2. LITERATURE REVIEW

2.1 Background of *A. excelsa*

2.1.1 Origin and Geographical Distribution

A. excelsa is native to Peninsular Malaysia (Penang, Perak, Selangor and Malacca), Sumatra (Bengkulu, Eastcoast), Borneo (North and East), Sulawesi, Celebes, Moloccas (Aru Islands), West New Guinea and the Philippines (Mindoro, Masbate, Samar, Palawan, Basilan). The tree thrives in moist tropical rain forests where rainfall is usually above 1600 mm/annum. Generally, *A. excelsa* is found in the lowlands. However, in Sabah, it has been found at 200 m altitude at Pangi, and may extend even higher than this in Western Sabah (Burgess, 1966). *A. excelsa* is a strong light demander that usually dominates other moist evergreen species. It may attain a height of 45 m and a breast height diameter (Dbh) of about 200 cm (a specimen of this size is found on Scotland Road in Penang) (Kijkar and Boontawee, 1995).

In the northern and eastern states of Peninsular Malaysia, *i.e.* Kedah, Penang, Terengganu and Kelantan, and Southern Thailand, *A. excelsa* is grown as a village tree along roadsides and farm boundaries or as a mixed crop. The timber of *A. excelsa* trees has been used by the villagers for manufacture of furniture, carving, construction of houses, boat building, and cigar boxes (Jacobs, 1960; Wong, 1976; Corner, 1988). The young shoot is eaten as vegetable or salad while the intensely bitter old leaves have been used in traditional medicines for the treatment of dysentery and diarrhoea (Anon, 1995). In addition, the tree has also been found to contain azadirachtin, a plant secondary compound in the limonoids group which has

been shown to be a strong insect antifeedant and exhibits growth disrupting properties (Mordue and Blackwell, 1993). These multipurpose properties and its fast growth rate make *A. excelsa* suitable for use as a plantation species.

2.1.2 Taxonomy

Azadirachta comprises 2 species : *A. excelsa* (Sentang) and *A. indica* (Neem). *A. excelsa* was first described by Jack (1820) as *Melia excelsa* from specimens from Penang Island in Malaysia. Later, Jacobs (1960) placed it in the genus *Azadirachta* and named it, *Azadirachta excelsa* (Jack) Jacobs.

2.1.3 Botanical characteristics

An *A. excelsa* tree can grow up to 45 m in height. It can be either evergreen or deciduous; in Thailand, *A. excelsa* was reported to shed its leaves from January to Febuary, and the new leaves emerge immediately thereafter (Kijkar and Boontawee, 1995). The tree has a rounded but rather open and uneven crown. Leaves are usually pinnately compound and measure about 30-75 cm long with 7 to 11 pairs of leaflets. When young, leaves are pink, becoming green at maturity. The bark of the young trees is pink or brownish-grey and smooth which becomes brownish or greyish-buff with age. The mature tree has a fissured and shaggy bark with grey, fibrous and oblong flakes (Ng, 1989).

A. excelsa starts to flower and set fruit when the tree reaches about 6-7 years of age. Flowers are greenish-white, fragrant, with five petals, each measuring 5-6.5 mm long

and 1.5-2.5 mm wide. The actual pollination mechanism is poorly understood. However, insects such as the dipteran fly have been seen visiting flowers of *A. excelsa* (Pannell, 1992). The size of the fruit is about 2.5-3.5 cm long, these are oblong in shape and often with a constriction near one end. The fruit is green in color and turns yellow when ripe. Fruits of this species are garlic-scented when bruised. There are approximately 500 fruits/kg. Fruits are usually mature in May-June in Thailand and Peninsular Malaysia, and in August-September in Borneo. The seeds (see Plate 1) are recalcitrant therefore not suitable for long term storage. Seed of *A. excelsa* germinates in 6-23 days (Ng, 1992).

The timber of *A. excelsa* is moderately hard to moderately heavy. Its heartwood is reddish-brown which darkens on exposure. This property is similar to the true mahogany, *Swietenia macrophylla*, also belonging to the Meliaceae family. The density of the wood is 550-780 kg/m^3 at 15% moisture content (Lemens *et al.*, 1995). The heartwood of *A. excelsa* is rated as non-durable to moderately durable. The sapwood is susceptible to termites, beetles and fungal attack.

2.2 Background of forest tree improvement

Forest tree improvement programs were started in the 1820's with the provenance trials on Scotch pine (Wright *et al.*, 1966). In the experiment, trees were grown at several locations in North-central United States from seeds collected from 170 parts

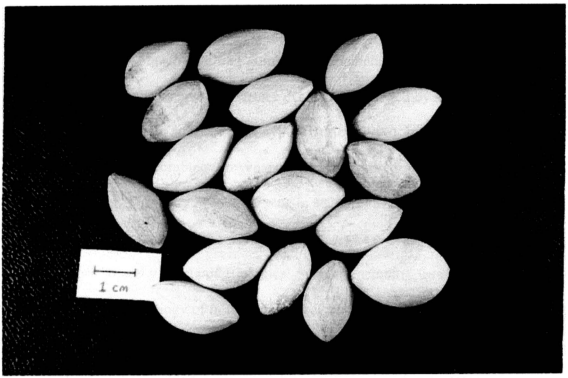

Plate 1 : Seeds of *A. excelsa* after removal of fruit coat and pulp

of Europe and Asia. In the 1900's, extensive provenance trials were carried out which included tree species such as loblolly pine, ponderosa pine and larch. Plus tree selection was practiced in the 1950's with the objectives to maximise survival, growth rate of stem volume (height, diameter and taper) and stem quality (straightness, branch size and number, and branching pattern). Breeding strategies were based on extensive provenance evaluation followed by simple selection, testing and propagation of superior phenotypes (Burley, 1996). In the same period there was a tendency to combine provenance and half-sib progeny testing in the same experiment. The study of loblolly pine in Mississippi by Wells and Switzer (1971) is one such example. In the 1980's, the emergence of biotechnology has brought changes to many facets of Biological Science including forestry tree improvement work. The technology of *in vitro* micropropagation can be deployed to produce large volumes of clonal plant materials. The recent availability of molecular marker technology has enabled genetic variability to be assessed at the DNA level.

2.2.1 Strategy of provenance trials

A "provenance trial" is an experiment in which seeds are collected from a number of widely scattered stands (usually natural), and the seedlings are grown under similar conditions (Wright, 1976). The experiment is necessary because large genetic differences associated with place of origin have been recorded in many species under study (Burley and Wood, 1976; Falkenhagen, 1991). It is very important to do provenance trials with exotic species due to the large genetic differences often existing between different races of the same species. However, even for native

species, some genetic differences exist for some species due to incomplete adaptation which therefore require provenance trials to screen for the best race.

2.2.2 Strategy of progeny trials

A "progeny trial" is the growing of the offspring of different species or races under similar conditions in a replicated experiment. The aim of the experiment is to study the genetic component of the selected trait assuming minimum environmental variability among the progeny trees. Sometimes a bulk progeny trial is carried out which means seeds were collected from trees categorized into different grades according to selected traits. The seeds are then grown in replicated experiments for assessment of traits later. Such experiments are useful for species with no available inheritance data and in which future intensive breeding work may be carried out. After progeny trials are established and measured, the trees can be thinned to leave only the best families. After thinning, the progeny trials are considered as seed orchards or as breeding arboreta.

2.2.3 Strategy of clonal trials

A "clonal trial" is an experiment in which clonal planting stocks such as tissue culture plantlets, rooted cuttings or grafted trees are planted according to a replicated design in which measurements can be made periodically. In clonal selection, clones are obtained from phenotypically superior trees selected from provenance or progeny trials. Selected clones are planted out in the field and measurements made after the trees have established. Future planting stock can then be obtained from the trees that

yield the best performance. Clonal trials offer the prospect of greater genetic gain through more intensive selection. However, they involve an extra test rotation per generation. Therefore the decision to carry out clonal trials depend on factors such as operational costs, time costs and the benefits that can be derived from the trials.

2.2.4 Limitations to forest tree improvement

Tree improvement programs have been carried out for a wide variety of commercially important forest tree species throughout the world (Burley, 1987). Progress in numerous aspects of forest tree productivity, ranging from growth and form improvements to increased disease resistance have been reported (Wright, 1976 ; Zobel and Talbert, 1984). Despite the progress, there are a number of limitations in forest tree improvement programs. A breeding program basically involves the regulation of variation through control over the reproductive system. In forest trees, the major limitation is the long generation time required for the trees to reach reproductive maturity. Other limitations related to time include evolutionary time, time to harvest and time to achieve phenotypic stability (e.g. long juvenile phase) (Cheliak and Rogers, 1990). Tree improvement involves three phases namely conservation, selection & breeding and propagation. The objective of conservation is to retain maximum genetic variability in the germplasm collection for future use. In breeding for desirable traits, directional selection to maximize genetic gain will result in reduced genetic diversity. Once a genetically superior stock is identified, multiplication of the stock will be performed through clonal propagation to produce a large number of genetically identical individuals. Therefore as we progress from

conservation phase through propagation phase, decreasing amounts of variation are being managed. The challenge of forest tree improvement programs is to achieve a balance between continuous genetic gain and maintenance of an adequate amount of genetic variability present in each phase for a given species. Early studies of forest genetic diversity were based on adaptive quantitative traits (e.g. growth) and they were found to associate closely with geographical variations (Wright and Baldwin, 1957; Callaham, 1964). Adaptive quantitative traits were calculated from the formula

Phenotype = Genotype (G) + Environment (E) + Genotype x Environment (GE)

Thus there is a need to perform testing across a range of environments. However problems such as lack of statistical efficiency due to huge micro-site variation and long duration of the tests due to poor juvenile-mature correlation may arise (Yeh, 2000). In addition, the effect of individual genes on quantitative traits is still not known. The evolutionary forces such as the mating system, gene flow and random genetic drift acting on allelic and genotypic distribution in time and space are also difficult to quantify. Therefore, a more sophisticated technology is needed to help make batter progress in forest tree improvement.

2.3 Genetic marker technology

A wide range of genetic markers has been used since the early 1900s. These markers include the following:

(a) Morphological markers - changes in the phenotype of the organism

(b) Protein markers – changes that occur as a result of modification of the polypeptide protein and enzyme product of gene action

(c) Molecular markers – changes in DNA sequence within the genome

2.3.1 Morphological markers

The concept that some morphological differences are due to simple heritable traits was demonstrated by Mendel in his classic work using the garden pea (Mendel, 1865). Mendel studied seven traits in the garden pea *Pisum sativum*. He self-fertilized the plants for several generations until he obtained strains that "bred-true" *i.e.* genetically stable strains that resemble the parents for a particular trait under study. He then crossed these "bred-true" plants with contrasting appearances. For example in one of the crosses, he chose round- and wrinkle-seeded peas. When plants with these phenotypes were intercrossed, Mendel found that all the F1 progenies were round-seeded peas. When Mendel crossed his F1 hybrid plants, he observed that three-quarters round and one-quarter wrinkle-seeded peas were obtained in the F2 generation. Mendel deduced that round seed is the dominant allele and wrinkled seed the recessive allele. These seed phenotypes have been characterized at the molecular level (Morell *et al.*, 1995). The seed wrinkling is the result of high sugar level found in the developing seeds which causes an increase in water accumulation in the embryonic cotyledons and the gene involved has been identified as one of the starch-branching enzymes (SBE-1).

The use of morphological markers in plant breeding is not a new concept. As early as 1923, Sax has reported the association between seed size and alleles influencing seed color in *Phaseolus*. Phenotype describes the morphological appearance of an

organism. A combination of genotype and the effect of environment on that genotype determine the external appearance of an organism. Environment can alter gene expression of an individual, thus an interaction between environment and genotype will produce a wide range of variability among individuals which form the basis for the genetic study.

Genes are not always readily categorized as dominant or recessive whereby the phenotype of the hybrid is identical to that of one parent. They may exhibit an "additive" relationship when the presence of one allele elicits an effect, and two alleles produce twice the effect of one. Thus the phenotype of the hybrid can be intermediate between those of the parents. For example in the plant *Antirrhinum* spp (snapdragon), when a red-flowered cultivar is crossed with a white-flowered cultivar, all the F1 plants are pink-flowered. The F2 progenies however segregate in the ratio of 1 red: 2 pink : 1 white. In this case, the heterozygous progenies have only one red-flower allele which could not produce enough red pigment for the red coloration of the flower. Another form of gene effect is called overdominancy. In this case the presence of one copy of allele has a greater effect than two (Paterson *et al.*, 1991). This could be demonstrated in the breeding of maize. When two unrelated maize inbreds of low productivity were crossed, heterozygous F1 progeny obtained were larger and more fruitful than either parent. This phenomenon is also known as hybrid vigor or heterosis. Sometimes a single gene can affect several traits. The multiple effect of a single gene is known as pleiotropy (Gruneberg, 1938). Alternatively, the collective actions of many different genes can influence the phenotype of an

individual. Much evidence has shown that interaction between different genes, or epistasis, determines many phenotypes of individuals (Wright, 1968; Allard, 1988).

Some morphological characters do not segregate monogenically into discrete phenotypes but are determined by the segregation of multiple loci. These characters exhibit continuous variations and are referred to as quantitative traits. Individual loci controlling a quantitative trait are referred to as polygenes or quantitative trait loci (QTL) (Tanksley, 1993). In agricultural crops, many traits of agronomic importance such as plant height, size, stress tolerance, maturity and yield are quantitative characters. Evidence (East and Hayes, 1911; Emerson and East, 1913) has shown that Mendelian principles apply to quantitative genetic trials as well. Genes controlling quantitative traits of major effect were found to segregate according to Mendelian ratios. The multigenic trait inheritance can be investigated using the classical biometrical approach. However the biometrical statistical method has its limitations in that some of the important information such as the number of loci influencing the expression of the traits, the chromosomal location of these loci and the relative size of contribution of individual loci to trait expression, are not generally available. Such information can be obtained with the advent of marker-based technology (see details below) which has fast become a powerful tool for identifying and mapping quantitative trait loci (QTL). A better understanding of genetic phenomena such as epistasis, pleiotropy and heterosis will also be possible.

There are several limitations associated with using morphological markers in crop improvement. Only a few morphological mutants are available and many of them tend to be associated with undesirable phenotypic effects. Alleles at morphological loci interact in a dominant-recessive manner that limits the identification of heterozygous genotypes. In addition, the phenotype of most morphological markers can only be determined at the whole plant level.

2.3.2 Protein markers

The most widely used protein markers in plant breeding are the isozymes that emerged in the 1960s. Isozymes are enzymes that share a common substrate but differ in electrophoretic mobility (Markert and Moller, 1959). The location of a particular enzyme on a gel can be detected with the use of an appropriate substrate and cofactors to produce colored end products. The colored product is deposited at the site of enzymic action to form a visible band that usually corresponds to a single-polypeptide gene product. Allelic variants of isozymes, due to differences in the amino-acid composition which in turn cause a change in the net charge, molecular weight or isoelectric point of the molecule, will result in changes in electrophoretic mobility of the polypeptides. Several bands can thus be visualized depending on the number of loci, their state of homo- or heterozygosity and the enzyme molecule configuration.

Isozyme marker technology has been successfully applied to many organisms from bacteria to numerous animal and plant species since the 1960s (May, 1992). The

studies have encompassed various fields (e.g. physiology, biochemistry, genetics, breeding) and purposes (population structure, mating system, hybridization, polyploidy, systematics) (Harris, 1969 ; Lewontin and Hubby, 1966 ; Murphy *et al.*, 1990).

Isozyme markers are codominant. They generally have 2-10 alleles/locus and provide 5-40 marker loci originating from low-copy regions of the genome. Isozyme analysis is relatively straightforward and easy to carry out. It involves low cost materials such as starch or polyacrylamide gel, hence it can be cheaply and efficiently applied to a genetic study. The main disadvantage is the relatively low level of polymorphism at loci identified by isozymes. In addition, isozyme expression is found to be dependent on which tissue is used and on the developmental stage of the tissue.

2.3.3 Molecular markers

Morphological and protein-based marker systems only sample actively expressed regions of the genome. However, the discovery of restriction endonucleases and the availability of cloned DNA technology have allowed a much greater portion of the plant genome to be assayed for genetic markers. The DNA or molecular marker technology detects natural DNA sequence polymorphism present between individuals in a population. If these differences represent alternatives at a single position in the genome, they are equivalent to alleles of wild-type and mutant genes. Molecular markers are simply inherited Mendelian characters and are therefore

suitable for use in genetic studies. Besides, molecular markers provide a large reservoir of loci for study since any DNA sequence can be the source of allelic differences between individuals without having to know the product for which it codes and even including non-coding regions of the genome. Furthermore, since molecular markers are stable attributes of DNA itself, DNA for analysis can be isolated from any part of the plant at any stage of the life cycle. They do not show pleiotropic effects on economic characters. Due to these advantages, molecular markers have potential wide applications in plant improvement programs. These applications include : varietal identification, identification and mapping of quantitative trait loci, screening genetic resource strains for useful quantitative trait loci alleles and their marker-assisted introgression from resource strain to commercial variety, and marker-assisted early selection of recombinant inbred lines.

Various types of molecular marker technology have been developed over recent years. Some of the more widely used molecular marker systems are outlined and discussed below.

2.3.3.1 RFLP (Restriction fragment length polymorphism)

RFLP detect DNA polymorphisms through restriction endonuclease digestions, coupled with DNA blot hybridizations. Restriction endonucleases are enzymes that recognize specific nucleic acid sequences in DNA and cleave the DNA at or proximal to these targeted sites (Zabeau and Roberts, 1979). A large piece of DNA will thus be reduced to a series of smaller fragments of defined size by digestion with

a restriction enzyme. The number of fragments produced and the sizes of each fragment will reflect the distribution of restriction sites in the DNA. The DNA fragments so formed can be separated by gel electrophoresis. Smaller fragments will migrate more rapidly through the pores of the gel matrix than larger fragments. When genomic DNA from a higher organism is digested by restriction enzymes, many different-sized fragments are produced and a continuous smear is formed on the gel. To resolve these bands, cloned DNA probes and DNA hybridization are used. The DNA pattern is transferred from the gel to a solid support, such as a nitrocellulose filter (Southern, 1975), and exposed to a radioactively (or other) labeled probe under conditions that promote DNA-DNA hybridization. The unhybridized probe is washed away and the filter is dried and placed against photographic film for autoradiographic exposure. The specific DNA fragment that hybridized with the probe will be visualized as a band on the film or directly viewed using an electronic imager. A diagrammatic representation of the RFLP technique is outlined in Figure 2.1.

RFLP variability in plants can be caused by (1) base sequence changes that add or eliminate restriction sites; (2) rearrangements such as insertions or deletions; or (3) unequal crossing over or replication slippage (Schlötterer and Tautz, 1992). Studies have found that most RFLP variability in plants is caused by genome rearrangement

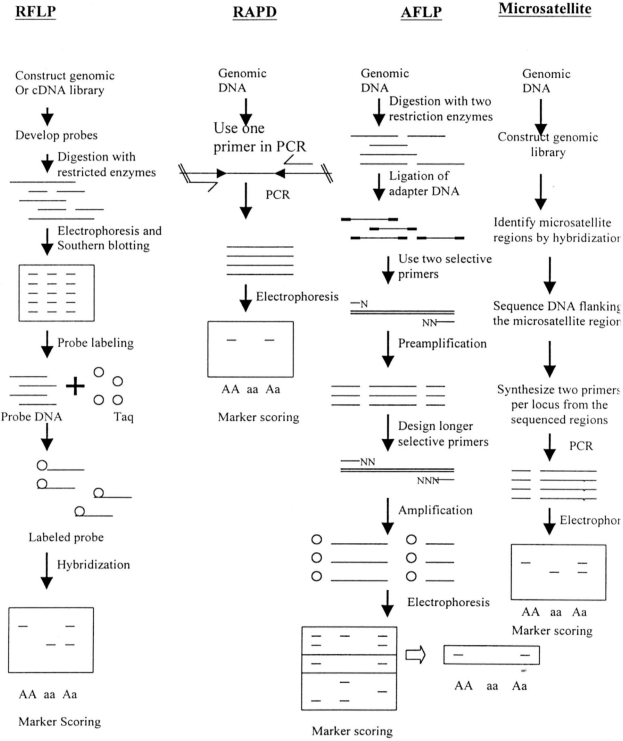

Figure 2.1 : Comparisons of various molecular marker technologies (adapted from Staub *et al.*, 1996)

rather than nucleotide sequence change (Landry *et al.*, 1987; Apuya *et al.*, 1988; McCouch *et al.*, 1988; Miller and Tanksley, 1990). Conventional RFLP analysis therefore often detects length polymorphisms. Larger fragments of the genome (*i.e.* up to 20 kb or more) are assayed by RFLP probes.

RFLPs have been found to be abundant in most organisms, and a virtually unlimited number can be mapped in any one cross. In genetic analysis, low-copy RFLP markers are defined as codominant because in segregating progenies, both homozygous and heterozygous genotypes are clearly differentiated. However, RFLP analysis is limited by the relatively large amount of DNA (5-10 μg) required for restriction digestion, Southern blotting and hybridization plus the requirement for radioactive isotopes and autoradiography or other labeling or detection systems. These factors make RFLP analysis time consuming, labour intensive and incompatible with applications requiring high throughput.

2.3.3.2 RAPD (Random Amplified Polymorphic DNA)

Over the last few years, polymerase chain reaction (PCR) technology has led to the development of several novel genetic marker assays based on DNA amplification (Krawetz, 1989; Innis *et al.*, 1990). These new assays are amenable to automation and relatively simple to perform. They are therefore suitable to be used for determining genotypes of a large number of individuals. One such assay, more popularly called RAPD, was developed independently by two different laboratories (Welsh and McClelland, 1990; Williams *et al.*, 1990). This procedure uses only a

single primer of arbitrary nucleotide sequence to detect nucleotide sequence polymorphism in a DNA amplification-based assay. In this reaction, a single species of arbitrary primer (usually 10mers) anneals to the genomic DNA at two different sites on opposite strands of the DNA template. If these priming sites are within an amplifiable distance of each other (between 500 to 5000bp), a discrete DNA product is produced through thermocyclic amplification (Figure 2.1). Unlike RFLP, this technology does not require labeled probes, sequence information, blotting or any hybridization steps. RAPDs are technically less demanding, cheaper, faster and easier to perform. They require only small amounts of DNA and the short random primers may readily be available commercially or easily synthesized. RAPD technology allows application of high throughput automation thus making it popular in map construction or genetic analysis studies. However, there have been contrasting opinions and reports on the reproducibility of RAPDs. Lin *et al.* (1996), Karp and Edwards (1996) and Staub *et al.* (1996) agreed that there is a problem in the reproducibility of RAPDs while Chaparro *et al.* (1995) and Plomion *et al.* (1995) reported otherwise. The reproducibility of RAPDs markers can be improved by utilizing strictly consistent PCR conditions (Karp and Edwards, 1996) or longer primers (e.g. 24 nt) or by cloning and characterizing flanking regions to synthesise specific probes for specific amplification (Sequence Characterised Amplification Regions - SCARs) (Staub *et al.*, 1996; Rafalski and Tingey, 1993).

Being a dominant marker system, RAPDs are less informative than other systems such as allozymes and RFLPs. However, this limitation can be compensated by using

higher numbers of markers and/or bigger sample sizes. Other approaches to improve the power of this marker system include the utilization of double-haploid individuals (Lu *et al.*, 1995), near-isogenic lines (NIL) (Waugh and Powell, 1992) or bulk segregant analysis (BSA) (Wang and Paterson, 1994; Waugh and Powell, 1992).

2.3.3.3 AFLP (Amplified Fragment Length Polymorphism)

A more recent PCR-based marker technology, AFLP, developed by Vos *et al.* (1995) has shown to be robust and reliable because of stringent reaction conditions used for primer annealing. AFLP marker system is powerful in detecting polymorphisms originating from base substitution, inversions, insertions and deletions. The technique combines restriction enzyme digestion, as in RFLP, and PCR amplification as in RAPDs. AFLP involves digestion of genomic DNA with two restriction endonucleases and ligation of oligonucleotide adapters to the resultant fragments to generate template DNA for PCR (Zabeau and Vos, 1993). The primers utilized are complementary to the adapter and restriction site sequences with the addition of extra nucleotides at their 3' end, thus, allowing selective amplification of only fragments in which the primer extension match the nucleotide flanking the restriction sites. Analysis of DNA fragments on denaturing polyacrylamide gels typically results in the production of 50 to 100 bands per individual sample depending on the detection system. The ability of AFLP technology to generate a large number of markers with minimum primer testing, and the system's high resolution on DNA fragment detection are features that make AFLP attractive as genetic markers. In addition, AFLP reactions are readily automated using robotics

and high throughput sequencer or capillary electrophoresis based genetic analyzers (Weller *et al.*, 1995).

AFLPs, like RAPDs are predominantly dominant markers with only two alleles – the presence and absence of a given band. However, heterozygous and homozygous genotypes may sometimes be differentiated by the intensity of the amplified bands (Staub *et al.*, 1996). AFLP is relatively more technologically demanding as high quality DNA is needed to allow reproducible restriction endonuclease digestion and ligation of adaptor oligonucleotides. Since AFLP can efficiently generate a large number of markers with high polymorphism, it has been applied to many mapping studies e.g. in *Oryza* (Zhu *et al.*, 1998), *Zea* (Xu *et al.*, 1999) and *Solanum* (Bradshaw *et al.*, 1998). AFLP has also been used in population genetic and diversity studies (Aggarwal *et al.*, 1999; Rieseberg *et al.*,1999; Beismann *et al.*, 1997).

2.3.3.4 Microsatellites

Microsatellites, also known as simple sequence repeats (SSRs), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs) are tandem repeats of sequence units generally less than 5 bp in length e.g. $(TG)_n$ or $(AAT)_n$ (Bruford and Wayne, 1993). Microsatellites were first developed in humans largely for use in forensics and medical diagnostics (Weber and May, 1989). Later they were found to also be abundant in plants as well (Morgante and Olivieri, 1993). Microsatellite variation results from differences in the number of repeat units. These differences can be caused by errors during DNA replication (Moxon and Willis, 1999; Jarne and

Lagoda, 1996) whilst larger changes in repeat number are due to processes such as unequal crossing over (Strand *et al.*, 1993). Microsatellites are analysed by PCR amplification of a short genomic region containing the repeated sequence, and size estimation of the repeated length by gel separation. Only a small quantity of DNA is required and it is amenable to automation technology. Microsatellites have the potential of revealing a large number of polymorphisms with multiple alleles per locus. For example one locus in soybean (*Glycine max*) is reported to have 26 alleles (Cregan *et al.*, 1994). Microsatellites are co-dominant markers.

One major drawback on microsatellites is the effort and expense required for their development which involves screening of genomic libraries for repeat motifs, sequencing and testing for selection of suitable PCR primer synthesis (Rafalski *et al.*, 1996; White and Powell, 1997; Ciofi *et al.*, 1998). The recent development of methods for the production of microsatellite enriched DNA libraries have helped to reduce the effort needed for microsatellite development (Kijas *et al.*, 1994; Edwards *et al.*, 1996). Applications of microsatellites in plants include diversity studies (e.g. Rossette *et al.*, 1999), gene-flow and mating systems (Chase *et al.*, 1996b) and paternity analysis (Streiff *et al.*, 1999).

2.3.4 Marker system selection

From the previous discussion, it is rather clear that there are advantages and disadvantages associated with each marker system. Therefore, numerous factors have to be taken into careful considerations when selecting a particular marker system for a certain project. Areas that need to be evaluated before utilisation of certain markers are :

- Objectives of project
- Previous work done and available information
- Technical expertise/facilities/equipment available
- Cost per unit information
- Time requirement

Since different markers serve different purposes, the decision as to which marker system to adopt depends very much on the aims to achieve in the project. For example, if the purpose is just to determine the genetic relationships between two populations, then a collection of morphological and/or isozyme markers coupled with some RAPDs markers may be sufficient. Vicario *et al.* (1995) showed a clear differentiation between *Abies alba* and *A. nebrodensis* and managed to classify them into two different taxonomic groups by using isozymes in combination with RAPDs. On the other hand, if the objective of utilisation of a marker system is to construct a genetic linkage map of a species, the marker system of choice should produce a large number of polymorphic markers which can be positioned according to their genetic distance with each other across the whole genome.

Availability, selection or design of mapping populations and the range of diversity present in a species population are of crucial in choosing the types of marker system to be employed. Previous work done on a particular species or related species may provide information on the availability of pre-designed probes or primers thereby saving cost and time.

RFLPs have been popularly used in map construction because they provide a large amount of reproducible information. However, the requirements for relatively large amounts of DNA, Southern blotting and detection probe (see previous discussion) are the limitations of this technique of analysing a large number of samples. Many researchers have thus opted for PCR-based marker systems such as RAPD or AFLP that are more amenable to automation. Lin *et al.* (1996) compared RFLPs, RAPDs and AFLPs in genetic mapping of soybean and concluded that AFLP markers are highly reproducible, rapid in generation and high in identifiable polymorphisms making it the most attractive technique of choice. Nevertheless, AFLPs do require some degree of technical expertise to carry out restriction digestion and pouring of sequencing gels. In some instances, researchers may choose a combination of marker systems. Faure *et al.* (1993) constructed a linkage map on *Musa acuminata* using RFLPs, isozymes and RAPDs while Philipp *et al.* (1994) utilised all the three marker systems in addition to morphological and physiological markers to construct a map of the rye genome.

PERPUSTAKAAN UNIVERSITI MALAYA

A510824021

Technical expertise, facilities, equipment, time and funds available for a project are other factors that need consideration. Some marker systems are more time or cost consuming than others but provide vast amounts of potentially useful genetic information (Table 2.1). No one marker system will be suitable for all purposes. To select the optimum marker system to be employed, the resources available must be closely examined against the potential of their output. These are the complex issues that have to be worked out for each individual project, taking into consideration that the objectives, funds, time and resources (which include information, expertise, equipment etc) available differ with every project.

2.3.5 Application of genetic marker system to forest trees

2.3.5.1 Tree Genome

The size of tree genomes generally ranges from 10^8 to 10^{10} base pairs per haploid genome. Gymnosperm nuclear genomes are among the largest eucaryotic genomes known. Their size generally falls within the range of 6-30 x 10^9 base pairs per haploid genome (Dhillon, 1987; Price *et al.*, 1974; Wakamiya *et al.*, 1993). Angiosperm trees have smaller genomes with sizes range from 250 x 10^6 to 15 x 10^9 base pair per haploid genome (Grattapaglia and Bradshaw, 1994; Arumuganathan and Earle, 1991). The bulk of the tree genome consists of repetitive DNA sequences, with most of it being of unknown function (Nowak, 1994). For example, the proportion of the eastern white pine (*Pinus strobus*) genome occupied by actual genes has been estimated to be a mere 0.1 % (Kriebel, 1984).

Table 2.1. Comparisons of molecular marker systems

	Non – PCR based	PCR based		
	RFLP	RAPD	AFLP	Microsatellites
Principle	RE digestion, Southern blotting and probe hybridization	PCR amplification with random primers	RE digestion, selective PCR amplification with designed primers	Screening of genomic library, primers sequencing and PCR amplification of simple sequence repeats
Type of polymorphism	Single base insertion or deletions	Single base insertion or deletions	Single base insertion or deletions	Repeat length
Level of polymorphism	Medium	Medium	Very high	High
Dominance	Codominant	Dominant	Dominant	Codominant
Amount of DNA required	2-10 µg	10-20 ng	50-200 ng	25-50 ng
Cloning and sequencing	Yes	No	No	Yes
Development cost	Medium	Low	Medium/High	High
Start-up cost	Medium/High	Low	High	High
Automation	+	++	+++	+++
Proprietary rights status	NA	LC	LC	NA

RE = restriction enzyme, NA = not applicable, LC = license required
For automation : + = least potential, ++ = high potential, +++ = most potential
(Adapted from Rafalski and Tingey, 1993; Staub *et al.*, 1996)

Trees, contain within their cells mitochondria and chloroplasts. Each of these organelles contains small, circular and bacterial-like genomes. The organellar genomes are present in multiple copies within each cell as there are many mitochondria or chloroplasts within a single cell. The size of the chloroplast genome in trees falls within the range of 120 to 160 kb (Strauss *et al.*, 1988; Palmer 1990). Mitochondrial genomes are not well characterised in trees. The size typically ranges from 200 to 2400 kb. A large proportion of the DNA in plant organellar genomes code for some of the protein required for the function of the organelle (Wakasugi *et al.*, 1994; Palmer, 1997).

2.3.5.2 Applications to the field of forestry

The advent of isozymes as genetic markers in the early 1970's has provided a useful technology for genetic studies of forest tree species. Isozymes have been used to determine genetic inheritance of the haploid megagametophytic tissue of conifer seeds without having to make controlled crosses (Conkle *et al.*, 1982; Moran *et al.*, 1983; Adams *et al.*, 1991). The megagametophytic tissue is genetically equivalent to the maternal contribution to the embryo of the seed. Isozymes still have wide applicability in studies of genetic resources in forest trees because they are cheap and technically simple. Although isozymes may give lower absolute levels of genetic diversity, the patterns of population genetic structure were found to be similar to that from nuclear DNA markers (Byrne *et al.*, 1998; Butcher *et al.*, 1998).

RFLP has been employed in trees mostly for the purposes of genetic linkage mapping. The ultimate aim is to accelerate tree breeding programs via early marker assisted selection for desirable traits influenced by genes that are linked to some of the markers (Devey *et al.*, 1994; Bradshaw *et al.*, 1994; Mukai *et al.*, 1995). RFLP has also been used to analyse chloroplast DNA in conifers. It was found that chloroplast DNA is paternally inherited in all three conifer families studied, including Pinaceae (Wagner *et al.*, 1987), Taxodiaceae (Neale *et al.*, 1989) and Cupressaceae (Neale *et al.*, 1991). However, mitochondrial DNA was found to be maternally inherited in Pinaceae (Neale and Sederoff, 1989) but paternally inherited in Taxodiaceae (Neale *et al.*,1989) and Cupressaceae (Neale *et al.*, 1991). It is therefore possible to trace both maternal and paternal lineages back to the Pinaceae species' recent evolutionary history by examining mitochondrial and chloroplast DNA polymorphism respectively.

The technical ease of RAPD markers and easy adaptation of their application to new species have led to their employment in a large number of studies in forest trees, both in genetic linkage mapping (Carlson *et al.*, 1991; Grattapaglia and Sederoff, 1994; Nelson *et al.*, 1994) and population genetic applications (Chalmers *et al.*, 1994; Bucci and Menozzi, 1995; Schierenbeck *et al.*, 1997). RAPD has proven to be very useful in the molecular genetic study of trees. One major setback is the degree of reproducibility of this marker system as mentioned in the previous discussion.

AFLP has shown a high degree of reproducibility in contrast to RAPDs (Akerman *et al.*, 1996) and is also highly reproducible across laboratories (Jones *et al.*, 1997a). Even though AFLPs are a relatively recent development, there are already several examples of their applications in forest trees (Cervera *et al.*, 1996; Akerman *et al.*, 1996; Beismann *et al.*, 1997; Gaiotto *et al.*, 1997; Remington *et al.*, 1998). AFLP is a powerful technique of great promise that could displace the RAPD technique in future for many applications (Glaubitz and Moran, 1999).

The first microsatellites developed in trees were from radiata pine (Smith and Devey, 1994). Recently, microsatellites have been developed from many forestry species including oaks (Dow *et al.*, 1995), eucalypts (Byrne *et al.*, 1996), eastern white pine (Echt *et al.*, 1996), Norway spruce (Pfeiffer *et al.*, 1997) and several tropical trees (Terauchi, 1994; Chase *et al.*, 1996a; White and Powell, 1997; Dawson *et al.*, 1997; Steinkellner *et al.*, 1997). The wide applications of microsatellites in various forest tree species indicates the potential usefulness of this marker technique, and as more microsatellite sequences become available to the research community, their use can be expected to increase.

2.4 Micropropagation

Like most of the other indigenous forest tree species, *A. excelsa* tends to bear fruits irregularly. The seeds that formed have short viability and each tree bears an average of about 200 to 300 fruits. Currently, seeds supply could not meet demands from the

plantation. One alternative way of providing quality uniform planting stock is through micropropagation or tissue culture technique.

2.4.1 Principles of the technology

Micropropagation refers to systems of tissue culture in which asexual or clonal propagation occurs from explants of the original plant. The main aim of micropropagation is to produce large numbers of genetically identical individuals. The source of explants could be seeds, buds, leaves, cotyledons, stem sections or cells. These explants could be obtained either from field-grown or sterile tissue-cultured plants. Micropropagation technologies have been established in many tree species over recent years (Durzan, 1988; Thorpe *et al.*, 1991; Le Roux and van Staden, 1991; Tartorius *et al.*, 1991).

2.4.2 Organogenesis

Organogenesis is a developmental process whereby organ primordia such as buds are initiated on an explant in response to the application of exogenously applied plant growth regulators (Cheliak and Rogers, 1990). Organogenesis can occur directly from explants (axillary budding) or indirectly by differentiation of shoot or root meristems in the callus tissue (adventitious budding).

Axillary budding system has been used to multiply a large number of plant species on a commercial scale (Bajaj, 1986; Le Roux and van Staden, 1991). Although multiplication rates are generally only five to ten propagules per culture cycle (Lutz

et al., 1985), such a multiplication factor can produce up to millions of plantlets per year for many plant species (Wang and Charles, 1991). In addition, the axillary budding system is usually associated with high genetic fidelity.

The adventitious budding system has been well established for poplar (McCown *et al.*, 1988) and radiata pine (Aitken-Christie *et al.*, 1988). The success of *in vitro* culture decreases dramatically with age of explant, particularly for the coniferous species (Bonga, 1982; Thorpe *et al.*, 1991). Multiplication rates are commonly substantially higher through adventitious budding than the axillary budding route (Wang and Charles, 1991). For example in radiata pine, it was estimated that an embryo of one of the more responsive clones could yield up to 260 000 plants (ready for the field) in 2.5 years (Aitken-Christie *et al.*, 1988).

2.4.3 Embryogenesis

Somatic embryogenesis refers to a developmental process where nonmeristematic cells can be cultured to produce an organised bipolar structure displaying shoot and root poles connected by a functional vascular tissue. Embryogenesis can be either (i) direct – where somatic embryos originate directly from an explant without a callus phase, or (ii) indirect, where somatic embryos originate after a proliferation of callus tissue (Tulecke, 1987). Regeneration through embryogenesis has been reported for over 50 woody species which include 20 angiosperm families and at least a dozen conifer species (Wann, 1988; Tartorius *et al.*, 1991; Attree and Fowke, 1991). Multiplication rates are high, particularly from cell suspension culture. For example,

in *Picea*, an estimation of 100 embryos per ml of culture was obtained with a doubling time of about 48 hours (Becwar *et al.*, 1988). However, the rate of conversion to plantlets was very low with highest mean efficiency of 4% reported in *Picea* (Becwar *et al.*, 1988) as an example.

One objective of research on somatic embryogenesis is the production of dormant embryos capable of being encapsulated to form artificial seeds (Redenbauch *et al.*, 1986). These artificial seeds can then be treated like sexually derived seeds which can be stored for future use (Gray, 1987). Since the rates of production are potentially high, somatic embryogenesis may provide an attractive method of producing commercial quantities of clonal propagules for use in forest regeneration programs.

2.4.4 Application in forest tree improvement

Micropropagation is now operating at a large commercial scale involving hundreds of laboratories around the world. For example, in Holland alone, 65-70 laboratories produced over 53 million plants in 1988 (Wang and Charles, 1991). Most of the commercially micropropagated plants are herbaceous ornamentals (Haines, 1994). Even though many forest tree species can be micropropagated, the application of the techniques in commercial plantation establishment is still not yet widely practiced. One good example is the successful commercial micropropagation of radiata pine by Tasman Forestry in New Zealand (Gleed, 1992). Tapping the benefit of

micropropagation techniques to capture more genetic gain will further speed up progress in forestry improvement work.