MOLECULAR MARKERS: APPLICATION IN TREE IMPROVEMENT PROGRAMMES

Aarif Ali Gattoo

Faculty of Forestry, Sher e Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar-191121
aarifali123@gmail.com

Abstract : A molecular marker is a nucleotide sequence corresponding to a particular physical location in the genome. Molecular markers are important tools for forest tree improvement. The most important markers are restriction fragment length polymorphisms (RFLPs), polymerase chain reaction- (PCR) based markers such as random amplified polymorphic DNA (RAPD), and fingerprinting markers. DNA markers can supplement isozyme markers for monitoring tree improvement activities such as estimating genetic diversity in breeding populations, germplasm identification, verifying controlled crosses, and estimating seed orchard efficiencies. Isozyme markers have been applied extensively during the past 15 years and have contributed significantly to tree breeding programs. Isozymes generally provide ample genetic information and are relatively inexpensive, rapid, and technically easy to apply, thus they should continue to play an important role in forest tree improvement.

Keywords : Molecular markers, RAPD, Tree improvement programmes

INTRODUCTION

The use of traits in plant as markers for their genetic relationship predates genetics itself. In the 18th century, Carl Linnaeus used the number and arrangement of plants sexual organs to determine their systematic relationship. Gregor Mendel derived his principles of inheritance by following visible traits in the progeny of sexual crosses, and the use of morphological markers has continued to the present day. Markers play an essential role today in the study of variability and diversity, in the construction of linkage maps, and in the diagnosis of individuals or lines carrying certain linked genes. Within this context, the limitations of morphological markers became quickly apparent. They tend to be restricted to relatively few traits, display a low degree of polymorphism, are often environmentally variable in their manifestation, and can depend on the expression of several unlinked genes. Furthermore, some may affect plant viability or seed set, distorting gene frequencies in the progeny. The emergence of marker systems has, for the last 30 years, closely tracked developments in biochemistry and molecular biology. Morphological markers were largely supplemented by biochemical markers, particularly isoenzymes that could be easily scored by electrophoresis (Ganapathy and Scandalios 1973; Tanksley 1983). The limitations of isoenzmyes as markers, in particular both the limited number of polymorphic enzymes that can be conveniently stained and the environmental effects on expression pattern, were apparent already twenty years ago (Tanksley 1983). The shortcomings drove the development of markers based on DNA polymorphisms. These marker types generate ‘‘fingerprints,’’ distinctive patterns of DNA fragments resolved by electrophoresis and detected by staining or labeling. A molecular marker is in essence a nucleotide sequence corresponding to a particular physical location in the genome. Its sequence needs to be polymorphic enough between plant accessions to allow its pattern of inheritance to be easily followed.

Genetic markers

A genetic marker is a measurable character that can detect variation in either a protein or DNA sequence. A difference, whether phenotypic or genotypic, may act as a genetic marker if it identifies characteristics of an individual’s genotype and/or phenotype, and if its inheritance can be followed through different generations. A genetic trait may not have necessarily observable consequences on an individual’s performance. Sometimes, however, this trait may be linked to, or correlated with, other traits that are more difficult to measure and do affect the individual’s performance. In such cases, these unobservable genetic traits may be used as genetic markers for the linked traits because they indirectly indicate the presence of the characteristics of interest. The two measures can be correlated, using an analysis of inheritance and studying the distribution of the characteristics in both parents and offspring. Molecular markers specially RAPD are being applied to a greater extent in forest trees to study genetic diversity and contributes about 25% of the total molecular markers (Fig. 1) used in forest biotech activities (FAO, 2004). Molecular markers assess variations in the nucleotide sequence of DNA of different individuals. Molecular markers are numerous and therefore a large genome can be easily assayed for existence of any variation, such genetic markers are easy to score. Use of molecular markers therefore provides an objective assessment of genetic diversity in a

plant species and enables unequivocal identification of elite genotypes.

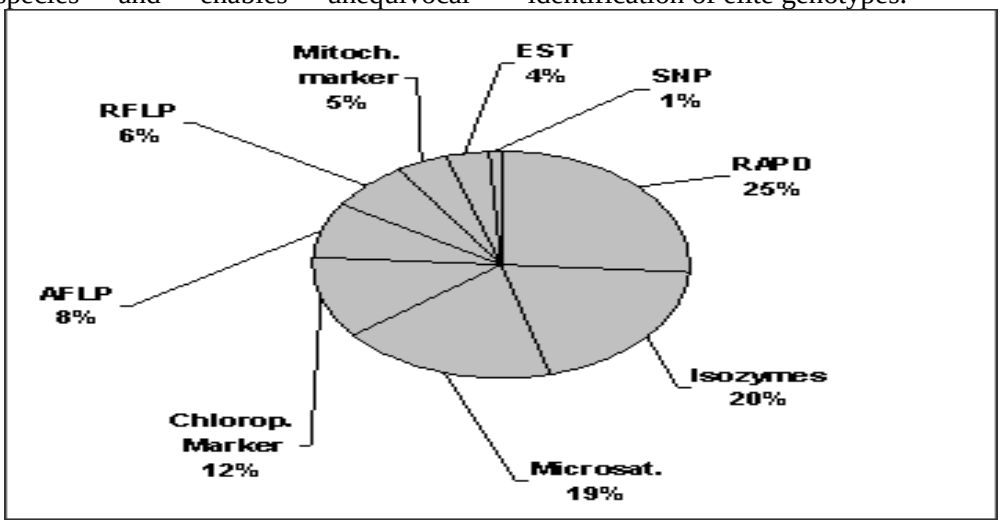

Fig 1: Distribution of molecular markers used in forest biotechnology activities

Types of Genetic markers

1. Morphological markers

Traditionally, diversity within and between populations was determined by assessing differences in morphology. These measures have the advantage of being readily available, do not require sophisticated equipment and are the most direct measure of phenotype, thus they are available for immediate use, an important attribute. However, morphological determinations need to be taken by an expert in the species, they are subject to changes due to environmental factors and may vary at different developmental stages and their number is limited.

2. Protein (biochemical) markers

To overcome the limitations of morphological traits, other markers have been developed at both the protein level (phenotype) and the DNA level (genotype). Protein markers are usually named biochemical markers' but, more and more; they are mistakenly considered as a common class under the so-called 'molecular markers'.

Protein markers (seed storage proteins and isozymes) are generated through electrophoresis, taking advantage of the migrational properties of proteins and enzymes, and revealed by histochemical stains specific to the enzymes being assayed.

Detecting polymorphisms i.e. detectable differences at a given marker occurring among individuals in protein markers is a technique that shares some of the advantages of using morphological ones. However, protein markers are also limited by being influenced by the environment and changes in different developmental stages. Even so, isozymes are a robust complement to the simple morphometric analysis of variation.

3. DNA (molecular) markers

DNA polymorphisms can be detected in nuclear and organelle DNA, which is found in mitochondria and chloroplasts. Molecular markers concern the DNA molecule itself and, as such, are considered to be objective measures of variation. They are not subject to environmental influences; tests can be carried out at any time during plant development; and, best of all, they have the potential of existing in unlimited numbers, covering the entire genome.

Classification of molecular markers based on the basic strategy

1. **Non PCR based approaches**
   - RFLP (Restriction fragment length polymorphism)
2. **PCR-based approaches**
   - RAPD (Random Amplified Polymorphic DNA)
   - SSR (Simple Sequence Repeat)
   - SCAR (Sequence characterized amplified regions)
   - CAPS (Cleaved amplified polymorphic sequence)
   - ISSR (Inter-simple sequence repeats)
   - AFLP (Amplified fragment length polymorphism)

DNA Sequencing

DNA sequencing is the most fundamental measure of diversity because it detects polymorphisms within the DNA's building blocks themselves. DNA sequencing is done by breaking the DNA into fragments, which are then subcloned. Each short piece is used as a template to generate a set of fragments that differ in length from each other by a single base. Fragments are separated by gel electrophoresis. The base at the end of each fragment is identified ('base-calling').

The original sequence of As, Ts, Cs and Gs is recreated for each short piece generated in the first step. The short sequences are assembled into one long sequence.

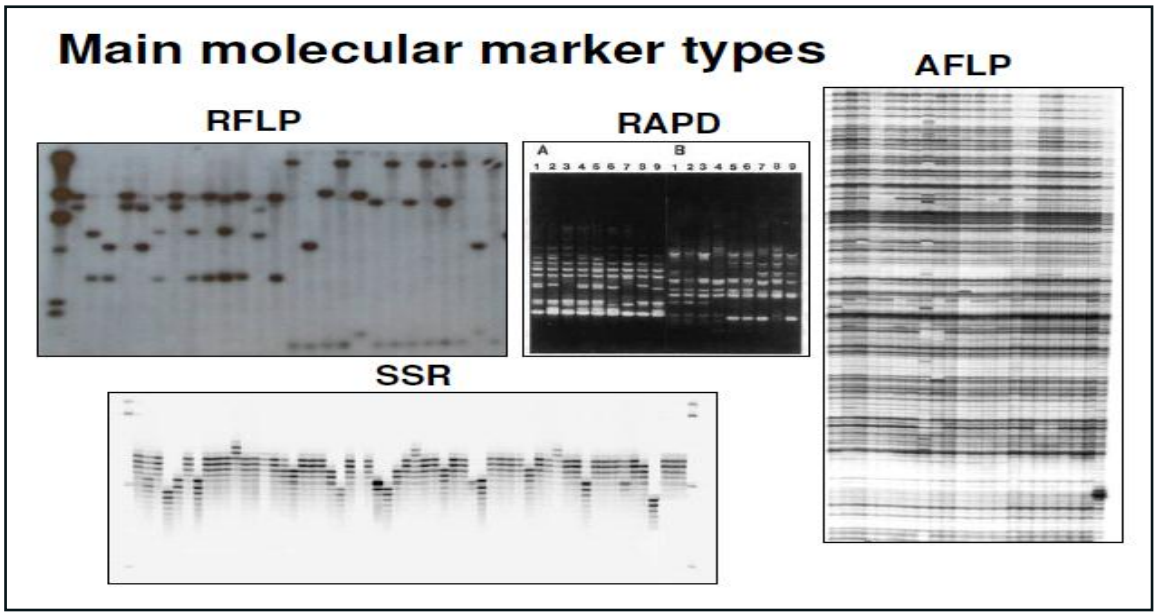

Fig: 2: Comparison of banding patterns generated by different markers

Application of molecular markers in forest tree improvement programmes

1. **Molecular markers for identification of subspecies: rapid PCR-based diagnostic test**
   Using five informative microsatellite markers and a Bayesian statistical approach, we developed an efficient polymerase chain reaction-based diagnostic tool for the rapid identification of individuals and populations of the *Acacia saligna* species complex of Western Australia. 189 individuals from 14 reference populations previously characterized based on morphology and used these data to investigate population structure in the species complex. High total genetic diversity (HT=0.729) and high population differentiation (¸=0.355) indicated strong intra- specific structuring. With the provision of prior population information, the reference data set was optimally resolved into four clusters, each corresponding to one of the four main proposed subspecies, with very high membership values (Q>97%). The reference data set was then used to assign individuals and test populations to one of the four subspecies. Assignment was unequivocal for all test individuals from two populations of subsp. *lindleyi* and for all but one individual of subsp. *stolonifera*. Individuals from populations of subsp. *saligna* and subsp. *pruinescens* showed a degree of genetic affinity for the two subspecies in their assignments, although the majority of individuals were correctly assigned to subspecies. The diagnostic tool will assist in characterizing populations of *A. saligna*, especially naturalized and invasive populations of unknown origin (Millar *et al*., 2008).

2. **Molecular markers for genome sequence of black cottonwood (*Populus trichocarpa*)**
   The genome sequence of *Populus trichocarpa* was screened for genes encoding cellulose synthases by using full-length cDNA sequences and ESTs previously identified in the tissue specific cDNA libraries of other poplars. The data obtained revealed 18 distinct CesA gene sequences in *P. trichocarpa*. The identified genes were grouped in seven gene pairs, one group of three sequences and one single gene. Evidence from gene expression studies of hybrid aspen suggests that both copies of at least one pair, CesA3-1 and CesA3-2, are actively transcribed. No sequences corresponding to the gene pair, CesA6-1 and CesA6-2, were found in Arabidopsis or hybrid aspen, while one homologous gene has been identified in the rice genome and an active transcript in *Populus tremuloides*. A phylogenetic analysis suggests that the CesA genes previously associated with secondary cell wall synthesis originate from a single ancestor gene and group in three distinct subgroups. The newly identified copies of CesA genes in *P. trichocarpa* give rise to a number of new questions concerning the mechanism of cellulose synthesis in trees (Djerbi *et al., 2005*).

3. **Genetic linkage map of european chestnut (*Castanea sativa* MILL.) based on RAPD, ISSR and isozyme markers**
   A genetic linkage map of European chestnut (*Castanea sativa* Mill.) based on RAPD, ISSR and isozyme markers was constructed using the two-way pseudo- testcross strategy. A total of 96 individuals from a F1 full-sib family were genotyped with 381 molecular markers (311 RAPDs, 65 ISSRs, 5 isozymes). Markers in

testcross configuration, segregating 1:1, were used to establish two separate maternal and paternal maps including 187 and 148 markers, respectively. The markers identified 12 linkage groups based on the haploid number of chestnut. The female and male framework maps reached a total length of 720 and 721 cM (Kosambi), respectively, representing a 76% and 68% coverage of the overall genome. A total of 46 markers, found in intercross configuration, segregating 3:1 and 1:2:1, were used to identify homologous linkage groups between parental maps; out of 12 linkage groups 11 could be joined. RAPD and ISSR markers showed a good and comparable reliability, allowing for the first time the establishment of a saturated linkage map for European chestnut. These maps will be a starting point for studies on the structure, evolution and function of the chestnut genome. Identification of QTLs for adaptive traits in chestnut will be the primary target (Casasoli *et al*., 2001).

4. **Molecular markers for testing genetic fidelity:**

More recently, molecular markers have also been used for testing the genetic fidelity during micropropagation/*ex situ* conservation on the one hand, and for characterization of plant genetic resources on the other. This aspect of the use of molecular markers has received attention in recent years due to the significance that is being attached to micropropagation of elite genotypes and to the *in situ* and *ex situ* conservation of plant genetic resources (PGRs). Molecular markers have particularly been suggested to be useful for confirmation of genetic fidelity in micropropagated tree species, where life span is quite long and performance of micropropagated plants could only be ascertained after their long juvenile stage in field conditions. a study on *Picea* the genetic integrity during somatic embryogenesis has been studied using RAPDs. In India also, an extensive study on genetic fidelity and molecular diagnostics in micropropagation systems was carried out where several molecular markers including RFLPs (using rDNA probes and mtDNA probes), RAPDs, MP-PCR and oligonucleotide in-gel hybridization were used in micropropagated clones of 4 tree species namely *Populus deltoides, Eucalyptus tereticornis, E. camaldulensis* and *Coffea Arabica*(Rani *et al* 1998*).* RFLPs (using nDNA and cpDNA probes) and RAPDs were also used for characterization and identification of genetic resources of perennial crops like *Musa* and to solve problems related to plant genetic diversity conservation (Bhat, 1997).

5. **DNA fingerprinting and classification of geographically related genotypes**

A reliable and reproducible method to detect RAPD and AP-PCR polymorphisms, using DNA from olive-tree (*Olea europaea* L.) leaves was developed. Starting from their natural orchards, fifty-six olive-tree cultivars throughout Málaga province, including oil and table olive cultivars, were screened and grouped into 22 varieties. A total of 62 informative polymorphic loci that provide 601 conspicuous bands were enough to differentiate the varieties. Clustering analyses managing 3 different pairwise distances, as well as phylogenetic analyses, led to the same result: olive-trees in Málaga can be divided into three main groups. Group I (90% of certainty) contains wild type and two introduced varieties, group II (83% of certainty) covers some native olive-trees, and group III (58% of certainty) is a heterogeneous cluster that includes varieties originating and cultivated in a number of Andalusian locations. Geographic location seems to be the first responsible of this classification, and morphological traits are needed to justify the group III subclustering (Claros *et al.*, 1997*).*

6. **Molecular markers for establishing distinctness in vegetatively propagated crops**

Distinctness, uniformity and stability (DUS) testing of varieties is usually required to apply for Plant Breeders' Rights. This exam is currently carried out using morphological traits, where the establishment of distinctness through a minimum distance is the key issue. The possibility of using microsatellite markers for establishing the minimum distance in a vegetatively propagated crop (grapevine) has been evaluated. A collection of 991 accessions have been studied with nine microsatellite markers and pair-wise compared, and the highest intra-variety distance and the lowest inter-variety distance determined. The collection included 489 diVerent genotypes, and synonyms and sports. Average values for number of alleles per locus (19), Polymorphic Information Content (0.764) and heterozygosities observed (0.773) and expected (0.785) indicated the high level of polymorphism existing in grapevine. The maximum intra-variety variability found was one allele between two accessions of the same variety, of a total of 3,171 pair-wise comparisons. The minimum inter-variety variability found was two alleles between two pairs of varieties, of a total of 119,316 pair-wise comparisons. In base to these results, the minimum distance required to set distinctness in grapevine with the nine microsatellite

markers used could be established in two alleles (Ibáñez *et al*., 2009).

7. **Molecular markers for ex situ conservation**
The enormous losses suffered by the European elms during recent Dutch elm disease outbreaks led to concern over the conservation of elm genetic resources, and the subsequent establishment of a series of ex situ collections. However, as ex situ collections are inevitably finite in size, some consideration needs to be given to selecting which samples to include in them. To contribute towards this process for European ex situ elm collections we have undertaken genetic studies on a Europe-wide sample of 535 individuals. A major aim has been to use genetic markers to clarify the identification of samples to ensure that the ex situ collections contain a representative spread of taxonomic diversity. This is important given the paucity of mature elms in the landscape due to Dutch elm disease. The lack of mature material (critical for identification) compounds identification problems in what was already a taxonomically difficult group. Our data (derived from random amplified polymorphic DNA and inter-simple sequence repeats) have provided a useful supplement to morphology in undertaking such sample identifications. The molecular data served to highlight mis-identified samples and led to extensive revisions of sample identities within individual countries. Our results were less useful in detecting regional intra-specific genetic structure, and do not provide sufficient information for prioritizing within-species sample selections (Copestake *et al*., 2005).

8. **Molecular markers for genetic variation and putative hybridization**
Genetic variability was estimated by enzyme electrophoresis in 239 Belgian clones from the *Salix alba-S, fragilis* complex. Most of the allozyme differentiation was between the species and less between the regions. The goal was mainly to estimate the size of a 'population' at the level of a location, and to identify possible hybrid populations (or with a relevant proportion of hybrid and introgressed individuals). Objective of this study further was to investigate as well the use of allozyme variation in *Salix* clones from the field in order to detect the amount and hierarchical distribution of the genetic diversity. This morphological complex suggested a high frequency of hybrids. The clones were pooled as a single coadapted species complex and secondly as belonging to either species, i.e. being *S. alba-like* or *S. fragilis-like.* The standard genetic variability measures showed higher values for the complex than for the separate species (Triest, 1999)

REFERENCES

Bhat, K. V., Lakhanpaul, S., Chandel, K. P. S. and Jarret, R. L. (1997). In Molecular Genetic Techniques for Plant Genetic Resources (eds. Ayad, W. G., Hodgkin, T., Jaradat, A. and Rao, V. R.), *IPGRI*, pp. 107–117.

Claros, M., Brunak, S. and von Heijne, G. (1997). Prediction of N-terminal protein sorting signals. Curr. Opin. Struct. Biol. 7, 394-398

Copestake, W.P., Hollingsworth, M. L., Hollingsworth, P. M., Jenkins G. I. and Collin, E . (2005). Molecular markers and ex situ conservation of the European elms (Ulmus spp.). *Biological Conservation* 122 : 537 – 546

Casasoli M. C. Mattioni M. Cherubini · F. Villani. (2002). A genetic linkage map of European chestnut (*Castanea sativa* Mill.) based on RAPD, ISSR and isozyme markers. *Theor Appl Genet* 102: 1190–1199

FAO. (2004). *The state of food and agriculture (SOFA) 2003–2004*. *Agricultural biotechnology: meeting the needs of the poor?* FAO Agriculture Series No. 35. Rome.

Djerbi S, Lindskog M, Arvestad L, Sterky F, Teeri T.T. (2005). The genome sequence of black cottonwood (Populus trichocarpa) reveals 18 conserved cellulose synthase (CesA) genes. *Plant* 221: 739–746

Ganapathy P.S, Scandalios J.G. (1973). Malate dehydrogenase isozymes in haploid and diploid Datura species. Their use as markers in somatic cell genetics. *J Hered* 64:186–188

Ibáñez, J., A. Vargas, M. Palancar, J. Borrego, and M.T. de Andrés. (2006). A genetic study on table grape varieties through microsatellite analysis. Abstr. 9th International Conference on Grape Genetics and Breeding, Udine, Italy. International Society for Horticultural Science, Leuven, Belgium.

Millar M. A., M. Byrne, I. Nuberg, M. Sedgley. (2008). A rapid PCR-based diagnostic test for the identification of subspecies of *Acacia saligna*. *Tree Genetics & Genomes* 4:625-635

Tanksley, S.D. (1983). Molecular markers in plant breeding. *Plant Mol Biol Rep* 1:3–8

Triest L., B. De Greef, S. Vermeersch, J. Van Slycken, And E. Coart. (1999). Genetic variation and putative hybridization in *Salix alba* and *S. fragilis* (*Salicaceae*): evidence from allozyme data. *Pl. Syst. Evol*. 215:169-187.