Invited presentation

Marker assisted selection in plantation forestry: Principles, practices, problems and prospects

K.K.Vinod
Indian Agricultural Research Institute
Rice Breeding and Genetics Research Centre, Aduthurai - 612 101
email: kkvinodh@gmail.com

Background

The term ‘plantation forestry’ by and large refers to organized planting of tree species for industrial use in large scale plantations. Most of the plantation forests are even-aged monoculture of trees with the purpose of producing wood, fuels, oils, tannins, resins, rubber and fodder. Although conventional breeding has still much to offer, breeding progress has been particularly slow in most of the plantation forestry species, mainly due to very long breeding cycle. Presently, the application of DNA based molecular markers in conventional breeding programmes seems to be the most useful application of biotechnology in plantation crops. This will ensure progress through conventional breeding, and marker assisted selection (MAS) can help to reduce the time scale.

Molecular markers and development of linkage maps

DNA based molecular markers are small regions of target genome, which can be selectively amplified using a pair of primers, in a polymerase chain reaction (PCR). They are unique, ubiquitous and distributed throughout the genome and follow Mendelian inheritance. They may be either dominant or co-dominant and may either be random or linked. Random markers are non-specific amplified bands and arise mostly from non-coding DNA that are useful in ‘fingerprinting’ and in studies of genetic diversity. Linked markers are those lie proximally to functional genes and are linked to useful characters. Linked markers are essential for developing efficient MAS strategies.

Molecular mapping is based on the simple genetic principles, of linkage and
recombination. Based on the segregation within marker loci that can be scored as parental types and recombinants, the recombination fraction (frequency) can be estimated as the map distance between the loci, lying on a linkage group (chromosome) from a segregating population developed from two genetically contrasting individuals. A linkage map is then constructed by comparing and ordering the distances between several loci in a linear order.

**Populations and mapping strategies**

Ideally, linkage mapping requires populations derived by intercrossing two genetically divergent homozygous individuals (purelines or inbreds). These mapping populations are structured because their range of genetic variability is confined between two inbreds. Most basic level population suitable for linkage mapping is a $F_2$ population. However $F_2$ populations are unstable on breeding, and cannot be replicated on experimentation unless cloned. Moreover, the level of genetic dispersion is low in $F_2$ families. A recombinant inbred line (RIL) population is advanced decedents of $F_2$ individuals that are homozygous. RILs are ideal mapping populations because they are immortal and stable on breeding. Their genetic dispersion is high because of many cycles of meioses while on development.

Plantation tree species are predominantly outbred (cross-pollinated) producing heterozygous and heterogeneous population. They exhibit inbreeding depression, making them highly unsuitable for RIL development. Molecular marker based linkage analysis and map construction is therefore difficult and complicated in outbred species. However, extensive efforts have been devoted in establishing marker rich linkage maps in major economically important genera such as *Eucalyptus*, *Acacia*, *Hevea*, *Pinus*, *Picea*, *Populus* and many more. In a heterozygous outbred species, markers may be dominant or co-dominant and linkage phase (coupling or repulsion) between two adjacent marker pair is unknown. Therefore most feasible approach is to use the populations derived from heterozygous parents of contrasting nature. Segregating populations can be derived in three ways, a) inbreeding or selfing b) full sib outcross and c) half-sib outcross. Either half-sibs or full-sibs are more useful in mapping then inbred lines. Full-sib allow more accurate mapping than half-sibs. By determining maximum likelihood estimators of recombination frequencies of different marker combinations, molecular linkage maps of both parents can be separately constructed and compared. Alternate mapping strategies such as (a) haploid tissue genotyping (b) bulked segregant analysis
(c) two-way pseudo testcross (d) double pseudo backcross and (e) half-sib mapping are commonly used in perennial species.

**Haploid tissue genotyping**

In conifers, the megagametophyte is derived from repeated mitotic divisions of a single meiotic product, and has the same maternal genetic complement as the embryo contained in the same seed. The megagametophyte is haploid of maternal origin, and therefore, segregation and recombination can be evaluated in a sample of seeds from a single tree without the need for controlled pollinations. Sufficient DNA for several reactions can be extracted from one megagametophyte, making this an excellent system for constructing genetic linkage maps. However, this system is highly inefficient in QTL analysis as only one parent is involved in segregation analysis.

**Bulked segregant analysis**

Bulked segregant analysis helps in identifying genetic differences in the form of identifiable amplicon variation from a set of segregants that prominently differ in a phenotypic trait. Here DNA from contrasting phenotype group segregants are bulked and analysed for the identifiable variation between a random set of markers. Once any such marker is identified it is used for validation among individual sergeant to ascertain its power to distinguish phenotype classes. In case of absolute identification, further, such markers can be used in marker assisted selection without the need of developing an elaborate map.

**Two-way pseudo testcross**

Most common and widely used approach is a ‘two-way pseudo testcross’ mapping method, first used in *Eucalyptus*. A ‘pseudo testcross’ is a full-sib method based on dominant markers which segregate in the test-cross ratio of 1:1 between two heterozygous parents. Dominant markers segregate in three different fashions, viz., (i) paternally inherited test cross of 1:1 ratio, (ii) maternally inherited test cross of 1:1 ratio and (iii) both parents inherited intercross of 3:1 ratio. Test cross markers based on the parental source of inheritance are scored separately for both pollen and pistil parents, and individual linkage maps are constructed simultaneously for both.

The term “pseudo test-cross” is used because mating configuration of the markers in the testcross fashion was not known *a priori* as in a conventional testcross. Rather, the configuration was inferred *a posteriori* after analyzing the segregation pattern of a marker in the progenies with respect to parental configuration for that marker, in a cross between highly heterozygous parents.
with no prior genetic information. As this inference is done for both the parents simultaneously, the term “two-way” was appropriately prefixed.

**Double pseudo-backcross**

Double pseudo-backcross is an extension of the pseudo-testcross model in which, instead of backcrossing $F_1$s to original parents to avoid inbreeding depression, $F_1$s are crossed to unrelated individuals of the same species of pollen and seed parents separately to mimic a backcross to original parents, hence the name “pseudo backcross”. The pseudo-backcross produced second-generation hybrid progenies offered advantages over first generation progenies such as (a) increased level of heterozygosity (b) more number of recombination resulting in a higher proportion of segregating (c) increased proportion of shared polymorphisms, which facilitate higher resolution, and (d) the opportunity to study the fertilization ability of recombinant gametes produced by the $F_1$. The pseudo-backcross approach is suitable for outbred species, wherein RIL development of conventional backcrossing is not feasible due to inbreeding depression or self-incompatibility.

**Half-sib mapping**

Populations derived from open pollinated mother trees constitute standard half-sib populations in plantation forest trees. Here only markers belonging to maternal genome can be recognized, while pollen parents remain obscure. Nevertheless, success of this strategy remains in two key factors, (i) a large marker diversity and (ii) availability of mother tree specific rare alleles. Rare alleles are those which are present in extremely low frequency in the pollen pool, but present in the mother plant. By analyzing the segregation of these dominant alleles in test cross pattern, maternal linkage map can be constructed. It has been demonstrated that, if sufficient genetic diversity and marker polymorphisms are available, the amount of rare alleles from a mother tree is sufficient enough to produce a reliable map. Even if marker alleles are rare in the population (<2%), by increasing the number of segregating progenies, map construction is still possible (Liu 1988). Microsatellite markers throw rare alleles in outcross tree species which can also be used in test cross like fashion.

**Mapping of quantitative trait loci**

Most of the economically important traits in plantation forestry have quantitative inheritance and low heritability. Therefore, mapping of genes using simple linkage analysis is seldom successful. However, with the formidable advancements in the analysis of molecular marker segregation data it is now possible to estimate marker-trait
segregation among segregating progenies, either singly or in combination of markers. Reliable associations with marker and trait segregation can be identified, that could lead to identification of the genes present near such loci. Popularly known as ‘quantitative trait loci’ (QTL), these genomic regions that influence a quantitative trait can be location or near location of a gene or a cluster of linked genes, a part of structural genes, regulators of gene expression or non-coding regions that affect gene expression. QTLs are therefore necessary to carry out MAS for a particular trait they map to.

The fundamental aspect of detecting a QTL is the understanding of the marker-trait association. There are three different levels of marker-trait association (a) direct markers, i.e. loci that code for the functional domain, (b) linkage disequilibrium (LD) markers, loci that are in population-wide LD with the functional domain, and (c) linkage equilibrium (LE) markers, loci that are in population-wide LE with the functional domain. Among these, direct markers can be used for MAS straightaway, while LD markers can be used identifying the gene(s) and thereby direct markers. LE markers segregate irrespective of the genes of interest making them unsuitable for mapping and MAS. There are two types of QTL mapping, (a) linkage mapping and (b) linkage disequilibrium mapping. In linkage mapping or linkage analysis a structured population is used to detect the phenotype variation homogenous with a marker variation to claim a QTL. Therefore, there is a prerequisite of a structured population such as RIL is necessary. In linkage disequilibrium (LD) mapping (also called association mapping) an unstructured population is used to detect QTLs using the advantage of evolutionary conservation of haplotype blocks. An unstructured population can be a natural population having wide variation of phenotype classes. Such techniques are highly useful in plantation forestry where development of a structured population is cumbersome, time consuming, expensive and may be genetically unviable.

Marker assisted selection – problems and prospects

Having identified QTL linked markers, logical next step is to carry out MAS. MAS provides a radical shift towards genotype-based selection from environment influenced phenotype-based selection. The linked markers facilitate selection for a trait even before the trait is expressed, such as disease resistance when there is no disease or mature traits at juvenile stage, thereby significantly accelerating the breeding progress. And
in case of other important commercial traits like wood quality, the major is to improve the rate of genetic gain by reducing the long generation interval. In plantation forests species, molecular maps have been used to locate markers associated with traits of commercial interest, such as wood properties, leaf oil composition, vegetative propagation, growth, frost tolerance and disease resistance.

Besides, aiding selection for traits that are expressed late in the life of the tree species in juvenility, molecular markers can aid in screening for traits that are extremely difficult, expensive or time consuming to score phenotypically such as quantitatively inherited or environmentally sensitive traits such as root morphology, resistance to quarantined pests or to specific races or biotypes of diseases or insects, tolerance to certain abiotic stresses such as drought, salt and mineral deficiencies or toxicities. Furthermore, markers help in distinguishing the homozygous from the heterozygous condition of many loci in a single generation without the need for progeny testing which is very important in marker assisted backcross breeding (MABC). Simultaneous MAS for several characters at the same time is also possible.

Although very popular in annual crops, MABC is not always a viable solution in plantation forestry. Alternatively, the simple and best logical method is to advance those lines which contain those alleles with a positive effect on the quantitative trait, the selection criterion being the summation index showing maximum accumulation of the desired QTL. A collateral advantage of such an approach is that it offers a true validation of the putative genes (QTL) for the traits of interest. The associated response to marker selection namely indicates the presence of true genes; in particular in the vicinity of markers strongly affected by the selection imposed.

MAS is effective when the genetic base of the participating genotypes is broad, so that identification of specific QTL related to target traits is easy. Ideally, the QTLs used for MAS should have been validated across different environments and genetic backgrounds. This again may not be feasible in all plantation species. Keeping this in view, restraint may be applied in formulating programs for MABC and QTL pyramiding in perennial species. Furthermore, QTLs based on LD mapping may be more viable in plantation species than that of QTLs identified by linkage mapping, because the former is precise and powerful on account of evolutionary conservation of haplotypes, while the later may be
confined to the founders of the structured population associated with the QTL. Although important in breeding context, understanding the genetic basis of QTL by QTL interaction (epistasis) and QTL by environment interaction (QEI) is rather difficult in perennial species.

**Major challenges**

a) Temporal variation in phenotypes, age x age trait correlations, and late expressing phenotypes;

b) Issue of duplicates and narrow genetic base in introduced and recently domesticated species;

c) Breeding schemes that involve altering the frequencies of favourable alleles through recurrent selection in large populations;

d) Variable heritabilities in breeding populations with a heterogeneous genetic base and in linkage equilibrium;

e) Use of relatively small progeny sizes in QTL detection;

f) Use of poor repeatable markers systems such as RAPD, AFLP for construction of linkage maps, which makes use of such maps unviable.

g) Large QTL intervals, rendering them unsuitable for MAS.

h) The effect of a QTL may vary with changing biotic and abiotic factors

i) The cost of mapping and MAS is soaring high.

**References**


**Keywords:** Genetic mapping, marker assisted breeding, QTL, double pseudo backcross, BSA.