

**Introduction**

Genome of each organism is the blue print that determines the ultimate life. Genes in each genome are identified based on the phenotypes produced by them. However, many of the genes are not identified because of complex phenotypic outputs. Each of these loci can be placed in order to construct linkage map of any organism. Construction of linkage maps is not new. In the past, mostly morphological markers were used to construct the linkage maps of different plant and animal genomes. However, the availability of less markers slowed down the approach of linkage map construction in several species. The advent of molecular markers - polymorphic and locus characterised by a number of variable length restriction fragments - paved the way for easy and faster linkage map construction. Demonstration of restriction fragment length polymorphisms (RFLPs); and the construction of RFLP maps have been completed in major crop species such as tomato (Bernatzky and Tanksley, 1986), maize (Helentzaris et al., 1987), potato (Bonierbala et al., 1988) and rice (McCouch et al., 1988). The recent development in molecular biology simplified many aspects in the genome mapping based on the position of certain landmarks (often referred to as genetic markers) in the genome. The present review highlights the various steps involved in linkage map construction and the possible utilities of mapping in crop breeding.

**Genetic map**

Genetic maps identify the linear arrangement of genes on a chromosome and are assembled from meiotic recombination data. These are theoretical maps which give the order of genes on a chromosome. They cannot pinpoint the physical whereabouts of genes or determine how far apart they are. Distances on this map are not directly equivalent to physical distances. The unit of measurement is the centimorgan (cM).

**Physical map.**

Physical maps identify the actual physical position of genes on a chromosome. Distances are measured in base pairs (bp), kilobases (kb = 1000 bases) or megabases (mb = 10,000,000 bases).

**Construction of a Genetic Map**

The construction of a genetic map involves the linkage analysis of genes segregating among the progenies of a sexual cross. According to the Mendel's Law of Independent Assortment, members of different pairs of alleles assort independently of each other when the germ cell is formed. This is only true if the genes are on different chromosomes. If they are on the same chromosome, they are said to be linked. The frequency of recombination of a pair of linked gene is constant and characteristic for that pair of genes. The strength of linkage and recombination observed, is a function of the distance on the chromosome between the genes in question. The greater the distance recombination is expected between genes. In other words, the frequency of recombinants (not similar to both the parents) determines the distance between genes. When multiple genes are considered for their recombination frequency, one can group all the genes which are linked and the resultant group is called as linkage group. The number of linkage groups equals the haploid number of chromosomes.
The early work on linkage map construction was mainly based on the segregation of easily observable morphological traits among the progenies of a cross. There are several examples of linkage maps resulted from the morphological traits, otherwise called as morphological markers.

Morphological markers are easily observable traits with discrete phenotype. Though these markers are highly advantageous, their number is the limitation. Further, these markers are influenced by the environment and the genetic background of the individual. Whereas the molecular markers lack the above mentioned negative points and the differences between the morphological and molecular markers are given in the following table.

<table>
<thead>
<tr>
<th>Morphological markers</th>
<th>Molecular markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less in number</td>
<td>More in number</td>
</tr>
<tr>
<td>Interact with the environment</td>
<td>No environmental influence</td>
</tr>
<tr>
<td>Show differential expression across genotypes</td>
<td>Phenotypically neutral</td>
</tr>
<tr>
<td>Not distributed throughout the genome</td>
<td>Distributed throughout the genome</td>
</tr>
<tr>
<td>Expression depends on the ontogeny of the individual</td>
<td>Possibility of using these markers at any growth stage of the individual</td>
</tr>
<tr>
<td>Dominant-recessive expression</td>
<td>Codominant or dominant</td>
</tr>
</tbody>
</table>

To construct a complete genome map of plant species, one should have a set of markers at DNA level and a suitable mapping population. Construction of genetic maps using molecular markers generally involves the following steps: viz.,

1) Parental survey to identify polymorphic probes,
2) Establishing the segregation of polymorphic probes among the progenies,
3) Linkage analysis to establish linkage groups based marker-segregation data and
4) Assigning chromosome numbers to linkage groups. However, mapping strategies vary according to the nature of markers and the mapping population (McCouch et al., 1988; Causse et al., 1994; Maheswaran et al., 1997)

Markers at DNA level

The markers at DNA level include short DNA sequences, whole genes or even longer sequences of DNA. These markers may be of different types.

Restriction Fragment Length Polymorphisms (RFLPs)

RFLPs are genetic differences observable at the DNA level characterised by a number of variable length restriction fragments. Original documentation of RFLPs came from the findings of Grodzicker et al., (1974). Jeffreys (1979) elegantly demonstrated the presence of RFLPs in globin genes of man. The RFLP analysis involves DNA cleavage by restriction enzymes, electrophoresis of resulting fragments, Southern transfer of separated fragments to a membrane support, radioactive labelling of suitable probes, hybridisation of probes to membrane supported fragments and detection as a banding pattern on X-ray film. RFLPs are codominant markers, inherited in a simple Mendelian fashion (Botstein et al 1980). RFLPs can be of various types: single copy sequences, multiple-copy sequences and repeated sequences. Among these, single copy sequences are highly amenable for linkage map construction and to some extent, the multiple copy sequences.

Plate 1. a) Autoradiogram revealing RFLPs between various rice accessions when probed with RG 64, a clone from rice genomic library of McCouch et al (1988)
amplification as an alternative to Southern blot analysis. Informative primers are synthesized from known sequences, or end sequencing the anonymous clones to amplify the regions of interest. This approach is called Sequenced Tagged Site (STS) approach (Traasoung et al., 1992). Recently, most of the genome mapping projects are underway to convert all the RFLP and other clones into informative primer sequences to amplify specific sequences by PCR. STSs are also codominant markers as RFLPs.

Randomly Amplified Polymorphic DNAs (RAPDs)

The generation of RAPDs involves the use of single short random oligonucleotides. When these random primers are mixed with genomic DNA and thermostable DNA polymerase and subject to PCR will prime the amplification of several DNA fragments. The DNA amplification with random primers expose polymorphisms distributed throughout the genome (Williams et al., 1990). The use of RAPD markers in genome mapping and gene tagging has been exploited by Martin et al., (1991), Paran et al., (1991), Michelmore et al., (1991) and Reiter et al., (1992). RAPDs do have their drawbacks. Among them, repeatability and stability are potential snags. Further, RAPDs are usually dominant markers thus preventing the accurate detection of heterozygotes (Williams et al., 1990). Considering the difficulties, dominant RAPD markers are converted into codominant RFLP markers for a stable performance.

Sequence Characterised Amplified Regions (SCARs)

Sequence Characterised Amplified Regions (SCARs) are PCR based markers that represent a single genetically defined loci that are identified by PCR amplification of genomic DNA with pair of specific primers. Amplified RAPD products are cloned and sequenced. The sequences are used to design 20-24 mer oligonucleotide primers. These markers are more reliable and PCR based. Paran and Michelmore, (1993) succeeded in producing SCARs. Thus, the polymorphism revealed by the RAPD markers as the presence or absence of specific band amplified by PCR which are dominant can be converted into codominant SCAR markers.
Variable Number Tandem Repeats (VNTRs)

Variable Number Tandem Repeats (VNTRs) include microsatellites, minisatellites and hyper variable regions. Microsatellites are arrays of tandemly repeated DNA sequences which occur dispersed throughout the genome (Jeffreys et al., 1985). These are otherwise called as "Sequence Tagged Microsatellite Site (STMS)". Microsatellites consist of around 10-50 copies of motifs from 1 to 5 bp that can occur in perfect tandem repetition, as in perfect repeats or together with another repeat type. Microsatellites occur frequently and randomly in all eukaryotic DNAs and are named analogy with the larger minisatellite arrays. The minisatellites are highly polymorphic but are less common than microsatellites and have larger sequence motifs extending over more than 1 kb, making them less amenable to PCR analysis. The VNTRs can be analysed by conventional RFLP analysis and also by using PCR.

Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphisms (AFLPs) technology is a technique of recent origin having a potential impact on genome fingerprinting and mapping. It combines both classical restriction-based and recent PCR-based approaches. AFLPs are obtained by various means and Selective Restriction Fragment Amplification (SRFA) is one among them. SRFA is a method by which a selective restriction fragments of a total genomic digest are detected by amplification using PCR. SRFA involves three major steps viz., i) cutting genomic DNA with restriction enzyme(s) ii) ligation double stranded adapters to the restriction fragments and iii) amplifying selective restriction fragments using universal primers. SRFA may be performed with a single enzyme, but the best results are achieved using two different enzymes, a rare cutter and a frequent cutter. AFLP is a powerful, reliable, stable and rapid assay with potential application in genome mapping. DNA fingerprinting and marker-assisted breeding (Vos et al., 1995). This PCR-based technique permits inspection of polymorphism at a large number of loci within a very short period of time and requires very small amounts of DNA. The reproducibility of AFLPs is ensured by using restriction site specific adapters and adapter-specific primers with a variable number of selective nucleotides under stringent amplification conditions (Vos et al., 1995). AFLPS are currently considered as the molecular markers of choice within the genome mapping community.
Apart from the development of new types of markers, various technological simplifications are made to resolve the problems associated with molecular marker technology. The recent PCR-based approach, gel free visualisation of PCR products and automation at various steps are boons to the molecular marker approaches adopted for genome mapping. Further simplifications are
expected, as and when new findings in the molecular biology come to the surface.

Mapping populations

The genome map of an organism summarises much of the genetic information available for that species and can serve as a reference for the development and testing of additional genetic hypotheses. However, generation of a complete linkage map remains a daunting task, for want of suitable population. Various options for generating a suitable mapping populations are available.

Segregating populations

In most of the map construction, F2 segregating populations are used. These populations are the result of selfing F1s of two homozygous inbred lines. Most of the molecular maps to date are based on segregation data from F2 progenies. In some cases, the segregating progenies of F1s backcrossed to recurrent parent were also used to construct linkage maps. The saturated RFLP map of rice, is of one such origin from the backcross progenies of Oryza sativa/ O. longistaminata cross (Causse et al., 1994).

Recombinant Inbred lines (RILs)

Developing a population of recombinant inbred lines (RILs) is an alternative strategy in mapping projects. RILs are developed by continuous selfing of F2 individuals until the homozygosity is achieved. Several permanent populations of RILs are developed to construct maps (Burr et al., 1988; Burr and Burr, 1991; Wang et al., 1994). The advantage of using a RILs as a mapping populations is that the resulting map shows higher resolutions of closely linked loci as compared F2 populations (Burr and Burr, 1991). The major disadvantage is that constructing RILs is a time consuming process.

Doubled Haploid Lines (DHLs)

Generation of doubled haploids from anther or microspore culture is an established technique to attain homozygosity in a single step in most of the crop species. Moreover, among doubled haploids recombination events are fixed resulting in stable recombinational values. Doubled haploids were used for linkage map construction in wheat (Chao et al., 1990), and rice (McCouch, 1990; Huang et al., 1994; Maheswaran et al., 1997). Though the strategy sounds good, evolving diploid population is a genotype-dependent process to the in vitro culture conditions.

Mapping the genes

Molecular marker technology remains as a tool for locating genes governing agronomically important traits via linkage to mapped DNA sequences. Phenotypic evaluation at the whole plant level or at the cellular level provides information which can be used to determine the chromosomal location of the genes that confer the phenotype of interest. This is accomplished by analysing linkage between mapped molecular markers and expression of the target phenotype in a range of related individuals. Markers linked to the genes of interest function as "genetags" facilitating selection of favourable alleles in a breeding programme.

As in linkage map construction, gene tagging component also needs a suitable population in which the trait to be tagged with molecular markers shows clear-cut segregation with a higher level of polymorphism for the molecular markers. The process of gene tagging involves two steps: 1) surveying parents with molecular markers for their level of polymorphism, and 2) surveying the polymorphic markers on progenies with an aim to tag the trait of interest with a molecular marker(s).

Tagging major genes

Establishing associations between molecular markers and simply inherited traits is comparatively easier. To date, several major genes have been tagged with molecular markers and among them genes conferring resistance to pest and diseases are more common. The strategies followed to tag major genes are:

1. For most of the traits, the genetics is not an established phenomenon. As far as crop species are concerned, the genetics of agronomically important traits has not been studied, except for the traits such as resistance to various insects and diseases. Under these circumstances, if one wishes to tag genes of a trait, the general process of parental survey
followed by progeny survey can be followed. This is a time consuming process.

2. When a breeder understands the genetics of a trait, the next step is the construction of Near-isogenic lines (NILs). Near isogenic lines are constructed via the back cross breeding method. These NILs possess constant donor parent (DP) DNA introgressed into the recurrent parent (RP). The strategy involves survey of RP, NIL and DP with molecular markers to identify putative positives (markers present in the DP and NIL..
but absent in RP). Then the putative positives are surveyed on a segregating population of RP and NIL to establish linkage between a marker and the gene of interest. Thus, NILs remain as a potential resource in the molecular-marker approach (Mushibauer et al., 1988).

3. Construction of NILs is a laborious and time-consuming process. An alternative strategy called, Bulked Segregant Analysis (BSA) has been proposed by Michelmore et al., (1991). In a population segregating for a trait of interest, individuals are sorted into two groups based on the expression of that trait. The two groups will differ at loci linked to the trait, but will be randomly segregating for unlinked regions. DNA samples from each group are pooled separately, and the two DNA pools are screened for markers' polymorphism. Markers polymorphic between the two pools are expected to be linked to the trait of interest. This is a viable strategy to a variety of simply inherited traits to tag their genes with molecular markers.

Tagging genes controlling quantitative traits

In crop breeding, breeders are mostly concerned with those traits that are polygenically controlled. Location of polygenes in individuals by conventional analysis is difficult. The advent of molecular marker technology provides the geneticists with powerful new tools for identifying the component Mendelian loci of those complex by inherited traits. The main practical limitation to localizing Quantitative Trait Loci (QTLs) seems to be the availability of suitable markers (Thoday, 1961). This limitation was remedied by the construction of complete and saturated linkage maps permitting systematic searches of an entire genome for QTLs influencing a trait (Paterson et al., 1983).

In general traits influenced by several genes, the effects of any one gene are partly masked by other genes/or by environment. Thus it is difficult to discern the effect of any one gene by merely looking at the appearance (phenotype) of the individual. Using DNA markers, QTLs can be described by their chromosomal location, dosage effect, phenotypic effects and sensitivity to the environment (Paterson et al., 1991). The identification and recombination of non-allelic polygenes can encourage breeders to accumulate the genes with like effect distributed in genetic materials to produce transgressive variation from which one can obtain the merit of true breeding genotypes (Xu, 1989). This has been demonstrated using molecular markers in an interspecific cross of tomato (de Vicente and Tanksley, 1993). Several analytical approaches have been developed to mapping QTLs with molecular markers. These include single marker analyses such as comparison of marker means, by ANOVA, regression analysis, and likelihood approach and many marker analysis using the methods of interval mapping. Lander and Botstein (1986) and Paterson et al. (1988) discussed new methods to map complex traits using a complete RFLP linkage map.

As far as QTL mapping is concerned, the phenotype must be evaluated in well replicated trials in different environments. Considering this point, the use of RILs or DHLs can be a good strategy for evaluating identical genotypes in different environments and simultaneously scoring them for molecular markers (McCouch, 1993). The early segregating generations of a cross may not represent the optimal strategy for the complete analysis of QTL-marker association.

Marker Aided Selection (MAS)

The development of molecular markers promises to overcome most of previous limitations associated with morphological markers. Tight linkage of a marker to a gene can be exploited for indirect selection of traits in a breeding programme. Two prerequisites for adopting marker aided selection in breeding programmes are:

1) a tightly linked marker to the gene concerned and

2) a population which is polymorphic for the marker and the gene which are in extreme linkage disequilibrium

Several aspects regarding marker-aided selection have been discussed by Beekman and Soller (1989), Stuber (1989) and Melchingar (1990). In plant breeding, two distinct methods of
selections are followed - one is for germplasm improvement (recurrent selection) and other for cultivar or hybrid development. These two applications are separated because recurrent selection usually is applied to random mating populations possibly at or near linkage equilibrium, whereas cultivar or hybrid development typically begins with populations derived by crossing elite inbred lines at or near maximum linkage disequilibrium.

Lande and Thompson (1990) and Lande (1992) investigated the efficiency of MAS for both individual and mass selection in random mating populations. There are three approaches to applying MAS to plant breeding.

a. selection based on markers alone with no measurement of phenotype,
b. simultaneous selection on markers and phenotype, and
c. two stage selection, the first stage involving use of markers to select among seedlings and second involving phenotypic selection among surviving individuals.

The potential efficiency of MAS depends upon the heritability of the trait, the proportion of genetic variance explained by the markers, and the selection method. A major practical problem in using MAS is that recombination will reduce linkage disequilibrium between the markers and genes, thus diminishing effectiveness of the selection. The successful application of MAS will require very tight linkages between marker and the trait.

Molecular Markers in Crop Breeding

Plant breeding efforts always include an element of chance, because numbers of genotypes and environments and consequent phenotypes to be evaluated are limitless (Wallace, 1985). The molecular markers may complement efforts to reduce the burden to a greater extent, especially in the selection of parents (based on molecular marker diversity), improving the screens for the selection of qualitative and quantitative traits, and understanding the architecture of the trait. Paterson et al. (1991) and Lamkey and Lee (1993) discussed all the details about the molecular markers and their potentials in crop improvement. Though the molecular marker approach has its own potential, several questions remain unanswered.

a. many major genes have been tagged with markers at a distance of 5-15 cM. In such cases, fine mapping is needed to use markers for predicting the phenotype. If the genes are fine mapped, whether these markers are useful to predict the phenotype of the same trait in other crosses or in an array of germplasm?

b. some traits are controlled by two or more major genes interacting with each other manifesting a phenotype. Under such conditions, marker based approach complicates the issue of selection. How can we partition the phenotypes of each gene to have individual markers? Will it be possible to use markers of different genes to select a common phenotype?

c. for most of the traits, phenotyping remains to be a difficult process and needs multidisciplinary approach. Detection of major or minor genes depends on the ability to phenotype individuals in clear-cut way. Under such circumstances, is it advisable to go for molecular marker approach for all the traits?

d. how to tackle the problems associated with QTL mapping? Several issues such as population size, nature, test locations etc. are to be settled.

e. Marker Aided Selection: Though several genes have been tagged, still markers are not put into use. Where does the gap exist? If so, how to bridge the gap?

f. for many traits, the genetics is not established which makes gene mapping tedious. Under these circumstances, can the QTL approach be of any help to know about the major genes? Can the major genes be fixed by differential phenotyping and QTL analysis?

g. in plant breeding, allelism test to identify individual genes is a difficult process. Can we
use the molecular markers to overcome this problem?

h. germplasm survey is a vital component in crop improvement. Can the germplasm survey of different ecotypes of each species with molecular markers and associating these markers with phenotype(s) be used as alternative strategy for gene tagging?

i. identifying cross specific markers for each trait is very difficult, considering the volume of crosses breeders make. To overcome this problem, can a strategy be established to trace the genes from germplasm at first and then confirming at crosses level (bottom up approach) instead from a cross to germplasm (top down approach)?

The objective of using molecular markers in plant breeding is to try and predict a plants’ performance on the basis of patterns of marker alleles.

Conclusion

Complexity of the techniques makes the molecular marker approach more labour intensive and skill oriented one. The molecular genetic expertise, though available, cost involved is a factor which cannot be comparable with the existing breeding strategies. Further, the gap existing between the molecular agriculturists and plant breeders is enormous - molecular agriculturists are not aware of breeding strategies and breeders do not have either time or skills for this work. Other aspect is that measuring phenotypes in the field is generally something plant breeders enjoy. The constant problems plant breeders face with their thousands of germplasm materials and hundreds of crosses they make with clear-cut goals are enormous. They need ways and means to overcome their problems and all plant breeders are heavy users of technology and will adopt rapidly any new technology that can be proved to augment and to improve the efficiency and the cost effectiveness of their breeding programmes. Adaptation and application of molecular markers to plant improvement will depend on the efficiency and more effectiveness of the technique to be improved in coming years.

REFERENCES


