



Agri search with a human touch

Molecular Markers in Crop Improvement

Easy & rapid

Reproducible genotyping

Accurate phenotyping

Authentic mapping population

Choice of parents



Indian Institute of Pulses Research
Kanpur - 208 024



Molecular Markers in Crop Improvement

D. Datta

Sanjeev Gupta

S.K. Chaturvedi

N. Nadarajan



**Indian Institute of Pulses Research
Kanpur - 208 024**



Printed : September, 2011

Published by : Dr. N. Nadarajan, Director
Indian Institute of Pulses Research, Kanpur

Edited by : Mr. Diwakar Upadhyaya

Correct citation : Datta, D., Gupta, Sanjeev, Chaturvedi, S.K. and Nadarajan, N. (2011): Molecular Markers in Crop Improvement. Indian Institute of Pulses Research, Kanpur - 208 024

Preface

With the advent of marker-assisted selection (MAS), a new breeding tool is available to make more accurate and useful selections in breeding populations. MAS allows heritable traits to be linked to the DNA segments that are responsible for controlling that trait. These segments of DNA or QTLs (Quantitative Trait Loci) can be detected through specific laboratory techniques.

The most commonly used method is Polymerase Chain Reaction (PCR) that amplifies segments of DNA linked to heritable traits such as yield or disease resistance. This method is useful because the DNA that we amplify is different (polymorphic) between cultivars. It is this difference that we use to determine whether the plant has the desired trait or not. The process in which the differential DNA sites (or primer sites) are explored, comes from genetic mapping techniques, i.e. RAPD, microsatellites etc. With a marker assisted selection breeding program the simpler methods are necessary since they are time and cost effective. PCR is an effective method for generating large quantities of a specific DNA sequence from a small amount of starting DNA. This technique is useful for a MAS breeding program because the results are reliable.

To learn how MAS works, basic molecular biology principles need to be understood. The present bulletin “Molecular markers in crop improvement” has been designed to provide a basic understanding with regards to use of molecular markers in crop improvement. This bulletin describes basic concepts used in marker assisted breeding programme, different applications of MAS and basic principles underlying DNA extraction, PCR, running of gel and data analysis. The help rendered by Shri Diwakar Upadhyaya in editing the manuscript is duly acknowledged. We hope that this bulletin will be of immense use for students and trainees of molecular breeding.

Authors

Contents

●	Preface	
1.	Basic Concepts	1
2.	Use of Molecular Markers in Breeding Programmes	24
3.	Principles of Basic Techniques	34
4.	Literature Cited	49

Basic Concepts

During the past 20 years there has been rapid growth in the relatively new field of plant biotechnology and its associated techniques. These have application, not only for the manipulation of biological systems for the benefit of mankind, but also to undertake studies for better understanding of the fundamental life processes. Consequently, it has become the fastest and most rapidly growing technology in the world. Biotechnology is defined as “any *technique that uses living organisms (or parts of organisms) to make/ modify products, to improve plants and animals or to develop microorganisms for specific uses*”. It offers efficient and cost-effective means to produce an array of novel, value-added products and tools. It has the potential to increase food productivity, reduce the dependency of agriculture on chemicals, lower the cost of raw materials and reduce the negative environmental impacts associated with traditional production methods.

Conventional breeding is a dynamic area of applied science. It relies on genetic variation and uses selection to gradually improve plants for traits and characteristics that are of interest for the grower and the consumer. Another important way of improvement is the introduction of new genetic material (*e.g.*, genes for biotic and abiotic stress resistance) from other sources, such as gene bank accessions and related plant species. Although, current breeding practices have been very successful in producing a continuous range of improved varieties, recent developments in the field of molecular biology can be employed to enhance plant breeding efforts and to speed up cultivar development. Modern biotechnology provides new tools that can facilitate development of improved plant breeding methods and augment our knowledge of plant genetics. The discovery of restriction enzymes by Smith and Wilcox, and the polymerase chain reaction (PCR) by Kerry Mullis and his group has created opportunity to understand the composition of organisms at the DNA level, and obtain a so-called genetic fingerprint. These studies are routinely done by the separation of DNA-fragments on a gel that results from a selective digestion of DNA with enzymes or from a selective amplification of DNA using PCR. DNA fragments that result in different gel patterns between samples or individuals are called polymorphic markers. The visible differences on the gel result from differences at the DNA level. Not all types of markers are the same; the information content depends on the method that is used to obtain the marker data and the population in which the markers were ‘scored’. Advanced tools for the retrieval of marker data and the subsequent analysis have been developed that allow quick and reliable results in most plant species.

Molecular (DNA) markers are segments of DNA that can be detected through specific laboratory techniques. For detection of markers, either restriction enzymes or Polymerase Chain Reaction (PCR) or their combination are used to generate/amplify the DNA sequences that are linked to a heritable trait such as yield or disease resistance. With the advent of marker-assisted selection (MAS), a new breeding tool is now available to make more accurate and useful selections in breeding populations. The objective of this section is to introduce genetic terminologies and concepts associated with molecular markers.

Allele

One alternative form of a given allelic pair - tall and dwarf are the alleles for the height of a pea plant. More than two alleles can exist for any specific gene, but only two of them will be found within any diploid individual. In terms of molecular marker the variant of a DNA sequence is referred as an allele. An allele defined by molecular means should have exactly the same genetic properties as a phenotypically defined allele. Molecular alleles should segregate by the same Mendelian principles as phenotypic alleles. Mostly, molecular alleles are selectively neutral. Theories of population genetics apply to molecular alleles as well.

Backcross

It is the cross of an F_1 hybrid to any one of the homozygous parents.

Testcross

It is the cross of any individual to a homozygous recessive parent. It is used to determine if the individual is homozygous dominant or heterozygous.

Pure line

An individual that breeds true to type for a particular trait. This was an important innovation because any non-pure (heterozygous parents) would confuse the segregation ratio in genetic experiments.

Homozygote

An individual which contains only one allele at the allelic pair, for example 'TT' is homozygous dominant and 'tt' is homozygous recessive. Pure lines are homozygous for the gene of interest.

Heterozygote

An individual which contains one of each member of the gene pair, for example the 'Tt' heterozygote.

Dominance

It is the ability of one allele to express its phenotype at the expense of an alternate allele. It represents the major form of interaction between alleles.

Epistasis

The interaction between two or more genes to control a single phenotype. Epistasis is the interaction between different genes. If one allele or allelic pair masks the expression of an allele at the second gene, that allele or allelic pair is epistatic to the second gene.

Suppressor

A genetic factor that prevents the expression of alleles at a second locus. This is an example of epistatic interaction.

Modifier genes

Genes that have small quantitative effects on the level of expression of another gene. Instead of masking the effects of another gene, a gene can modify the expression of a second gene.

Genotype

In general, genotype is the genetic architecture of an individual. Genotype is also used to refer to the pair of alleles present at a single locus. For a diploid organism, with alleles 'A' and 'a' three possible genotypes are AA, Aa and aa. The expression of the genotype is affected by the genetic constitution of the individual and the environment. With reference to molecular markers, genotype is used to refer to the alternate DNA sequences (molecular alleles) of an individual.

Phenotype

In classical terms phenotype is any detectable characteristic of an organism determined by $G + E + G \times E$ interactions and inherited as Mendelian factor. Phenotype is used to describe a trait/morphology of an individual *viz.*, flower colour, whereas phenotypic value is the mean of the measurements of a trait *viz.*, height of a plant, weight of a fruit.



Fig. 1: Wild pigeonpea sowing leaf and pod phenotypes

Phenotyping

The term ‘phenotyping’ of mapping population refers to scoring (disease resistance, yield or any other morphological observation) of individuals of the mapping population.

Genotyping

It is a rather loose terminology to describe the DNA profiling (RFLP, RAPD or PCR profiles) of individuals of the mapping population or the breeding population for detecting presence/absence of molecular markers. Thus, genotyping is done to generate information on the status of individuals with respect to presence or absence of a set of specific molecular markers.

Genetic marker

Any easily scoreable phenotype that is linked with a trait of interest intended to be marked. They are used to ‘flag’ the position of a particular allele or the inheritance of a particular character. Phenotypes for which the variation observed in the population of interest is partially or entirely explained by a single “Mendelian” factor. Three properties that define a genetic marker are:

- It should be locus-specific
- It should be polymorphic in the studied population
- It should be easily phenotyped.

The quality of a genetic marker is typically measured by its:

- Heterozygosity in the population of interest
- Polymorphism Information Content (PIC).

Polymorphism Information Content is defined as the probability of identifying one homologue of a given parent that transmitted an allele to a given offspring, the other parent being genotyped as well.

PIC= probability that the parent is heterozygous x probability that the offspring is informative

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n \sum_{j=i+1}^n 2p_i^2 p_j^2$$

Morphological marker

A morphological marker is expressed as a specific and distinct morphological trait. Morphological marker may be affected by environment. Generally it is incompletely linked with the gene of interest. Its phenotypic expression may be dependent on growth stage. These markers are rare in a natural population and show extremely low level of polymorphism.

Molecular markers

Molecular markers are specific fragments of DNA that can be identified within the whole genome. Molecular markers are found at specific locations of the genome. They are used to ‘flag’ the position of a particular gene or the inheritance of a particular character. Molecular markers are phenotypically neutral.

Marker categories

Depending on the technique used for detection and amplification of markers there can be different classes of markers. Restriction fragment length polymorphism (RFLP) is based on restriction site changes in the target DNA and subsequent hybridization with probe DNA. Random amplified polymorphic DNA (RAPD), Sequence characterized amplified region (SCAR) and Sequence tagged sites (STS) are based on mutation at primer annealing site in the target DNA. Cleaved amplified polymorphic sequence (CAPS) and Amplified fragment length polymorphism (AFLP) are based on both restriction site changes and mutation at primer annealing site in the target DNA. Additionally, there are Simple sequence repeat (SSR), Inter simple sequence repeat (ISSR) and Single nucleotide polymorphism (SNP) markers. Highly specialized techniques are required to detect SNP's.

Dominant marker

A marker is called dominant if only one form of the trait (which is targeted to be marked) is associated with the marker, whereas the other form of the trait is not associated with any marker. Such markers can not discriminate between heterozygote and homozygote marker allele (Fig.2).

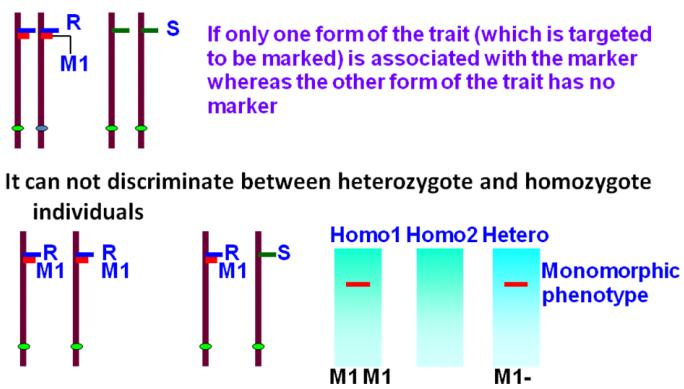


Fig. 2: Illustration of dominant marker

Co-dominant marker

A marker is designated as co-dominant if both forms of the trait (which is targeted to be marked) are associated with the marker. It can discriminate between heterozygote and homozygote marker allele (Fig.3).

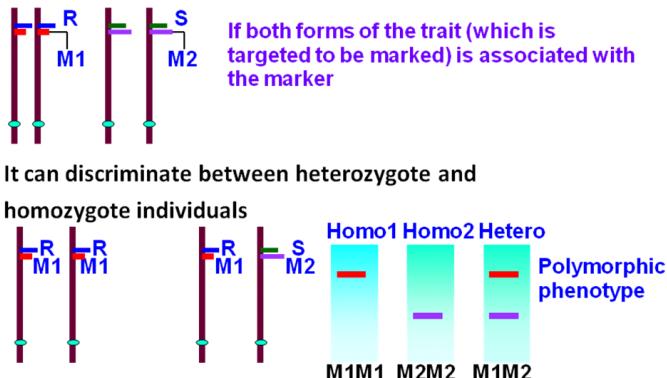


Fig. 3: Illustration of co-dominant marker

Random amplified polymorphic DNA (RAPD)

RAPD is a PCR based method, which employs single primers of arbitrary nucleotide sequence with 10 nucleotides to amplify anonymous PCR fragments from genomic template DNA. In RAPD analysis, the target sequence(s) (to be amplified) are unknown. In RAPD, PCR is generally carried out with arbitrary primers. The amplifications are visualized through agarose gel electrophoresis. For amplification to occur it is essential that primers anneal in a particular orientation (such that they point towards each other) and the primers must anneal within a reasonable distance to one another.

Advantages

- No prior knowledge of DNA sequences is required
- Random distribution throughout the genome
- The requirement for small amount of DNA (5-20 ng)
- Easy and quick to assay
- The efficiency to generate a large number of markers
- Commercially available decamer primers are applicable to any species
- The potential automation of the technique
- RAPD bands can often be cloned and sequenced to make SCAR (sequence-characterized amplified region) markers
- Cost effectiveness as compared to other markers.

Limitations

- Dominant nature (heterozygous individuals cannot be separated from dominant homozygous)
- Sensitivity to changes in reaction conditions, which affects the reproducibility of banding patterns
- Co-migrating bands can represent non-homologous loci
- The scoring of RAPD bands is open to interpretation
- The results are not easily reproducible between laboratories.

Applications of RAPD

- Measurements of genetic diversity
- Genetic structure of populations
- Germplasm characterisation
- Verification of genetic identity
- Genetic mapping
- Development of markers linked to a trait of interest
- Cultivar identification
- Identification of clones (in case of soma-clonal variation)
- Interspecific hybridization
- Verification of cultivar and hybrid purity
- Clarification of parentage

Restriction fragment length polymorphism (RFLP)

RFLP is a molecular marker based on the differential hybridization of cloned DNA to DNA fragments in a sample of restriction enzyme digested DNAs (Fig. 4). RFLPs involve digestion of genomic DNA with restriction enzymes (bacterial enzymes that cut DNA at specific sequences known as restriction sites). The resulting DNA fragments are size fractionated on gel electrophoresis, transfer of fractionated DNA

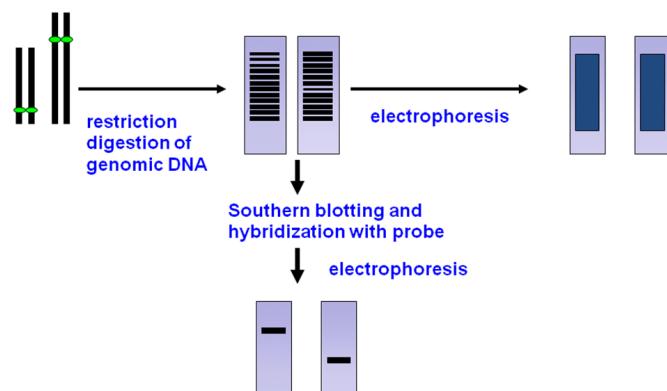


Fig. 4: Illustration of RFLP development

fragments on Nylon membranes (a process known as Southern blotting) and finally hybridization with labeled probe to visualize DNA polymorphisms. The first step in RFLP analysis is to derive a set of clones that can be used to identify RFLPs. The two primary sources of these clones for RFLP mapping of plants are cDNA clones and PstI-derived genomic clones. RFLP markers are defined by a specific enzyme-probe combination. This technique is highly reproducible, and the markers are co-dominant in their inheritance therefore, allows the differentiation of heterozygotes from homozygotes. RFLP procedure is time consuming and expensive but they have been used to generate saturated genetic map. RFLPs behave like any other Mendelian trait. Each band seen in a Southern blot indicates the presence of one or more restriction sites in a sequence. The sequence containing a restriction site is one allele, while the corresponding sequence missing the restriction site is the other allele. The “phenotypes” of these alleles are the differences in banding patterns, due to presence or absence of bands. RFLP loci are co-dominant (twice as much informative in a genetic cross as compared to dominant markers like RAPDs).

Amplified fragment length polymorphism (AFLP)

Amplified fragment length polymorphisms (AFLPs) are polymerase chain reaction (PCR)-based markers for rapid screening of genetic diversity. AFLPs are DNA fragments with different nucleotide sequence of which large number of copies have been amplified *via* PCR. This technique is a combination of the RFLP and PCR techniques. Like RFLP, the AFLPs are highly heritable and polymorphic. The technique involves restriction digestion of DNA with two different enzymes and ligation of two adaptors, selective amplification of sets of restriction fragments and gel analysis of amplified fragments. The amplified products are generally separated on a denaturing polyacrylamide gel and visualized using autoradiography. The technique is more skill demanding than RAPD and also requires more amount of DNA. The reproducibility of AFLP is ensured by using site specific adaptors. AFLP method rapidly generates hundreds of highly replicable markers from DNA of any organism, and thus, they allow high resolution genotyping of fingerprinting quality. The time and cost efficiency, replicability and resolution of AFLPs are superior or equal to those of other markers [allozymes, random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), microsatellites], except that AFLP methods primarily generate dominant rather than co-dominant markers. Because of their high replicability and ease of use, AFLP markers have emerged as a major new type of genetic marker with broad application in systematics, pathotyping, population genetics, DNA fingerprinting and quantitative trait loci (QTL) mapping.

Cleavage amplification polymorphisms (CAPs)

The scoring of this type of marker is dependent on the variation of size of fragments following the digestion of the PCR product by a restriction enzyme. A completely new set of CAPs markers would be generated from a different restriction enzyme.

Simple sequence repeat (SSR) marker

Microsatellite or Simple sequence repeats (SSRs) provide fairly comprehensive genomic coverage. They are amenable to automation, they have locus identity and they are multi-allelic. Many agronomic and quality traits show quantitative inheritance and the genes determining these traits have been quantified using Quantitative trait locus (QTL) tools. SSR markers have wide applicability for genetic analysis in crop improvement strategies. They are widely used in plants because of their abundance, hyper-variability, and suitability for high throughput analysis.

Inter-simple sequence repeat (ISSR) marker

Inter-simple sequence repeat (ISSR) are semi-arbitrary markers amplified by PCR in the presence of one primer complementary to a target microsatellite. Amplification in presence of non-anchored primers also has been called microsatellite-primed PCR, or MP-PCR. Such amplification does not require genome sequence information and leads to multi-locus and highly polymorphic patterns. Each band corresponds to a DNA sequence delimited by two inverted microsatellites. Like RAPDs, ISSRs markers are quick and easy to handle, but they seem to have the reproducibility problem because of the longer length of their primers.

STS and SCAR markers

Random amplified polymorphic DNA (RAPD) is an application of PCR where arbitrarily chosen 10 base primers are used to search for variation in DNA. RAPD data can contain artifacts and are not fully reproducible. However, RAPDs have been used to generate large number of genetic markers useful for linkage mapping quickly and cheaply. RAPD fragments can be separated on agarose gels. The excised bands from the gel can be re-amplified in to individual bands from gel slices using the original RAPDs primer. The fragments can be cloned and sequenced. The sequence data can be used to design PCR primers specific to RAPDs fragments, and use PCR to produce specific RAPD's fragments from genomic DNA, which then can function as sequence tagged sites (STSs) or Sequenced characterized amplified region (SCAR). This method allows for rapid generation of STSs derived from RAPD fragments and eliminates the problems associated with reproducibility.

Factors influencing efficiency of a marker

Efficiency of markers depends on their closeness to the linked trait; how the phenotype of marker is affected by environment; consistency in phenotypic expression; how easy is to score the phenotype; and level of polymorphism. Ideally, a marker should be polymorphic, tightly linked with the trait of interest, highly heritable, co-dominant, easily scoreable and it should not affect the fitness of the individual. DNA markers have many advantages over the morphological markers. DNA markers are phenotypically neutral which is a significant advantage compared to traditional

phenotypic markers, highly polymorphic, abundant, usually randomly distributed throughout the genome, easily scoreable and as such DNA markers are not affected by environment but the gene of interest may be sensitive to GxE and hence its (DNA marker) association with phenotype of the gene may vary with change of environment.

Bulk segregant analysis

Often a geneticist is not interested in developing a molecular map, but would rather find a few markers that are closely linked to a specific trait. Identification of these markers is often achieved by a procedure called bulk segregant analysis. The essence of this procedure is the creation of a bulk sample of DNA for analysis by pooling DNA from individuals with similar phenotypes. For example, you may be interested in finding a molecular locus linked to a disease resistance locus. You would create two bulk DNA samples, one containing DNA from plants or lines that are resistant to the disease and a second bulk containing DNA from plants or lines that are susceptible to the disease. Each of these bulk DNA samples will contain a random sample of all the loci in the genome, except for those that are in the region of the gene upon which the bulking occurred. Therefore, any difference in RFLP or RAPD pattern between these two bulks should be linked to the locus upon which the bulk was developed.

Population

In classical genetics, population is defined as a group of potentially interbreeding individuals. Whether haploid or diploid, a population has two basic attributes: gene frequencies and gene pool. Gene (allele) frequency is the proportion of different alleles of a gene in a population, whereas gene pool is the sum total of genes in the reproductive gametes of a population. If external forces do not apply then gene frequencies remain unchanged from one generation to the next generation in a random mating population. It may be noted that gene frequency in particular generation is dependent upon the gene frequencies of previous generation and frequencies of different genotypes depend on gene frequency alone. After one generation of random mating and in the absence of external forces the genotypic frequencies remain stable which said to be at equilibrium. In terms of molecular breeding, a population may be defined as a group of individuals originating from a cross combination that is capable of representing frequencies of alternate alleles and which allow to calculate gene frequencies in a predictable manner.

Mapping population

Mapping population consists of individuals of one species, or in some cases they are derived from crosses among related species. It is a group of individuals on which genetic analysis is carried out. It can be either segregating for traits under study or a set of near homozygous lines representing a F₂ variation. In both situations mapping population is generally derived from a single cross whose parents were polymorphic for the trait of interest.

Different types of mapping populations

Mapping population may comprise F_2 , backcross, recombinant inbred lines (RIL), doubled haploid lines (DHL), F_2 derived F_3 ($F_2 \times F_3$) populations and near-isogenic lines (NILs). Specialized populations are required in cases of cross pollinated species that can't tolerate inbreeding, perennials and the trees.

F_2 , backcross, and recombinant inbred are the three primary types of mapping populations used for molecular mapping. A F_2 population is developed by selfing (or intermating for cross pollinated species) among F_1 individuals. These F_1 individuals are developed by crossing two parents that show significant polymorphism for whichever type of loci you are going to score. Backcross populations are developed by crossing the F_1 with one of the two parents used in the initial cross. Advantage of F_2 population is that it is the most suitable population for preliminary mapping. It requires less time and effort for development. The major drawback in using F_2 or backcross populations is that the populations are not eternal. Therefore, the source of tissue to isolate DNA or protein will be exhausted at some point of time then its mapping has to be undertaken in another population. Populations of recombinant inbred lines can be a powerful solution to this problem. Recombinant inbred lines are developed by single-seed random selection from individual plants of a F_2 population (Fig.5). Because of this procedure, these lines are also called F_2 -derived lines. Single-seed descent is repeated for several generations. At this point, all of the seeds from an individual plant are bulked. For example, a $F_{3,4}$ RI population underwent single-seed descent through the F_3 generation, and was bulked to develop the F_4 . This population of seed can then be grown to obtain a large quantity of seeds of each individual line. Importantly, each of the lines is fixed for many recombination events. These lines have several uses. First, they can be used to derive a map because it is essentially an eternal F_2 population with unlimited mapping possibilities. Additionally, these lines can be scored for morphological traits (such as disease resistance or flower colour) or quantitative traits (such as yield or maturity). This morphological trait data can then be compiled and those traits can be placed on the developing molecular map. These lines are especially powerful for analyzing quantitative traits because replicated trials can be analyzed using identical genetic material. The quantitative trait data can then be used to determine if any molecular markers are closely associated with those traits.

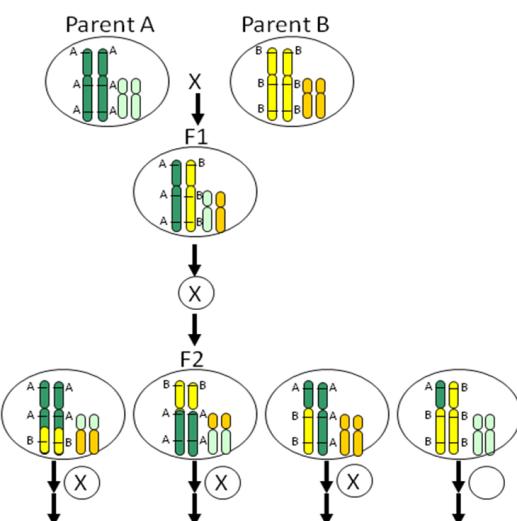


Fig. 5: Illustration of development of RIL population

Doubled haploid population

Doubled haploids (DH) are also the products of one meiotic cycle, and hence comparable to F_2 in terms of recombination information. DHs are permanent mapping population and hence can be replicated and evaluated over locations and years and maintained without any genetic change. These are useful for mapping both qualitative and quantitative characters. It provides opportunity to induce homozygosity in single generation and instant production of homozygous lines. But in DH lines recombination from the male side alone is accounted. Since it involves *in-vitro* techniques, relatively more technical skills are required in comparison with the development of other mapping populations. It is often suitable culturing methods / haploid production methods are not available for a number of crops, and different crops differ significantly for their tissue culture response. In addition, tissue culture induced variation should be taken care.

Near isogenic lines (NIL)

NILs can be generated through two different breeding procedures. It is developed through repeated selfing and selecting heterozygous individuals until sufficient homozygosity is attained for all traits except for the trait of interest. NILs can also be generated by backcrossing the F_1 plants to the recurrent parents and selecting the trait of interest in each generation. NILs developed through backcrossing are similar to recurrent parent except for the gene of interest, whereas the NILs generated through selfing are produced in pairs of near identical individuals (identical for all traits except for the loci of interest). Like DHs and RILs, NILs are also immortal mapping population. NILs are quite useful in functional genomics. NILs have disadvantages too. They require many generations for development. These are directly useful only for molecular tagging of the concerned gene but not for linkage mapping. Linkage drag is a potential problem in constructing NILs.

Size of the mapping population

The size of mapping population depends on type of mapping population, genetic nature of the target traits, objectives of the experiment, resources available for handling a sizable population. Depending on the need, the mapping population may vary from 100 to 3000 individuals. Generally 200 to 300 individuals would be suffice.

Choice of parents for deriving a mapping population

Parents should be polymorphic for the trait under study. It is desirable to choose parents which are adapted to the conditions where its progenies will be phenotyped. Unadapted and exotic parents may pose difficulties in phenotypic evaluation. Interspecific crosses are required if contrasting parents (which are distinct for the traits under study) are not available in the same species.

Efficiency of mapping population

Efficiency of mapping population for co-dominant markers in F_2 population ranges between that of a completely classified F_2 and a backcross, depending on distance between markers because, with an F_2 individual, two meiotic products are observed simultaneously, and some ambiguity occurs in that Ab/aB (two recombinant gametes) cannot be distinguished from AB/ab (two non-recombinant gametes) without progeny testing. The efficiency approaches that of a completely classified F_2 population as the linkage distance between markers decreases. Efficiency of mapping population for dominant markers: mapping efficiency is less in a F_2 population; efficiency increases as the linkage distance decreases; markers in repulsion phase are not informative; backcrosses, doubled haploid (DH) and recombinant inbreds (RIs) are more informative. The information content of these population types with dominant markers is unaffected by linkage phase. Dominant markers can be used for linkage estimation, if it is closely linked in coupling with the trait. Molecular markers need to be validated if intended to use it in different population in which a set of lines are tested for the marker-trait association. If one to one association between the marker and the trait, it can be utilized in MAS breeding. A mapping population should be at maximum linkage disequilibrium with respect to the gene of interest and markers in vicinity. F_2 individuals completely classified with respect to linkage phase provide, on an average, twice as much information as backcross individuals. A backcross population is more informative when greater genetic distances are involved. A DH population is genetically equivalent to a backcross population derived from backcrosses to a completely recessive parent: one meiotic event is analyzed per individual. DH and RI mapping populations possess an additional advantage in that once constructed, they represent a practically inexhaustible ‘immortal’ population.

Linkage

When two genes lie in vicinity of each other on a chromosome they tend to inherit together. It can be defined as the tendency of certain loci or alleles to be inherited together. The closer the two genes the more tight will be linkage between them and the more often they will be inherited together. A marker can be linked with an allele of interest either in coupling or repulsion phase as depicted in Fig.6. Markers that co-segregate (are always present or absent together) must be linked, *i.e.*, they must be located in each other’s vicinity on the genome. In some cases however, due to recombination events, the linkage between the markers may be lost. The frequency

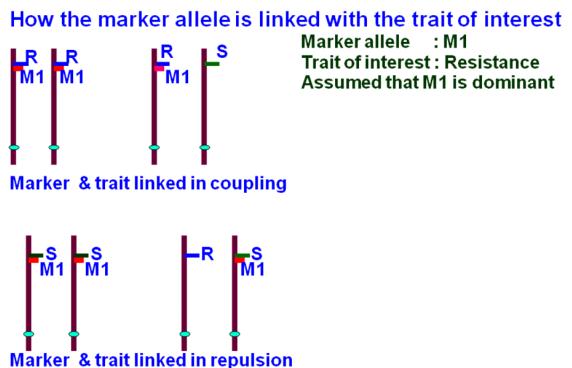


Fig. 6: Depiction of coupling and repulsion linkage

with which the linkage between co-segregating markers is broken is an indication of the genetic distance between the markers. An extensive analysis of the linkage between a large number of molecular markers yields information on their arrangement on the genome. Such analysis can finally result in the construction of a genetic map, on which all markers are arranged in separate linkage groups or chromosomes. On such a map, the distances between markers reflect the degree of observed linkage.

Linkage disequilibrium (LD)

Two alleles at different loci that occur together on the same chromosome (or gamete) more often than would be predicted by random chance is known as linkage disequilibrium. It is a measure of co-segregation of alleles in a population. LD is the non-independence, at a population level, of the alleles carried at different positions in the genome. Consider genotypes with two genes and two alleles per locus. When extreme genotypes are mated ($AABB \times aabb$), only two types of gametes are produced (AB and ab) and equilibrium for all genotypes can not be reached in the next generation since many genotypes are missing (e.g., $AAbb$, $aaBB$ etc.). At equilibrium, the gene frequency of Ab in repulsion gamete (Ab and aB) will be equal to the gene frequency in coupling gamete (AB and ab). The product of gene frequencies of gametes at repulsion should be equal to the product of gene frequency of gametes at coupling [$(Ab) \times (aB) = (AB) \times (ab)$]. The difference between the coupling and repulsion product is known as linkage disequilibrium ' d ' [$(Ab) \times (aB) - (AB) \times (ab) = d$]. If the two segregating loci in repulsion are linked on the same chromosome, attainment of equilibrium will be delayed further based on the closeness of linkage distance. However, it should be noted that ' d ' depends only on the gametic frequencies and not on linkage distance. Thus, once equilibrium is attained there is no way of distinguishing linked and unlinked genes except for departures from independent assortment. Therefore, linkage disequilibrium is the basis for detection of linkage between a gene and a marker.

Establishment of linkage between marker and the trait

Mapping populations are required for establishing linkage between molecular marker and the trait of interest. It is prepared by studying segregation of markers in the mapping population and their association with the trait of interest.

Estimation of linkage distance

The recombination fraction between two loci is the proportion of meiotic products which are non-parental (recombinant) at the loci. Recombination fractions can be determined by examining the DNA of a large number of meiotic products at or very near the loci, to see if parental origins differ at them. When two loci lie on the same chromosome, parental origins differ if the loci are separated by an odd number of crossovers. In a typical experimental cross, or pedigree analysis,

knowledge concerning the genotype of the diploid cells undergoing meiosis is used together with the genotypes of many meiotic products, to count or estimate the proportion of recombinants. Genetic linkage has to be determined essentially either from backcross or F_2 segregation data. Under specific condition some other population may serve the purpose, but the F_2 /BC populations will give the best estimate. Linkage can not be detected in F_1 . It may be noted that mere association of traits does not qualify them to be linked traits. Linkage has to be established through genetic studies by applying statistical tests. Two aspects are embedded in linkage estimation. First step is detection of linkage and second step is estimation of linkage in terms of centi morgan (cM). The more the linkage the less will be cM distance. Once linkage is detected then its estimate is calculated. Generally, Chi square test is applied for detection of linkage and maximum likelihood method is applied for estimating linkage.

Let us assume gene ‘A’ governs a trait and gene ‘B’ governs another trait. We want to know whether ‘A’ and ‘B’ are linked *i.e.*, they are located on the same chromosome.

Example 1 with Back cross (BC) data : *i.e.*, AaBb/ aabb

Four classes of phenotype (association of trait in question) from the BC will be : AaBb, Aabb, aaBb, aabb and expected number of each phenotypic class may be represented as $m_1=m_2=m_3=m_4=1/4$. The expected frequency of each class (phenotype) will be 1:1:1:1

Let the observed number of each phenotypic class be a_1, a_2, a_3 and a_4 , where $a_1+a_2+a_3+a_4=n$ (Total number of individuals phenotyped).

Now, in order to detect linkage, a null hypothesis is formulated which is as follows: “that two traits (genes ‘A’ and ‘B’) are segregating independently” (*i.e.*, they are not linked).

If the null hypothesis is correct (genes ‘A’ and ‘B’ are segregating independently) then the observed number of each phenotypic class will not vary from the expected number of phenotypic classes and it will prove that the traits assorting independently. But significant variation of observed number from the expected number will indicate that the traits (genes) in question are linked.

The best statistical test to accept or reject the null hypothesis is Chi square test.

For a two point data (study of two traits), joint deviation of all observed frequencies from the expected is given as:

$$\chi^2 \text{ for joint segregation (linkage)} = \sum [(a_{ij} - m_{ij})^2] / m_{ij}$$

i.e.

$$[(a_{11} - m_{11})^2] / m_{11}$$

$$[(a_{12} - m_{12})^2] / m_{12}$$

$$[(a_3 - m_3 n)^2] / m_3 n$$

$$[(a_4 - m_4 n)^2] / m_4 n$$

sum of the above four values will give $\Sigma [(a - mn)^2] / mn$

The above χ^2 has 3 df. That is the above χ^2 value is due to 3 independent χ^2 . These are due to :

1df for deviation of Aa segregation from 1:1

1df for deviation of Bb segregation from 1:1

1df for joint segregation (linkage) of 'A' and 'B'

Formula for calculating χ^2 for the above 3 dfs are given below:

$$\chi^2_A = [(a_1 + a_2 - a_3 - a_4)^2]$$

$$\chi^2_B = [(a_1 - a_2 + a_3 - a_4)^2]$$

$$\chi^2_{\text{Linkage}} = [(a_1 - a_2 - a_3 + a_4)^2]$$

deviation due to χ^2_{Linkage} is significant then it indicates presence of linkage.

Linkage can also be detected from the F_2 data. The formulae will be different because the F_2 segregation ratio is different from the BC segregation ratio.

For two traits each governed by single dominant gene the BC segregation ratio is 1:1:1:1 whereas for F_2 the segregation ratio is 9:3:3:1.

Estimation of linkage distance may be calculated by 4 different methods. Maximum likelihood (ML) method is the best method because the efficiency of other three methods are dependent on specific situation.

LOD scores

LOD means the logarithm of odds. However, it is not the logarithm of the odds for linkage *per se*, but the logarithm of the likelihood ratio for a particular value of the recombination fraction versus free recombination ($q = 0.5$). It is calculated for validating the linkage statistically. LOD determines the likelihood of obtaining the observed results based on an assumption of linkage as compared to the likelihood of obtaining the same results by pure chance. In other words LOD score serve as a test of the null hypothesis of free recombination versus the alternative hypothesis of linkage. In terms of molecular marker, it is a statistical measure of the likelihood that two genetic markers occur together on the same chromosome and are inherited as a single unit of DNA. LOD score has nothing to do with linkage disequilibrium. LOD is expressed as a logarithm to the base 10 to accommodate the wide range of values.

Calculation of LOD score

The ordered probability of obtaining the observed association between two phenotypes based on the assumption of linkage is divided by the ordered probability of obtaining association between two phenotypes based on the assumption that there is no linkage. The LOD score (Z) is calculated as follows:

$$Z = \log_{10} P(\text{observed data assuming linkage}) / P(\text{observed data assuming no linkage}).$$

Confirmation of linkage

A LOD score of 3.0 (1000:1 odds) or more is considered to be strong evidence of linkage, while lower positive values are considered suggestive of linkage. Negative values suggest absence of linkage.

Linkage map

It is a map of the genes on a chromosome based on linkage analysis. In terms of molecular markers, it is the relative placement of markers on a chromosome based on genetic linkage study involving a segregating population. A genetic linkage map of *Medicago truncatula* is illustrated in Fig. 7.

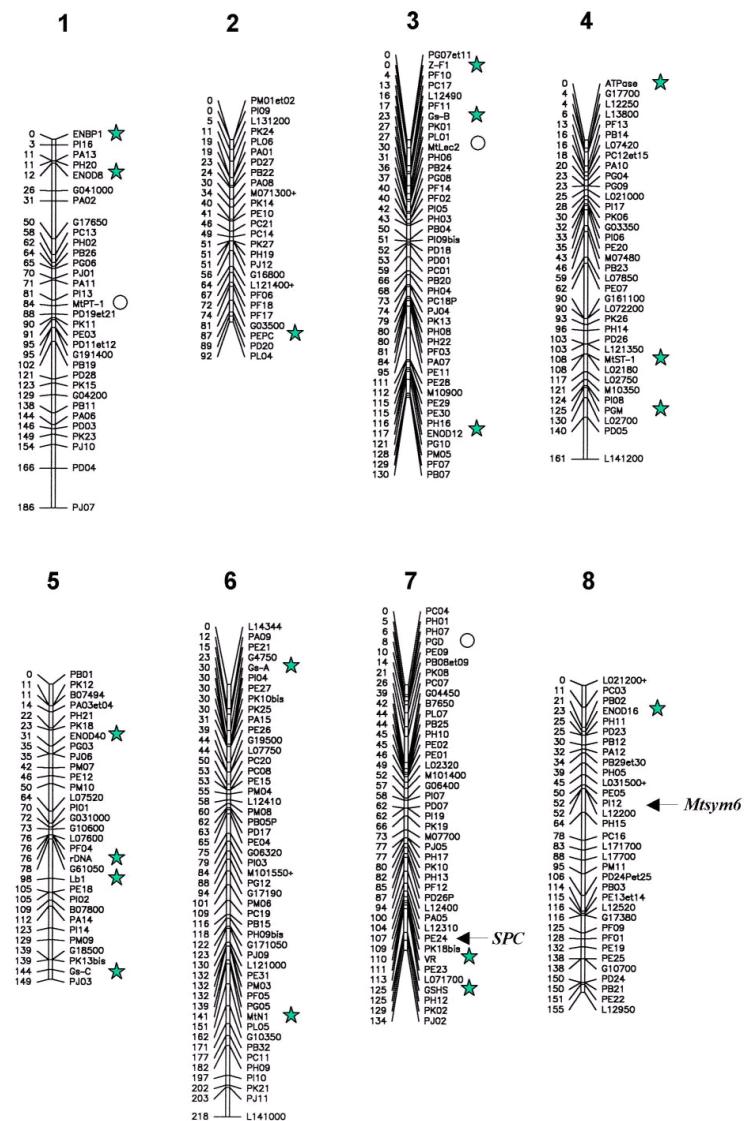


Fig. 7. The molecular genetic linkage map of the model legume *Medicago truncatula* (source: Thoquet et al. 2002. BMC Plant Biology. 2(1) doi: 10.1186/1471-2229-2-1)

Construction of linkage maps

Construction of genetic map can be very interesting as during map construction one can gather data, which is useful in systematic or evolutionary studies. Segregation analysis can be applied with segregating population that is derived from a common set of ancestors. Genetic linkage maps should not be confused with physical genomic maps, which can be obtained by determining the DNA sequence of chromosomes. Linkage maps and physical maps are related, but this relation is usually non-linear. The molecular maps are not important by itself in plant breeding. It is only useful when it is used in conjunction with analysis of conventional markers. Few examples of linkage maps in pulses are discussed hereunder.

Bharadwaj *et al.* (2010) developed a chickpea genetic linkage map by using sequence tagged microsatellite markers from a *desi* x *kabuli* F₂ population. Thirty three loci were distributed over a distance of 471.1cM with an average marker density of 14.2 cM. A microsatellite enriched library of chickpea was constructed from putative SSR clones (Gaur *et al.* 2011). Total 254 STMS primers were screened in a RIL population derived from ICCV 2 × JG 62 cross for generating new markers which improved the marker density and saturation of linkage in the vicinity of *sfl* (double podding) gene by integrating newly identified markers with that of previously developed chickpea intra-specific map. A genetic linkage map of the *Lathyrus sativus* was developed with 92 backcross individuals derived from a cross (ATC 80878 × ATC 80407) using 47 RAPD primers, 7 sequence-tagged microsatellite site and 13 STS/CAPS markers (Skiba *et al.* 2004). Two QTLs were associated with ascochyta blight resistance.

Gene mapping

One of the recent applications of new techniques of molecular biology is the rapid development in gene mapping. Use of DNA based markers is allowing researchers to determine the sequence of genes along chromosomes and the distances between them. These techniques are providing methods to mark, and in some cases, sequence genes that are related to genetic traits such as disease resistance or fruit colour. If genes can be identified, sequenced and cloned, gene transfer techniques can be used to transfer them to other species. With the available techniques it is possible to link markers on a genetic map with traits of interest in a particular species. Although much of the initial research effort has been applied to important cereal species, these techniques will have a significant impact on pulses breeding. If a genetic marker can be obtained for a trait of interest, plants could be selected at the seedling stage thereby saving many years at each step in a breeding programme. In constructing genetic maps, the amount of information generated depends on three factors: completeness of detection of recombinational events, linkage distance between loci and number of individuals assayed. The first two factors are influenced by selection of parents for population construction and markers used. More polymorphism between parents and the utilization of more informative markers increases the number of loci that can be mapped. Generally, the

selection of parents for genetic map construction is optimized for maximum polymorphism between the parents. However, for specific applications such as gene tagging, where a specific population is used, the level of polymorphism may not be as high as for the initial mapping population.

Genetics of mapping molecular loci

Each of the mapping populations will give a specific segregation ratio at each locus. The knowledge of these ratios is important to determine if the population is expressing a skewed segregation ratio at any locus. The ratios that you would expect at each locus for co-dominant and dominant makers segregating in the five different types of populations are given in Table 1. To score a dominant marker in a backcross population, you must cross the recessive parent with the F₁ plant. Therefore to score RAPD loci you would need to create two populations, each one developed by backcrossing to one of the two parents. For this reason, backcross populations have not been used for mapping RAPD loci. Once you have analyzed your segregating population by RFLP, RAPD or isozyme makers and have determined that the segregation ratio of each locus does not deviate from the expected ratio, you are ready to begin developing the map.

Table 1. Segregation ratio at markers in different populations

Population	Co-dominant loci	Dominant loci
F ₂ population	1:2:1	3:1
Backcross population	1:1	1:1
RILs	1:1	1:1
DH	1:1	1:1
NILs	1:1	1:1

Uses of DNA markers

a. DNA fingerprinting

- To establish distinctness among biological entities
- Genetic diversity studies
- Evolutionary studies

b. Molecular mapping

- To prepare saturated genetic map
- Chromosome identification

c. Map based cloning of genes

d. Marker Assisted Selection

- QTLs
- Disease resistance

e. Construction of genetic maps

Materials required for construction of a genetic map

Mapping population and molecular markers are required for constructing a genetic map. Linkage estimation is based on segregation of markers. Huge data are generated but software packages available to calculate linkage. Saturation of genetic map depend on completeness of detection of recombination events; linkage distance between loci; number of individuals assayed. Polymorphic parents and more informative markers increase the number of loci that can be mapped

Quantitative trait loci (QTL)

QTLs are short segments of DNA (locus) that have some contribution towards the phenotypic value of quantitative traits. Such locus may carry single or group of genes that are tightly linked and mostly inherited together. Many such loci determine the total phenotypic value of the trait (*e.g.*, yield). Each of these loci are called QTLs. Major QTLs are those loci that have major impact on the phenotypic value, whereas minor QTLs have minor impact on the phenotypic value. In QTL-mapping, association between observed trait values and presence/absence of alleles of markers, that have been mapped onto a linkage map is analysed. When it is significantly clear that the correlation that is observed did not result from some random process, it is proclaimed that a QTL is detected. In addition, the size of the allelic effect of the detected QTL can be estimated.

Basic steps in QTL analysis

- Make cross and generate mapping population
- Identify markers that are polymorphic between the parents
- Generate marker data
- Generate linkage maps of molecular markers
- Collect phenotypic measurements of QTL trait
- Map QTLs (Association of QTL with marker).

QTLs analysis

Scoring individuals of a random segregating population for a QTL trait is done by growing the segregating population in replicated multi-location trials. Determination of the molecular genotype/DNA marker data of each member of the segregating population is done by analyzing the DNA with specific markers. Specific programmes are prepared (with certain theoretical assumptions) for determining association between any of the markers and the quantitative trait. Therefore, experimental procedure must comply with these assumptions.

Methods of determining association of QTL and trait

- Single marker test
- Interval mapping
- Composite interval mapping
- Multiple interval mapping

Statistical procedures in marker assisted selection

Statistical applications in biotechnology have increased tremendously in recent years. With the help of statistical software a researcher may identify statistically significant differences in a single variable; between two or more groups and design efficient single factor experiments, which are fit for purpose and economical with scarce experimental resource. Statistical procedures are also required to fit simple calibration curves to biological data, produce informative data summaries for experiments with measurements on multiple variables, compare proportions between two or more groups, identify sources of variability in an experiment and design efficient experiments with appropriate choice of the number of replicates and level of replication, analyze data from experiments with multiple factors and to compare the response profiles on multiple variables between groups.

Correlation

Correlation usually refers to the degree to which a linear predictive relationship exists between random variables, as measured by a correlation coefficient. Correlation is a statistical procedure used to determine the degree to which two (or more) variables vary together. Correlation does not suggest a cause-effect relationship but only the degree of parallelism or concomitance between the variables, the cause of which may be unknown. The *Pearson product-moment correlation (r)* is the most frequently used, and this coefficient is used unless another is specified. Correlation may be positive or negative.

Cluster analysis

Given a set of varieties each of which belongs to one set of class, the statistical/computational techniques used to place the objects in the class to which they belong is called cluster analysis. If the classes are known it is mere classification. If the classes are not known then it is a problem. for example, given a set of DNA finger print profiles, along with the information that they belong to different species of rice except that the origin of some of the finger prints is not known, a cluster analysis can be used to classify the finger print according to species and possibly discover the species identity of the unlabelled finger print. In order to find the clustering relationships among a

set of objects it is crucial that we must have the dissimilarities or similarities among them. So the measurement of similarity of finger printing is of very important in population genetics. When DNA profiles of two individual plants are compared, a certain number of bands will be common (shared) between the two profiles, even by chance. The number or proportion of shared bands is expected to be larger if the two individuals are biologically related. It is, therefore, important to objectively measure the expected degree of similarity due to chance or relatedness.

Scoring of markers and development of molecular data set

Scoring of bands is done on a scale of 1 and 0 for presence or absence of bands, respectively.

Example of a Binary band matrix

	Band 1	Band 2	Band 3	Band 4
Plant A	1	0	0	1
Plant B	0	1	0	1
Plant C	1	1	1	0
Plant D	1	1	0	1
Plant E	0	1	0	1
Plant F	1	0	0	1
Plant G	1	0	1	0

Softwares used for interpreting molecular data

NTSYSpc (Numerical taxonomy and multivariate analysis system) can be used to discover pattern and structure in multivariate data. For example, one may wish to discover that a sample of data points suggests that the samples may have come from two or more distinct populations or to estimate a phylogenetic tree using the neighbour-joining or UPGMA methods for constructing dendograms. Of equal interest is the discovery that the variation in some subsets of variables are highly inter-correlated (clustered). NTSYSpc can then be used to compute various measures of similarity or dissimilarity between all pairs of objects and then summarize this information either in terms of nested sets of similar objects (*cluster analysis*) or in terms of a spatial arrangement along one or more coordinate axes (*ordination analysis* or various types of *multidimensional scaling analysis*). The software available for data analyses are summarized in Table 2.

Table 2. Softwares and their application in molecular breeding

MAPMANAGER	Application	Interval mapping, multiple QTL modeling
	Population	F ₂ backcross, RIL, DH
	Language	Unix
QTLSTAT	Application	Interval mapping using non linear regression
	Population	F ₂ backcross, RIL, DH
	Language	Unix
PGRI	Application	t-test, conditional t-test, linear regression
	Population	F ₂ backcross, RIL, DH, F ₁ , OP
	Language	Unix
QTL Cartographer	Application	t-test, Composite Interval mapping, permutation test, bootstrap, jackknife
	Population	F ₂ backcross, RIL, DH
	Language	Unix/Mac/PC Windows
MAPQTL	Application	Interval mapping, MQM
	Population	F ₂ backcross, RIL, DH, F ₁
	Language	Vax/Unix/Mac/ PC Windows
Map Manager QT	Application	Interval mapping using regression, MQM
	Population	F ₂ backcross
	Language	MAC OS
QGENE	Application	linear regression
	Population	F ₂ backcross
	Language	MAC

Use of Molecular Markers in Breeding Programmes

Marker aided selection is a tool for breeding, wherein genetic marker(s) tightly linked with the desired trait/gene(s) are utilized for indirect selection for that trait in segregating/non-segregating generations. In its simplest form it can be applied to replace evaluation of a trait that is difficult or expensive to evaluate. When a marker is found that co-segregates with a major gene for an important trait, it may be easier and cheaper to screen for the presence of the marker allele linked to the gene, than to evaluate the trait. From time to time the linkage between the marker and the gene should then be verified. When more complex, polygenic controlled traits are concerned, the breeder is faced with the problem how to combine as many as possible beneficiary alleles for the QTLs that were detected. In this case, the breeding material can be screened for markers that are linked to QTLs. Based on such an analysis specific crosses can be devised for creation of an optimal genotype by combining QTL alleles from different sources. Marker assisted selection, when applied within the current breeding material to enhance a breeding programme, does not solve the problem of limited genetic variability that is often seen in breeding stocks. A different application of marker assisted selection could contribute to a genetic enrichment of breeding material. Marker assisted selection may be used to facilitate a controlled inflow of new genetic material. The wild species often carries desired components that may be missing in cultivated material. Such components can be transferred to elite cultivated material by repeated backcrossing. However, breeders are often reluctant to apply this method because of unpredictable linkage drag. These are caused by other genes, which are unintentionally transferred along with the genes that control the target trait. It may take considerable effort and screening to get rid of the unwanted genes and return the material to an acceptable agronomic value. Markers can be used to pinpoint the genetic factors that are responsible for the desired characteristics in the unadapted material. In a backcross programme, the presence of the desired QTL alleles can be verified continuously by observing linked markers.

Marker assisted selection (MAS)

MAS is most useful for traits that are difficult to select e.g., disease resistance, salt tolerance, drought tolerance, heat tolerance, quality traits (aroma of *basmati* rice, flavour of vegetables). The approach involves selecting plants at early generation with a fixed, favourable genetic background at specific loci, conducting a single large scale marker assisted selection while maintaining as much as possible the allelic segregation in the population and the screening of large populations to achieve the objectives of the scheme. No selection is applied outside the target genomic regions, to maintain as much as possible the Mendelian allelic segregation among the selected genotypes. After selection with DNA markers, the genetic diversity at un-selected loci may allow breeders to generate new varieties and hybrids through conventional breeding in response to targets set in breeding programme.

Material required for MAS

Molecular markers, a set of authentic lines carrying trait of interest and a population to validate the markers to be used e.g., F_2 or BCF_2 for each of the individual traits/genes. Following are the basic pre-requisites for MAS :

- Search of molecular markers that are linked to the trait of interest
- Validate the available markers in parents and breeding population
- If markers are not available, it has to be designed and validated before use (if mapping populations are not available in hand it may take 2-4 years to generate and validate markers)
- Design a selection scheme and breeding strategy
- Fix the minimum population to be assayed to capture all beneficial alleles
- Meticulous record keeping
- Progeny testing for fixation of traits.

Steps involved in MAS

1. Validation of molecular markers. Extract the DNA from test individuals and find out whether there is one to one relationship with marker and the trait.
2. Extract the DNA of breeding population at the seedling stage and apply MAS. Select the individuals on the basis of presence of desired molecular markers for the concerned trait. For other traits, selection is based on classical breeding methods. Minimum individuals to be assayed should be as per the defined strategy and statistical considerations.

Limitations of MAS

- Cost factor
- Requirement of technical skill
- Automated techniques for maximum benefit
- *Per se*, DNA markers are not affected by environment but traits may be affected by the environment and show $G \times E$ interactions. Therefore, while developing markers, phenotyping should be carried out in multiple environments and implications of $G \times E$ should be understood and markers should be used judiciously.
- DNA marker has to be validated for each of the breeding population. Any apriori assumption regarding the validity of markers may be disastrous.

Marker assisted backcross breeding

A backcross breeding programme is aimed at gene introgression from a “donor” line into the genomic background of a “recipient” line. The potential utilization of molecular markers in such

programmes has received considerable attention in the recent past. Markers can be used to assess the presence of the introgressed gene (“foreground selection”) when direct phenotypic evaluation is not possible, or too expensive, or only possible late in the development. Markers can also be used to accelerate the return to the recipient parent genotype at other loci (“background selection”). It is assumed that the introgressed gene can be detected without ambiguity, and the theoretical study was restricted to background selection only. The use of molecular markers for background selection in backcross programmes has been tested experimentally and proved to be very efficient. Introgressing the favourable allele of QTL by recurrent backcrossing can be a powerful mean to improve the economic value of a line, provided the expression of the gene is not reduced in the recipient genomic background. Yet, recent results show that for many traits of economic importance QTLs have rather small effects. In this case, the economic improvement resulting from the introgression of the favourable allele at a single QTL may not be competitive when compared with the improvement resulting from conventional breeding methods over the same duration. Marker assisted introgression of superior QTL alleles can then compete with classical phenotypic selection only if several QTLs could be manipulated.

Selection scheme for MAS breeding

The approach involves selecting plants at early generation with a fixed, favourable genetic background at specific loci, conducting a single large scale marker assisted selection while maintaining as much as possible the allelic segregation in the rest of the genome. First, the identification of elite lines presenting high allelic complementarity and being outstanding for traits of interest is required to capture favourable alleles from different parental lines. Second, after identification of the most favourable genomic regions for each selected parental line, those lines are intercrossed to develop segregating populations from which plants homozygous for favourable alleles at target loci are selected. One objective of the scheme is to conduct the marker assisted selection only once, and it requires the selection of a minimum number of plants to maintain sufficient allelic variability at the unselected loci. Therefore, the selection pressure exerted on the segregating population is quite high and the screening of large populations is required to achieve the objectives of the scheme. No selection is applied outside the target genomic regions, to maintain as much as possible the Mendelian allelic segregation among the selected genotypes.

Application of DNA markers in crop improvement

QTL mapping

Some of the most difficult tasks of plant breeders relate to the improvement of traits that show a continuous range of values. Genetic factors that are responsible for a part of the observed phenotypic variation for a quantitative trait are called quantitative trait loci (QTLs). The term QTL was coined by Gelderman. Conceptually it can be a single gene or may be a cluster of linked genes of the trait. Although similar to a gene, a QTL merely indicates a region on the genome comprised

of one or more functional genes. Among such quantitative traits like yield, plant length and days to flowering etc., are important ones. Selection for quantitative traits is difficult, because the relation between observed trait values in the field (the phenotype) and the underlying genetic constitution (the genotype) is not straight forward. Quantitative traits are typically controlled by many genes, each contributes only a small part to the observed variation. The environmental variations resulting from differences in growing conditions, further create the problem to understand the relation between phenotype and genotype. In practice, this problem is typically dealt with by evaluating large and replicated trials, which allow identification of genotypic differences through statistical analysis. Plant breeders would like to utilize the quantitative traits for genetic factors that are responsible for the observed variability in quantitative traits. In a process called QTL mapping, association between observed trait values and presence/absence of alleles of markers, that have been mapped onto a linkage map is analysed. When it is significantly clear that the correlation that is observed did not result from some random process, it is proclaimed that a QTL is detected. Also the size of the allelic effect of the detected QTL can be estimated. Identification of molecular makers associated with QTLs involves three basic steps namely, scoring individuals of a random segregating population for a QTL trait; determination of the molecular genotype of each member of the population and determination of association between any of the markers and the quantitative trait. The first step is to make cross and generate marker data. In the next step generate linkage maps of molecular markers. Subsequently collect phenotypic measurements of QTL trait across the environments in replicated trials. Finally, mapping of QTL is done. The most common method of determining the association between marker and QTL is done by analyzing phenotypic observation of trait and scoring of molecular data by one-way analysis of variance and regression analysis. For each marker, presence of a specific fragment of DNA is considered a marker class, and all individuals (in a segregating population) possessing that marker class are considered to be positive for that class. If the variance due to a particular class is significant, then the molecular marker used to define, that class is considered to be associated with a QTL. Regression values are calculated for all the markers which have shown association with the quantitative trait which reflect the amount of total genetic variation that is explained by the specific molecular marker. There are very few examples of QTL mapping in pulse crops.

In chickpea, two RIL populations were used to construct a composite linkage map with the help of RAPD, ISSR, RGA, SSR and ASAP markers (Radhika *et al.* 2007). Marker trait association was observed among three yield related traits: double podding, seeds per pod and seed weight. The double podding gene was tagged by the markers NCPGR 33 and UBC 249z. Seeds per pod was tagged with TA 2x and UBC 465 markers. Eight QTLs were identified for seed weight.

Duran *et al.* (2002) developed a QTL map for plant height, pod dehiscence, number of shoots and seed diameter from inter-subspecific population of *Lens culinaris* ssp *culinaris* x *Lens culinaris* ssp *orientalis*. Tullu *et al.*, (2008) identified QTLs in lentil for earliness and plant height with RAPD, SSR and AFLP markers. Kahraman *et al.* (2004) identified five independent QTLs

for winter hardiness in a population of RILs derived from a cross between lentil accessions WA 8649090 x Precoz. One IISR marker Ubc 808-12 was found to be useful in MAS for predicting winter survival in segregating populations. Rubeena *et al.* (2003) identified eight QTLs for ascochyta blight resistance gene in lentil through composite interval mapping. Five QTLs were identified in F_2 population of ILL 5588/ILL 7537 whereas three QTLs were detected in F_2 of the cross ILL 7537/ILL 6002.

Young *et al.* (1993) identified three QTL associated with powdery mildew resistance in mungbean, while Chaitieng *et al.* (2002) and Humphry *et al.* (2003) found one QTL responsible for powdery mildew resistance in *Vigna* species. Sholihin and Hautea (2002) identified six AFLP derived putative QTLs associated with two traits (leaf relative water content and leaf stress rating) used to measure draught tolerance.

Tagging of disease resistance genes

DNA based markers have shown great promise in expediting plant breeding methods. The identification of molecular markers closely linked with resistance genes can facilitate expeditious pyramiding of major genes into elite background, making it more cost effective. Once the resistance genes are tagged with molecular marker the selection of resistant plant in the segregating generations becomes easy. A chickpea linkage map was established with help of 354 molecular markers (118 STMSs, 96 DAFs, 70 AFLPs, 37 ISSRs, 17 RAPDs, eight isozymes, three cDNAs, two SCARs and three markers linked to fusarium wilt resistance) surveyed among 130 recombinant inbred lines derived from a *C. arietinum* \times *C. reticulatum* (Winter *et al.* 2000). The fusarium wilt resistant genes for race 4 and 5 were placed on the linkage group that also contained STMS and a SCAR marker previously shown to be linked to fusarium wilt race 1. This is an indication of clustering of several fusarium wilt resistance genes. These markers will pave the way for MAS and searching other useful genes.

DNA markers associated with two closely linked genes for resistance to fusarium wilt race 4 and 5 in chickpea were identified from a population of 131 recombinant inbred lines derived from a wide cross between *Cicer arietinum* and *Cicer reticulatum* (Benko-Iseppon *et al.* 2003). With the aid of bulk segregant analysis nineteen new markers were identified in the vicinity (4.1-9 cM) of fusarium wilt resistance genes on linkage group 2, R-2609-1 showed closest linkage (2cM) with race 4 resistance locus. Gowda *et al.* (2009) identified flanking markers for chickpea fusarium wilt resistance genes in a recombinant inbred line population. H3A12 and TA101 SSR flanked the *Foc* 1 resistance gene whereas *Foc* 2 was mapped between TA96 and H3A12. The H1B06y and TA194 markers flanked the *Foc* 3 locus.

Reddy *et al.*, (2009) performed bulk segregant analysis on a segregating population of ICPL 7035 x ICPL 8863 for identification of RAPD markers associated with pigeonpea sterility mosaic disease (PPSMD) resistance. The primer OAP18 revealed polymorphism between the

parents and the resistant and the susceptible bulks. The OAP18 marker was converted into SCAR marker for identification of PPSMD resistant plants in the segregating population.

Dhanasekhar *et al.* (2010) identified two RAPD markers OPF04₇₀₀ and OPA09₁₃₇₅ were linked with the open and tall plant type gene in pigeonpea F₂ population of the cross between TT44-4 and TDI2004-1 through bulk segregant analyses. These markers were validated in 15 genotypes with open-tall plant type. Kotresh *et al.* (2006) used bulk segregant analysis with 39 RAPD primers which led to identification of two markers (OPM03₇₀₄ and OPAC11₅₀₀) that were associated with Fusarium wilt susceptibility allele in a pigeonpea F₂ population derived from GS1 x ICPL87119.

Taran *et al.* (2003) identified two molecular markers associated with Ascochyta blight resistance in lentil *viz.*, UBC 2271290 linked with ral1 gene and RB18680 linked with AbR1 and a marker (OPO61250) linked with Anthracnose resistance gene were utilized for identifying lines that possessed pyramided genes in a population of 156 recombinant inbred lines (RILs) developed from a cross between 'CDC Robin' and a breeding line '964a-46'. These markers can be converted into more robust SCAR markers for routine use in marker assisted selection. Tullu *et al.* (2003) tagged anthracnose resistance gene LCt-2 of lentil cultivar PI 320937 with RAPD and AFLP markers.

Basak *et al.* (2004) developed molecular marker linked to yellow mosaic virus (YMV) resistance gene in *Vigna* sp. from a population segregating for YMV disease resistance. Maiti *et al.* (2010) identified molecular markers CYR1 and YR4 in a F₂ population for screening of MYMV resistance genes. CYR1 co-segregated with MYMV resistance gene in F₂ plants and F₃ progenies. These two markers can be used simultaneously with the help of a multiplex PCR reaction.

Katoch *et al.* (2009) identified a powdery mildew resistance gene er2 in pea that was associated with a RAPD marker OPX-17_1400, exhibiting cis phase linkage (2.6 cM) in a F₂ population derived from Lincoln/JI2480. The reproducibility of RAPD marker was enhanced by converting it to a sequence characterized amplified region (SCAR) marker. Ek *et al.* (2005) used bulk segregant analysis on a F₂ population derived from the cross 955180 x Majoret for screening of SSR markers linked with powdery mildew resistance gene in pea. Out of 315 markers, only five showed linkage with the PM resistance gene. It was noted that none of single marker was tightly linked with the gene that can be considered optimal for inclusion in a MAS program. Therefore, a combination of two markers can be utilized for selecting PM resistant plants which would result in only 1.6% false positives.

Nguyen *et al.* (2001) converted a RAPD marker into a SCAR(SCARW19) for selecting ascochyta blight resistance gene of lentil accession ILL5588. Rubeena *et al.* (2003) identified QTLs for ascochyta blight resistance in lentil. Further validation is required to use these markers for MAS.

Hamwieg *et al.* (2005) mapped microsatellite markers identified from a genomic library of lentil. The linkage spanning about 751cM, consisting of 283 marker loci was derived from 86

recombinant inbred lines derived from the cross ILL 5588 × L 692-16-1(s) using 41 microsatellite and 45 amplified fragment length polymorphism markers. The average marker distance was 2.6 cM. Two flanking markers (SSR marker SSR59-2B at 8.0 cM and AFLP marker p17m30710 at 3.5 cM) were linked with fusarium resistance.

Saxena (2010a) assessed the DNA polymorphism in a set of 32 pigeonpea lines screened with 30 SSR markers. Based on polymorphism of marker alleles, higher genetic dissimilarity coefficient and phenotypic diversity for *Fusarium* wilt and sterility mosaic disease resistance data, five parental combinations were identified for developing genetically diverse mapping populations suitable for the development tightly linked markers for *Fusarium* wilt and sterility mosaic disease resistance.

Tagging of male sterility genes

A cytoplasmic male sterile system is desirable for use in hybrid seed production, as it eliminates the need for hand emasculation. CMS is a maternally inheritable trait characterized by the inability to produce viable pollen but without affecting the female fertility, and it is often associated with mitochondrial DNA rearrangements, mutations and editing. Several restorer locus have been identified using RAPD and STS in different crop and DNA markers linked to these locus enable the molecular study of the CMS system. These co-dominant markers are useful in identifying the homozygous restorer genotypes after the backcrossing for production of restorer lines. In this way, the restorer lines could be produced in a shorter period than by conventional methods. Soufmanien *et al.* (2003) identified RAPD marker linked with male sterility gene. Primer OPC-11 produced a unique amplicon of 600bp in male sterile (A) lines 288A (derived from *C. scarabaeoides*) and 67A (derived from *C. sericeus*), which was absent in their respective, maintainers and putative R lines (TRR 5 and TRR 6). Genetic distance based on similiaraty index revealed considerable genetic variation between male sterile lines, two putative R lines and donors of male sterility genes.

Diversity evaluation

Stability and identity of crop variety has assumed great importance for predicting plant breeder's right/farmers right. Traditionally, evaluation and conservation of bio-diversity/genetic variability is based on comparative anatomy, morphology, embryology, physiology, etc., which provide informative data but of low genetic resolution. Recent advances in molecular biology have provided powerful genetic tools, which can provide rapid and detailed genetic resolution. Molecular marker based genotyping involves the development of marker profile unique to an individual. This unambiguous pattern of crop varieties obtained using DNA a marker is termed as "DNA Fingerprinting". The technique was developed by Alec Jeffery in 1985 in human and was used first time in crop (rice) in 1988 by Dallas for cultivar identification. The choice of molecular marker to be used for DNA fingerprinting usually depends on technical expertise, available funds as well as the requirements of the experiment. At the same time the most important considerations in the use

of molecular techniques are discrimination power and reproducibility. RAPD markers have shown to be of low discrimination power as compared to SSR and AFLP. Now days, microsatellites are the method of choice for varietal identification due to their abundance, high polymorphism, and simple protocol etc. Odeny *et al.* (2007) deduced DNA polymorphism in pigeonpea by using 113 primers (designed from genomic SSR) and 220 soybean primers.

Sivaramakrishnan *et al.* (2002) showed that RFLP of mtDNA can be used for the diversity analysis of pigeonpea. They assayed restriction enzyme digested fragments of 28 accessions representing 12 species of the genus, *Cajanus* arranged in 6 sections including 5 accessions of the cultivated species *C. cajan* and 4 species of the genus *Rhyncosia* with maize mtDNA probes. In addition to inter-specific variability, intra-specific diversity was observed between the accessions of wild species (*C. scarabaeoides*, *C. platycarpus*, *C. acutifolius*) and cultivated species of *C. cajan*.

Saxena *et al.* (2010b) designed 23 primer pairs from 36 SSR enriched genomic library of pigeonpea. Sixteen primer pairs produced expected amplification fragments, of which 13 were polymorphic amongst 32 cultivars and 8 wild accessions representing six species. The average polymorphic information content was 0.32 per marker which varied from 0.05 to 0.55 for these 13 primer pairs.

Ratnaparkhe *et al.* (1995) developed DNA fingerprints for cultivated and wild pigeonpea accessions with the help of RAPD markers. The polymorphism among the cultivated species was low whereas high level of polymorphism was observed among the wild species. All pigeonpea accessions including cultivars under study were distinguishable from each other which demonstrated utility of RAPD in the genetic fingerprinting of pigeonpea. Ganapathy *et al.* (2011) generated 561 amplified fragment length polymorphism (AFLP) loci for clustering cultivated and wild pigeonpea accessions. Jaccard's similarity index indicated greater diversity within wild species which clustered into several groups. Most of the cultivated accessions were grouped into one major cluster. Among the cultivated lines, BRG 3, ICP 7035, TTB 7 and ICP 8863 were selected on the basis of morphological and molecular diversity for generating mapping population for identification of markers linked to sterility mosaic disease.

Hamwieh *et al.* (2009) developed new set of microsatellite markers in lentil for delineating the molecular diversity. Souframanien and Gopalkrishna (2004) used RAPD and IISR markers for deducing the genetic diversity among 18 blackgram cultivars.

Heterosis breeding

Another important application of DNA markers is their prediction of heterosis in hybrids. Evaluation of hybrids for heterosis or combining ability in field is expensive. Molecular markers have been used to correlate genetic diversity and heterosis in several cereal crops (like rice, oat, and wheat). It has been reported that measures of similarity based on RFLP and pedigree knowledge

could be used to predict superior hybrid combinations. However, both low and high correlations between heterosis and DNA based genetic distance have been observed.

The correlation between heterosis and molecular (isozyme and RFLP) diversity between the parents was investigated in three maturity groups in soybean (Cern *et al.* 1997). Parental RFLP diversity was not significantly correlated with mid parent and better parent heterosis indicating that heterosis in yield may not be associated with genetic diversity at the molecular level as determined by RFLPs. Though isozyme diversity in parents was associated with yield heterosis, it is of limited importance because of the reduced number of assayable isozyme loci in soybean.

Hybrid seed purity testing

To determine the hybrid seed quality it is to be verified that the designated cross has occurred, the number of self pollination between the female parents meet the necessary purity and the product has adequate quality. For years, the only method to check the hybrid seed purity has been the grow out test. Now the RAPD and RFLP are markers are used to test the purity of F₁ hybrids. Kumar *et al.* (2011) used male specific markers SSR 218, SSR 306 and Ty2 gene CAPs gene marker testing F₁ purity of Pbc × EC 538408, Pbc × EC 520061 and H 86 × EC 520061 tomato hybrids.

Gene pyramiding

It is essentially a way of determining and introducing multiple genes which impart resistance to an independent insect/microbial pest, or impart resistance to a single pest through independent host pathways. One strategy for increasing the durability of resistance is to incorporate multiple resistance genes into a single variety. If the pyramided genes have never been deployed as single genes, then it is estimated that the durability of resistance could be extended for 50 years. However, during the process of constructing a cultivar, it is difficult to verify the number of resistance genes that have been successfully pyramided. Plants with only one resistance gene are just as resistant as plants that contain three resistance genes, although the resistance of the latter is probably more durable.

Li *et al.* (2010) pyramided phytopthora tolerance QTLs in soybean. A recombination inbred line population was generated by crossing two phytopthora tolerant soybean cultivars Conrad and Hefeng 25. With the help of 161 SSR markers, 7 environmentally stable QTLs (QPRR-1 and QRR-2 Conrad and QRR3 to QRR7 from Hefeng 25) were identified. The more the number of QTLs the higher was the level of tolerance.

Pyramiding of Bt genes

The insecticidal *cry* (crystal) genes from *Bacillus thuringiensis* (*Bt*) have been used for insect control both as biopesticides and in transgenic plants. Discovery of new insecticidal genes is of importance for delaying the development of resistance in target insects. The diversity of *Bt*

strains facilitates isolation of new types of *cry* and *vip* (vegetative insecticidal protein) genes. PCR is a useful technique for quick and simultaneous screening of *Bt* strains for classification and prediction of insecticidal activities. PCR together with other methods of analysis such as RFLP, gene sequence determination, electrophoretic, immunological and chromatographic analysis of Cry proteins and insect bioassays for evaluation of toxicity have been employed for identification of new insecticidal proteins. Some other new approaches have also been devised. Many *Bt* strains with novel insecticidal genes have been found. A desired combination of Cry proteins can be assembled via site-specific recombination vectors into a recipient *Bt* strain to create a genetically improved biopesticide. For better pest control, the *cry* genes have been transferred to plants. Stacking of more than one insecticidal gene is required for resistance management in transgenic crops. Modification of Cry proteins through protein engineering for increasing the toxicity and/or the insecticidal spectrum is also a promising approach, but requires detailed understanding of the structure and function of these proteins and analysis of toxin-receptor interactions. Stringent implementation of resistance management is needed for maintaining the efficacy of *Bt* transgenic crops and deriving maximum economic and environmental benefit. *cry1Ac* and *cry1C* *Bt* genes were incorporated in the Broccoli line (Cao *et al.* 2002). *cry1Ac* and *cry1C* mRNAs were detected in the hybrid lines which also produced stable Cry1Ac and Cry1C proteins at levels comparable to the parental plants carrying individual genes. Plants with pyramided *cry* genes did not show leaf damage and killed the DBM larvae resistant to Cry1A or Cry1C.

Map based cloning of genes

The first step of map-based or positional cloning is to identify a molecular marker that lies close to the gene of interest. Initially gene search is done with a small mapping population and subsequently, a highly saturated genetic map is prerequisite to clone a gene. With reference to an initial marker, the region around that original molecular marker is saturated with high density markers. A large number of individuals are screened to find a marker which rarely recombines with the gene of interest. The next step is to screen a large insert genomic library (BAC or YAC) with the previously identified closely linked molecular markers to isolate clones that hybridize with the marker. Once two flanking markers that show linkage with the target gene are identified then the target gene is searched with the help of chromosomal walking. The goal is to find clones harbouring a set of flanking markers that co-segregate with the gene of interest. Such putative clones are introduced into the individuals devoid of the target gene. If a transgenic is shown to rescue the mutant phenotype, then detailed molecular and biochemical analysis of the newly cloned gene is undertaken to characterize the gene.

Map based cloning technique was employed to clone *Pto* (resistance to bacterial speck disease of tomato) gene of tomato (Martin *et al.* 1993). Initially a genetic population of 251 F₂ plants was screened with DNA probes. The locus TG 538 cosegregated with *Pto* gene. A YAC library was screened with the TG 538 probe and the clone PTY538-1 was identified. Susceptible plants transformed with PTY 538 recovered resistant phenotype.

Principles of Basic Techniques

Genomic DNA extraction

The method of DNA extraction from plant tissue generally varies depending on the material used for isolation of DNA. Any mechanical means of breaking down the cell wall and membranes to allow access to nuclear material, without its degradation is essential. For this purpose, generally an initial grinding with liquid nitrogen is employed to break down cell wall material and allow access to DNA while harmful cellular enzymes and chemicals remain inactivated. Once the tissue has been sufficiently ground to a fine powder, it can then be resuspended in a suitable buffer, such as CTAB. Later, insoluble particulates (cell debris) are removed through centrifugation, while soluble proteins and other material are separated through repeated mixing with chloroform and isoamyl alcohol and centrifugation. DNA is then precipitated from the aqueous phase and washed thoroughly to remove contaminating salts. The purified DNA can be resuspended and stored in TE buffer or in sterile distilled water. The CTAB method has been shown to give intact genomic DNA from plant tissues. To check the quality of the extracted DNA, a sample is run on an agarose gel, stained with ethidium bromide, and visualised under UV light.

Quantification and quality check of DNA

UV quantification of DNA

Add 15 µl of each sample to 735 µl TE, mix well, and read OD 260 and OD 280 to determine purity. After UV quantification, adjust the concentration of each DNA sample to 0.3 µg/µl or a concentration of your choice with TE, and store at 4°C.

The ratio OD 260/OD 280 should be determined to assess the purity of the sample. If this ratio is in between 1.8-2.0, the absorption is probably due to nucleic acids. A ratio of less than 1.8 indicates there may be proteins and/or other UV absorbers in the sample, in that case it is advisable to re-precipitate the DNA. A ratio higher than 2.0 indicates the samples may be contaminated with chloroform or phenol and should be re-precipitated with ethanol.

Approximate DNA quantification through gel electrophoresis

Make a 0.8% agarose gel with 1x TAE and 0.1µl of Ethidium bromide (10 mg/ml) per 10 ml solution. Load one undiluted sample and dilutions, with 3µl loading buffer. Also include a λ DNA ladder cut with *Hind*III and *Eco*RI. This contains 100 ng/µl of DNA and use as follows:

1µl ladder + 4µl water + 2µl loading buffer

2µl ladder + 3µl water + 2µl loading buffer

The different bands of the ladder are of known molecular weight and known DNA concentration. Match the brightness of samples with those of the two dilutions of the ladder. Remember that although the ladder concentrations are absolute, you have loaded 5 μ l of sample and also diluted some of them. This must be taken into account when calculating the strength of the samples in $\text{cg}/\mu\text{l}$.

Assessment of DNA quality

- Prepare 0.8% solution of agarose by melting 1.0 g of agarose in 100 ml of 0.5x TBE buffer for approximately 2 min. Allow to cool for a couple of minutes then add 2.5 μ l of ethidium bromide, stir to mix.
- Cast a gel using the tray and comb. Allow the gel to set for a minimum of 20 min at room temperature on a flat surface.
- Load 10 μ l lkb ladder and 5 μ l sample + 5 μ l water + 2 μ l 6x loading buffer into separate wells
- Run the gel for 30 min at 100 V
- Expose the gel to UV light and photograph
- Confirm DNA quality: presence of a highly resolved high molecular weight band indicates good quality DNA, presence of a smear in the lane indicates DNA degradation.

DNA digestibility test

This step is essential before setting up large scale digestion experiments. A small amount of DNA is digested with restriction endonucleases.

- Put 2 μ g of each DNA sample in a 0.5 ml micro-centrifuge tube.
- Prepare a bulk digestion mix based on the recipe given below and keep it on ice.
- Add 8 μ l of this to each of the tubes containing the DNA. Mix thoroughly but gently and spin down the tube contents.
- Incubate at 37°C for 1.5 to 3 h. Stop the reactions with 3 μ l of 5X SGB.
- Load samples in a 0.8% agarose gel and run the gel at 40 mAmp for 2-3 h.
- Use Lambda DNA digested with *HindIII* as a molecular weight marker.

Polymerase chain reaction (PCR)

Preparation of reaction mixture

To perform several parallel reactions, it is recommended to prepare a master mix containing water, buffer, dNTPs, primers and *Taq* DNA polymerase in a single tube, thereafter, different aliquots can be made for individual tubes. MgCl₂ and template DNA solutions are then added. This

method of setting reactions minimizes the possibility of pipetting errors and saves time by reducing the number of reagent transfers.

Reaction mixture set up

1. Gently vortex and briefly centrifuge all solutions after thawing.
2. Add, in a thin-walled PCR tube kept on ice:

Reagent	Final concentration	Quantity, for 25µl of reaction mixture
Sterile deionized water	-	Variable
10X PCR buffer	1X	2.5µl
25mM MgCl ₂	1-4mM	Variable
2mM dNTP mix	0.2mM of each	2.5µl
Primer I	0.1-1µM	Variable
Primer II	0.1-1µM	Variable
<i>Taq</i> DNA polymerase	0.5-1U/25µl	Variable
Template DNA	10pg-0.5µg	Variable

3. Gently vortex the sample and briefly centrifuge to collect all drops from walls of tube.
4. Overlay the sample with half volume of mineral oil or add an appropriate amount of wax. This step may be omitted if the thermal cycler is equipped with a heated lid.
5. Place samples in a thermocycler and start PCR.

Composition of the reaction mixture

a. Template DNA

Usually the template DNA amount is in the range of 0.01-1.0 ng for plasmid or phage DNA and 0.1-1.0 µg for genomic DNA, for a total reaction mixture of 50 µl. High template DNA concentration generally increases the yield of non-specific PCR products, but if the fidelity of synthesis is crucial, maximal allowable template DNA quantities in conjunction with limiting number of PCR cycles should be used to increase the percentage of “correct” PCR products. Nearly, all routine methods are suitable for template DNA purification. Even trace amounts of agents used in DNA purification procedures (phenol, EDTA, Proteinase K, etc.) strongly inhibit *Taq* DNA Polymerase activity. Ethanol precipitation of DNA and repetitive treatments of DNA pellets with 70% ethanol is usually effective in removing traces of contaminants from the DNA samples.

b. Primers

PCR primers are usually 10-30 nucleotides in length. Longer primers provide sufficient specificity. The GC content should be more than 50%. The C and G nucleotides should be distributed uniformly within the full length of the primer. More than three G or C nucleotides at the 3'-end of the primer should be avoided, as non-specific priming may occur. The primer should not be self-complementary or complementary to any other primer in the reaction mixture as this avoids primer-dimer and hairpin formation. The melting temperature of flanking primers should not differ by more than 5°C, so the GC content and length must be chosen accordingly. All possible sites of complementarity between primers and the template DNA should be noted. If primers are degenerate, at least 3 conservative nucleotides must be located at the primer's 3'-end.

Estimation of the melting and annealing temperatures of primer: If the primer is shorter than 25 nucleotides, the approx. melting temperature (T_m) is calculated using the following formula:

$$T_m = 4(G + C) + 2(A + T)$$

Annealing temperature should be approximately 5°C lower than the melting temperature. If the primer is longer than 25 nucleotides, the melting temperature should be calculated using specialized computer programmes, where the interactions of adjacent bases, the influence of salt concentration, etc., are evaluated.

Primer quantity

- a. Primer in pico-moles = $\frac{\text{weight in } \mu\text{g} \times 1000,000}{\text{Length of primer} \times 327}$
- b. Micromolar concentration of primer = pico-moles/ μl

Example 1: Calculate the primer quantity if, 20 pmoles of primer is mixed in 100 μl PCR mixture

$$= 20 \times 100 \times 10^{-4} \mu\text{M}$$

$$= 0.20 \mu\text{M}$$

Example 2: How much 20 nt primer (con. 25 $\eta\text{g}/\mu\text{l}$) is required for a 25 μl reaction mixture requiring 0.5 μM primer?

At first convert 25 $\eta\text{g}/\mu\text{l}$ in to pico-moles and then in to μM concentration

$$\begin{aligned} \text{pmoles in } 1 \mu\text{l} &= \frac{25 \times 10^{-3} \times 10^6}{20 \times 327} \\ &= 3.82 \text{ pmoles}/\mu\text{l} \end{aligned}$$

pmoles per μl is equivalent to μM concentration

Therefore, the concentration of primer is $3.82\mu M$

Now, calculate the required quantity by $N_1V_1 = N_2V_2$

$$3.82 \times X = 0.5 \times 25$$

$$X = 3.27\mu l$$

Therefore, the required quantity of primer is $3.27\mu l$

Example 3: What is the micromolar concentration of a primer of 26 nt length dissolved in 0.2ml of water?

The spectrophotometer reading of a $20\mu l$ aliquot diluted to 1ml is $A_{260} = 0.82$. The primer base composition is A=6, T=6, C=7, G=7.

The stock solution is diluted 100 times ($10\mu l$ made to 1ml)

Therefore, the stock solution has an absorbance at 260 nm (A_{260}) of 82

The Molar Extinction Coefficient at 260 nm for the primer = $k (15,200) + I (12,010) + m (7,050) + n (8,400)$

Where, k = number of A's; m = number of C's; I = number of G's; n = number of T's

The Molar Extinction of the Primer = $6 \times 15,200 + 7 \times 12,010 + 7 \times 7,050 + 6 \times 8,400 = 275020$

The Molar Concentration = $82/275020 = 298 \times 10^{-6}$ molar = $298\mu M$

c. $MgCl_2$ concentration

Since Mg^{2+} ions form complexes with dNTPs, primers and DNA templates, the optimal concentration of $MgCl_2$ has to be selected for each experiment. Less quantity of Mg^{2+} ions results in a low yield of PCR product, and too many ions increase the yield of non-specific products and promote misincorporation. Lower Mg^{2+} concentrations are desirable when fidelity of DNA synthesis is critical. Under the standard reaction conditions specified for most of the experiments, the recommended range of $MgCl_2$ concentration is 1-4 mM. In general, at a final dNTP concentration of 0.2mM, a $MgCl_2$ concentration ranges of 1.5 ± 0.25 mM (in traditional PCR buffer) and of 2.0 ± 0.5 mM (in PCR buffer with $(NH_4)_2SO_4$) are suitable in most cases. If the DNA samples contain EDTA or other chelators, the $MgCl_2$ concentration in the reaction mixture should be raised proportionally.

Example 4: How much 25 mM $MgCl_2$ is required to make 2 mM of 25 μl reaction mixture?

Apply $N_1V_1 = N_2V_2$

Units in RHS and LHS must be same

$$25 \times X = 2 \times 25$$

$$X = \frac{2 \times 25}{25}$$

$$X = 2$$

Therefore, 2 μ l MgCl₂ is required.

d. dNTPs

The concentration of each dNTP in the reaction mixture is usually 200 μ M. It is very important to have equal concentrations of each dNTP (dATP, dCTP, dGTP, dTTP), as inaccuracy in the concentration of even single dNTP dramatically increases the misincorporation. When maximum fidelity of the PCR process is crucial, the final dNTP concentration should be 10-50 μ M. In addition, the concentration of MgCl₂ should be selected empirically, starting from 1mM and increasing in 0.1mM steps, until a sufficient yield of PCR product is obtained.

Example 5: How to make 1.0 ml of 1 mM dNTP from a stock of 100 mM dNTP?

$$100 \times X = 1 \times 1000$$

$$X = 10$$

Therefore, 10 μ l of each dNTP (dATP, dTTP, dCTP, dGTP) added to 960 μ l of H₂O.

Example 6: How much 1 mM dNTP is required to make 200 μ M of 25 μ l reaction mixture?

$$1000 \times X = 200 \times 25$$

$$X = 5$$

Therefore, 5 μ l dNTP is required.

e. Taq DNA Polymerase

Usually 1-1.5 U of *Taq* DNA polymerase is used in 50 μ l of reaction mix. Higher *Taq* DNA polymerase concentration may cause synthesis of non-specific products. However, if inhibitors are present in the reaction mix (e.g., if the template DNA used is not highly purified), higher amount of *Taq* DNA polymerase (2-3 U) is helpful in obtaining a better yield of amplification products.

Example 7: How much *Taq* DNA polymerase (5U/ μ l) is required in a reaction requiring 0.6 unit of *Taq*?

5 U ————— 1 μ l

For 0.6U ————— $\frac{1 \times 0.6}{5}$

Therefore, 0.12 μ l *Taq* DNA polymerase is required.

f. Temperature cycling

i. Initial denaturation step

The complete denaturation of the DNA template at the start of the PCR reaction is of key importance. Incomplete denaturation of DNA results in the inefficient utilization of template in the first amplification cycle and in a poor yield of PCR product. If the GC content is 50% or less, an initial denaturation should be performed over an interval of 1-3 min at 95°C. This interval could be extended up to 10 min for GC rich templates. If the initial denaturation is no longer than 3 min at 95°C, *Taq* DNA polymerase can be added into the initial reaction mixture. If longer initial denaturation or a higher temperature is necessary, *Taq* DNA Polymerase should be added only after the initial denaturation, as the stability of the enzyme generally, decreases at temperatures over 95°C.

ii. Denaturation step

Usually 30 sec to 2 min denaturation at 94/95°C is sufficient, since the PCR product synthesized in the first amplification cycle is significantly shorter than the template DNA and is completely denatured under these conditions. If the amplified DNA has a very high GC content, denaturation time may be increased up to 3-4 min. Alternatively, additives facilitating DNA denaturation [e.g., glycerol (10-15 vol.%), DMSO (up to 10%), or formamide (up to 5%)] can be used. In the presence of such additives, the annealing temperature should be adjusted, since the melting temperature of the primer-template DNA duplex decreases significantly. If additives are used, the amount of *Taq* DNA polymerase in the reaction mix should be increased, because DMSO and formamide, at the suggested concentrations, reduces (inhibit) the enzyme activity by 50%. Alternatively, a common way to decrease the melting temperature of the PCR product is to substitute dGTP with 7-di-aza-dGTP in the reaction mix.

iii. Primer annealing step

Usually, the optimal annealing temperature is 5°C lower than the melting temperature of primer-template DNA duplex. Incubation for 30 sec to 2 min is sufficient. However, if non-specific PCR products are obtained in addition to the expected product, the annealing temperature should be optimized by increasing it stepwise by 1-2°C.

iv. Extension step

The extension step is generally performed at 70-75°C. The rate of DNA synthesis by *Taq*

DNA polymerase is highest at this temperature (2-4 kb/min), and a 1 min extension time is sufficient for the synthesis of PCR fragments of ~2 kb. When larger DNA fragments are to be amplified, the extension time is increased by 1 min for each 1000 bp.

v. Final Extension step

After the last cycle, the samples are usually incubated at 72°C for 5-15 min to fill-in the protruding ends of newly synthesized PCR products. Also, during this step, the terminal transferase activity of *Taq* DNA polymerase adds extra A nucleotides to the 3'-ends of PCR products. Therefore, if PCR fragments are to be cloned into TA cloning vectors, this step can be extended for up to 30 min.

vi. Number of cycles

The number of PCR cycles depends on the amount of template DNA in the reaction mix and on the expected yield of the PCR product. For less than 10 copies of template DNA, 40 cycles should be performed. If the initial quantity of template DNA is higher, 25-35 cycles are usually sufficient.

How to avoid contamination during PCR?

- DNA sample preparation, reaction mixture assemblage, and the PCR process, in addition to the subsequent reaction product analysis, should be performed in separate areas.
- A Laminar air-flow cabinet equipped with a UV lamp is recommended for preparing the reaction mixture.
- Fresh gloves should be worn for each PCR setup.
- The use of dedicated vessels and positive displacement pipettes or tips with aerosol filters for both DNA sample and reaction mixture preparation, is strongly recommended.
- The reagents for PCR should be prepared separately and used solely for this purpose. Autoclaving of all solutions, except dNTPs, primers and *Taq* DNA polymerase is recommended. Solutions should be aliquoted in small portions and stored in designated PCR areas. Aliquots should be stored separately from other DNA samples.
- A control reaction, omitting template DNA, should always be performed, to confirm the absence of contamination.

Optimization of RAPD reaction

RAPD reactions require optimization for getting better bands. For optimization, the reactions are first carried out at an optimal PCR conditions. By varying the different variables one by one the amplification conditions are standardized. Reactions without DNA were used as negative controls. The standardization the reactions are generally carried out in a 25 µl reaction volume. Generally, a

standard reaction require about 50 ng of template DNA, 1x PCR Buffer (10 mM Tris HCl pH 8.3; 50 mM KCl), 3 mM MgCl₂, 0.2 mM dNTP Mix, 0.5 μM of single primer, 0.4 U of *Taq* DNA polymerase. The thermocycler may be programmed for an initial denaturation step of 3 min at 94°C, followed by 30 cycles of 45 s at 94°C, 1 min at 37°C, extension was carried out at 72°C for 1 min and final extension at 72°C for 7 min and a hold temperature of 4°C at the end. PCR products to be electrophoresed on 2% (w/v) agarose gels, in 1x TBE Buffer at 50 V for 3 h and then stained with ethidiumbromide (0.5 μg/ml). Gels with amplification fragments can be visualized and photographed under UV light. Lambda DNA *EcoR1-Hind III* double digest is used as molecular marker.

PCR Parameter and Optimum conditions

PCR Parameter	Remarks
DNA concentration (ng)	At low concentration there will be no amplification and at higher concentration there will be smearing of bands
Magnesium chloride (mM)	Excess/lower concentration increases the non specificity and yield of the product
Deoxynucleotide triphosphates (dNTPs) (mM)	Increased concentration reduces the free Mg ²⁺ , interfering with the enzyme.
Primer concentration (M)	Lower and higher concentrations lead to absence of amplification and primer dimer formation, respectively.
<i>Taq</i> polymerase (units)	Lower concentration do not show proper amplification. High concentration show decreased specificity
Initial denaturation time interval (min) at 94°C	Higher/lower time intervals (from optimum) leads to reduction in amplification, loss of <i>Taq</i> polymerase activity and lack of reproducibility
Annealing temperature (°C)/ Time intervals (s)	Higher/lower annealing temperatures (from optimum) results in difference in specificity
Reaction volume (l)	Influences the cost of the PCR ingredients
Number of cycles	Higher/lower cycles (from optimum) effects the amplification

Agarose gel electrophoresis

Electrophoresis is method used in biochemistry and molecular biology experiments to separate DNA or RNA molecules by size. This is achieved by moving negatively charged nucleic acid molecules through a matrix (generally made of agarose) in an electric field (electrophoresis). Shorter molecules move faster and migrate farther than longer ones. Agarose gel electrophoresis can be used for the separation of DNA fragments ranging from 50 base pair to several megabases (millions

of bases) using specialized apparatus. Increasing the agarose concentration of a gel increases the gel strength and reduces the pores resulting in reduced migration of large fragments; this enables separation of smaller DNA molecules. The distance between DNA bands of a given length is determined by the percent agarose in the gel. In general lower concentrations of agarose are better for larger molecules because they result in greater separation between bands that are close in size. The disadvantage of higher concentrations is the long run times (sometimes days). Instead high percentage agarose gels should be run with a pulsed field electrophoresis (PFE), or field inversion electrophoresis.

Materials required :

- Molecular Weight Marker (DNA Ladder)
- Agarose
- Tank buffer
- Sample DNA
- Sample loading dye
- Electrophoresis apparatus

Reagents required

50X TAE Buffer

Make a 50x stock and dilute to 1x before use

- 242 g Tris base
- 57.1 ml glacial acetic acid
- 100 ml 0.5M EDTA (pH 8.0)

Dissolve the ingredients in a small quantity of distilled water and bring up to 1000 ml with distilled water

TBE (Tris Borate EDTA)

Make a 10x stock.

For making 2000 ml of 10xTBE, add the following :

- 218.0 g Tris base
- 110.0 g Boric acid
- 9.3 g EDTA

Dissolve the ingredients in a small quantity of distilled water. Adjust the pH to 8.3 using NaOH and make up to 2000 ml.

Agarose gel (1%)

Weigh out 1.0 g of Analytical Grade (Ultra Pure DNA Grade) Agarose. Add to 100 ml of autoclaved 1x TAE. Heat the solution in order to dissolve the agarose. Let the solution cool to ~60°C. Add 1-2 µl of 10mg/ml ethidium bromide.

Gel loading dye (10 ml)

- Bromophenol Blue : 0.025 g
- Xylene cyanol : 0.025 g
- Glycerol : 3.0 ml

Make up the volume up to 10 ml with sterilized distilled water.

Procedure of agarose gel electrophoresis

- Pour the melted agarose gel into the gel casting tray (kept on a flat surface) along with comb.
- Remove any bubbles in the gel
- Allow the gel for ~40 minutes to set ensuring that the gel is undisturbed
- When the gel has set remove the combs and transfer to the gel tank
- Pour the appropriate amount of 1x TAE buffer into the gel tank so that it covers the gel. (It is advisable to use the same batch of TAE in the tank which was used for making the agarose gel)
- Mix 5 volumes of PCR product with 1 volume of 6x gel loading buffer. Load samples into the wells in the gel
- Load 1 kb ladder into at least one well
- Run the gel at the appropriate voltage
- Photograph the gel under ultraviolet light

Purpose of electrophoresis

- To look at the DNA, to quantify it or to isolate a particular band.
- Estimation of the size of DNA molecules
- Analysis of PCR products
- Separation of restricted genomic DNA prior to Southern analysis, or of RNA prior to Northern analysis.

Factors affecting gel electrophoresis

- **DNA or RNA Molecular Weight:** The length of the DNA molecule is the most important factor, smaller molecules move farther.
- **Voltage:** The higher the voltage, the faster the DNA moves. But voltage is limited by the fact that it heats and ultimately causes the gel to melt. High voltages also decrease the resolution (above about 5 to 8 V/cm)

What are the most common agarose gels?

- 1% gels are common for many applications
- 0.7% gels are good for separation of large 5-10 kb DNA fragments
- 2.0% gels allow good resolution for small 0.2-1 kb fragments.
- Up to 3% can be used for separating very tiny fragments but a vertical polyacrylamide gel is more appropriate in this case.
- Low percentage gels are very weak and may break when we try to lift them. High percentage gels are often brittle and do not set evenly.

The most common buffers for agarose gel

TAE: Tris-acetate-EDTA

TBE: Tris-borate-EDTA

SB: Sodium borate.

TAE has the lowest buffering capacity but provides the best resolution for larger DNA. This means a lower voltage and more time, but a better product. The most common dye used to make DNA or RNA bands visible for agarose gel electrophoresis is ethidium bromide (EtBr). It intercalates into DNA/ RNA and fluoresces under UV light. By running DNA through an EtBr-treated gel and visualizing it with UV light, any band containing more than ~20 ng DNA becomes distinctly visible.

Visualization of the DNA fragments after electrophoresis

The patterns of DNA-bands are visualized under UV light. The DNA is visualised in the gel by addition of ethidium bromide. This binds strongly to DNA by intercalating between the bases and is fluorescent meaning that it absorbs invisible UV light and transmits the energy as visible orange light. A color marker dye containing a low molecular weight dye such as “bromophenol blue” (to enable tracking of the electrophoresis) and glycerol (to make the DNA solution denser so it will sink into the wells of the gel).

Gel tank

Small 8 x 10 cm gels (minigels) are very popular and give good photographs. Larger gels are used for applications such as Southern and Northern blotting. The volume of agarose required for a minigel is around 30–50ml, for a larger gel it may be 250ml.

Quantity of DNA loaded

A band is easily visible if it contains about 20 ng of DNA. Too much DNA loaded onto a gel is a bad thing. The band appears to run fast (implying that it is smaller than it really is) and in extreme cases can mess up the electrical field for the other bands, making them appear the wrong size also. Too little DNA is only a problem in that it is difficult to visualize the smallest bands because they are too faint.

Comb size

This depends on the volume of DNA is to be loaded and the number of samples. Combs with many tiny teeth may hold 10 µl. This is no good if one wants to load 20 µl of restriction digest plus 5 µl of loading buffer. When deciding whether a comb has enough teeth, remember the need to load at least one or two marker lane.

Addition of ethidium bromide

The reason for allowing the agarose to cool a little before adding EtBr is to minimise production of ethidium bromide vapour. Ethidium bromide is a mutagen and should be handled with extreme caution. Dispose of the contaminated tip into a dedicated ethidium bromide waste container. 10 mg/ml ethidium bromide solution is made up using tablets (to avoid weighing out powder) and is stored at 4°C in the dark.

Preparing the samples

Transfer an appropriate amount of each sample to a fresh micro-centrifuge tube. It may be 10 µl of a 50 µl PCR reaction or 5 µl of a 20 µl restriction enzyme digestion. If you are loading the entire 20 µl of a 20 µl PCR reaction or enzyme digestion then there is no need to use fresh tubes, just add in the loading buffer into the PCR tubes. Add an appropriate amount of loading buffer into each tube and leave the tip in the tube. Add 0.2 volumes of loading buffer (2 µl into a 10 µl sample).

Marker lane

The markers may be ready-mixed with loading buffer at 4°C.

Avoid using the end wells because they are likely to run aberrantly.

Running the gel

If the electrodes are 10 cm apart then run the gel at 50V. It is fine to run the gel slower than this but do not run any faster. Above 5V/cm the agarose may heat up and begin to melt with disastrous effects on gel's resolution. Some people run the gel slowly at first (2V/cm for 10 minutes) to allow the DNA to move into the gel slowly and evenly, and then speed up the gel later. This may give better resolution. The best way is to look at the electrodes and check that they are evolving gas (bubbles). If not then check the connections, that the power-source is plugged in etc. Stop the gel when the bromophenol blue has run 3/4 the length of the gel. Switch off and unplug the gel tank and carry the gel to the dark-room to look at on the UV light-box. Some gel holders are not UV transparent so you have to carefully place the gel onto the glass surface of the light-box. UV is carcinogenic and must not be allowed to shine on naked skin or eyes. So wear face protection, gloves and long sleeves.

Loading buffers

The loading buffer gives color and density to the sample to make it easy to load into the wells. Also, the dyes are negatively charged in neutral buffers and thus move in the same direction as the DNA during electrophoresis. This allows to monitor the progress of the gel. The most common dyes are bromophenol blue and xylene cyanol. Density is provided by glycerol or sucrose. The exact amount of dye is not important. Dye should be stored at 4°C to avoid growth of mould/fungus in the sucrose. 10 ml of loading buffer will last for years. Bromophenol blue migrates at a rate equivalent to 200-400 bp DNA. If any fragments smaller than 600 bp has to be visualized then other dye should be used because the bromophenol blue will obscure the visibility of the smaller fragments. Xylene cyanol migrates at approximately 4 kb equivalence; thus it should not be used if fragments of 4 kb have to be visualized.

Measurement of DNA fragment

There are lots of different kinds of DNA size markers. In the old days the cheapest defined DNA was from bacteriophage so a lot of markers are phage DNA cut with restriction enzymes. Many of these are still very popular eg, lambda HindIII, lambda PstI, PhiX174 HaeIII. These give bands with known sizes but the sizes are arbitrary. Choose a marker with good resolution for the fragment size you expect to see in your sample lanes. More recently, companies have started producing ladder markers with bands at defined intervals, eg. 0.5, 1, 1.5, 2, 2.5 kb and so on up to 10 kb. With the total amount of DNA loaded into a marker lane, and the sizes of all the bands, one can calculate the amount of DNA in each band visible on the gel. It is always good to load two markers lanes, flanking the samples.

Faint or no bands on the gel

- There was insufficient quantity/concentration of DNA loaded on the gel. Adjust the amount of DNA 50 ng/band.
- The DNA was probably degraded. Avoid nuclease contamination.
- The DNA was electrophoresed off the gel.

Smeared DNA bands

- The DNA was degraded. Avoid nuclease contamination.
- Too much DNA was loaded on the gel. Decrease the amount of DNA.
- Improper electrophoresis conditions were used. Do not allow voltage to exceed ~20 V/cm. Maintain a temperature <30° C during electrophoresis.
- There was too much salt in the DNA. Use ethanol precipitation to remove excess salts, prior to electrophoresis.
- The DNA was contaminated with protein. Use phenol extractions to remove protein prior to electrophoresis.
- Small DNA bands diffused during staining. Add the ethidium bromide during electrophoresis.

Anomalies in DNA band migration

- Improper electrophoresis conditions were used. Do not allow voltage to exceed ~20 V/cm. Maintain a temperature <30° C during electrophoresis. Check that the electrophoresis buffer used had sufficient buffer capacity and adequate quantity.
- The DNA was denatured properly.

Literature Cited

- Basak, J, Kundagrami S, Ghoose T K and Pal A. 2004. Development of yellow mosaic virus (YMV) resistance linked DNA marker in Vigna mungo from populations segregating for YMV-reaction. *Mol. Breed.* **14**: 375-382.
- Benko-Iseppon A, Winter M, Huettel P, Staginnus B, Muehlbauer C and Kahl F J G. 2003. Molecular markers closely linked to fusarium resistance genes in chickpea show significant alignments to pathogenesis-related genes located on Arabidopsis chromosomes 1 and 5. *Theor. Appl. Genet.* **107**:379-386
- Bharadwaj S K, Chauhan S K, Yadav S, Satyavati C T, Singh R, Kumar J, Srivastava R and Rajguru G 2010. *Indian J. Agric. Sci.* **81**: 116-118
- Cao J, Zhao J Z, Tang D, Shelton M, Earle D. 2002. Broccoli plants with pyramided cry1Ac and cry1C Bt genes control diamondback moths resistant to Cry1A and Cry1C proteins. *Theor. Appl. Genet.* **105**: 258-264
- Cern F J, Cianzio SR, Rafalski A, Tingey S and Dyer D. 1997. Relationship between seed yield heterosis and molecular marker heterozygosity in soybean *Theor. Appl. Genet.* **95**: 460-467
- Chaitieng B, Kaga A, Han O K, Wang Y W, Wongkaew S, Laosuwan P, Tomooka N and Vaughan D A. 2002. Mapping a new source of resistance to powdery mildew in mungbean. *Plant Breed.* **121**: 521-5.
- Dhanasekar P, Dhumal K N, and Reddy K S. 2010. Identification of RAPD markers linked plant type gene in pigeonpea. *Indian J. Biotech.* **9**: 58-63.
- Duran Y and Vega P M. 2004. Assessment of genetic variation and species relationships in a collection of *Lens* using RAPD and ISSR. *Spanish J. Agri. Res.* **2**, 538-544.
- Ek M, Eklund M, Von Post R, Dayteg C, Henriksson T, Weibull P, Ceplitis A, Isaac P and Tuvesson S. 2005. Microsatellite markers for powdery mildew resistance in pea (*Pisum sativum* L.) *Hereditas*. **142**:86-91
- Ganapathy K N, Gnanesh B, Gowda N, Byre M, Venkatesha S C, Gomashe S S and Mallikarjuna V C. 2011. AFLP analysis in pigeonpea (*Cajanus cajan* (L.) Millsp.) revealed close relationship of cultivated genotypes with some of its wild relatives. *Genet. Res. and Crop Evo.* **58**: 837-847
- Gaur R, Sethy N J K, Choudhary S, Shokeen B, Gupta V and Bhatia S. 2011. Advancing the STMS genomic resources for defining new locations on the intraspecific genetic linkage map of chickpea (*Cicer arietinum* L.) *BMC Genomics*. 2011. **12**: 117
- Gowda S J M, Radhika P, Kadoo N Y, Mhase L B, and Gupta V S. 2009. Molecular mapping of wilt resistance genes in chickpea. *Mol. Breed.* **24**, 177-183.
- Hamwieh A, Udupa S M, Sarker A, Jung C, and Baum M. 2009. Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils. *Breed. Sci.* **59**: 77-86.
- Hamwieh A, Udupa S M, Choumane W, Sarker A, Dreyer F, Jung C, and Baum M. 2005. A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. *Theor. Appl. Genet.* **110**: 669-677.

- Humphry M E, Magner C J, McIntyr E T, Aitken E A B C L and Liu C J. 2003. Identification of major locus conferring resistance to powdery mildew in mungbean by QTL analysis. *Genome*. **46**: 738-44.
- Katoch V, Sharma S, Pathania S, Banayal D K, Sharma S K and Rathour R. 2009. Molecular mapping of pea powdery mildew resistance gene er2 to pea linkage group III. *Mol Breed*. DOI 10.1007/s11032-009-9322-7
- Kahraman A, Kusmenoglu I, Aydin N, Aydogan A, Erskine W, and Muehlbauer F J. 2004. QTL mapping of winter hardness genes in lentil. *Crop Sci*. **44**: 13-22.
- Kotresh H, Fakrudin B, Punnuri S M, Rajkumar B K, Thudi M, Paramesh H, Lohithaswa H and Kuruvinashetti M S. 2006. Identification of two RAPD markers genetically linked to a recessive allele of a Fusarium wilt resistance gene in pigeon pea (*Cajanus cajan* L. Millsp.). *Euphytica* 149: 113-120.
- Kumar A, Datta D and Singh M. 2011. SSR polymorphism among ToLCV resistant and susceptible lines for genotype identification and F_1 purity Testing. *African J. Biotech*. In press
- Li X, Han Y, Teng W, Zhang S, Yu K, Poysa V, Anderson T, Ding J, and Li W. 2010. Pyramided QTL underlying tolerance to *Phytophthora* root rot in mega-environments from soybean cultivars Conrad and Hefeng 25. *Theor. Appl. Genet.* **121**, 651-658.
- Martin G B, Brommonschenkel S H, Chunwongse J, Frary A, Ganap M W, Spivey R, Wu T, Earle E D and Tanksley S D. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science*. **262**: 1432-1436
- Nguyen T T, Taylor P W J, Brouwe J B. R, Pang E C K and Ford R. 2001. A novel source of resistance in lentil (*Lens culinaris* ssp. *culinaris*) to ascochyta blight caused by Ascochyta lentis. *Australasian Plant Pathol*. **30**: 211–215.
- Odeny D A, Jayashree B, Ferguson M, Hoisington D, Cry L J and Gebhardt C. 2007. Development, characterization and utilization of microsatellite markers in pigeonpea. *Plant Breed*. 126, 130-136
- Radhika, P, Gowda S J M, Kadoo N Y, Mhase L B, Jamadagni B M, Sainani M N, Chandra S, and Gupta V S. 2007. Development of an integrated intraspecific map of chickpea (*Cicer arietinum* L.) using two recombinant inbred line populations. *Theor. Appl. Genet.* **115**: 209-216.
- Ratanparkhe, M B, Gupta V S, Murthy M R V and Ranjekar P K. 1995. Genetic fingerprinting of pigeonpea [*Cajanus cajan* (L.) Millsp.] and wild relatives using RAPD markers. *Theor. Appl. Genet.* **91**: 893-898.
- Reddy L P, Reddy B V Ba, Rekha Rani K, Sivaprasad Y, Rajeswari T and Reddy K R. 2009. RAPD and SCAR marker linked to the sterility mosaic disease resistance gene in pigeonpea (*Cajanus cajan* L. Millsp.). *The Asian and Australasian J. Plant Sci. and Biotech*. **3**:16-20)
- Rubeena P, Taylor W J and Ades P K. 2003. QTL mapping of resistance in lentil (*Lens culinaris* ssp. *culinaris*) to ascochyta blight (*Ascochyta lentis*). *Plant Breed*. **125**: 506–512
- Saxena, R. K., C. Prathima, K. B. Saxena, D. Hoisington, N. K. Singh, and R. K. Varshney, 2010b: Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). *Plant Breeding* **129**: 142-148.

- Saxena, R. K., K. B. Saxena, R. V. Kumar, D. A. Hoisington, and R. K. Varshney, 2010a: SSR-based diversity in elite pigeonpea genotypes for developing mapping populations to map resistance to *Fusarium* wilt and sterility mosaic disease. *Plant Breed.* **129:** 135–141.
- Sholihin and Hautea D M. 2002. Molecular mapping of drought resistance in mungbean (*Vigna radiata*): 2: QTL linked to drought resistance. *Journal of Biotechnology Pertanian* **7:** 55-61
- Sivaramakrishnan S K Seetha, and Reddy L J. 2002. Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. *Euphytica* **125:** 21-28.
- Skiba B, Ford R and Pang E C K. 2004. Construction of a linkage map based on a *Lathyrus sativus* backcross population and preliminary investigation of QTLs associated with resistance to Ascochyta blight. *Theor. Appl. Genet.* **109:** 1726-1735.
- Souframanien J, and Gopalkrishna T. 2004. A comparative analysis of genetic diversity in blackgram genotypes using RAPD and ISSR markers. *Theor. Appl. Genet.* **109:** 1687-1693
- Souframanien J, Manjaya J G, Krishna T G, and Pawar S E. 2003. Random amplified polymorphic DNA analyses of cytoplasmic male-sterile and male fertile pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Euphytica*. **129:** 293-299.
- Maiti S, Basak J, Kundagrami S, Kundu A, Pal A. 2010. Molecular Marker-Assisted Genotyping of Mungbean Yellow Mosaic India Virus Resistant Germplasms of Mungbean and Urdbean. *Mol Biotechnol.* **2:** 95-104
- Taran B, Buchwaldt L, Tullu A, Banniza S, Warkentin T D, and Vandenberg A. 2003. Using molecular markers to pyramid genes for resistance to ascochyta blight and anthracnose in lentil (*Lens culinaris* Medik). *Euphytica* **134:** 223-230
- Tullu A, Buchwaldt L, Warkentin T, Taran B, and Vandenberg A. 2003. Genetics of resistance to anthracnose and identification of AFLP and RAPD markers linked to the resistance gene in PI320937 germplasm of lentil (*Lens culinaris* Medikus). *Theor. Appl. Genet.* **106:** 428-434.
- Tullu A, Taran B, Warkentin T and Vandenberg A. 2008. Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. *Crop Sci.* **48:** 2254-2264
- Winter P, Benko-Iseppon A M, Huttel B, Ratnaparkhe M, Tullu A, Sonnante G, Pfaff T, Tekeoglu M, Santra D, Sant V J, Rajesh P N, Kahl G, and Muehlbauer F J. 2000. A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for fusarium wilt race 4 and 5. *Theor. Appl. Genet.* **101:** 115-1163.
- Young N D, Danesh D, Menuncio-Hautea D and Kumar L. 1993. Mapping oligogenic resistance to powdery mildew in mungbean with RFLPs. *Theor. Appl. Genet.* **87**(1-2): 243-9.

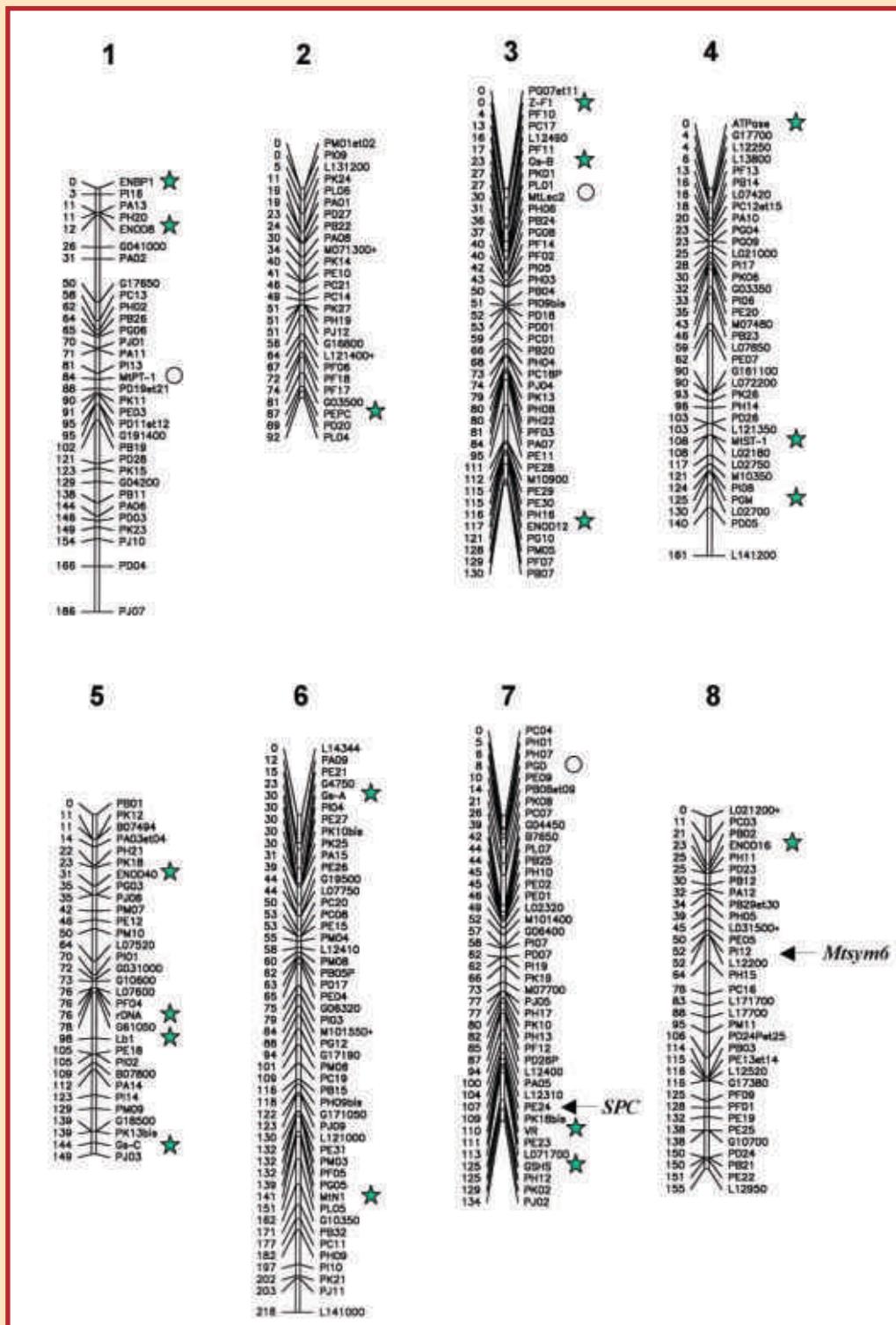
Suggested readings

- Domonique De Vienne. 2003. Molecular markers in Plant Genetics and Biotechnology. 248p. Science Publishers.
- Lorz Horst and Wenzel Gerhard. 2005. Molecular Marker Systems in Plant Breeding and Crop Improvement. Series: Biotechnology in Agriculture and Forestry, Vol. 55. XXII, 478 p. Springer.

- Phillips R L and Vasil I K. 1994. DNA based Markers in Plants. Springer.
- Robert J Henry. 2001. Plant Genotyping: The DNA Fingerprinting of Plants. 344 p. CABI Publishing.
- Yunbi Xu. 2010. Molecular Plant Breeding. 734p. CABI publishing.
- Burr B and Burr F A. 1991. Recombinant inbred lines for molecular mapping in maize. *Theor. Appl. Genet.* **85**: 55-60.
- Kaeppler S M, Philips R L and Kim T S. 1993. Use of near-isogenic lines derived by backcrossing or selfing to map quantitative traits. *Theor. Appl. Genet.* **87**: 233-237.
- Lefebvre V, Palloix A, Caranata C and Pochard E. 1995. Construction of an intraspecific linkage map of pepper using molecular markers and doubled-haploid progenies. *Genome* **38**: 112- 121.
- Michelmore R W, Paran I and Kesseli R V. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl Acad. Sci. USA* **88**: 9829- 9832.
- Simpson S P 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. *Theor. Appl. Genet.* **77**: 815-819.
- Tamulonis J P, Luzzi B M, Hussey R S, Parrot W A and Boerma H R 1997. DNA marker analysis of loci conferring resistance to root-knot nematode in soybean. *Theor. Appl. Genet.* **95**: 664-670.
- Vales M I, Schon CC, Capettini F, Chen X M, Corey A E, Mather D E, Mundt C C, Richardson K L, Sandoval-Islas J S and Utz H F. 2005. Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. *Theor. Appl. Genet.* **111**: 1260-1270.

Notes

Notes



Source: Thoquet *et. al.* 2002, BMC Plant Biology, 2