared to control. Highest peroxidase activity was observed in non-embryogenic cultures showing round, globular and oval cells. RAPD and ISSR analysis found no evidence of genetic variation either within or between the embryogenic lines established from three of these trees, or between these lines and the trees of origin, or between somatic embryos derived plantlets and the trees of origin.
In another report, Ginwal et al. (2010) used RAPD markers to reveal genetic variation in fifty-five geographically distinct populations of *P. roxburghii* of the Himalayan region, covering the entire natural range of distribution of the species in India. Genetic similarity analysis showed considerable variation among the populations ranging from highest similarity between the populations of Una (Himachal Pradesh; HP) and Jamta (HP); Una (HP) and Manikaran (HP) with 78% similarity. Highest dissimilarity was observed between the populations of Gaucher (Uttarakhand) and Akhnoor (Jammu and Kashmir) with only 36% similarity (64% dissimilarity). The results reveal that there is a fair amount of genetic variability in the existing forests of Himalayan chir pine (*P. roxburghii*) and as such the information is valuable for initiating conservation and genetic improvement efforts in this species.

### 2.3.2.3 Inter Simple Sequence Repeats (ISSR)

Inter simple sequence repeats (ISSR) technique is a PCR based method, which involves amplification of DNA segment present at an amplifiable distance between two identical microsatellite repeat regions oriented in opposite direction (Zietkiewicz et al., 1994). The technique uses microsatellite, usually 16-25 bp long, as primers in a single primer PCR reaction targeting multiple genomic loci to amplify mainly the inter-SSR sequences of different sizes. The microsatellite repeats used as primer can be di, tri, tetra or penta-nucleotides. ISSR markers overcome the problem of low reproducibility of RAPD markers and are highly polymorphic so they are used in studies on genetic diversity, phylogeny, gene tagging, genome mapping and evolutionary biology (Reddy et al., 2002).

ISSR markers have been used for genetic diversity studies in different species such as *P. densiflora* (Hong et al., 2004; Hong et al., 2007); *P. squamata* (Zhang et al., 2005); *P. sylvestris* (Labra et al., 2006); *P. strobos* and *P. monticola* (Mehes et al., 2007); *P. banksiana* and *P. resinosa* (Ranger et al., 2008); *P. pinea* (Cuesta et al., 2010) and *P. nigra* (Moraga et al., 2012).
Zhang et al. (2005) revealed very low genetic variation in *P. squamata*, one of the most endangered conifers in the world using two classes of molecular markers, RAPD and ISSR. Shannon’s index of phenotypic diversity ($I$) was 0.030; the mean effective number of alleles per locus ($Ae$) was 1.032; the percentage of polymorphic loci ($P\%$) was 6.45; and the expected heterozygosity ($He$) was 0.019 at the species level based on RAPD markers. The results of ISSR were consistent with those detected by RAPD but somewhat higher ($I = 0.048$, $Ae = 1.042$, $P\% = 12.3$, $He = 0.029$).

Hong et al. (2007) analyzed 8eighty ISSR variants from 192 individuals in eleven populations to monitor the level and distribution pattern of genetic diversity in countrywide populations of Japanese red pines in Korea. Relatively higher level of genetic diversity was observed in nineteen populations of Japanese red pines (mean of 0.453) than those in other tree species.

Ranger et al. (2008) determined the levels of genetic diversity in jack pine and red pine populations growing in metal contaminated and uncontaminated areas in northern Ontario (Canada) using ISSR markers. Newly introduced populations were compared to 40 to 60 old populations. For jack pine, the percentage of polymorphic loci ($P\%$) ranged from 14.6 to 45.8% with a mean of 31.6%. Nei’s gene diversity ($h$) varied from 0.046 to 0.169 with an average of 0.100, and Shannon’s index ($I$) ranged from 0.070 to 0.250 with an average of 0.153. The level of genetic variation was much lower in the red pine populations. For this species, the level of polymorphic loci varied from 4.55 to 27.27%. The mean for Nei’s gene diversity and Shannon’s information index, were 0.034 and 0.053, respectively. Overall, the genetic distance among the *P. banksiana* populations revealed that all the populations analyzed were genetically close to each other. There was no association between metal accumulation and genetic diversity for both species.

Apart from genetic diversity studies, ISSR markers have also been used to develop species specific molecular markers in pines. Smith et al. (2010) developed and physically mapped species-diagnostic and species-specific molecular markers in pine
and spruce. Five RAPD and one ISSR species-diagnostic or species-specific markers for *Picea mariana*, *Picea rubens*, *P. strobus*, and *P. monticola* were identified, cloned, and sequenced.

Moraga-Suazo *et al.* (2012) tested two ISSR markers and twelve selective amplifications of microsatellite polymorphic loci (SAMPL) primer combinations using a first-generation full-sib family of eighty-six individuals in order to increase the genetic molecular markers in *P. radiata*. A total of eighteen polymorphic fragments were found for two ISSR primers tested, with an average segregation distortion of 33%.

In *P. roxburghii*, there are only two reports on the use of ISSR markers. Malabadi and Nataraja (2007) highlighted the key role of peroxidases influencing somatic cells towards embryogenic pathway in *P. roxburghii*. Further the clonal identity of somatic seedlings was assessed using RAPD and ISSR markers. Recently, Parasharami and Thengane (2012) used ISSR markers to carry out inter-population genetic diversity analysis between four populations of *P. roxburghii* from India (two native/natural populations from sub-tropical forests and two provenance trial plantations in plains). The genomic DNA isolated from mature needles of a total of 72 accessions from these four populations was subjected to ISSR analysis. The per cent band polymorphism (P %) was found to be 100 per cent, observed number of alleles (nₒ) was 2, effective number of alleles (nₑ) was 1.30, Nei’s Gene diversity/ Heterozygosity (h) was found to be 0.19 and Shannon’s information index (I) was 0.31. For these 4 populations, Gene flow (Nₘ) was estimated to be 1.39. The highest genetic distance of 0.0904 was obtained between native and provenance trial plantation. The clusters in the dendrogram were found to be separated at 0.3 to 0.4 similarity values, indicating robust separations.

### 2.3.2.4 Amplified Fragment Length Polymorphism (AFLP)

AFLPs are PCR based markers for the rapid screening of genetic diversity. AFLP methods rapidly generate hundreds of highly replicable markers from DNA of any organism and can be manipulated by choosing the right kind of primers and changing
their selective bases (McGregor et al., 2000); thus they allow high-resolution genotyping of fingerprinting quality. The time and cost efficiency, replicability and resolution of AFLPs are superior or equal to those of other markers such as allozymes, random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) and microsatellites, except that AFLP methods primarily generate dominant rather than co-dominant markers. Vos et al. (1995) described this novel DNA fingerprinting technique. The AFLP technique was based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. The technique involved three steps: (i) restriction of the DNA and ligation of oligonucleotide adapters, (ii) selective amplification of sets of restriction fragments, and (iii) gel analysis of the amplified fragments. PCR amplification of restriction fragments was achieved by using the adapter and restriction site sequence as target sites for primer annealing. The selective amplification was achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotides flanking the restriction sites. The AFLP technique provided a novel and very powerful DNA fingerprinting technique for DNAs of any origin or complexity.

The first report of the use of AFLPs in trees was by Cervera et al. (1996) who used this marker system to genetically map a disease resistance gene in Populus. Because of their high replicability and ease of use, AFLP markers have emerged as a major new type of genetic marker with broad application in systematics, pathotyping and population genetics (Mueller and Wolfenbarger, 1999).

Cato et al. (1999) developed protocols to generate reproducible AFLP profiles in P. radiata and evaluated the inheritance and informativeness of AFLP markers in this important timber species. The large genome size of P. radiata necessitated increased levels of selection at both the pre-amplification and selective amplification steps of the AFLP protocol to generate reproducible AFLP profiles. Once optimized, approximately 41.3 scorable AFLP bands were resolvable through denaturing gels, of which 48.4%
were polymorphic in a screen of eight unrelated trees. This level of polymorphism is almost three times higher than that observed with RAPD markers.

Lerceteau and Szmidt (1999) analyzed the properties of AFLP markers in *P. sylvestris*. Using primers with three selective nucleotides, the AFLP protocol produced large number of amplified bands and could only be used with a restricted number of primer combinations. Replacement of the EcoRI+3 primer by an EcoRI+4 primer halved the number of bands, facilitating analysis. The inheritance of all but about 8.4% of the amplified bands was confirmed to be Mendelian.

In *Pinus* species, there are several reports on the use of AFLP markers for genetic diversity studies. Mariette *et al.* (2001) analyzed twenty-three populations (13 Aquitaine and 10 Corsican) of *P. pinaster* at three microsatellite loci and 122 AFLP loci. The aims of the study were: (i) to compare levels of within-population and among-population diversity assessed with both kinds of markers; (ii) to compare Aquitaine and Corsican provenances of *P. pinaster*; and (iii) to know if both markers gave the same information for conservation purposes. Even though microsatellites showed higher within-population diversity, they showed the same level of differentiation as AFLP markers. Moreover, both markers also showed a higher genetic diversity in the Aquitaine provenance and a higher differentiation among Corsican populations. AFLPs and microsatellites gave different population diversity rankings.

Diaz *et al.* (2001) used RAPD and AFLP markers to assess the genetic variation among ten populations of *P. oocarpa* from three geographical regions of Nicaragua. Both markers revealed high levels of diversity in these populations.

Ribeiro *et al.* (2002) compared the genetic variation of *Pinus pinaster* populations using AFLP and chloroplast simple-sequence repeat (cpSSR) loci.

gene diversity with cpSSR markers were found both at population and species level, while median differentiation and higher gene diversity was observed with AFLP data.

Xu et al. (2008) developed and used AFLP markers to study genetic diversity and its structure in shortleaf pine (P. echinata) and loblolly pine (P. taeda). After screening 48 primer pairs, 17 and 21 pairs were selected that produced 794 and 647 AFLPs in shortleaf pine and loblolly pine, respectively. High-AFLP-based genetic diversity existed within shortleaf pine and loblolly pine, and most (84.73% in shortleaf pine; 87.69% in loblolly pine) of this diversity was maintained within physiographic populations. For shortleaf pine, there was no significant correlation between geographic distance and genetic distance (r=0.28), while for loblolly pine there was a weak but significant correlation (r=0.51).


Parchman et al. (2011) investigated genetic structuring across twenty-two stands of lodgepole pine (P. contorta) in the central Rocky Mountains (USA) and in isolated peripheral populations that experience different seed predators and exhibit parallel divergence in cone traits using a set of nine simple sequence repeats and 235 AFLP loci. The analysis revealed high levels of genetic diversity within and low genetic differentiation among populations. Nonetheless, geographic and genetic distances were correlated, and isolated populations to the east and west of the Rocky Mountains had higher levels of differentiation than the populations in central part of the range.

Kim et al. (2011) used AFLP markers to assess the genetic diversity and structure among 15 populations across the native range of western white pine using 66 polymorphic AFLP loci. Nei's gene diversity (h) at the population level ranged from
0.187 to 0.316. Genetic differentiation ($G_{ST}$) indicated that 20.1% of detected genetic variation was explained by differences among populations.

In two other studies, Aquea and Arce-Johnson, 2008 used cDNA-AFLP analysis to study gene expression underlying the early embryogenic process in *P. radiata* and Richardson *et al.* (2009) analyzed the molecular (AFLP) and quantitative genetic data to demonstrate the existence of congruent climate-related patterns in western white pine (*P. monticola*). The results suggested that divergent climatic selection has influenced phenotypic traits associated with growth. Furthermore, the congruence of AFLPs suggested a linkage between some loci and genomic regions under selection and/or climatic influences on post mating reproductive success.

In *P. roxburghii*, till date there is no report on the use of AFLP markers for any genetic studies.

### 2.3.2.5 Microsatellites or Simple Sequence Repeats

Microsatellites, also known as Simple Sequence Repeats (SSR) were first developed for use in genetic mapping in humans (Litt and Luty, 1989; Weber and May, 1989). SSRs are short tandemly repeated sequence motifs consisting of a repeat unit of 1-6 bp in length (Tautz and Schlotterer, 1994). The di-nucleotide repeat AC is commonly found in *Pinus* genomes. Since there is a great variation in the number of tandem repeats at a locus, SSR markers tend to be amongst the most polymorphic genetic marker types. They are highly polymorphic DNA markers with discrete loci and co-dominant alleles. Because of their ubiquitous and uniform distribution within both eukaryotic and prokaryotic genomes (Field and Wills, 1996), microsatellites are currently the markers of choice for cultivar and variety identification (Faria *et al.*, 2000; Cantini *et al.*, 2001), genome mapping (Brondani *et al.*, 2002; Sharopova *et al.*, 2002), population genetic analysis (Mariette *et al.*, 2002b, Deutech *et al.*, 2002), genetic diversity study (Kisuki and Isagi, 2002), and evolutionary genetics (Zhang and Hewitt, 2003; Selkoe and Toonen, 2006; Ellis and Burke, 2007). They are also characterized with
high information content, ease of genotyping through PCR, high discrimination power and high degree of allelic variability which is presumably caused by a high mutation rate and replication strand slippage changing microsatellite array length.

Nuclear SSRs (nSSRs) have been developed for several forest trees, including the species of *Pinus* (Smith and Devey, 1994; Kostia *et al.*, 1995; Echt *et al.*, 1996; Echt and May-Marquardt, 1997; Pfeiffer *et al.*, 1997; Fisher *et al.*, 1998), *Quercus* (Dow *et al.*, 1995); *Picea* (Ven and McNicol, 1996) and *Populus* (Dayanandan *et al.*, 1998).

Since the complete DNA sequence of the entire chloroplast genome of *P. thunbergii* was known (Wakasugi *et al.*, 1994), therefore some of the SSR markers developed in pines were from the chloroplast genome (Powell *et al.*, 1995; Cato and Richardson, 1996; Vendramin *et al.*, 1996). It has been reported that the microsatellite flanking regions in which primers are placed evolve much more slowly, and therefore might permit cross-species amplification in species other than the one for which it has been designed (Peakall *et al.*, 1998; Karhu *et al.*, 2000; Turpeinen *et al.*, 2001). Cross species or trans-specific amplification is a valuable tool in pines as de novo development is difficult in the large, highly duplicated conifer genome (Kostia *et al.*, 1995; Soranzo *et al.*, 1998; Joyner *et al.*, 2001; Mariette *et al.*, 2001).

### 2.3.2.5.1 Development of SSRs in *Pinus* species

Microsatellites have been developed for different *Pinus* species including *P. thunbergii*, *P. taeda*, *P. resinosa*, *P. densiflora*, *P. merkusii*, *P. radiata* and *P. pinaster*. They have been used to analyze the level of genetic variation among the populations and also tested for cross species amplification.

Vendramin *et al.* (1996) designed twenty primer pairs flanking the mononucleotide stretches in *P. thunbergii*. The primers were then used to amplify DNA samples extracted from *P. leucodermis*. The amplification was successful using all primer pairs and the size of fragments was in expected size range. Sequences flanking the microsatellites were highly homologous between *P. thunbergii* and *P. leucodermis.*
This confirmed the highly conserved nature of the chloroplast genome in conifers and the usefulness of the chloroplast microsatellite approach in detecting the variation in the genome. The universality of the cpSSR markers was also tested by testing the twenty primer pairs in other *Pinus* species, belonging to different taxonomic classification (Farjon, 1984), as well as in species of Pinaceae belonging to different genera using the same experimental conditions. In the tested *Pinus* species, all the primer pairs produced amplification products similar in size to those detected in *P. leucodermis* (100% of successful amplification in *P. brutia* Ten, *P. halepensis* Mill, *P. pinaster* Ait, *P. pinea* L, *P. sylvestris* L.). The rate of successful amplification in other conifers was recorded to be 65% in *Abies alba* Mill, 75% in *C. atlantica* (Endl.) Maneti and 75% in *Picea abies* (L.) Karst. The results therefore indicated that cpSSR markers can be amplified and utilized as genetic markers in relatively distant species using heterologous primers, increasing the utility of these markers.

Echt and marquardt (1997) probed a large insert genomic library from eastern white pine (*P. strobus*) for the microsatellite motifs (AC)$_n$, and (AG)$_n$, ten tri-nucleotide motifs, and twenty-two of the thirty-three possible tetra-nucleotide motifs. For comparison with a species from a different subgenus, a loblolly pine (*P. taeda*) genomic library was also probed with the same set of di- and tri nucleotide repeats and eleven of the tetra-nucleotide repeats. The four most abundant microsatellite motifs in both species were (AC)$_n$, (AG)$_n$, (AAT)$_n$, and (ATC)$_n$, which as a group accounted for over half the microsatellite sites investigated. The two di-nucleotide repeats were the most abundant microsatellite motifs tested in both species, each at 24.5 sites/megabase pair (Mbp), but the two tri-nucleotide motifs were nearly as abundant and were considered good candidates for pine microsatellite marker development efforts. Eastern white pine had more than twice as many (AC)$_n$, as (AG)$_n$, loci, in contrast with loblolly pine and most other plant species in which (AG)$_n$, is more abundant. In both *Pinus* species, the minimum estimated genome density for all microsatellites, excluding (AT)$_n$, repeats, was 16 sites/Mbp.
Vendramin and Ziegenhagen (1997) identified and sequenced two polymorphic microsatellite loci in the genus *Abies*, using primer pairs derived from cp SSRs of *P. thunbergii*. PCR products exhibited considerable length variation among six different *Abies* species and within *Abies alba*. F₁ progeny of both interspecific and intraspecific reciprocal cross confirmed that the two SSRs were predominantly paternally inherited.

Fisher *et al.* (1998) isolated dinucleotide microsatellites from *P. radiata* using both a standard genomic library and libraries enriched for microsatellites. They designed locus-specific primers to amplify forty-three unique microsatellites. Thirty-two of these loci had interpretable PCR patterns, eleven of which were polymorphic in a screen of nineteen *P. radiata* individuals. All eleven polymorphic loci contained at least seventeen repeats in the sequenced plasmid. Six of the eleven primer pairs amplified multiple fragments per individual (3-8), suggesting that these loci were present in multiple copies in the genome. Genotyping a *P. radiata* production population consisting of forty-eight trees with seven of the most polymorphic microsatellites revealed an average of seventeen bands per locus (the multi-copy microsatellites were treated as one locus). When tested on known pedigrees, both single and multi-copy microsatellites exhibited co-dominant inheritance and Mendelian segregation. Two loci had null alleles and one locus had a high frequency of non-parental alleles, suggesting a high mutation rate. Eight of these microsatellites, including five multi-copy loci, were placed on a partially constructed *P. radiata* genetic map. Four of the five multi-copy microsatellites had two or more sets of alleles that mapped to the same locus, and the fifth mapped to two unlinked loci. All seven tested primer pairs amplified PCR products from other species of hard pine, three amplified products from soft-pine species, and one amplified bands in other conifers.

Devey *et al.* (2002) developed and characterized fifty microsatellite markers in *P. radiata*, and from among these, a subset of ten easily scored and highly polymorphic markers was selected for use in fingerprinting, quality control and breeding applications. The markers were characterized based on reliable and reproducible
amplification, observed and expected heterozygosities, number of alleles, low frequency of null alleles and lack of close linkage with other selected markers. Allele numbers and frequencies were estimated using twenty-four first generation breeding clones from Australia and New Zealand. Observed heterozygosities for the selected markers were all greater than 0.67, and there was an average of 10.5 alleles per locus. The ten markers were not closely linked to each other.

Zhou et al. (2002) used undermethylated (UM) DNA fragments to construct a microsatellite-enriched library to improve the efficiency of developing P. taeda microsatellites. A methylation-sensitive restriction enzyme, McrBC, was used to enrich for UM DNA before library construction. Digested DNA fragments larger than 9 kb were then excised and digested with RsaI and used to construct nine di-nucleotide and tri-nucleotide libraries. A total of 1,016 microsatellite-positive clones were detected among 11,904 clones and 620 of these were unique. Of 245 primer sets that produced a PCR product, 113 could be developed as UM microsatellite markers and 70 were polymorphic. Inheritance and marker informativeness were tested for a random sample of 36 polymorphic markers using a three-generation outbred pedigree. Thirty-one microsatellites (86%) had single-locus inheritance despite the highly duplicated nature of the P. taeda genome. Nineteen UM microsatellites had highly informative intercross mating type of configurations. Allele number and frequency were estimated for eleven UM microsatellites using a population survey. Allele numbers for these UM microsatellites ranged from 3 to 12 with an average of 5.7 alleles/locus. Frequencies for the 63 alleles were mostly in the low range; only 14 of the 63 were in the rare allele (q < 0.05) class. Enriching for UM DNA was an efficient method for developing polymorphic microsatellites from a large plant genome.

Guevara et al. (2005) developed 20 polymorphic nSSRs from P. pinaster using genomic libraries enriched with repeated motifs. They used a total of twenty-five individuals from five different populations to estimate genetic diversity parameters.
Boys et al. (2005) isolated and characterized thirteen nSSR loci by screening a partial genomic library with di-, tri-, and tetranucleotide repeat oligonucleotide probes. In an analysis of over 500 individuals representing seventeen red pine (P. resinosa Ait.) populations from Manitoba through Newfoundland, they identified five polymorphic microsatellite loci with an average of nine alleles per locus. The mean expected and observed heterozygosity values were 0.508 and 0.185, respectively. Significant departures from Hardy-Weinberg Equilibrium with excess homozygosity indicating high levels of inbreeding were evident in all populations studied. The population differentiation was high with 28–35% of genetic variation partitioned among populations. The genetic distance analysis showed that three northeastern (two Newfoundland and one New Brunswick) populations were genetically distinct from the remaining populations. The coalescence-based analysis suggested that “northeastern” and “main” populations likely became isolated during the most recent Pleistocene glacial period, and severe population bottlenecks might have led to the evolution of a highly selfing mating system in red pine.

Nurtjahjaningsih et al. (2005) isolated ten microsatellite markers from P. merkusii using a dual-suppression-polymerase chain reaction technique. Of these markers, five loci were co-dominant and polymorphic. The number of alleles per locus ranged from three to six and the expected heterozygosity ranged from 0.389 to 0.728. These microsatellite markers are available for analysis on population genetics and mating patterns.

Watanabe et al. (2006) developed seven microsatellite loci from P. densiflora using a dual polymerase chain reaction (PCR) technique. Of 186 clones from a library based on suppression PCR, 127 contained microsatellite sequences. Of these, forty-three candidates were determined to have sequences of both flanking regions, and sixteen regions from this group were chosen as development markers. Seven of these primer pairs successfully amplified polymorphic single loci among eighty-three resistant trees against pine wood nematode. The observed heterozygosity of the seven
microsatellite markers ranged from 0.247 to 0.843. Mendelian inheritance was confirmed using mega-gametophytes.

### 2.3.2.5.2 Genetic diversity studies in pines using SSRs

Provan *et al.* (1998) used polymorphic cpSSRs to analyze levels of genetic variation within and between seven native Scottish and eight mainland European populations of Scots pine (*P. sylvestris*). Diversity levels for the Scottish populations based on haplotype frequency were far in excess of those previously obtained using monoterpenes and isozymes and confirmed lower levels of genetic variation within the derelict populations at Glen Falloch. The diversity levels were higher than those reported in similar studies in other *Pinus* species. An analysis of molecular variance (AMOVA) showed that small (3.24-8.81%) but significant ($p \leq 0.001$) portions of the variation existed between the populations and there was no significant difference between the Scottish and the European mainland populations.

Provan *et al.* (1999) used cpSSRs to examine variation present in the chloroplast genome of *P. torreyana*. Analysis of the seventeen cpSSR loci showed no variation, which was consistent with previous cpRFLP work and confirmed that the species has descended from an original, highly monomorphic population following a bottleneck. The lack of biological variation in the chloroplast genome of *P. torreyana* allowed estimation of the mutation rates at cpSSR loci as between $3.2 \times 10^{-5}$ and $7.9 \times 10^{-5}$.

Elsik *et al.* (2000) tested eighteen low-copy and genomic microsatellite markers for Mendelian inheritance and then assayed in forty-one *P. taeda* samples drawn from five regions in the southern United States. The PCR products had multiple alleles, high levels of polymorphism, and little non-specific priming. Fifteen of the eighteen markers were informative for a *P. taeda* three generation RFLP pedigree, and a *P. taeda* population survey revealed 3 to 28 alleles per locus. The highest allele numbers and polymorphism information content (PIC) were associated with complex repeat sequences and (or) with sequences consisting of the longer strings of perfect repeats.
The abundance of low to rare frequency alleles also accounted for high PIC values in both types of markers. Low-copy microsatellites are useful for the large, complex pine genome, especially in the absence of entire gene sequences in public databases and with the low levels of polymorphism in markers developed from expressed sequence tags (ESTs).

Walter and Epperson (2001) examined ten cpSSR loci in 136 individuals from ten widespread populations of *P. resinosa*. Substantial variation for the cpSSR loci was observed in the populations studied. The amount of variation observed was lower than that generally found for cpSSR loci in other pine species. In addition, the variation exhibited a striking geographical pattern. Most of the genetic diversity observed was among populations, with little within populations, indicating substantial isolation and genetic drift within many populations in the southern half of the species distribution.

Derory *et al.* (2002) studied the genetic diversity, allelic richness and heterozygosity of forty-seven populations of *P. pinaster* using three nSSRs. Classical parameters of diversity (allelic richness and heterozygosity) and differentiation were estimated for forty-seven populations of *P. pinaster*. The populations were collected from France (40), Iberian Peninsula (6) and Morocco (1). Highest levels of diversity were recorded in Moroccan and Iberian populations. Levels of diversity tended to be higher in the west group of France than in the south east of France group than in other groups. Population identified by the nSSRs showing diversity reservoirs could thus be considered in conservation programmes.

Cloutier *et al.* (2003) studied variation at nuclear and chloroplast microsatellite loci among and within clonally propagated individuals of eastern white pine (*P. strobus*). Total DNA was extracted and assayed from gamete bearing tissue (mega gametophytes) located on six different branch positions on each of twelve individual genets. No within individual variation was observed among twelve loci studied. The numbers of mitotic cell divisions required to produce the tissue used as the source of genomic DNA were estimated by combining the tree growth and anatomical data. This
allowed for the calculation of upper bound estimates of numbers of mutations per locus per somatic cell division. The estimated somatic mutation rate was found substantially lower than those published for genomic microsatellite mutation rates in other plant species.

Walter and Epperson (2005) sampled and examined ten cpSSRs from nineteen populations of *P. resinosa*. Analysis of these nineteen populations and ten previously studied populations showed that the geographic distribution of genetic diversity over the range of *P. resinosa* was markedly nonuniform. Although the pattern exhibited little isolation by distance, a region centered in north eastern New England contained much higher chloroplast haplotype diversity than elsewhere. The geographic pattern indicated that *P. resinosa* is not at equilibrium, and that the species has had a more complex postglacial history than that typically purported for forest trees in eastern North America.

Hohn *et al.* (2005) studied diversity in Swiss stone pine populations (*P. cembrana*) native to the Carpathians with cpSSRs. Six loci were analyzed in four populations and a total number of 22 size variants and 41 combinations referred to as haplotypes were detected. Diversity within populations was found to be high, whereas divergence between the populations was low ($F_{ST} = 0.02$). The most variable population with the highest haplotype diversity ($h = 0.956$) originated from the Retezet mountains. The great haplotypic variation found in cpSSR loci made all populations a useful source for gene conservation purpose.

The genetic variation within and between thirteen populations (385 individuals) of *P. uncinata* was analyzed with ten cpSSR markers (Dzialuk *et al.*, 2009). High level of genetic diversity and low but significant differentiation among compared population were found. Three marginal populations, Sierra de Cebollera, Margaride Mountains and Sierra de Gudar were strongly differentiated from the rest. The southernmost populations from the Sierra Cebollera and the Sierra de Gudar were the most
genetically distinct which suggested a long period of spatial isolation and/or origin from different ancestral populations.

Hohn et al. (2009) characterized the genetic structure and diversity of *P. cembra* L. populations native to two disjunct geographical areas, the Alps and the Carpathians. They evaluated the rate of genetic differentiation among the populations. Twenty-eight populations were screened at three cpSSR loci for length variation. Haplotypic diversity was high throughout the natural range of *P. cembra*, with the mean value substantially higher in the Carpathians (h = 0.53) than in the Alps (h = 0.35). AMOVA revealed that only 3% of the total genetic variation derived from genetic differentiation between the two mountain ranges resided among populations. Differentiation among Carpathian populations was higher ($F_{ST} = 0.19$) than among Alpine populations ($F_{ST} = 0.04$). Low, but significant, correlation was found between the geographical and genetic distances among pairs of populations ($r = 0.286, P < 0.001$).

Mehes et al. (2009) studied the genetic diversity of *P. strobus* and genetic structure of *P. monticola* with the help of fifteen microsatellite primers. Seeds from ten populations of *P. monticola* were sampled from western Canada and ten populations of *P. strobus* from eastern Canada were sampled. It was observed that *P. monticola* and *P. strobus* exhibited moderate to high genetic diversity. The mean number of alleles per locus was 20 for *P. monticola* and 16 for *P. strobus*. Exact test for HWE revealed that nine *P. monticola* and ten *P. strobus* populations deviated from the expected HWE. Both the species showed low levels of inbreeding. Gene flow estimates were high Nm = 5.89 for *P. monticola* and Nm = 3.10 for *P. strobus* and population differentiation values were relatively low.

Gomez-Garay et al. (2010) evaluated the genetic diversity of *P. ayacahuite* from six stands at the Communal Forest of Totonicapan (Guatemala) with cpSSR markers. The analysis along an altitude cline showed a diverse pattern in both genetic diversity and genetic differentiation estimates. Three areas were established: 1) a lower area, located near human establishments that included a forest nursery, was characterized
by highest genetic diversity ($h = 0.7-0.8$) and differentiation ($D^2: 0.22-0.44$). Human impact and introduction of foreign plant material could have caused alterations in the genetic composition; 2) an intermediate area which was characterized by a moderate anthropogenic perturbation (familiar small agricultural farms), exhibited a reduction of haplotype richness; and 3) a higher area showed an elevated genetic diversity and differentiation values. The latter was revealed as a genetic diversity reservoir. The study showed that low perturbated stands with high genetic diversity indexes were critical to ensure effective sampling with afforestation purposes.

2.3.2.5.3 Cross species amplification of SSRs in pines

Development of a microsatellite marker system for a new species requires isolation, cloning, sequencing, and characterization of microsatellite loci. The development of microsatellite markers through the laboratory based screening of SSR libraries is highly time consuming and expensive. An alternative approach that could be used in a genome with no or little DNA sequence information is to test the SSR primers developed for other species for cross amplification in the species of interest. This approach offers a potential for low cost development of SSR markers for species with very little or no information on sequence, through the screening of primers from different sources. Cross-species amplification of SSR markers has been reported by many workers.

Cato and Richardson (1996) tested and amplified chloroplast sequence from $P.\ thunbergii$ in $P.\ radiata$, $P.\ elliottii$, $P.\ taeda$, $P.\ strobus$, $Pseudotsuga\ menziesii$, $Cupressus\ macrocarpa$, four New Zealand native conifer species ($Podocarpus\ totara$, $Podocarpus\ Hallii$, $Podocarpus\ nivalis$, $Agathis\ australis$), and four angiosperms ($Vitex\ lucens$, $Nestegis\ cunninghamii$, $Actinidia\ chinensis$, and $Arabidopsis\ thaliana$). A PCR product in the expected size range was amplified from all species and interspecific polymorphism was detected in $P.\ radiata$ with four of the five primer pairs. One of this polymorphic cpSSR was then used to determine the inheritance of chloroplast in 206 progeny from four control-pollinated, full-sibling $P.\ radiata$ families. Approximately 99%
of the progeny had the cpSSR variant of the pollen parent indicating that in *P. radiata*, like most other conifers, chloroplast are typically inherited from the paternal parent. These results suggest that polymorphic cpSSRs are a valuable tool for studying chloroplast diversity, cyto-nuclear disequilibrium, and plastid inheritance in a range of species, and for the analysis of gene flow via pollen and paternity in species with paternal transmission of chloroplasts.

Echt *et al.* (1999) evaluated twenty-one SSR loci from *P. strobus* and six from *P. radiata* to determine whether SSR marker amplification could be achieved in ten other conifer species. Eighty percent of SSR primer pairs for (AC)_n loci that were polymorphic in *P. strobus* also amplified SSR loci in two other soft pines of the subgenus *Strobus* but not in seven hard pines of the subgenus *Pinus*, nor in *Picea glauca* (Moench) Voss or *Pseudotsuga menziesii* (Mirb.) Franco. Six SSR primer pairs that were specific to monomorphic loci within *P. strobus* showed successful amplification in conifers other than soft pines. These six loci were also monomorphic within seven other species tested, but four of the loci were polymorphic among species. A comparison of allelic variation among the three soft pine species found only twenty-five shared alleles among a total of 122 alleles at eight loci. Primer pairs for di-nucleotide SSR loci that were polymorphic in *P. radiata* also specifically amplified loci from various other hard pines but not from the soft pines or from the other conifers tested.

Kutil and Williams (2001) tested fifteen triplet-repeat microsatellites from hard pine (*P. taeda*) for trans-specific amplification across seven hard pines (*P. palustris, P. echinata, P. radiata, P. patula, P. halepensis* and *P. kesiya*), a soft pine (*P. strobus*), and *Picea rubens*. Seven of fifteen microsatellites showed trans-specific amplification in both hard and soft pine subgenera. Two *P. taeda* microsatellites had conserved flanking regions and repeat motifs in all seven hard pines, soft pine (*P. strobus*) and *Picea rubens* using trinucleotide microsatellite improved trans-specific microsatellite recovery among hard and soft pine species.
Mariette et al. (2001) tested seventy-six SSR primer pairs from four *Pinus* species to amplify microsatellites in *P. pinaster*. Twenty-six primer pairs were stemmed from a microsatellite library on *P. pinaster* and the other primer pairs were obtained in other species of the same genus (*P. radiata*, *P. strobus* and *P. halepensis*). Only three out of the seventy-six SSR primer pairs amplified at a single polymorphic locus in *P. pinaster*. The Mendelian inheritance of those three primer pairs was studied and their genetic map position was determined. The number of alleles and the level of heterozygosity were assessed in an analysis of a sample of 196 trees. The development of microsatellites in *Pinus* species has been reported to be a difficult task because of the size and complexity of their genome. In this study, cross-species amplification was unsuccessful.

Gonzalez-Martinez et al. (2004) tested nineteen nSSR markers from *P. taeda* (subsection Australes) and three from *P. sylvestris* (subsection *Pinus*) on seven Eurasian hard pine species (*P. uncinata*, *P. sylvestris*, *P. nigra*, *P. pinaster*, *P. halepensis*, *P. pinea* and *P. canariensis*). They found that the transfer rates to species in subsection *Pinus* (36–59%) were slightly higher than those to subsections *Pineae* and *Pinaster* (32–45%). Half of the trans-specific microsatellites were found to be polymorphic over evolutionary times of approximately 100 million years (ten million generations). Sequencing of three trans-specific microsatellites showed conserved repeat and flanking regions. Both a decrease in the number of perfect repeats in the non-focal species and a polarity for mutation (higher substitution rate in the flanking sequence regions close to the repeat motifs) were observed in the trans-specific microsatellites. The transfer of microsatellites among hard pine species proved to be useful for obtaining highly polymorphic markers in a wide range of species, thereby providing new tools for population and quantitative genetic studies.

Chagne et al. (2004) screened two unigene datasets of *P. taeda* and *P. pinaster* to detect di, tri and tetranucleotide repeat motifs using the SSRIT script. A total of 419 SSRs were identified, from which only 12.8% overlapped between the two sets. Fifty-
three primer pairs amplifying a single PCR fragment in the source species mainly *P. taeda*, were tested for amplification in six other pine species. The amplification rate with other pine species was high and corresponded with the phylogenetic distance between species, varying from 64.6% in *P. canariensis* to 94.2% in *P. radiata*. Genomic SSRs were found to be less transferable. Fifty eight of the 107 primer pairs (54%) derived from *P. radiata* amplified a single fragment in *P. pinaster* linkage maps. The level of polymorphism of these cDNA-SSRs was compared to that of previously and newly developed genomic SSRs. It was observed that genomic SSRs tend to perform better in terms of heterozygosity and number of alleles. The study suggested that useful SSR markers can be developed from pine ESTs.

Xiang-Xiang *et al.* (2005) selected and identified two hundred and seventy-six primer pairs from seven species of *Pinaceae* for cross-species transferability to ten *Pinus* species (*P. massoniana, P. kesiya, P. tabulaeformis, P. densiflora, P. thunbergii, P. caribaea, P. taeda, P. yunnanensis, P. densata and P. sylvestris*) belonging to section *Pinus* by bulk segregant analysis method. The results showed that twenty-three of 276 (8.0%) markers showed successful amplification in ten species. Five out of twenty-three markers (21.7%) were polymorphic across species but lacked polymorphism within species. Eight out of ten *Pinus* species were identified by using one, two and more combinations of primers, but *P. kesiya* and *P. densata* could not be identified effectively by any of the SSR markers.

In *P. roxburghii*, till date there are only two reports on the use of SSR markers. Gauli *et al.* (2009) assessed the genetic structure of five population pairs each consisting of one natural population and one neighboring plantation of *P. roxburghii* in Nepal using nSSR and cpSSR loci. The mean number of alleles at nSSR loci in natural populations was 5.0 compared to 4.93 in neighboring plantations while the average observed heterozygositites were the same in both groups (*H₀ = 0.50*). Similarly, forty-seven haplotypes were observed in natural populations compared to fifty haplotypes in plantations. Mean haplotype diversities of natural populations (0.953) and plantations
(0.955) were very similar. Genetic diversity of *P. roxburghii* was relatively high with low or no evidence of inbreeding while genetic differentiation among all the populations was very low (about 1%). The very low differentiation among natural populations indicates efficient long-distance gene flow among populations resulting in homogenous genetic structures at least at selectively neutral loci.

Chauhan *et al.* (2010) tested eighty microsatellite markers (nuclear and chloroplast) developed for *P. thunbergii, P. sylvestris, P. taeda, P. resinosa, P. merkussi and P. densiflora* on Himalayan chir pine (*P. roxburghii*). From the tested eighty primers, forty-nine showed positive amplification and nineteen (fourteen cpSSRs and five nSSRs) were found polymorphic. The polymorphic primers were screened on 275 adult trees of *P. roxburghii* from a single large population in its natural range of distribution. The number of alleles in case of nSSRs was up to five per locus and the expected heterozygosities ranged from 0.409 to 0.562. Using fourteen cpSSR loci, a total of sixty-one variants in 275 individuals were found which ranged from 2 to 6 per loci. The total gene diversity revealed through cpSSRs ranged from 0.461 to 0.781. These sets of markers can further be used for population genetic studies and characterization of *P. roxburghii* germplasm.

2.4 DNA markers in pines for trait improvement

The primary goal of tree breeding is to increase the quantity and quality of wood products as well as NWFPs from plantations. Major gains have been achieved using recurrent selection in genetically diverse breeding populations to capture additive variation. However, the long generation time of trees and poor juvenile-mature trait correlations, have promoted interest in marker-assisted selection (MAS) to accelerate breeding through early selection. MAS rely on identifying DNA markers, which explain a high proportion of variation in phenotypic traits. The different approaches for MAS are linkage mapping, QTL mapping, association mapping and physical mapping.
2.4.1 Genetic linkage mapping studies in pines

Mapping and sequencing of plant genomes help to elucidate gene function, gene regulation and their expression. Molecular markers are used to identify and tag desired genes. Linkage analysis is one of the basic and indispensable methods in genetics that define the genetic distances between polymorphic traits which may be recognized as differences in appearance of enzyme activities, restriction fragment lengths or nucleotide sequences at an allelic locus. Linkage maps have been utilized for identifying chromosomal regions that contain genes controlling simple traits (controlled by single genes).

Genetic linkage maps have been developed for many Pinus species such as P. elliottii var elliottii (Nelson et al., 1993); P. sylvestris (Yazdani et al., 1995); P. palustris (Kubisiak et al., 1996); P. radiata (Devey et al., 1996). These maps can be used to locate chromosomal regions where DNA markers co-segregate with quantitative traits (quantitative trait loci, QTL). Devey et al. (1994) constructed a genetic linkage map using RFLP markers for P. taeda. Plomion et al. (1995) constructed genomic maps for P. pinaster using RAPD markers. Haploid linkage analysis of P. strobos, was carried out using mainly RAPD markers and SSR markers (Echt and Nelson 1997). Ninety one loci mapped to twelve linkage groups of three or more markers. The resulting framework genome map, the first for a soft pine species, contained 69 markers. The map covered 58% of the estimated genome length of 2,071 cM (K), with a 95% confidence interval of 1,828-2,242 cM (K). AFLPs were used to construct a genetic linkage map of P. edulis (Travis et al., 1998). A complete genetic map of an individual loxboly pine (P. taeda) was constructed by Remington et al. (1999) using AFLP markers. A two-way pseudo-testcross strategy was applied in an analysis of P. sylvestris for genetic mapping and detection of quantitative trait loci (QTLs) associated with economically important traits targeted in the Swedish tree-breeding program (Lerceteau et al., 2000). A genetic map of maritime pine was constructed based on AFLP, RAPD and protein markers (Costa et al., 2000). Linkage map for Japanese Black Pine was constructed based on AFLP and
RAPD markers by Hayashi et al. (2001). A high density linkage map of maritime pine (Pinus pinaster) was constructed by Chagne et al. (2002) based on AFLP markers using a three-generation outbred pedigree. A consensus map was obtained which covered 1,441 cM and comprised a total of 620 AFLP markers on twelve linkage groups.

Genetic maps for individual P. elliottii var. elliottii and P. caribaea var. hondurensis trees were generated using a pseudo-testcross mapping strategy using a total of 329 AFLP and twelve microsatellite markers (Shepherd et al., 2003). Based on estimated genome sizes for these species, the framework maps for P. elliottii var. elliottii and P. caribaea var. hondurensis covered 82% and 88% of their respective genomes.

Clustering has been reported for conifer genetic maps based on hypomethylated or low-copy molecular markers, resulting in uneven marker distribution. To test this, a framework genetic map was constructed from three types of microsatellites: low-copy, undermethylated, and genomic (Zhou et al., 2003). These P. taeda microsatellites were mapped using a three-generation pedigree with 118 progeny. The fifty-one microsatellites covered a map distance of 795 cM, an average distance of 21.8 cM between markers, roughly half of the estimated total map length. The minimum and maximum distance between any two bins was 4.4 and 45.3 cM, respectively. These microsatellites provided anchor points for framework mapping for polymorphism in P. taeda and other closely related hard pines. A genetic map of P. sylvestris was constructed using ESTP (expressed sequence tag polymorphism), AFLP and microsatellite markers (Komulainen et al., 2003).

Nearly complete linkage maps of P. sylvestris have been constructed using AFLP markers based on a two-way pseudo-testcross strategy in a full-sib family founded in an advanced breeding program (Yin et al., 2003). Their genome coverage was estimated to be more than 98% with a framework marker interval of 20 cM for both parents. Most of the female and male linkage groups were associated through the analysis of the intercross markers.
Kim et al. (2005) constructed an AFLP-based linkage map of Japanese red pine (\textit{P. densiflora} Siebold et Zucc.) using haploid DNA samples of 96 mega gametophytes from a single maternal tree, selection clone Kyungbuk 4. Twenty-eight primer pairs generated a total of 5,780 AFLP fragments. Five hundred and thirteen fragments were verified as genetic markers with two alleles by their Mendelian segregation. The resulting map provided crucial information for future genomic studies of the Japanese red pine, in particular for QTL mapping of economically important breeding target traits.

Recently the genetic linkage map of \textit{Pinus koraiensis} was constructed using an F1 progeny of eighty-eight individuals. One hundred and thirty (130) molecular markers were mapped onto six linkage groups, four triples and fifteen pairs at the linkage criteria LOD 4.0. Nine primer combinations were applied to map construction. The consensus map gained covers 620.909 cM, with an average marker spacing of 4.776 cM (Chen et al., 2010).

\textbf{2.4.2 QTL mapping studies in pines}

The process of constructing linkage maps and conducting QTL analysis to identify genomic regions associated with traits of interest is known as QTL mapping (McCouch and Doerge, 1995; Mohan et al., 1997; Paterson, 1996a, b). It is an effective technique for dissecting the genetic architecture of important quantitative traits, as it can reveal the nature of gene action, the number of genes involved and their interactions. QTL mapping has been reported in many \textit{Pinus} species.

Groover et al. (1994) reported the identification of QTL influencing wood specific gravity (WSG) in an outbred pedigree of loblolly pine (\textit{P. taeda}).

Emebiri et al. (1997) investigated QTLs affecting the expression of stem growth efficiency in radiata pine seedlings using a RAPD linkage map constructed from markers scored on haploid, megagametophytic DNA. Four putative QTLs were detected which accounted for 8.5-36.4\% of the population variance.
Kaya *et al.* (1999) estimated QTLs for annual height and diameter increment in each of the two three-generation loblolly pine pedigrees. An interval mapping-approach was used to estimate the number of QTLs, the magnitude of QTL effects, and their position on genetic linkage maps. Thirteen different height-increment and eight different diameter-increment QTLs were detected, suggesting that these traits are, at least in part, controlled by a few genes of large effect.

Sewell *et al.* (2000) analyzed physical wood property traits in *P. taeda*. They identified nine QTLs for wood specific gravity, five QTLs for volume percentage of latewood, and 5 QTLs for microfibril angle.

Hurme *et al.* (2000) examined the genetic basis of large adaptive differences in timing of bud set and frost hardiness between natural populations of *P. pinaster*. They found four potential QTLs for timing of bud set and seven for frost hardiness.

Weng *et al.* (2002) employed RAPD markers to map the genome and QTL controlling the early growth of a pine hybrid F1 tree (*P. palustris* Mill. x *P. elliottii* Engl.) and a recurrent slash pine tree (*P. elliottii* Engl.) in a (longleaf pine x slash pine) x slash pine BC 1 family consisting of 258 progeny. Of the 150 hybrid F1 parent-specific RAPD markers, 133 were mapped into seventeen linkage groups covering a genetic distance of 1,338.2 cM. Of the 116 slash pine parent-specific RAPD markers, eighty-three were mapped into nineteen linkage groups covering a genetic distance of 994.6 cM. A total of 11 different marker intervals were found to be significantly associated with 13 of the 20 traits on height and diameter growth using the software MAPMAKER/QTL. The amount of phenotypic variance explained by the putative QTLs ranged from 3.6% to 11.0%.

Yazdani *et al.* (2003) identified quantitative trait loci (QTLs) involved in the genetic control of growth rhythm and autumn cold acclimation in *P. sylvestris*. One family north x south with 108 individuals was used for the construction of one linkage map for the male and one for the female parent using 286 RAPD markers. The study confirmed that there are major QTL regions on different linkage groups controlling a
large part of the variation for growth rhythm and autumn cold acclimation in *P. sylvestris*.

Brown *et al.* (2003) identified forty-four QTLs for *P. taeda*, of which ten QTLs were identified for early wood specific gravity, eight for late wood specific gravity, twelve for percentage volume of latewood, four for latewood microfibril angle, five for earlywood cell wall chemistry, and five for latewood cell wall chemistry. Further, they have reported co-location of eighteen candidate genes including laccases in the verified QTLs.

In *P. pinaster*, fifty-four QTLs were detected for wood properties with eight QTLs for timber quality traits, three for Height/Growth ratio, seven for lignin, four for alfa cellulose, four for hemicelluloses, three for heterogeneity of density, three for kappa, one for pulpy and twenty-one for fibre properties. An attempt was made to identify the genes in the specific QTLs. Nine wood quality candidate genes involved in cell wall structure were located on the genetic map. Only one of them, Korrigan, a gene encoding for β-1, 4 endoglucanase, co-located with hemicelluloses content and fibre characters (Pot *et al.*, 2006).

QTL based marker assisted selection is most likely to be used for within family selection in a limited number of elite families that can be clonally propagated. Limitations of the approach include the low resolution of marker-trait associations, the small proportion of phenotypic variation explained by QTL and the low success rate in validating QTL in different genetic backgrounds and environments. This has led to a change in research focus towards association mapping to identify variation in the DNA sequence of genes directly controlling phenotypic variation.

2.4.3 Association mapping

The development and use of molecular markers for the detection and exploitation of DNA polymorphisms in plant and animal systems is one of the significant developments in the field of molecular biology and biotechnology. Two major
phenomena involved in the generation of DNA polymorphism detected by molecular markers are mutation and recombination. Therefore, detection of linkage and tracing the history of a DNA polymorphism has been central to the use of molecular markers for a variety of studies (Terwilliger and Weiss, 1998; Nordborg and Tavare, 2002; Gupta and Rustgi, 2004). For the study of linkage, one needs to perform suitably designed crosses, sometimes leading to the development of mapping populations or near isogenic lines (NILs). This is a serious limitation on the use of molecular markers in some cases, because the desired crosses cannot be made in all cases (e.g. in forest trees), and/or the mapping populations that are examined for this purpose are sometimes too small, with only two alleles at a locus sampled. In view of this, alternative methods have been developed and used to study the phenomenon of linkage and recombination on the one hand, and for the study of mutational history of a population on the other. One such method is linkage disequilibrium (LD) based association analysis. This approach has the potential not only to identify and map QTLs (Meuwissen and Goddard, 2000) but also to identify causal polymorphism that is responsible for the difference in two alternative phenotypes (Palaisa et al., 2003, 2004).

Originally developed for human genetics, this approach exploits the candidate gene sequence variation and relies on the existence of linkage disequilibrium (non-random association between alleles at linked loci) between detectable sequence polymorphism. The advantage of this approach over anonymous markers is that once a major effect gene is identified and validated, MAS can be practiced directly on the gene. LD mapping can be applied to wild, unstructured and unpedigreed populations. It is similar to QTL approaches but is more amenable to hardwoods, because it does not rely on a structured pedigree, but instead analyzes the variation within an entire population.

Human disease genetics was the first area for which association mapping methodology was developed and where successes were achieved (Lander and Schork 2004; Jorde 2000; Carlson et al., 2004). In animal genetics, most concern was about the
LD patterns in breeding populations to determine the extent to which LD holds and the marker density required to fine map genes, e.g. in cattle (Farnir et al., 2000), sheep (McRae et al., 2002) and pig (Nsengimana et al., 2004). In plant genetics, however, fewer reports are available on association mapping, except for model plant systems like Arabidopsis thaliana (Hagenblad and Nordborg 2002; Nordborg et al., 2002), maize (Remington et al., 2001; Tenaillon et al., 2001; Ching et al., 2002; Palaisa et al., 2003, 2004; Parisseaux and Bernardo 2004; Rafalski and Morgante 2004;) and rice (Garris et al., 2003; Lu et al., 2005; Semon et al., 2005), where association mapping is gaining importance due to the development of high throughput marker systems and the availability of genome sequences. Other plant systems have been investigated for LD patterns and associations as well. In barley, association analysis was performed for agronomical, resistance related and morphological traits (Kraakman et al., 2004, 2006). The status and prospects of association mapping in plants was discussed by Zhu et al. (2008).

LD patterns were examined deploying RFLPs in sugarcane (Jannoo et al., 1999); AFLPs in sugar beet (Kraft et al., 2000) and SNPs in soybean (Zhu et al., 2003). The mating system of the species (selfing versus out crossing), and phenomena such as population structure and recombination hot spots, can strongly influence patterns of LD (Garcia et al., 2003). Conifers were examined using the candidate gene approach to dissect complex traits (Neale and Savolainen 2004). In Lolium perenne, associations were found for heading date (Skot, 2005). A complete overview on association mapping and the status of this approach in plants has been published (Gupta et al., 2005). The LD pattern of grapevine was reported by Barnaud et al. (2006). Kernel size and milling quality were fine mapped applying association mapping in wheat (Breseghello and Sorrells 2006). Association analysis that exploits the natural diversity of a genome to map at very high resolutions is becoming increasingly important. In most studies, however, researchers must contend with the confounding effects of both population and family structure. The software TASSEL (Trait Analysis by Association,
Evolution and Linkage) implements general linear model and mixed linear model approaches for controlling population and family structure (Bradbury et al., 2007).

In association studies, a set of unlinked, selectively neutral background markers scaled to achieve genome-wide coverage are employed to broadly characterize the genetic composition of individuals. Background genetic markers are useful in assigning individuals to populations (Pritchard and Rosenberg, 1999), preventing spurious associations (if population structure and relatedness exist) (Pritchard et al., 2000b; Thornsberry et al., 2001; Yu et al., 2006), and estimating kinship and inbreeding (Lynch and Ritland, 1999). RAPD and AFLP markers can serve as background markers, but these markers are dominantly inherited and thus demand special statistical methods if used to estimate population genetic parameters (Ritland, 2005; Falush et al., 2007). Conversely, co-dominant microsatellites, or SSRs, and SNPs are more revealing than their dominant counterparts and, therefore, are more powerful in estimating population structure (Q) and the relative kinship matrix (K). It is recommended that a mixture of co-dominant and dominant markers should be used for better characterization of genetic structure of a population (Li et al., 2007).

A serious problem with association mapping is that population structure can lead to spurious associations between a candidate marker and a phenotype. One common solution has been to abandon case-control studies in favor of family-based tests of association, such as the transmission/disequilibrium test (TDT), but this comes at a considerable cost because of the need to collect DNA from close relatives of affected individuals. A novel, statistically valid, method for case-control association studies in structured populations was described by Pritchard et al. (2000). This method uses a set of unlinked genetic markers to infer details of population structure, and to estimate the ancestry of sampled individuals, before using this information to test for associations within subpopulations.

Pritchard and Rosenberg (1999) examined the issue of population stratification in association-mapping studies. In case-control studies of association, population
subdivision or recent admixture of populations can lead to spurious associations between a phenotype and unlinked candidate loci. Using a model of sampling from a structured population, they showed that if population stratification exists, it can be detected by use of unlinked marker loci. They showed that the case control–study design, using unrelated control individuals, is a valid approach for association mapping, provided that marker loci unlinked to the candidate locus are included in the study, to test for stratification. They suggested guidelines to describe the number of unlinked marker loci to be used.

Falush et al. (2003) described extensions to the method of Pritchard et al. (2000) for inferring population structure from multilocus genotype data. Most importantly, they developed methods that allow for linkage between loci. The new model accounted for the correlations between linked loci that arise in admixed populations (“admixture linkage disequilibrium”). This modification had several advantages, allowing (1) detection of admixture events farther back into the past, (2) inference of the population to which chromosomal regions belong, and (3) more accurate estimates of statistical uncertainty when linked loci are used. It was also of potential use for admixture mapping. In addition, they described a new prior model for the allele frequencies within each population, which allows identification of subtle population subdivisions that were not detectable using the existing method. They presented results applying the new methods to study admixture in African-Americans, recombination in Helicobacter pylori, and drift in populations of Drosophila melanogaster.

Evanno et al. (2005) tested the ability of the software STRUCTURE to detect the true number of clusters (K) in a sample of individuals when patterns of dispersal among populations were not homogeneous using various dispersal scenarios from data generated with an individual based model. They found that in most cases the estimated ‘log probability of data’ did not provide a correct estimation of the number of clusters (K). However, using an ad hoc statistic ΔK based on the rate of change in the log
probability of data between successive K values, they found that STRUCTURE accurately detected the uppermost hierarchical level of structure for the scenarios they tested. The results were sensitive to the type of genetic marker used (AFLP vs. microsatellite), the number of loci scored, the number of populations sampled, and the number of individuals typed in each sample.

2.4.3.1 Association analysis in pines

The availability of genes linked to a range of important traits in model plants have opened up new areas of investigation in association genetics in tree species. Conifers, and in particular pines, are well-suited for association studies due to (1) high levels of genetic variation, (2) low level of domestication, (3) large natural populations and outcrossing mating systems, resulting in a low level of population structure, (4) high population recombination rates and rapid decay of intragenic LD and (5) accurate evaluation of phenotypes due to the possibility of establishing clonal replicates or large progenies.

Association studies have been carried out in loblolly pine to examine the adaptive significance of sequence variation in monolignol biosynthesis genes (Peter and Neale, 2004) and other genes controlling wood properties (Brown et al., 2004). Similarly, another study was carried out for identifying genes controlling wood formation in maritime pine (Pot et al., 2004). Association studies have also been carried out to identify genes controlling pathogen resistance in Pinus (Ersoz et al., 2004; MacKay et al., 2005).

In P. tremula, high levels of nucleotide polymorphism and low linkage disequilibrium was observed (Ingvarsson, 2005) and it was suggested that it may be possible to map functional variation to very fine scales by using association mapping approaches.

Gonzalez-Martinez et al. (2007) tested genetic association between fifty-eight single nucleotide polymorphisms (SNPs) from twenty wood and drought related
candidate genes and some wood property traits such as earlywood and latewood specific gravity, percentage of latewood, earlywood microfibril angle and wood chemistry (lignin and cellulose content) in *P. taeda*. He used mixed linear models (MLMs) that account for relatedness among individuals by using a pair wise kinship matrix. Twenty-two nSSRs were used to assess the population structure. Different phenotype: genotype associations were found, some of them confirming previous evidence from collocation of QTL and genes in linkage maps (for example, 4cl and percentage of latewood), and two that involve non-synonymous polymorphisms (cad SNP M28 with earlywood specific gravity and 4cl SNP M7 with percentage of latewood).

The strongest genetic association found in this study was between allelic variation in α-tubulin (a gene involved in the formation of cortical microtubules) and earlywood microfibril angle. Intragenic LD decays rapidly in conifers, thus SNPs showing genetic association were likely to be located in close proximity to the causative polymorphisms.

This first multi gene association genetic study in forest trees showed the feasibility of candidate gene strategies to dissect complex adaptive traits, provided that genes belonging to key pathways and appropriate statistical tools are used. This approach was of particular utility in species, such as conifers, where genome-wide strategies are limited by their large genomes.

Abdurakhmonov and Abdukarimovonov (2008) provided the basic concept, merits, and simple description of existing methodologies for an association mapping with the recent improvements for plant populations. They discussed the details of some of the pioneer and recent studies on association mapping in various crop species to demonstrate the feasibility, success, problems, and future perspectives of the efforts in plants.

Semagn et al. (2010) provided an overview of the two most commonly used methods for QTL mapping, including mapping population type and size, phenotypic evaluation of the population, molecular profiling of either the entire or a subset of population. They discussed different statistical methods and softwares for carrying out
marker-trait association analysis as well as the future prospects of using markers in crop improvement.

Quesada et al. (2010) used a population genomic approach to identify genes in *P. taeda* that are associated with resistance to pitch canker, a disease incited by the necrotrophic pathogen *Fusarium circinatum*. A set of 498 largely unrelated, clonally propagated genotypes were inoculated with *F. circinatum* micro-conidia. The data for lesion length (a measure of disease resistance) were collected 4, 8, and 12 weeks after inoculation. Best linear unbiased prediction was used to adjust for imbalance in number of observations and to identify highly susceptible and highly resistant genotypes (“tails”). The tails were re-inoculated to validate the results of the full population screen. Significant associations were detected in ten SNPs (out of 3,938 tested). As hypothesized for genes involved in quantitative resistance, the ten SNPs had small effects and proposed roles in basal resistance, direct defense, and signal transduction. They also discovered associated genes with unknown function, which would have remained undetected in a candidate gene approach constrained by annotation for disease resistance or stress response.

Eckert et al. (2012) recently used the standard association genetic methods to correlate 3,563 SNPs to the concentrations of 292 metabolites measured in a single *P. taeda* association population. A total of twenty-eight single locus associations were detected, representing twenty-four and twenty unique SNPs and metabolites, respectively. Multilocus Bayesian mixed linear models identified 2,998 additional associations for a total of 1,617 unique SNPs associated to 255 metabolites. These SNPs explained sizeable fractions of metabolite heritabilities when considered jointly and had lower minor allele frequencies and magnitudes of population structure as compared with random SNPs. Modest sets of SNPs (*n* = 1-23) explained sizeable portions of genetic effects for many metabolites, thus highlighting the importance of multi-SNP models to association mapping, and exhibited patterns of polymorphism consistent with being linked to targets of natural selection.