**MOLECULAR BREEDING**

Molecular breeding is the application of molecular biology tools, often in plant breeding and animal breeding.

The areas of molecular breeding include:

¬ QTL mapping or gene discovery

¬ Marker assisted selection and genomic selection

¬ Genetic engineering

¬ Genetic transformation

Aspects of Molecular Breeding includes the following field:-

Marker assisted breeding

¬ Genotyping and creating molecular maps - genomics

¬ The commonly used markers include Simple sequence repeats (or microsatellites), single nucleotide polymorphisms (SNP).

The process of identification of plant genotypes is known as genotyping.

¬ Development of SNPs has revolutionized the molecular breeding process as it helps to create dense markers. Another area that is developing is genotyping by sequencing.

Phenotyping – Phenomics

¬ To identify genes associated with traits, it is important to measure the trait value - known as phenotype. "omics" for measurement of phenotypes is called phenomics. The phenotype can be indicative of the measurement of the trait itself or an indirectly related or correlated trait. QTL mapping or association mapping

¬ Genes (Quantitative trait loci (abbreviated as QTL) or quantitative trait genes or minor genes or major genes) involved in controlling trait of interest is identified. The process is known as mapping. Mapping of such genes can be done using molecular markers. QTL mapping can involve single large family, unrelated individuals or multiple families (see: Family based QTL mapping). The basic idea is to identify genes or markers associated with genes that correlate to a phenotypic measurement and that can be used in marker assisted breeding / selection Marker assisted selection or genetic selection

¬ Once genes or markers are identified, they can be used for genotyping and selection decisions can be made.

**Marker-assisted backcrossing (MABC)**

¬ Backcross is crossing F1 with its parents to transfer a limited number of loci (e.g. transgene, disease resistance loci, etc.) from one genetic background to another. Usually the recipient of such genes is good adapted cultivars otherwise except the gene that is to be transferred. So we want to keep genetic background of the recipient genotypes, which is done by 4-6 rounds of repeated backcrosses while selecting for the gene of interest. We can use markers from the whole genome to recover the genome quickly in 2-3 rounds of backcrossing might be good enough in such situation. Marker-assisted recurrent selection (MARS)

¬ MARS include identification and selection of several genomic regions (up to 20 or even more) for complex traits within a single population.

**Genomic selection**

¬ Genomic selection is novel approach to traditional marker-assisted selection where selection are made based on few markers. Rather than seeking to identify individual loci significantly associated with a trait, genomics uses all marker data as predictors of performance and consequently delivers more accurate predictions. Selection can be based on genomic selection predictions, potentially leading to more rapid and lower cost gains from breeding. Genomic prediction combines marker data with phenotypic and pedigree data (when available) in an attempt to increase the accuracy of the prediction of breeding and genotypic values.

**Genetic transformation or Genetic engineering**

¬ Transfer of genes make it possible for horizontal transfer of genes from one organism to another. Thus plants can receive genes from humans or algae or any other organism. This provides limitless opportunity in breeding crop plants.

**Development of sequence based molecular markers - SSRs and SNPs**

**INTRODUCTION**

Attempts to characterize current types of DNA markers generally yield several different classifications. This is explained by the fact that the development of an integral and more or less comprehensive classification of DNA markers is hindered by the vast diversity of the available methods of analysis of DNA polymorphism and numerous combinations of these methods.

In studies of the plant genome, the following markers are commonly used: RFLP, CAPS, STS, RAPD, SCAR, AFLP, SSAP, SSR, ISSR, and SNP. These markers can be classified in different groups, according to the approach employed in the classification, i.e., methodical principle, the aim of use, or the type of polymorphism.

A basic classification have been proposed of DNA markers, according to which they are divided into three classes:

(1) Hybridization-based DNA markers (e.g., RFLP);

(2) PCR-based DNA markers (e.g., CAPS, STS, RAPD, SCAR, AFLP, SSAP, SSR, ISSR);

(3) DNA chip and sequencing-based DNA markers (SNP).

In this classification, the markers are grouped on the basis of the principal method used for analysis. This classification also reflects the “evolution” of DNA markers. The first class represents the first generation of DNA markers, which were widely used in the 1980s; then in the 1990s, PCR markers (the second group) gained key importance; and since 2000, these were succeeded by a new generation—SNP markers, based on microarray technologies (the third group). In recent years, owing to the development of the programs on sequencing cDNA libraries, more than 16 000 loci of expressed DNA sequences (ESTs) have been mapped with the use of deletion and aneuploid wheat lines. The application of ESTs as genetic markers is based on Southern blot hybridization, which is costly and time-consuming. Because of this, more convenient STS, SSR, and SNP markers have been developed on the basis of ESTs. Presently researchers is devoted to various aspects of developing and using new-generation markers, i.e., SNPs. Analysis of this marker type is based on a number of modern highly effective approaches, in particular, application of DNA microarrays. Compared to other types of DNA markers, the use of SNPs allows to make this analysis automated and to enhance the efficiency of genotype analysis by an order of magnitude. Origin of SNPs SNP is defined as a nucleotide site, for which a high substitution rate has been shown among individual samples in a population. According to the more accurate definition given by Brookes, “SNPs are single base pair positions in genomic DNA, at which different sequence alternatives (alleles) exist in normal individuals in some population(s), herein the least frequent allele has an abundance of 1% or greater.” The restriction posed on frequency distinguishes SNPs from rare point mutations and implies the use of the former as genetic markers. Based on this definition, SNP markers formally do not include single-nucleotide insertions/deletions (indels), though some authors combine these two polymorphism types with single-nucleotide substitutions under the common term of SNPs (simple nucleotide polymorphisms). Theoretically, four alleles can occur at each nucleotide position (because of the four existing nucleotide types), but in practice only two variants occur, as a rule.

In other words, SNP markers are biallelic, which is explained by unequal occurrence of nucleotide transitions (A G, T C) and transversions (A C, A T, G C, G T). Although the number of variants of nucleotide transversion is two fold higher than that of transition, the frequency of transitions is, conversely, higher by a factor of 1.5–2.5, than the frequency of transversions. This is explained by a high spontaneous rate of deamination of 5-methylcytosine to thymidine in CpG dinucleotides. SNP Classification Several types of SNPs are distinguished, according to their assignment to the structural element of genomic DNA or their functional effect. Depending on the SNP position in a gene, SNPs are classified into exon (synonymous and nonsynonymous), intron, and promoter ones. Oligonucleotide substitutions, involving regulatory regions that control gene expression, are referred to as regulatory SNPs and designated rSNPs. They may include promoter SNPs and some inton SNPs. Depending on the presence of data on the functional effect of the single-nucleotide substitutions, they are classified into anonymous SNPs (functional effect is unknown), candidate SNPs (presumably having a functional effect), and protein SNPs (single-nucleotide substitutions, resulting in a change in the protein function or expression). Strategies of Developing SNP Markers In principle, analysis of DNA polymorphism at single-nucleotide sites is not something entirely new. For instance, restriction fragment length polymorphism (RFLP) is also based on single-nucleotide substitutions. Wide interest to SNPs as a novel promising class of genetic markers has appeared in the course of projects on complete sequencing of the genome of many model organisms and accumulation of nucleotide sequences of cDNA libraries of a number of organisms (http:/www.ncbi.nlm.nih.gov/dbEST/). Overlapping DNA sequences from genomic and cDNA libraries (EST sequences) provided a possibility for identifying single-nucleotide substitutions in silico. This possibility substantially facilitates large-scale development of SNP markers.

The difficulties, related to sequences errors in data bases, have been shown to be surmountable. In general, the development of SNP markers involves the following approaches:

(1) The development of SNP markers on the basis of overlapping genomic DNA sequences. This method does not require additional DNA sequencing, thus being advantageous for large-scale generation of SNP markers. However, its use at the moment is possible only for a limited number of model organisms, whose genome is completely or nearly completely sequenced, as, for example, rice (http:/plantgenome.agtec.uga.edu/snp/).

(2) The development of SNP markers on the basis of overlapping EST sequences. This approach has been employed for large-scale generation of SNP markers (without additional DNA sequencing) in a number of organisms. In some cases, ovelapping cDNA libraries are purposefully developed before generating EST-derived SNPs, as was done. The interest to this approach is also explained by the fact that its use makes possible identifying at high probability cSNPs (coding SNPs) among other SNPs. In rare cases, the singlenucleotide polymorphism detected by this method may result from a change in template, rather than genomic, DNA, which is a disadvantage of the method of ESTderived SNP generation.

(3) The development of SNP markers on the basis of unique (nonoverlapping) genomic and EST sequences. In contrast to the two previous approaches, this method cannot be used for rapid large-scale generation of SNP markers because it requires additional DNA sequencing on several cultivars. However, this method is helpful in the case, when the databases lack data on overlapping for some sequences, thus supplementing the two previous approaches. It is employed for developing SNPs for organisms, for which no projects of complete genome sequencing are underway. Using EST sequences permits selective detection of coding SNPs, rejecting vast number of anonymous SNPs, occurring in non-expressing genome regions. Unfortunately, regulatory SNPs, which are as valuable as cSNPs as potential diagnostic markers, can be lost with the anonymous markers.

(4) The development of SNP markers on the basis of “shotgun” sequencing a particular sample of fragments of genomic DNA, which simultaneously represents several genotypes (reduced representation shotgun, RRS). The RRS approach has been designed for developing a map of the human genome, uniformly saturated with SNP markers, which has been constructed by the SNP Consortium (http://snp.cshl.org). This approach allows one to avoid a nonuniform distribution of the detected SNPs, which is characteristic for the method of SNP generation by in silico analysis of overlapping genomic sequences, associated with the admittedly unequal number of overlapping BAC clones for a certain genome part. So far, this method has not been used for detecting SNPs in plant genomes.

(5) The development of SNP markers on the basis of RFLP, STS, CAPS, and other molecular markers. This approach is aimed at the targeted generation of SNP markers to tag particular genes. Single-nucleotide polymorphism underlies many known types of molecular markers (except those based on repeated sequence polymorphism).

For instance, single-nucleotide substitutions (or insertions/deletions) in restriction recognition sites result in such DNA polymorphisms as RFLP and CAPS, while substitutions of nucleotides adjacent to restriction sites lead to APLP polymorphisms. Singlenucleotide substitutions/insertions/deletions in sites of PCR primer annealing may generate randomly amplified polymorphic DNA (RAPD). Main Advantages of SNP Markers In contrast to multiallelic markers, analysis of biallelic SNP markers can be practically fully automated. Several thousands of SNPs can be analyzed simultaneously by application of DNA microarrays. Thus, using modern technologies, the effectiveness of SNP analysis can be many times higher than that of other methods of DNA analysis. Moreover, none of other types of DNA polymorphisms has such diverse and numerous methods of analysis, as SNP. Because of this, SNP markers can be used in projects of different financial level. Single-nucleotide polymorphism is the most common type of DNA polymorphisms. For instance, in the wheat genome, the SNP density ranges from 1 per 370 bp to 1 per 540 bp. High occurrence of SNP loci in the genome opens a possibility to develop on their basis molecular genetic maps of very high density, required for isolating and studying genes for resistance to various diseases as well as other agronomically valuable plant genes. In addition, owing to the advantages of SNP markers, they can be widely used for certification of cultivars and lines. High SNP frequency involves also coding and regulatory regions of genomic DNA (cSNP and rSNP). This feature opens a perspective for effective development of diagnostic SNP markers. Strategies of Developing SSR Markers Microsatellites (1 to 10 nucleotides) and minisatellites (> 10 nucleotides) are subcategories of tandem repeats (TRs) that, together with the predominant interspersed repeats (or remnants of transposable elements), make up genomic repetitive regions. TRs are evolutionarily relevant due to their instability. They mutate at rates between 103 and 106 per cell generation i.e., up to 10 orders of magnitude greater than point mutations. Microsatellites, Simple Sequence Repeats (SSR), Short Tandem Repeats (STR) and Simple Sequence Length Polymorphisms (SSLP) are found in prokaryotes and eukaryotes. They are widely distributed throughout the genome, especially in the euchromatin of eukaryotes, and coding and non-coding nuclear and organellar DNA. There is a lot of evidence to back up the hypothesis that SSRs are not randomly distributed along the genome.

In a comparative study, SSR distribution was found to be highly non-random and to vary a great deal in different regions of the genes of Arabidopsis thaliana and rice. Development of SSR markers, including de novo nucleotide sequences for finding SSRs The development of SSR markers can basically be divided into the following stages:

1. Prior knowledge of nucleotide sequences in which SSRs occur;
2. Design of oligonucleotides (or primers) complementary to the regions flanking the SSR;
3. Validation of primers by PCR and electrophoresis of the product of the reaction, and
4. Detection of polymorphisms among individuals.

A schematic workflow showing how an SSR marker can be obtained is given in Figure. Interestingly, the efficiency of SSR marker development was found to be associated with the microsatellite class.

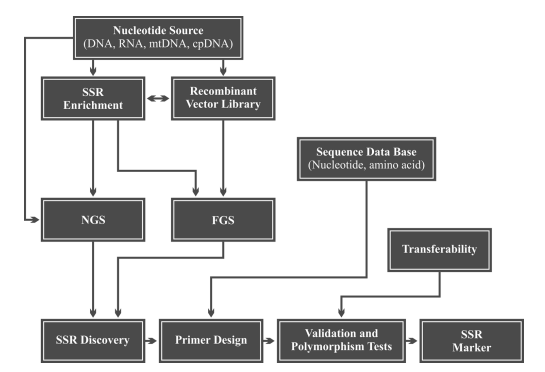


Figure - Workflow steps of SSR marker development

Microsatellites were originally developed from both coding and non-coding regions of plant genomes, and several sources were used to search for SSRs, including a variety of DNA libraries (genomic, genomic-enriched for SSR, bacterial artificial chromosome and cDNA libraries), as well as public databases, including expressed sequence tag (EST) databases. In prospecting for SSRs, the first step consists of constructing enriched genomic libraries and various enrichment methods have been successfully developed. To construct and sequence genomic libraries, the DNA is fragmented, ligated to adaptors and inserted into vectors for transforming Escherichia coli. Most protocols involve a stage of enrichment for repetitive sequences that can be achieved using selective hybridization, PCR or both techniques. In enrichment by hybridization, positive clones are detected using radioactively or chemically labeled SSR probes. Finally, these clones are selected by PCR amplification and sequencing. Another way of enriching a library is to use biotinized SSR probes that are captured by streptavidin-coated beads. The captured DNA is eluted, amplified, cloned and sequenced. The enriched libraries are screened to identify clones containing SSRs, producing the subsample of repetitive sequences that is intrinsic to this approach. PCR-based methods can bias the sampling of repetitive sequences in non-enriched libraries, since fragment selection and amplification are dependent on complementarity with specific primers for the SSR and cloning vector. However, non-enriched libraries and alternative methods derived from other molecular markers (e.g. RAPD and AFLP) have also been used to find SSRs. The advances made in Next Generation Sequencing (NGS) have provided a new scenario for detecting microsatellites. Various NGS-based projects have been developed over the last few decades, generating an enormous quantity of sequences made available in public databases and widely used for prospecting for microsatellites. Automation of the original sequencing method proposed by Sanger and Coulson (1975) has made it possible to sequence the complete genome of A. thaliana (Arabidopsis Genome Initiative, 2000). However, because of the high cost of the Sanger method when sequencing complete genomes, it has been replaced by NGS platforms or a combination of both methods. NGS has been very useful for various studies, including prospecting for new SSR markers. Successors of the Sanger sequencing method include the 454 FLX (Roche), Solexa (Illumina), SOLiD (Applied Biosystems) and HeliScope True Single Molecule Sequencing (Helicos) platforms. Third generation platforms are also currently available, including a platform developed by Pacific Biosciences (PacBio), based on a new sequencing technology, SMRT sequencing, which has the advantage of producing longer DNA reads. Each platform has specific characteristics in terms of the number and size of reads generated, run time, as well as the accuracy and cost of each base read, with both advantages and disadvantages compared to other platforms.

In order to advice researchers in sequencing technology choice. Control error analysis is one of the most important steps in sequencing data analysis, mainly in de novo sequencing projects, that lack a reference genome. Furthermore, sequences that contain repetitive regions are challenges to be overcome by error correction methods, due to their vulnerability to errors. Initiatives for sequencing the complete genomes of various species use combinations of different platforms with the aim of incorporating the best features of each and extracting the maximum amount of information. Currently, 454 and Illumina are the NGS platforms most widely used for developing SSR markers. However, the PacBio SMRT sequencing technology is being considered an economically viable alternative for discovering microsatellites.

**Advanced methods of genotyping**

**Introduction**

Genotyping is the process of determining differences in the genetic make-up (genotype) of an individual by examining the individual's DNA sequence using biological assays and comparing it to another individual's sequence or a reference sequence. It reveals the alleles an individual has inherited from their parents. Traditionally genotyping is the use of DNA sequences to define biological populations by use of molecular tools. It does not usually involve defining the genes of an individual.

Current methods of genotyping include restriction fragment length polymorphism identification (RFLP) of genomic DNA, random amplified polymorphic detection (RAPD) of genomic DNA, amplified fragment length polymorphism detection (AFLPD), polymerase chain reaction (PCR), DNA sequencing, allele specific oligonucleotide (ASO) probes, and hybridization to DNA microarrays or beads. Genotyping is important in research of genes and gene variants associated with disease. Due to current technological limitations, almost all genotyping is partial. That is, only a small fraction of an individual’s genotype is determined, such as with (epi) GBS (Genotyping by sequencing) or RADseq. New mass-sequencing technologies promise to provide whole-genome genotyping (or whole genome sequencing) in the future. Genotyping applies to a broad range of individuals, including microorganisms. For example, viruses and bacteria can be genotyped. Genotyping in this context may help in controlling the spreading of pathogens, by tracing the origin of outbreaks.

This area is often referred to as molecular epidemiology or forensic microbiology. When genotyping transgenic organisms, a single genomic region may be all that needs to be examined to determine the genotype.

1. Restriction fragment length polymorphism (RFLP) is a technique that exploits variations in homologous DNA sequences, known as polymorphisms, in order to distinguish individuals, populations, or species or to pinpoint the locations of genes within a sequence. The term may refer to a polymorphism itself, as detected through the differing locations of restriction enzyme sites, or to a related laboratory technique by which such differences can be illustrated. In RFLP analysis, a DNA sample is digested into fragments by one or more restriction enzymes, and the resulting restriction fragments are then separated by gel electrophoresis according to their size. Although now largely obsolete due to the emergence of inexpensive DNA sequencing technologies, RFLP analysis was the first DNA profiling technique inexpensive enough to see widespread application. RFLP analysis was an important early tool in genome mapping, localization of genes for genetic disorders, determination of risk for disease, and paternity testing. RFLP Analysis The basic technique for the detection of RFLPs involves fragmenting a sample of DNA with the application of a restriction enzyme, which can selectively cleave a DNA molecule wherever a short, specific sequence is recognized in a process known as a restriction digest. The DNA fragments produced by the digest are then separated by length through a process known as agarose gel electrophoresis and transferred to a membrane via the Southern blot procedure. Hybridization of the membrane to a labeled DNA probe then determines the length of the fragments which are complementary to the probe. A restriction fragment length polymorphism is said to occur when the length of a detected fragment varies between individuals, indicating non-identical sequence homologies. Each fragment length is considered an allele, whether it actually contains a coding region or not, and can be used in subsequent genetic analysis.

2. RAPD (pronounced as "rapid") stands for 'Random Amplification of Polymorphic DNA. It is a type of PCR, but the segments of DNA that are amplified are random. The scientist performing RAPD creates several arbitrary, short primers (8–12 nucleotides), then proceeds with the PCR using a large template of genomic DNA, hoping that fragments will amplify. By resolving the resulting patterns, a semi-unique profile can be gleaned from an RAPD reaction. No knowledge of the DNA sequence of the targeted genome is required, as the primers will bind somewhere in the sequence, but it is not certain exactly where. This makes the method popular for comparing the DNA of biological systems that have not had the attention of the scientific community, or in a system in which relatively few DNA sequences are compared (it is not suitable for forming a cDNA databank). Because it relies on a large, intact DNA template sequence, it has some limitations in the use of degraded DNA samples. Its resolving power is much lower than targeted, species-specific DNA comparison methods, such as short tandem repeats. In recent years, RAPD has been used to characterize, and trace, the phylogeny of diverse plant and animal species. RAPD markers are decamer (10 nucleotides long) DNA fragments from PCR amplification of random segments of genomic DNA with a single primer of arbitrary nucleotide sequence and which are able to differentiate between genetically distinct individuals, although not necessarily in a reproducible way. It is used to analyze the genetic diversity of an individual by using random primers. Due to problems in experiment reproducibility, many scientific journals do not accept experiments merely based on RAPDs anymore. RAPD requires only one primer for amplification. How it works Unlike traditional PCR analysis, RAPD does not require any specific knowledge of the DNA sequence of the target organism: the identical 10-mer primers will or will not amplify a segment of DNA, depending on positions that are complementary to the primers' sequence.

For example, no fragment is produced if primers annealed too far apart or 3' ends of the primers are not facing each other. Therefore, if a mutation has occurred in the template DNA at the site that was previously complementary to the primer, a PCR product will not be produced, resulting in a different pattern of amplified DNA segments on the gel.

Limitations

• Nearly all RAPD markers are dominant, i.e. it is not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies). Codominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are detected only rarely.

• PCR is an enzymatic reaction, therefore, the quality and concentration of template DNA, concentrations of PCR components, and the PCR cycling conditions may greatly influence the outcome. Thus, the RAPD technique is notoriously laboratory dependent and needs carefully developed laboratory protocols to be reproducible.

• Mismatches between the primer and the template may result in the total absence of PCR product as well as in a merely decreased amount of the product. Thus, the RAPD results can be difficult to interpret, unlike traditional PCR analysis.

3. AFLP-PCR or just AFLP is a PCR-based tool used in genetics research, DNA fingerprinting, and in the practice of genetic engineering. Developed in the early 1990s by Keygene. AFLP uses restriction enzymes to digest genomic DNA, followed by ligation of adaptors to the sticky ends of the restriction fragments. A subset of the restriction fragments is then selected to be amplified. This selection is achieved by using primers complementary to the adaptor sequence, the restriction site sequence and a few nucleotides inside the restriction site fragments (as described in detail below). The amplified fragments are separated and visualized on denaturing on agarose gel electrophoresis , either through autoradiography or fluorescence methodologies, or via automated capillary sequencing instruments. Although AFLP should not be used as an acronym it is commonly referred to as "Amplified fragment length polymorphism", the resulting data are not scored as length polymorphisms, but instead as presence-absence polymorphisms. AFLP-PCR is a highly sensitive method for detecting polymorphisms in DNA. The technique was originally described by Vos and Zabeau in 1993. In detail, the procedure of this technique is divided into three steps:

1. Digestion of total cellular DNA with one or more restriction enzymes and ligation of restriction half-site specific adaptors to all restriction fragments.

2. Selective amplification of some of these fragments with two PCR primers that have corresponding adaptor and restriction site specific sequences.

3. Electrophoretic separation of amplicons on a gel matrix, followed by visualisation of the band pattern.

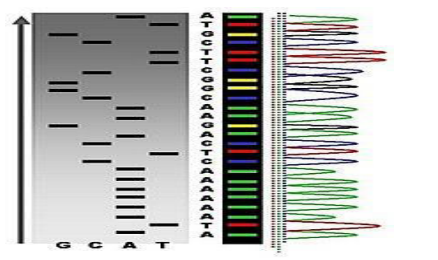
**Applications**

The AFLP technology has the capability to detect various polymorphisms in different genomic regions simultaneously. It is also highly sensitive and reproducible. As a result, AFLP has become widely used for the identification of genetic variation in strains or closely related species of plants, fungi, animals, and bacteria. The AFLP technology has been used in criminal and paternity tests, also to determine slight differences within populations, and in linkage studies to generate maps for quantitative trait locus (QTL) analysis.

There are many advantages to AFLP when compared to other marker technologies including randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), and microsatellites. AFLP not only has higher reproducibility, resolution, and sensitivity at the whole genome level compared to other techniques, but it also has the capability to amplify between 50 and 100 fragments at one time. In addition, no prior sequence information is needed for amplification. As a result, AFLP has become extremely beneficial in the study of taxa including bacteria, fungi, and plants, where much is still unknown about the genomic makeup of various organisms.

4. Polymerase chain reaction (PCR) is a method widely used in molecular biology to rapidly make millions to billions of copies of a specific DNA sample allowing scientists to take a very small sample of DNA and amplify it to a large enough amount to study in detail. PCR was invented in 1983 by Kary Mullis. It is fundamental to much of genetic testing including analysis of ancient samples of DNA and identification of infectious agents. Using PCR, copies of very small amounts of DNA sequences are exponentially amplified in a series or cycles of temperature changes. PCR is now a common and often indispensable technique used in medical laboratory and clinical laboratory research for a broad variety of applications including biomedical research and criminal forensics. The majority of PCR methods rely on thermal cycling. Thermal cycling exposes reactants to repeated cycles of heating and cooling to permit different temperature-dependent reactions – specifically, DNA melting and enzyme-driven DNA replication. PCR employs two main reagents – primers (which are short single strand DNA fragments known as oligonucleotides that are a complementary sequence to the target DNA region) and a DNA polymerase. In the first step of PCR, the two strands of the DNA double helix are physically separated at a high temperature in a process called Nucleic acid denaturation. In the second step, the temperature is lowered and the primers bind to the complementary sequences of DNA. The two DNA strands then become templates for DNA polymerase to enzymatically assemble a new DNA strand from free nucleotides, the building blocks of DNA. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the original DNA template is exponentially amplified. Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase, an enzyme originally isolated from the thermophilic bacterium Thermus aquaticus. If the polymerase used was heat-susceptible, it would denature under the high temperatures of the denaturation step. Before the use of Taq polymerase, DNA polymerase had to be manually added every cycle, which was a tedious and costly process. Applications of the technique include DNA cloning for sequencing, gene cloning and manipulation, gene mutagenesis; construction of DNAbased phylogenies, or functional analysis of genes; diagnosis and monitoring of hereditary diseases; amplification of ancient DNA; analysis of genetic fingerprints for DNA profiling (for example, in forensic science and parentage testing); and detection of pathogens in nucleic acid tests for the diagnosis of infectious diseases.

5. DNA sequencing is the process of determining the nucleic acid sequence – the order of nucleotides in DNA. It includes any method or technology that is used to determine the order of the four bases: adenine, guanine, cytosine, and thymine. The advent of rapid DNA sequencing methods has greatly accelerated biological and medical research and discovery. Knowledge of DNA sequences has become indispensable for basic biological research, and in numerous applied fields such as medical diagnosis, biotechnology, forensic biology, virology and biological systematics. Comparing healthy and mutated DNA sequences can diagnose different diseases including various cancers, characterize antibody repertoire, and can be used to guide patient treatment. Having a quick way to sequence DNA allows for faster and more individualized medical care to be administered, and for more organisms to be identified and cataloged. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of complete DNA sequences, or genomes, of numerous types and species of life, including the human genome and other complete DNA sequences of many animal, plant, and microbial species.



**An example of the results of automated chain-termination DNA sequencing.**

The first DNA sequences were obtained in the early 1970s by academic researchers using laborious methods based on two-dimensional chromatography. Following the development of fluorescence-based sequencing methods with a DNA sequencer, DNA sequencing has become easier and orders of magnitude faster.

6. An allele-specific oligonucleotide (ASO) is a short piece of synthetic DNA complementary to the sequence of a variable target DNA. It acts as a probe for the presence of the target in a Southern blot assay or, more commonly, in the simpler Dot blot assay. It is a common tool used in genetic testing, forensics, and Molecular Biology research. An ASO is typically an oligonucleotide of 15–21 nucleotide bases in length. It is designed (and used) in a way that makes it specific for only one version, or allele, of the DNA being tested. The length of the ASO, which strand it is chosen from, and the conditions by which it is bound to (and washed from) the target DNA all play a role in its specificity. These probes can usually be designed to detect a difference of as little as 1 base in the target's genetic sequence, a basic ability in the assay of single-nucleotide polymorphisms (SNPs), important in genotype analysis and the Human Genome Project. To be detected after it has bound to its target, the ASO must be labeled with a radioactive, enzymatic, or fluorescent tag. The Illumina Methylation Assay technology takes advantage of ASO to detect one base pair difference (cytosine versus thymine) to measure methylation at a specific CpG site.

7. Sequencing by hybridization is a class of methods for determining the order in which nucleotides occur on a strand of DNA. Typically used for looking for small changes relative to a known DNA sequence. The binding of one strand of DNA to its complementary strand in the DNA double-helix (known as hybridization) is sensitive to even single-base mismatches when the hybrid region is short or if specialized mismatch detection proteins are present. This is exploited in a variety of ways, most notably via DNA chips or microarrays with thousands to billions of synthetic oligonucleotides found in a genome of interest plus many known variations or even all possible single-base variations. The type of sequencing by hybridization described above has largely been displaced by other methods, including sequencing by synthesis, and sequencing by ligation (as well as pore-based methods). However hybridization of oligonucleotides is still used in some sequencing schemes, including hybridizationassisted pore-based sequencing, and reversible hybridization.

8. A DNA microarray (also commonly known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface. Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles (10−12 moles) of a specific DNA sequence, known as probes (or reporters or oligos). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA (also called anti-sense RNA) sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target. The original nucleic acid arrays were macro arrays approximately 9 cm × 12 cm and the first computerized image based analysis was published in 1981.[2] It was invented by Patrick O. Brown.

**- QTL mapping using structured populations; AB-QTL analysis;**

**Introduction**

The objective of genetic mapping is to identify QTL responsible for natural phenotypic variation. Two strategies have been widely applied to genetic mapping in plants:

(1) linkage mapping

(2) association or linkage disequilibrium (LD) mapping.

Linkage mapping, a conventional mapping method, depends upon genetic recombination during the construction of mapping populations. Over the past two decades, linkage mapping has been commonly used in various plant species, and many QTL have been cloned or tagged. However, linkage mapping has the disadvantages of relatively low mapping resolution, low allele richness, and low speed. Association mapping, as a complement to linkage mapping, takes advantage of historic recombination events accumulated over hundreds of generations, thus providing higher resolution and greater allele numbers Since human diseases were successfully dissected, association mapping has been applied to crops. Following its introduction to crops association mapping has attracted increased attention in genetic studies. Owing to the dramatic reduction in costs of sequence technologies, association mapping has been conducted in plants from the model plant Arabidopsis thaliana to many major crops. The key distinction between association and linkage mapping lies in whether recombination events occur in populations or families. However, both of these methods share a consistent strategy for identifying molecular markers that are linked to QTL. As we step into the era of complete genome sequencing, the difference between the two methods will disappear . For organisms whose genomes are known, one might now try to exclude genes in the identified region whose function is known with some certainty not to be connected with the trait in question. If the genome is not available, it may be an option to sequence the identified region and determine the putative functions of genes by their similarity to genes with known function, usually in other genomes. This can be done using BLAST, an online tool that allows users to enter a primary sequence and search for similar sequences within the BLAST database of genes from various organisms. It is often not the actual gene underlying the phenotypic trait, but rather a region of DNA that is closely linked with the gene. Another interest of statistical geneticists using QTL mapping is to determine the complexity of the genetic architecture underlying a phenotypic trait.

For example, they may be interested in knowing whether a phenotype is shaped by many independent loci, or by a few loci, and do those loci interact. This can provide information on how the phenotype may be evolving. In a recent development, classical QTL analyses were combined with gene expression profiling i.e. by DNA microarrays. Such expression QTLs (eQTLs) describe cis- and trans-controlling elements for the expression of often disease-associated genes. Observed epistatic effects have been found beneficial to identify the gene responsible by a cross-validation of genes within the interacting loci with metabolic pathway- and scientific literature databases.

The process of QTL mapping involves the following major steps, such as:

a) Development of appropriate mapping population [F2 or Fx derived families, backcross (BC), double haploids (DH), recombinant inbred lines (RILs) and near isogenic lines (NILs)],

b) Precise phenotyping for related traits (morphological characters, agronomic traits, disease and pest scores, drought resistance, etc.)

c) Deciding the type of marker(s) and identification of polymorphism,

d) Genotyping of the mapping populations with polymorphic markers,

e) Construction of genetic maps using statistical programs

f) QTL mapping using both genotypic and phenotypic data Analysis of variance

The simplest method for QTL mapping is analysis of variance (ANOVA, sometimes called "marker regression") at the marker loci. In this method, in a backcross, one may calculate a t-statistic to compare the averages of the two marker genotype groups. For other types of crosses (such as the intercross), where there are more than two possible genotypes, one uses a more general form of ANOVA, which provides a so-called F-statistic.

The ANOVA approach for QTL mapping has three important weaknesses. First, we do not receive separate estimates of QTL location and QTL effect. QTL location is indicated only by looking at which markers give the greatest differences between genotype group averages, and the apparent QTL effect at a marker will be smaller than the true QTL effect as a result of recombination between the marker and the QTL. Second, we must discard individuals whose genotypes are missing at the marker. Third, when the markers are widely spaced, the QTL may be quite far from all markers, and so the power for QTL detection will decrease.

1. Interval mapping Lander and Botstein developed interval mapping, which overcomes the three disadvantages of analysis of variance at marker loci. Interval mapping is currently the most popular approach for QTL mapping in experimental crosses. The method makes use of a genetic map of the typed markers, and, like analysis of variance, assumes the presence of a single QTL. In interval mapping, each locus is considered one at a time and the logarithm of the odds ratio (LOD score) is calculated for the model that the given locus is a true QTL. The odds ratio is related to the Pearson correlation coefficient between the phenotype and the marker genotype for each individual in the experimental cross. The term 'interval mapping' is used for estimating the position of a QTL within two markers (often indicated as 'marker-bracket').

Interval mapping is originally based on the maximum likelihood but there are also very good approximations possible with simple regression. The principle for QTL mapping is:

1) The Likelihood can be calculated for a given set of parameters (particularly QTL effect and QTL position) given the observed data on phenotypes and marker genotypes.

2) The estimates for the parameters are those where the likelihood are highest.

3) A significance threshold can be established by permutation testing.

Conventional methods for the detection of quantitative trait loci (QTLs) are based on a comparison of single QTL models with a model assuming no QTL. For instance in the "interval mapping" method the likelihood for a single putative QTL is assessed at each location on the genome. However, QTLs located elsewhere on the genome can have an interfering effect. As a consequence, the power of detection may be compromised, and the estimates of locations and effects of QTLs may be biased (Lander and Botstein 1989; Knapp 1991). Even nonexisting so-called "ghost" QTLs may appear (Haley and Knott 1992; Martinez and Curnow 1992). Therefore, multiple QTLs could be mapped more efficiently and more accurately by using multiple QTL models. One popular approach to handle QTL mapping where multiple QTL contribute to a trait is to iteratively scan the genome and add known QTL to the regression model as QTLs are identified. This method, termed composite interval mapping determine both the location and effects size of QTL more accurately than single-QTL approaches, especially in small mapping populations where the effect of correlation between genotypes in the mapping population may be problematic.

2. Composite interval mapping (CIM) In this method, one performs interval mapping using a subset of marker loci as covariates. These markers serve as proxies for other QTLs to increase the resolution of interval mapping, by accounting for linked QTLs and reducing the residual variation. The key problem with CIM concerns the choice of suitable marker loci to serve as covariates; once these have been chosen, CIM turns the model selection problem into a single-dimensional scan. The choice of marker covariates has not been solved, however. Not surprisingly, the appropriate markers are those closest to the true QTLs, and so if one could find these, the QTL mapping problem would be complete anyway.

3. Family-based mapping

3.1. Biparental populations

The first and most important step in family-based mapping is constructing experimental populations, which may be biparental populations such as F2, backcrosses (BC), doubled haploids (DH), recombinant inbred lines (RIL), and nearisogenic lines (NIL). These commonly used populations with their strengths and weaknesses.

The general process of biparental mapping includes:

(1) collection of parental strains that differ for traits of interest,

(2) selection of molecular markers such as RFLP, SSR and SNP that distinguish between the two parents,

(3) development of a mapping population,

(4) genotyping and phenotyping of the mapping population

(5) detection of QTL using a suitable statistical method.

The power of QTL detection is affected by QTL effects, allele frequencies, and the type and size of the mapping population. Biparental mapping has proven to be useful in crop breeding. The main limitation of a biparental population is that only a few recombination events occur during the development of the population, allowing the localization of QTL to 10–20 cM intervals. Additionally, detection of QTL in biparental populations depends on the phenotypic diversity of the two parents, which may account for only a small part of the genetic variation in the species.

3.2. Multiparent mapping populations Multiparent mapping populations have been constructed to overcome the limitations of biparental populations. The genetic diversity of multiple parents leads to a population with large phenotypic diversity, making it suitable for high-resolution QTL mapping. Increasing in popularity are two experimental designs of multiparent populations that include nested association mapping (NAM) and multiparent advanced generation intercrosses (MAGIC).

NAM is an excellent multiparent population design for dissecting the genetic architecture of maize flowering time. A NAM population was created by crossing 25 diverse inbred maize lines to the B73 inbred, chosen as a reference line, resulting in 5000 RILs from 25 families, with 200 RILs per family. As a combination of several high-resolution biparental populations in one large population, the NAM population affords very high resolution and power for detecting QTL. In maize, a NAM population has been used for large-scale genetic mapping for several important traits including leaf architecture and disease resistance. Compared with other multiparent populations, MAGIC populations involve intermating multiple inbred founders for multiple generations prior to the construction of inbred lines, considerably improving the precision of QTL detection. Undoubtedly, MAGIC populations offer great opportunities for dissecting complex traits and improving breeding populations. Statistical approaches for QTL mapping in MAGIC populations have become available, some of them based on the general linear model (GLM) used in biparental populations .

4. Natural population-based mapping With the advantages of high resolution, high allelic richness, and absence of need of the tedious development of a mapping population, natural population-based mapping has become a powerful tool for detection of natural variation underlying complex traits in more than a dozen crops since 2001. They consist of first, collection of a sample population including elite cultivars, landraces, wild relatives, and exotic accessions; second, phenotyping traits, estimating broad-sense heritability of traits of interest and determining the genotypes of the population entries, either for candidate genes or genome-wide; third, quantification of the LD extent of the selected population; fourth, identification of the influence of population structure and kinship; and fifth, testing the associations between genotypes and phenotypes using appropriate statistical approaches. Subsequent experimental validations such as mutagenesis and gene expression analysis are required. The statistical power of natural populationbased mapping is strongly dependent upon the extent of LD and population structure, as well as the sample size and minor allele frequency (MAF).

5. Statistical analysis in genetic mapping 4.1. Statistical methods for family-based mapping Single-marker analysis is used for initial QTL mapping in biparental populations and identifies QTL according to the difference between the average phenotypes of different genotype groups without using information about genetic distances in the linkage map. To improve the power of mapping, Lander and Botstein [38] proposed interval mapping (IM) based on maximum-likelihood parameter estimation, which efficiently estimates the effect and position of a QTL within two flanking markers. A regression version of IM developed by Haley and Knott is used to simplify computation of IM. IM assumes that only one QTL affects quantitative traits of interest and ignores the effects of other QTL. However, it is well known that quantitative traits are usually controlled by several loci, and therefore QTL can be mapped more accurately by analysis of multiple QTL simultaneously. To overcome the limitation of IM, composite interval mapping (CIM), combining regression and interval mapping, selects a subset of markers as covariates that can account for the effects of linked QTL and reduce residual error. The key question in CIM is how to choose appropriate marker covariates, and this question can be addressed by stepwise regression or preliminary interval mapping. A modified algorithm called inclusive composite interval mapping (ICIM) better controls sampling variance and avoids the complicated selection of marker covariates while retaining all the advantages of CIM. Multiple interval mapping (MIM), an extension of interval mapping to multiple QTL, tends to be more powerful and precise than CIM in identifying QTL and allows the simultaneous estimation of multiple QTL with epistasis. A large number of software packages implementing the above methods are available for biparental mapping, including QTL Cartographer, QTL Network, and R/qtl. Although many dense marker maps have been constructed, QTL interval mapping remains useful today because complete genomic information and high-density marker maps are not available for all crop species. AB-QTL analysis Advanced Backcross QTL Advanced backcross (AB) QTL analysis, integrates QTL analysis with variety development, by identifying and transferring the valuable QTL alleles from un adapted to cultivated germplasm in a single process. In this approach, QTL and marker analyses are performed in advanced generations, like BC2 or BC3. The power of detection of different types of QTL differs in selfing populations and backcross populations. Backcross populations are less powerful in detecting a donor QTL having some degree of recessive gene action. Nevertheless, the power of detection in backcross population increases with the degree of dominance. The completely dominant or over-dominant QTLs can be detected with efficiency equivalent to or greater than that of selfing populations. Backcross population does not prove to be efficient in detecting QTLs requiring epistatic interactions. However, they are more competent and useful than selfing populations in detecting the recessive and epistatic QTLs as these QTLs contribute very slightly to the phenotypic variance of background populations. Consequently, detection of additive, dominant and overdominant QTLs is boosted in backcross generations in comparison to selfing generations where recessive or epistatic QTLs are segregating simultaneously in the population. The objective of QTL analysis is to detect the valuable QTL alleles from a donor genome and subsequently their transfer into elite breeding lines. In this context, one must have the idea concerning number of additional generations of backcrossing required and number of individuals needed to be sampled, to attain the lines having segment of the donor chromosome with the valuable QTL in the background of recurrent parent genome.

Those lines are referred as QTL/nearly isogenic lines or QTL-NILs. Tanksley and Nelson, 1996 propounded two factors that are important in determining this, which are

(1) maximum size (in centimorgans) of the donor segment designating the QTL

(2) the amount of residual donor genome (unlinked to the targeted QTL) still present in the genome.

In this context, the backcross populations get skewed toward recurrent parent alleles which make them superior over selfing generations. QTL-NILs can be derived from BC1 or BC2 derived populations, but for this screening of large number of individuals i.e. around 5,000 or 10,000 is required respectively. However, selection can be exerted to eliminate non-targeted donor segments by screening a smaller number of individuals over two sequential generations (e.g., a backcross followed by a selfing). Thus, in contrast to the BC1 and BC2, QTL-NILs can be derived directly from BC3 - BC5 selections from a comparatively small number of individuals. In the other words, we can say that the more advanced the backcross population, the simpler it will be to derive a desired QTL-NILs. Based on the above considerations, “Advanced Backcross QTL Analysis” can be applied effectively to detect and interogress QTLs simultaneously in which QTL mapping is delayed to either BC2 or BC3 generations. Use of statistical tools will assist in detection of additive, dominant and over-dominant QTLs in the advanced backcross generations. Selection can be implemented to recognize and isolate individual plants that are sufficiently similar to the recurrent parent for advancing to extract QTL-NILs which may serve as a parent in breeding program or directly as improved varieties. They can also be used to further confirm and fine map selected QTLs. The potential applications and limitations of AB-QTL analysis with respect to crop improvement through plant breeding are discussed below. Application of AB-QTL analysis in crop improvement The utility of the AB-QTL approach has been tested in numerous crops species and proved to be a feasible method in crop breeding. AB-QTL in self pollinated crops Single elite inbred variety is initially crossed to an unrelated donor line to generate BC1 progeny (around 100 plants). Plants with desirable characters are selected from the BC1 progeny and this selected BC1 are crossed again with the recurrent parent to produce the BC2 progeny of around 200 individuals. Then the BC2 /BC3 generation plants are evaluated in replicated trials and genotyped for marker trait loci and selfed to produce BC2 S1 /BC2 S2 progeny. The genotype and phenotype data are subjected to QTL analysis to identify donor genome regions containing favourable QTL alleles.BC2 S1 or BC2 S2 families assist in detection of some recessive QTL donor alleles in addition to the expected dominant and additive donor QTL alleles. Ultimately QTLNILs are extracted from the superior BC2 S1 /BC2 S2 which is used to confirm the findings from the QTL mapping or to fine map the detected QTLs. The outperforming QTL-NILs can be used as parent in future breeding program or as new varieties. AB-QTL in cross pollinated crops Advanced backcross QTL analysis could be applied to cross pollinated crops through a slight modification in the strategy presented above. In case of the cross pollinated crops the elite inbred parent e.g. inbred A is used as the recurrent parent which is being involved in the production of commercial hybrid derived from the single cross A x B. Hybridization and backcrossing of donor with the recurrent parent is performed to produce the BC2 population. The selected plants from the BC2 /BC3 are genotyped for marker loci. Instead of selfing the BC2 /BC3 are crossed with the inbred B to produce BC2 F1 families and later the phenotyping is performed for this. The marker and the phenotypic data are used for the QTL analysis. On the basis of the QTL analysis of this data, favourable QTLs from the donor parent are identified, eventually the QTL-NILs could be generated. Advantages and limitations of AB-QTL Approach The potential applications and limitations of AB-QTL analysis with respect to plant breeding are discussed below. Advantages Since the smaller number of genes from donor parent will be present in BC2 or BC3, the undesirable effect of wild species on the elite germplasm is reduced and the effect of individual QTLs is measured more precisely. As the phenotypic selection is delayed for advanced generation, the frequency of deleterious or undesirable alleles from the donor is further reduced. Therefore, the problems with deleterious effects which are associated experienced with balanced population (F2, BC1, or RILs) are minimized in this method. In addition, to have a better prediction of the effect of QTL, epistatic effect should be minimized. MAS performed in advanced generation is more effective than in F2 or BC1 asaccumulation of the donor alleles is minimised in advanced generation due to breaking of assembly of favourable epistatic gene combination through recombination. Due to this, AB population is skewed more towards the recurrent parent genome and the epistatic interaction between the favourable epistatic interactions between the recurrent parent and tester alleles are comparatively less distorted. In addition, QTL-NILs (lines which differ from its parent in only one genomic location, ideally a single favourable QTL) can be created by one or two additional backcrossing with selected QTL. Also, the opportunity for useful meiotic recombination occurs as the generation is advanced, the deleterious effect due to linkage drag is also minimized in the advanced generations. In ABQTL, the mean performance of the advanced backcross families is skewed towards the elite parent that provides opportunity to detect the subtle pleiotropic effect of introgressed QTLs. Limitations Despite of having so many advantages of this method, there are still some cases where the effortless application of this method is limited. Advanced backcross QTL analysis is not likely to be useful in the crops with relatively longer generation time (> 2 years). Use of this method in longgeneration perennial crops are limited because intricacy in development and exploitation of the necessary population and QTL-NILs. As the success of this method in production of any commercial or hybrid varieties depends on the availability of the inbreds, this method is strenuous to apply in the crops for which inbred lines do not exist. In addition, this method is also difficult to apply in highly heterozygous, outcrossing crops like alfalfa or clonally propagated crops like potato for which inbred lines are not commonly employed in breeding programs.

**Association mapping of QTL; Fine mapping of genes/QTL**

**Introduction**

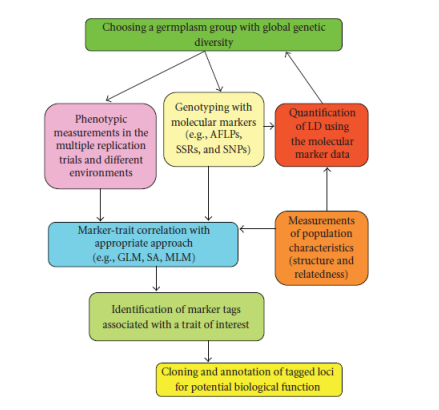
In genetics, association mapping, also known as "linkage disequilibrium mapping", is a method of mapping quantitative trait loci (QTLs) that takes advantage of historic linkage disequilibrium to link phenotypes (observable characteristics) to genotypes (the genetic constitution of organisms), uncovering genetic associations. Theory Association mapping is based on the idea that traits that have entered a population only recently will still be linked to the surrounding genetic sequence of the original evolutionary ancestor, or in other words, will more often be found within a given haplotype, than outside of it. It is most often performed by scanning the entire genome for significant associations between a panel of SNPs (which, in many cases are spotted onto glass slides to create “SNP chips”) and a particular phenotype.

These associations must then be independently verified in order to show that they either

(a) contribute to the trait of interest directly

(b) are linked to/ in linkage disequilibrium with a quantitative trait locus (QTL) that contributes to the trait of interest.

Association mapping seeks to identify specific functional genetic variants (loci, alleles) linked to phenotypic differences in a trait to facilitate detection of trait causing DNA sequence polymorphisms and selection of genotypes that closely resemble the phenotype. In order to identify these functional variants, it requires high throughput markers like single nucleotide polymorphisms (SNPs).



**Figure- The scheme of association mapping for tagging a gene of interest using germplasm accessions**.

Note that the outlined scheme may vary based on population characteristics and methodology chosen for association study LINKAGE DISEQUILIBRIUM (LD) Concept of LD Genetic linkage generally refers to coinheritance of different loci within a genetic distance on the chromosome. There are two terms used in population genetics, linkage equilibrium (LE), and linkage disequilibrium (LD) to describe linkage relationships (co-occurrence) of alleles at different loci in a population. LE is a random association of alleles at different loci and equals the product of allele frequencies within haplotypes, meaning that at random combination of alleles at each locus its haplotypes (combination of alleles) frequency has equal value in a population. In contrast, LD is a nonrandom association of alleles at different loci, describing the condition with nonequal (increased or reduced) frequency of the haplotypes in a population at random combination of alleles at different loci. LD is not the same as linkage, although tight linkage may generate high levels of LD between alleles. Usually, there is significant LD between more distant sites or sites located in different chromosomes, caused by some specific genetic factors. Genome-wide association (GWAS) Genome-wide association (GWAS) is another method that takes advantage of a number of historic recombination events that have occurred within natural populations, and can overcome the limitations of linkage analysis. Advent of high-throughput genotyping methods, genome-wide association (GWA) is the emerging tool for studying the genetics underlying natural phenotypic variation. GWA studies have been applied across a wide range of species where it enabled fine-scale genetic mapping.

**MAPPING POPULATION DEVELOPMENT, GENOTYPING, AND VARIANCE COMPONENTS ESTIMATION APPROACHES FOR GWAS**

Mapping population derived from a cross with clear pedigrees with known ancestry. Genome-wide association study (GWAS) and Linkage mapping are the suitable methods to identify the QTLs for complex traits. Linkage analysis in plants has been typically conducted with experimental population which is derived from bi-parentail cross with a clear ancestry, while a collection of cultivars with unobserved ancestry are often adopted for association. Genetic recombinations among alleles are the genetic basis for linkage mapping and association mapping in a plant. Association mapping exploits historical and evolutionary recombinations at a natural population level whereas, Linkage mapping exploits the functional polymorphisms and adjacent markers within families or pedigrees with known ancestry. Although based on the same fundamental principles of genetic recombination as a linkage analysis, association mapping examines this shared inheritance for a collection of individuals often with unobserved ancestry. By exploring deeper population genealogy rather than family pedigree, association mapping offers advantages over linkage analysis: Much higher mapping resolution, greater allele number and broader reference population and less research time in establishing an association. Linkage analysis and association mapping, however, are complementary to each other in terms of providing prior knowledge, cross validation and statistical power. Both linkage analysis and association mapping studies rely on coinheritance of neighboring DNA variants with the phenotype. In both analyses, recombination is the main force eliminating linkages over generation. The difference is that in linkage analysis, there are only few opportunities allowed for recombination to occur (one or two generation), resulting in comparatively low resolution of the map. In association analysis, which exploit natural variation in the population, recombination events over many generation are expected to eliminate linkage between a mutated gene and molecular markers, except the markers that are located close to gene. Only very close markers are in linkage disequilibrium (LD) with the mutated gene and form the basis for association mapping .

**LINKAGE DISEQUILIBRIUM AND GENOME WIDE ASSOCIATION ANALYSIS**

Genome wide association mapping refers to the significant association of a marker locus with a phenotype trait while LD refers to the non randomassociation between two markers or two genes/QTLs. Genetic linkage between two loci (gene or molecular markers) on a genome is measured by the fact that they are in association and the involved loci are very closely associated and are located very close to each other. A complementary approach is to analyze linkage Disequilibrium (LD) between the marker and target gene in a natural population. LD between two loci in natural population is affected by all the recombination events that happened. Thus, LD declines as the number of generation increases, so that in old population LD is limited to small distance. Both QTL and association mapping approaches are therefore based on LD between molecular markers and functional loci. In QTL mapping, LD is generated by the mating design while in association mapping, LD is a reflection of the germplasm collection under study. In a mapping population, LD is influenced only by recombination in the absence of segregation distortion. In AM, LD may also be influenced by other forces such as those mentioned above as well as by recombination. In plants usually another candidate gene association mapping is also used to dissect out the genetic control of complex traits, with biologically relevant candidates selected and ranked based on the available results from genetic, biochemical or physiological studies in model and non-model plants species. candidate association mapping requires the identification of SNP between lines and within specific genes. Identifying candidate gene SNPs relies on resequencing of amplicons from several genetically distinct individuals of a larger association population. SNPs offer highest resolution for mapping QTL and are potently in LD with the causative polymorphism they are the preferential candidate-gene variant to genotype in association studies.

**GENOME WIDE ASSOCIATION STUDIESPROCEDURES**

All qualitative and quantitave trait for crop improvement can be assess and identified their genetic determinants. Selection of a genotyping platform, sample size and collection, statistical analysis plans, statistical power, correction for multiple testing and population structure should be considered for study the GWAS. Association analysis involves assessment of linkage disequilibrium (LD) based association. A generalized scheme of genome wide association mapping for tagging gene of interest using germplasm accession. Recently genotype data for GWAS are usually produced with microarray technology allowing the detection of polymorphisms within a population. Correlation of microarray data with phenotypic trait and interpretation will describe the thinking idea for crop improvement. Several microarrays or “chips” commonly available for plant species. Some chips are designed to test as many SNPs as practically possible–currently up to about five million. Some chips are specifically designed to test SNPs in coding regions of genes, which make up about 2% of the genome. Other chips may test relatively small numbers of SNPs that have been carefully selected to efficiently represent worldwide haplotype diversity. Some chips are designed for specific ethnic groups or may be enriched with SNPs from genes implicated in particular diseases. In selecting a genotyping chip, it is important to consider the goals of the current project, compatibility with data from past or planned future studies, and the budget available.

**STATISTICAL APPROACHES USES FOR ASSOCIATION MAPPING**

Recent developments in statistical methodologies make it possible to properly interpret the results of association tests. Pritchard et al., (2000) have developed an approach that incorporates estimates of population structure directly into the association test statistic. The essential idea of this method is to decompose a sample drawn from a mixed population into several unstructured subpopulations and test the association in the homogeneous subpopulations. The methods have been applied to association analyses in humans and crop plants, with modified test statistics being used to deal with quantitative traits. LD between a single marker and a QTL can be measured by regression analysis, where the data on the trait is regressed on the individual marker genotypes, so that significant regressions will identify the markers associated with the phenotype. Nowadays several software use to assess the association of marker loci with traits.

The most commonly used statistics include logistic regression with the possibility of structured associations implemented in TASSEL General Linear Model (Yu and Buckler, 2006, TASSEL: http://www.maizegenetics.net), a multiple regression model combined with the estimates for the false discovery rate. Advantages The advantages of population based association mapping, utilizing a sample of individuals from the germplasm collections or a natural population, over traditional QTL-mapping in biparental crosses, primarily are due to availability of broader genetic variations with wider background for marker and trait correlations. The advantage of association mapping is that it can map quantitative traits with high resolution in a way that is statistically very powerful. The resolution of the mapping depends on the extent of LD, or non-random association of markers, that has occurred across the genome. Association mapping offers the opportunity to investigate diverse genetic material and potentially identify multiple alleles and mechanisms of underlying traits. It uses recombination events that have occurred over an extended period of time. Association mapping allows the possibility of exploiting historically measured trait data for association, and lastly has no need for the development of expensive and tedious biparental populations that makes approach timesaving and cost-effective. Limitations A major issue with association studies is a tendency to find false positives. Populations showing a desired trait also carry a specific gene variant not because the variant actually controls the trait, but due to genetic relatedness. In particular, indirect associations that are not causal will not be eliminated by increasing the sample size or the number of markers. The main sources of such false positives are linkage between causal and noncausal sites, more than one causal site and epistasis. These indirect associations are not randomly distributed throughout the genome and are less common than false positives arising from population structure. Likewise, population structure has always remained a consistent issue. Population structure leads to spurious associations between markers and the trait. This generally is not a problem in linkage analysis because researchers know the genetic structure of the family they created. But in association mapping, where relationships between diverse populations are not necessarily well understood, marker–trait associations arising from kinship and evolutionary history can easily be mistaken for causal ones. This can be accounted for with mixed models MLM. Also called the Q+K model, it was developed to further reduce the false positive rate by controlling for both population structure and cryptic familial relatedness.

**FINE MAPPING OF GENES/QTL**

Fine-mapping from the variant perspective Fine-mapping variants in GWAS loci require an understanding of the underlying mechanism by which a variant can contribute to a trait. Overcoming LD and identifying the context-specific variants that are causal to a trait is imperative for understanding disease mechanisms and confidently identifying which downstream genes and pathways are affected. Many functional and computational (high-throughput) fine-mapping methods have been developed and applied for this purpose.

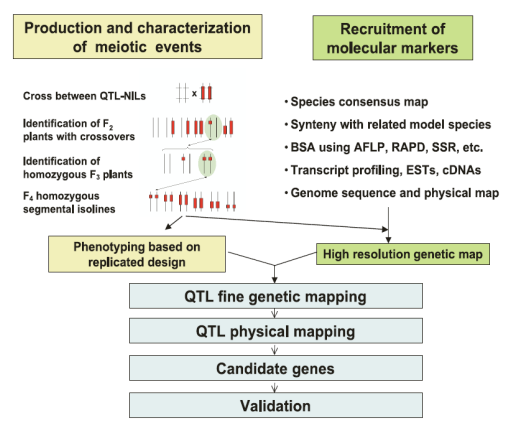
Below several fine-mapping methods according to their increasing ability to describe the complex role of variants in GWAS traits and diseases.

1. Identifying overlap with functional elements The most straightforward fine-mapping approach is to overlap GWAS variants in high LD with functional elements such as promoters and enhancers. Currently, the best resource for functional elements has been compiled by the NIH Roadmap Epigenomics Mapping Consortium , which used to measure histone marks to determine the location of functional elements in 127 different cell and tissue types. Fine-mapping of GWAS variants from 21 autoimmune diseases using the NIH Roadmap and similar data estimated that approximately 60% of candidate causal variants map to immune cell enhancers, and another approximately 8% to promoters. This was also reflected in the tissue-specific enrichment of type 1 diabetes susceptibility variants in lymphoid gene enhancers.

**Map based gene/QTL isolation and development of gene based**

Introduction Map-based positional cloning has been used to isolate a large number of genes that are responsible for Mendelian trait differences. The availability of numerous molecular markers, large-insert genomic libraries, and suitable mapping populations that segregate for traits of interest has made map-based cloning of genes with discrete effects theoretically straightforward, if laborious in practice, for a number of organisms. However, map-based cloning has been considered problematic for quantitative trait loci, or QTLs . In map-based cloning, the marker interval containing the gene is refined by genetic mapping to a region small enough to be physically mapped, by scoring the phenotypes of individuals that are recombinant between flanking markers. Additional markers in this interval are derived from the physical map and used to delimit the recombinational breakpoints. Physical contigs or even individual largeinsert clones spanning the smallest marker interval containing the gene are identified, and known or predicted candidate genes within the region are tested using complementation or other methods. The process is tractable even in organisms with relatively large genomes if the requisite resources are available. Quantitative trait variation, however, is controlled by an unknown number of loci as well as environmental factors, so the phenotypes of individual recombinants do not provide reliable genotypic information. Moreover, even if this problem can be overcome, there is no guarantee that an identified QTL represents a single gene. These complications have led to doubts about the utility of map-based cloning for QTLs, and even to suggestions that QTL analysis in general is an inferior tool for understanding the basis of genetic variation. Currently, QTL mapping is a well-established procedure in quantitative genetics. QTL mapping usually begins with the collection of genotypic (based on molecular markers) and phenotypic data from a segregating population, followed by a statistical analysis to reveal all possible marker loci where the allelic state correlates with the phenotype. Since this procedure only allows for an approximate mapping of the QTL, it is usually referred to as primary (or coarse) QTL mapping. More recent technical progress in the area of molecular biology and genomics have made possible the cloning of QTLs, i.e. the identification of the DNA sequences (coding or non-coding) responsible for QTLs. Herein, we present a critical review of the results obtained in QTL cloning in plants and discuss its perspectives, with emphasis on a number of major limitations and promising novel approaches.

The comparison of the number of research papers reporting new QTLs mapped in plants with the number of reports dealing with QTL cloning in plants clearly indicates that QTL cloning is still in its infancy. Positional cloning of QTLs Among the handful of QTLs isolated so far in plants, the majority have been cloned via positional cloning. Positional cloning implies a number of steps (Figure) that enable us, by creating a large mapping population and by use of molecular markers, to assign a QTL to the shortest possible genetic interval (QTL fine genetic mapping) and to identify the corresponding interval on the DNA sequence (QTL physical mapping). Candidate genes genetically and physically co- segregating with the QTL are then identified and/or selected for evaluation. The increase in mapping resolution required to accomplish QTL positional cloning is substantial, since after primary mapping a QTL is positioned within a chromosome interval of ca. 10-30 cM which usually includes several hundred genes. Eventually, independent proof is required to validate the role of the identified allelic polymorphism on the observed phenotypic effect.



**Figure-Flow chart depicting the major experimental steps involved in cloning a QTL using a map-based (positional cloning) approach.**

Genetic materials A widely-adopted strategy to more accurately estimate the position and effect of a oarsely mapped QTL is to create a new experimental population by crossing nearly isogenic lines differing only for the allelic constitution at the short chromosome segment (usually varying from ca. 10 to 30 cM in length) harbouring the QTL (QTL-NILs). The population should provide enough crossover events in order to genetically distinguish the gene responsible for the target QTL from other nearby genes. Factors like the physical vs. genetic distance ratio and gene density have to be duly considered when planning the size of the population. Figure 3 shows the number of chromosomes to be screened depending on the resolution needed in genetic terms. Such estimates should be considered with caution due to the variability of the kb/cM ratio on any plant chromosome. In a nearly isogenic population, due to the absence of other segregating QTLs, the target QTL becomes the only genetic source of variation and, accordingly to the heritability of the trait, a major source of the total phenotypic variation, thus enabling the phenotypic means of the QTL genotypic classes (+/+, - /- and, when present and in case of prevailing additive effects, -/+) to be statistically differentiated. Appropriate replication and/or progeny testing are generally implemented based upon the heritability of the trait considered. Under such conditions, the QTL is considered Mendelized (Alonso-Blanco and Koornneef 2000) and cM distances between a QTL and the nearby molecular markers can be more precisely estimated. Starting from the same population where primary mapping was carried out, QTL-NILs can be produced by (i) marker-assisted backcross introgression (i.e. substitution) of one QTL allele into one or both parental genetic backgrounds or (ii) iteratively identifying and selfing individuals which are heterozygous at the QTL region. QTL-NILs can also be efficiently identified within introgression libraries (ILs), i.e. collections of lines where each line is isogenic to a background parental line with the exception of a single short chromosome segment introgressed from a donor. Within an IL, the donor genome should be completely represented among the different IL lines so that a QTL-NIL will potentially exist for any segregating QTL. Remarkably, the same IL of the wild tomato Lycopersicon pennellii within the cultivated tomato genetic background provided the source of the QTL-NILs utilized for the cloning of three tomato QTLs. NILs suitable for positional cloning are also produced by the advanced backcross QTL analysis (ABQA) method, which combines backcrossing chromosome segments from a wild accession within an elite line with some level of phenotypic selection against extreme phenotypes. Critical aspects to be considered are the time and effort required for the development of collections of IL or ABQA lines as well as the limitation of the genetic variability provided by the use of only two parental lines: no matter how carefully the parental lines are chosen, the collections will only segregate for a fraction of the many more QTLs segregating for the same trait in other populations, which should be recognized. This important limitation can be partially overcome through the use of multiparental intercrossed populations. These populations are generated by crossing a carefully chosen set of parental lines capturing much of the genetic variation of the species, and then by performing several cycles of intermating. However, while this approach promises to increase the efficiency of QTL mapping both in terms of detection (segregation is expected at many loci) and genetic resolution (many rounds of meiosis), a recent report from the same authors highlighted unexpected analytical complexities. It should be noted that a substantial increase in genetic resolution can also been obtained by intercrossing standard biparental populations. During the fine mapping step, the resolution of the target QTL in two or more linked loci may bring positional cloning projects to an end when the proportion of phenotypic variability explained by each QTL is too small to be revealed with a realistically manageable number of replications. QTL clusters have indeed been observed in plants. On the other hand, cloning was accomplished when one of the linked QTLs retained most of the effect. Genotyping The recruitment of polymorphic markers required for the fine mapping of a QTL is rather simple for Arabidopsis and rice, where the whole genome has been sequenced, and for species such as maize or tomato for which genomic sequencing is under way or information is available in terms of sequenced ESTs. However, in species for which detailed sequence information is not available or cannot be deduced from syntenic relatives, a large number of molecular markers (e.g. AFLPs) need to be screened in genotypes contrasted at the target region (e.g. pair of QTL-NILs). As to genotyping techniques, microarray-based platforms appear particularly promising for a high-throughput identification of polymorphisms (SNPs and insertion-deletions) at thousands of loci. Physical mapping and candidate sequences When the genetic resolution approaches the cM level, the markers closest to the QTL are used for anchoring the genetic map to the physical map, i.e. the genomic sequence or a BAC (bacterial artificial chromosome) contig covering the QTL region. An early transfer of the information to the physical map allows for the efficient generation of new single-copy markers useful for refining the genetic mapping and searching candidate genes. Even if only a BAC contig is available, sequenced BAC ends can often be transformed in genetic markers and low-pass, shot-gun sequencing can provide a glimpse of local gene content. In this phase, bioinformatics provides an important contribution in terms of gene prediction and annotation, and exploitation of syntenic relationships. Validation of a candidate sequence The functional testing of a candidate gene/s can be performed by overexpressing or down-regulating the target gene through genetic engineering or RNAi, by genetic complementation of a known mutant or by rescuing and phenotypically and molecularly characterizing mutants at the candidate gene. If available within the species under investigation, reverse genetics tools such as TDNA or transposon-tagged populations and/or TILLING can also be exploited. As compared to transposon tagging, TILLING and RNAi are appealing alternatives for their almost universal applicability and for providing subtle changes of gene functionality comparable to those observed naturally. Gene replacement, still in its infancy but already reported in rice can be considered the ultimate tool for validating candidate genes. Cloning QTLs by association mapping As an alternative to positional cloning, QTLs can be molecularly resolved through association mapping, i.e. by identifying, within a set of genotypes (e.g. germplasm accessions, cultivated varieties, etc.), a statistical association between allelic variants at marker or candidate loci and the mean of the analyzed trait.

The analysis evaluates the trait mean change caused by the substitution of one allele with another.

For QTL cloning in plants, the interest lies in

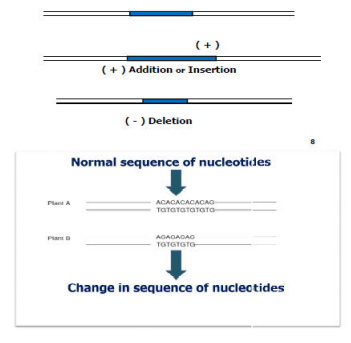
1. the possibility of finding chromosome regions important for controlling quantitative traits without the costly and time-consuming production of large experimental populations,
2. the potentially high genetic resolution provided by the many meiotic events which occurred during past generations and
3. the possibility of surveying a large number of functionally diverse alleles per locus. A major factor to be considered in association mapping is the level of linkage disequilibrium (LD; level of non-random assortment of alleles at different loci) among the tested accessions. In plants, the extensive LD analyses conducted in Arabidopsis and maize (Flint-Garcia et al. 2003) have indicated that while LD persists over hundreds of kb in Arabidopsis, in maize LD decays after a few kb, although it can extend significantly farther in collections of elite germplasm. Functional genomics and QTL cloning Functional genomics is contributing to many aspects of QTL analysis and cloning. Transcriptional profiling between contrasting QTL genotypes can quickly provide a list of genes differentially expressed; subsequently, those genes functionally related to the target trait and mapping at the QTL region can be selected as candidates. Unfortunately, the number of QTLs cloned so far in plants is too small to test the validity of this approach. Indeed, when the QTL was shown to involve a difference in gene expression level between alleles, those differences were either too low (ca. two-fold) or showed too strong of a spatial and/or temporal pattern to allow for their identification with a standard microarray-based transcriptome analysis. Other profiling platforms, such as SAGE (serial analysis of gene expression; and MPSS (massively parallel signature sequencing; are better suited to detect subtle differences in gene expression. Transcript profiling can reach the sub-tissue level of resolution if carried out in combination with lasercapture microscopy.

**Allele mining by TILLING and Eco-TILLING; Use of markers in plant breeding**

**Introduction**

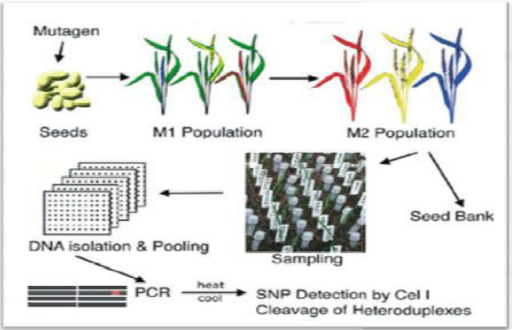
Progress in plant breeding in terms of development of superior and high yielding varieties of agricultural crops is possible by accumulation of beneficial alleles from vast plant genetic resources existing worldwide. But still, a significant portion of these superior alleles cannot be used, because those alleles are left behind during evolution and domestication. This genetic variation existing in wild relatives and land races of crop plants and it can be used for development of agronomically superior cultivars. Introgressions of novel alleles occur from wild crop plants into cultivated varieties. It clearly demonstrated that certain alleles and their combinations potentially make dramatic changes in character expression when it moved to a suitable genetic background. Hence, the vast germplasm resources need to for novel alleles to further enhance the genetic potential of crop varieties for various agronomic traits. Enormous progress has been made in the last 15 years in depositing an amount of sequence information into GenBank. With rapid accumulation of sequence and expression data in various genomic databases accelerate discovery of new genes. It is easy to development of allelespecific markers. Based on gene and genome sequences, Polymerase Chain Reaction (PCR) strategies are use to isolate Superior alleles of genes from a wide range of species. This key alleles imparting resistance to biotic and abiotic stresses, greater nutrient use efficiency enhanced yield and improved quality. Using novel genomic tools, similar alleles responsible for a given trait and their variants in other genotypes can be identified. This is often referred to as ‘dissection of naturally occurring variation at candidate genes/loci’ or simply ‘allele mining’. Identification of allelic variants from germplasm collections not only provides new germplasm for delivering novel alleles to targeted trait improvement but also categorizes the germplasm entries for their conservation. Evolution of new alleles It is alternative forms of gene. In other words, genes occur at similar loci of homologous chromosomes. evolutionary driving force which causes existing allelic diversity in any crop species. For creation of new alleles or causing variations in the existing allele and allelic combinations, generally mutations will occur in the genic regions of the genome either as Single Nucleotide Polymorphism (SNP) or as Insertion and Deletion important. The mutations in coding regions and/or re regions may have tremendous effect on the phenotype by changing the encoded protein structure and/or function while those that occur in noncoding regions of a gene. It has been silent without any effect on the phenotype. Even though most of the mutations are deleterious, in genera l0.1% of the mutations are vital leading to changes will occur in gene function which may be highly necessary for the survival of the plant.

a) How mutation does these variations occur?



**Figure- Insertion and deletion in nucleotide sequence**

Mutation is defined as sudden heritable changes occur in DNA segment of individuals. This change will occur in normal sequence of nucleotides in individuals. For example, the long growth habit of wild rice is controlled by a single gene PROG1. The wild type allele is replaced by mutant allele (prog1) in most of the Oryza sativa cultivars, which inactivate the PROG1, gene function and grain number and higher grain yield in cultivated rice. This phenomenon is observed as the wild rice has an allele their expression or its effect inactivate in cultivated rice due to mutation. In other words change will occur in nucleotides. b) How new alleles are formed? Mutations will happen in both genomic region such as coding region as well as non-coding regions of gene. That creates heritable change in individuals. What is allele mining? Mining is extraction of any non resources such as Petroleum, natural gas, minerals, or even water etc. It is a finding of superior allele from the natural population. In other words introgression of novel or superior allele from wild relatives into cultivated one. Approaches for allele mining There are two major approaches. They are: TILLING -based allele mining: It is nothing but a (Targeting Induced Local Lesions In Genomes), to determine variation in individual through artificially changed mutation 2004). It is a powerful reverse genetics tool for functional genomics where knockout methodologies cannot be applied. Tilling allows the identification of allelic variation of target gene in a high-throughput manner. The use of the Tilling technique to survey natural variation in genes is called Ecotilling. Tilling make use of chemical mutagens to introduce random mutation. Seeds are mutagenized with EMS, which causes G/C-to-A/T point mutations (Nagy, seeds are selfed to produce M2 seeds. M2 progeny from single seed descent are used for screening. For screening, DNAs are pooled eightfold to maximize the efficiency of mutation detection. PCR is performed using 5’-end specific primers to target the desired locus, a are formed by heating and cooling the PCR products. CEL I nuclease is used to cleave at base mismatches, and products representing induced mutations are visualized with denaturing polyacrylamide gel electrophoresis. The detailed procedure has been given by Till, et al. (2003).



**Figure-TILLING based allele mining**

Eco-TILLING EcoTILLING is a molecular technique that is similar to TILLING, except that its objective is to uncover natural ge-netic variation as opposed to induced mutations Many species are not amenable to chemical mutagenesis; therefore, EcoTILLING can aid in the discovery of natural variants and their putative gene function. This approach allows one to rapidly screen through many samples with a gene of interest to identify naturally occurring SNPs and / or small INDELS. The method has proven to be successful to detect DNA polymorphisms including variations in satellite repeat number Furthermore, in highly heterozygous outcrossing species, EcoTILLING can be used to determine heterozygosity levels within a gene fragment. EcoTIL-LIING reduces the time and effort for SNP discovery gener-ally required by weeding out identical haplotypes. Therefore, this method does not require one to sequence all individuals in a population to identify polymorphisms, which can be a burdensome expense and time consuming. It also has the advantage of detecting multiple polymorphisms in a single fragment because CEL I will digest only a small proportion of the heteroduplexes at a single position Advantages and disadvantages of tilling and Eco-tilling TILLING is a non-transgenic, high throughput reverse genetic approach. This technique unlike other SNP detection methods, provides the approximate location within a few base pairs of the induced mutation, which allows targeted sequencing in the area of the induced mutation opposed to sequencing the entire fragment. Since chemical mutagensis produces a range of various mutations throughout the genome such as nonsense, splice site, and missense, all of which potentially can affect the protein structure and the resulting phenotype, it has been used for decades to obtain mutants for genetic studies. Therefore, through mutagensis one can obtain partial loss or complete loss of function and new novel functions, which can provide valuable insight into the true role of a gene in a species of interest. As discussed previously, using chemical mutagensis and TILLING to pinpoint these mutations has been highly effective in the elucidation of gene function in plants and animals without the production of transgenic material. TILLING has been demonstrated to be sensitive enough to detect induced mutations and naturally occurring SNPs as well as the detection of heterozygotes.

EcoTILLING, which has been less frequently employed in the current literature, can also be a valuable tool for mining for SNPs in germplasm, assessing heterozygosity, uncovering variants for disease resistance, or ascertaining the function of a gene or regulatory element by detecting natural variants. EcoTILLING can be a good technique to employ especially when working with a well established population with thoroughly characterized morphological data. Use of markers in plant breeding When using molecular markers to study the genetics of a particular crop, it must be remembered that markers have restrictions. It should first be assessed what the genetic variability is within the organism being studied. Analyze how identifiable particular genomic sequence, near or in candidate genes. Maps can be created to determine distances between genes and differentiation between species. Genetic markers can aid in the development of new novel traits that can be put into mass production. These novel traits can be identified using molecular markers and maps. Particular traits such as color, may be controlled by just a few genes. Qualitative traits (requires less than 2 genes) such as color, can be identified using MAS (marker assisted selection). Once a desired marker is found, it is able to be followed within different filial generations. An identifiable marker may help follow particular traits of interest when crossing between different genus or species, with the hopes of transferring particular traits to offspring. One example of using molecular markers in identifying a particular trait within a plant is, Fusarium head blight in wheat. Fusarium head blight can be a devastating disease in cereal crops but certain varieties or offspring or varieties may be resistant to the disease. This resistance is inferred by a particular gene that can be followed using MAS (Marker Assisted Selection) and QTL (Quantitative Trait Loci). QTL's identify particular variants within phenotypes or traits and typically identify where the GOI (Gene of interest) is located. Once the cross has been made, sampling of offspring may be taken and evaluated to determine which offspring inherited the traits and which offspring did not. This type of selection is becoming more beneficial to breeders and farmers because it is reducing the amount of pesticides, fungicides and insecticides. Another way to insert a GOI is through mechanical or bacterial transmission. This is more difficult but may save time and money.

Applications of markers in cereal breeding

1. Assessing variability of genetic differences and characteristics within a species.

2. Identification and fingerprinting of genotypes.

3. Estimating distances between species and offspring.

4. Identifying location of QTL's.

5. Identification of DNA sequence from useful candidate genes

**Application**

applications in fisheries and aquaculture:

1. Species Identification

2. Genetic variation and population structure study in natural populations

3. Comparison between wild and hatchery populations

4. Assessment of demographic bottleneck in natural population

5. markers assisted breeding

**Marker assisted selection (MAS) in backcross and heterosis breeding**

**Introduction**

Marker assisted selection or marker aided selection (MAS) is an indirect selection process where a trait of interest is selected based on a marker (morphological, biochemical or DNA/RNA variation) linked to a trait of interest (e.g. productivity, disease resistance, abiotic stress tolerance, and quality), rather than on the trait itself. This process has been extensively researched and proposed for plant and animal breeding, nevertheless, as of 2013 "breeding programs based on DNA markers for improving quantitative traits in plants are rare". For example, using MAS to select individuals with disease resistance involves identifying a marker allele that is linked with disease resistance rather than the level of disease resistance. The assumption is that the marker associates at high frequency with the gene or quantitative trait locus (QTL) of interest, due to genetic linkage (close proximity, on the chromosome, of the marker locus and the disease resistance-determining locus). MAS can be useful to select for traits that are difficult or expensive to measure, exhibit low heritability and/or are expressed late in development. At certain points in the breeding process the specimens are examined to ensure that they express the desired trait. Features of Marker Assisted Selection (MAS): The main features of MAS are briefly presented below: Marker assisted selection (MAS) is also termed as marker aided selection and marker assisted breeding (MAB).

It differs from gene assisted selection (GAS) which refers to the selection which is based on QTLs (quantitative trait locus or loci). ii. Pre-Requisites: There are two pre-requisites for marker assisted selection.

These are:

1. a tight linkage between molecular marker and gene of interest, and
2. high heritability of the gene of interest.

Application:

MAS is applicable for genetic improvement of plants as well as animals. In plants, it is equally applicable in both self-pollinated and cross pollinated species.

iv. Markers Used: MAS makes use of various types of molecular markers. The most commonly used molecular markers include amplified fragment length polymorphisms (AFLP), restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (RAPD), simple sequence repeats (SSR) or micro satellites, single nucleotide polymorphisms (SNP), etc. The use of molecular markers differs from species to species also.

v. Efficiency: The relative efficiency of MAS is greatest for characters with low heritability, if a large fraction of the additive genetic variance is associated with the marker loci.

In other words, MAS is useful when the heritability of the trait is low. Moreover, MAS is more efficient than purely phenotypic selection in quite large populations. It has been found by some workers that MAS may become less efficient than phenotypic selection in the long term. This is because the rate of fixation of unfavourable alleles at QTLs with small effects is higher under MAS than under phenotypic selection. It may be a consequence of the strong selection applied to QTLs with large effects under MAS in early generation. However, such problem comes after a long period.

vi. Accuracy: Molecular markers have very high accuracy. They are not affected by environmental conditions. MAS is a new breeding tool which is available to make more accurate and useful selections in breeding populations. MAS allows heritable traits to be linked to the DNA which is responsible for controlling that trait.

vii. Speed of Progress: MAS is a rapid method of crop improvement. For example, in conventional breeding when we transfer a recessive character through backcross, one selfing is required after every backcross for identification of recessive character. MAS permits identification of recessive alleles even in heterozygous condition and thus speeds up the progress of crop improvement work.

viii. Traits Improved: MAS can be used for improvement of both oligogenic and polygenic traits. In the past, MAS has been mostly used for the genetic improvement of oligogenic traits and little progress has been made with polygenic traits.

ix. Material Developed: MAS leads to development of non-transgenic genotypes or cultivars. In other words, MAS is used for development of non-transgenic cultivars. The transgenic cultivars face public resistance. On the other hand, cultivars developed by MAS are acceptable by consumers.

x. Cost: MAS is very costly as compared to phenotypic selection. In MAS, the costly items include equipment’s, consumables, infrastructure, labour and DNA extraction process.

MAS requires sophisticated and well equipped laboratory.

3. **Steps in Marker Assisted Selection (MAS):**

In the marker aided selection, RFLP markers are widely used for genetic improvement of crop plants for various economic characters.

The marker aided selection consists of five important steps, viz:

1. Selection of parents,
2. Development of breeding population,
3. Isolation of DNA from each plant,
4. Scoring RFLPs,
5. Correlation with morphological traits.

These are briefly discussed below:

1. Selection of Parents:

Selection of suitable parents is an important step in marker aided selection. The parents should be such so that we can get usable level of polymorphism (variation) in the RFLP markers. In other words, parents with contrasting characters or divergent origin should be chosen. This will help in identification of DNA of both the parents and also their segments in F2 generation in various recombination’s. For selection of parents, we have to screen germplasm and select parents with distinct DNA. The parents that are used for MAS should be pure (homozygous). In self- pollinated species, plants are usually homozygous. In cross-pollinated species, inbred lines are used as parents.

1. Development of Breeding Populations:

This is the second important step for application of marker aided selection. The selected parents are crossed to obtain F1 plants. F1 plants between two pure-lines or inbred lines are homogeneous (alike phenotypically) but are heterozygous for all the RFLPs of two parents involved in the F1. The F2 progeny is required for the study of segregation pattern of RFLPs. Generally 50-100 F2 plants are sufficient for the study of segregation of RFLP markers.

1. Isolation of DNA:

The third important step is isolation of DNA from breeding population. The main advantage of MAS is that DNA can be isolated even from the seedlings and we need not to wait for flowering or seed development stage. The DNA is isolated from each plant of F2 population. Standard procedures are available for DNA isolation. The isolated DNA is digested with specific restriction enzyme to obtain fragments of DNA. The DNA fragments of different sizes are separated by subjecting the digested DNA to agarose gel electrophoresis. The gel is stained with ethidium bromide and the variation in DNA fragments can be viewed in the ultraviolet light. The DNA of chloroplasts, when digested with specific enzyme, produces about 40 fragments of different sizes. The nuclear DNA of higher plants, when digested with specific restriction enzymes, produces millions of fragments in a continuous range of sizes. It is a tedious job to identify individual DNA fragment in such cases.

1. Scoring RFLPs:

The polymorphism in RFLPs between the parents and their involvement in the recombinants in F2 population is determined by using DNA probes. The labelled probes are used to find out the fragments having similarity. The probe will hybridize only with those segments which are complementary in nature. Generally 32P is used for radioactive labelling of DNA probe. Now nonradioactive probe labelling techniques are also available. In this way RFLPs are determined.

1. Correlation with Morphological Traits:

The DNA marker (say RFLPs) are correlated with morphological markers and the indirect selection through molecular markers is confirmed. Once the correlation of molecular markers is established with morphological markers, MAS can be effectively used for genetic improvement of various economic traits. 4. Applications of Marker Assisted Selection (MAS): In crop improvement programmes MAS can be used in various ways.

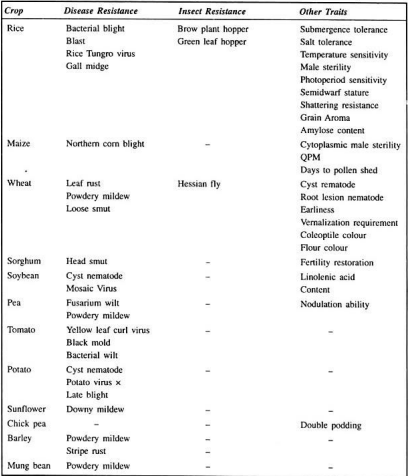
In other words, MAS has several useful applications in plant breeding. Important applications of MAS in plant breeding are briefly presented below:

1. MAS is very effective, efficient and rapid method of transferring resistance to biotic and abiotic stresses in crop plants.
2. It is useful in gene pyramiding for disease and insect resistance.
3. It is being used for transfer of male sterility and photo period insensitivity into cultivated genotypes from different sources.
4. MAS is being used for improvement of quality characters in different crops such as for protein quality in maize, fatty acid (linolenic acid) content in soybean and storage quality in vegetables and fruit crops.
5. MAS can be successfully used for transferring desirable transgene (such as Bt gene) from one cultivar to another.
6. MAS is very effective in introgression of desirable genes from wild into cultivated genotypes. vii. MAS is equally effective in genetic improvement of plants and animals.
7. MAS is useful in genetic improvement of tree species where fruiting takes very long time (say 20 years) because for application of phenotypic selection we have to wait for such a long time.
8. MAS has wide application for genetic improvement of oligogenic traits as compared to polygenic traits.

5. **Achievements of Marker Assisted Selection (MAS):**

MAS has been used for genetic improvement of different field crops such as maize, barley, rice, wheat, sorghum, soybean, chickpea, pea, sunflower, tomato, potato and some fruit crops for various economic characters. MAS has been mainly used for developing disease resistant cultivars in different crops (Table). Some notable examples of the use of MAS are given below:

1. Rice: In rice MAS has been successfully used for developing cultivars resistant to bacterial blight and blast. For bacterial blight resistance four genes (Xa4, Xa5, Xa13 and Xa21) have been pyramided using STS (sequence tagged site) markers. The pyramided lines showed higher level of resistance to bacterial blight pathogen. In Indonesia, two bacterial blight resistant varieties of rice viz Angke and Conde have been released through MAS. For blast resistance, three genes (Pil, Piz5 and Pita) have been pyramided in a susceptible rice variety Co 39 using RFLP and PCR based markers.
2. Maize: In maize, normal lines have been converted into quality protein maize (QPM) lines through MAS using opaque 2 recessive allele. This work has been done at CIMMYT (international centre for wheat and maize improvement, Mexico). Three SSR markers (Umc 1066, Phi 057 and Phi 112) present within opaque 2 gene have been used for this purpose. The MAS used for conversion of normal maize lines into QPM is simple, rapid and accurate.



**Table-Characters being used for MAS in different crops**

1. Soybean: In soybean cyst nematodes pose serious problem and most of the varieties are susceptible to this parasite. The resistant gene (rhg 1) is available. In soybean, nematode resistant lines have been developed through MAS using SSR marker (Sat 309). MAS has been used for genetic improvement of various characters in different crops. Important characters which have been improved through MAS in different crops include disease resistance, insect resistance, salinity resistance, shattering resistance. It has also been used for transfer of various characters such as male sterility, photoperiod insensitivity, earliness, and improvement of protein contents in some crop plants (Table 34.2).

In MAS various types of DNA markers have been used in different crop plants. Molecular markers which have been widely used in MAS in different crops include Restriction fragment length polymorphisms (RFLP), random amplified polymorphic DNA (FAPD) and simple sequence repeats (SSR) or microsatellite. Other markers which have been used in some crops include amplified fragment length polymorphism (AFLP), sequence tagged site (STS), expressed sequence tags (EST), and SCAR markers. Single nucleotide polymorphism (SNP) is also used. SNPs have been identified for all the major cereal crops. MAS is also being used for genetic improvement of forage and fruit crops. Among fruit crops MAS has been extensively used in pomegranate, apple and pear. In fruit crops MAS is based on RFLP, RAPD, SSR and AFLP markers. In these fruit crops MAS is being used for higher fruit production, better keeping quality for storage and disease resistance. In vegetable crops MAS is being used in tomato and potato based on RFLP, RAPD and AFLP markers mainly for disease resistance. MAS has been found useful for genetic improvement of tree crops such as coconut and rubber. In majority of field crops genes and linked markers for various important traits have been identified which are being used for MAS.

6. **Advantages of Marker Assisted Selection (MAS):**

MAS has several advantages over phenotypic selection and other breeding techniques. Some important advantages of MAS are briefly discussed below:

1. Accuracy: The accuracy of MAS, is very high because molecular markers are not affected by environmental conditions. It is very effective even with the characters having low heritability.
2. Rapid Method: MAS is a rapid method of crop improvement. It takes 3-5 years for developing a new cultivar against 10-15 years taken by the conventional method of breeding.
3. Non-transgenic Product: MAS leads to development of non-transgenic cultivars which are acceptable to everybody. In other words, it does not involve transgene. Hence there is no question of gene silencing.
4. Identification of Recessive Alleles: MAS permits identification of recessive alleles even in heterozygous condition and thus speeds up the progress of crop improvement programmes. In other words, it is equally effective for the genetic improvement of recessive characters.
5. Early Detection of Traits: MAS permits early detection of traits that are expressed late in the life of plant. For example characters such as grain or fruit quality, flower colour, male sterility, photoperiod sensitivity that express late in the life of a plant can be screened in the seedling stage. In other words, DNA tested at seedling stage can through light about the trait which are expressed later on.
6. Screening of Difficult Traits: MAS permits screening traits that are extremely difficult expressive and time consuming to score phenotypically.

For example, screening for traits such as root morphology and resistance to biotic (insects and diseases) and abiotic stresses (drought, salinity, heat, frost etc.) is very easy through MAS.

1. Gene Pyramiding: MAS is very effective method in accumulating multiple genes for resistance to specific pathogens and pests within the same cultivar. This process is called gene pyramiding. Maker assisted backcrossing is routinely applied in breeding programmes for gene introgression. MAS can provide an effective and efficient breeding tool for detecting, tracking, retaining, combining and pyramiding genes for disease resistance.
2. Small Sample for Testing: MAS requires only a small amount of plant tissue for DNA testing. In other words, MAS can be carried out with small breeding populations. Moreover, MAS can be applied at any stage of plant growth.
3. Permits QTL Mapping: MAS permits mapping or tagging of quantitative trait loci (QTL) which is not possible by conventional method.
4. Highly Reproducible: The MAS is based on DNA fingerprinting technique and the results of DNA fingerprinting pattern are highly reliable and reproducible.

**Limitations of Marker Assisted Selection (MAS):**

However, it has some limitations or drawbacks which are briefly below:

1. MAS is a costly method. It requires well equipped laboratory viz. expensive equipment’s, glassware and chemicals.
2. MAS requires well trained manpower for handling of sophisticated equipments, isolation of DNA molecule and study of DNA markers.
3. The detection of various linked DNA markers (AFLP, RFLP, RAPD, SSR, SNP etc.) is a difficult, laborious and time consuming task.
4. MAS sometimes involves use of radioactive isotopes in labelling of DNA, which may lead to serious health hazards.

This is a major disadvantage of RFLP based markers. The PCR, markers are safe in this regard.

1. It has been reported that MAS may become less efficient than phenotypic selection in the long term.
2. The use of MAS is more difficult for QTL because they have minor cumulative effects and are greatly influenced by environmental conditions and genetic background.

**MARKER-ASSISTED BACKCROSSING (MABC)**

MABC procedure and theoretical and practical considerations Marker-assisted or marker-based backcrossing (MABC) is regarded as the simplest form of marker-assisted selection, and at the present it is the most widely and successfully used method in practical molecular breeding. MABC aims to transfer one or a few genes/QTLs of interest from one genetic source (serving as the donor parent and maybe inferior agronomically or not good enough in comprehensive performance in many cases) into a superior cultivar or elite breeding line (serving as the recurrent parent) to improve the targeted trait. Unlike traditional backcrossing, MABC is based on the alleles of markers associated with or linked to gene(s)/QTL(s) of interest instead of phenotypic performance of target trait. The general procedure of MABC is as follow, regardless of dominant or recessive nature of the target trait in inheritance: Select parents and make the cross, one parent is superior in comprehensive performance and serves as recurrent parent (RP), and the other one used as donor parent (DP) should possess the desired trait and the DNA markers allele(s) associated with or linked to the gene for the trait. Plant F1 population and detect the presence of the marker allele(s) at early stages of growth to eliminate false hybrids, and cross the true F1 plants back to the RP. Plant BCF1 population, screen individuals for the marker(s) at early growth stages, and cross the individuals carrying the desired marker allele(s) (in heterozygous status) back to the RP. Repeat this step in subsequent seasons for two to four generations, depending upon the practical requirements and operation situations as discussed below. Plant the final backcrossing population (e.g. BC4F1), and screen individual plants with the marker(s) for the target trait and discard the individuals carrying homozygous markers alleles from the RP. Have the individuals with required marker allele(s) selfed and harvest them. Plant the progenies of backcrossing-selfing (e.g. BC4F2), detect the markers and harvest individuals carrying homozygous DP marker allele(s) of target trait for further evaluation and release.

Application of MABC Success in integrating MABC as a breeding approach lies in identifying situations in which markers offer noticeable advantages over conventional backcrossing or valuable complements to conventional breeding effort. MABC is essential and advantageous when:

• Phenotyping is difficult and/or expensive or impossible;

• Heritability of the target trait is low;

• The trait is expressed in late stages of plant development and growth, such as flowers, fruits, seeds, etc.;

• The traits are controlled by genes that require special conditions to express;

• The traits are controlled by recessive genes; and

• Gene pyramiding is needed for one or more traits.

Among the molecular breeding methods, MABC has been most widely and successfully used in plant breeding up to date. It has been applied to different types of traits (e.g. disease/pest resistance, drought tolerance and quality) in many species, e.g. rice, wheat, maize, barley, pear millet, soybean, tomato, etc.

**Transgenic breeding; Foreground and background selection**

**Introduction**

Agricultural production has increased worldwide through the development of varieties with improved yield traits or stress-tolerant characteristics. To this end, crop breeding programs are devised to accumulate alleles from within the gene pool and from cross-compatible wild relatives. Advances in genetic engineering through recombinant DNA technology and spatial and temporal targeted expression of genes facilitated

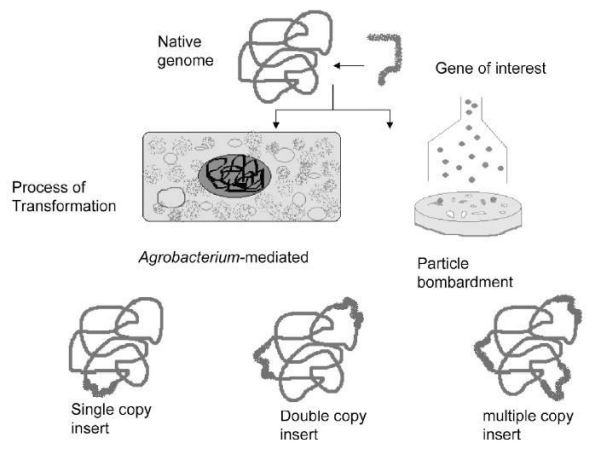
1. The transfer of precise gene sequences and
2. The transfer of genes across gene pools.

Asexual techniques of gene transfer help to engineer new characters that are otherwise very difficult to introduce by conventional breeding. Plants that receive genes using recombinant DNA technology are called genetically modified (GM) crops. A number of GM varieties and hybrids are developed and cultivated globally. Stable crop systems across broad agroecosystems under a variety of stresses can be addressed by GM crops in conjugation with plant breeding programs. Genetically modified crop varieties can benefit large acreage cropping system in the developed world as well as small farms in the developing world with the support and implementation of complementary national biotechnology policies that safeguard humankind, national trade, and the environment. Transfer of genes by crossing involves recombination of genetic material, while genetic transformation allows the transfer of only the desired and specific gene sequences. Development of a new cultivar is a sequential and cyclic process of creating new diversity, recombination, selection for superior recombinants, and testing followed by commercialization. Through transformation, a new genetic variation is created, but it has to be performance-tested like its classically bred counterparts before release and commercialization. In transgenic breeding, the breeder looks for a defined trait phenotype and then introgresses the transgenic event into a broad range of desirable genetic backgrounds. But in the case of transgenics, the site of integration of the transgene is usually random, sometimes resulting in alteration of the DNA sequence to be inserted and sometimes disrupting the recipient genome.

**PRODUCTION AND EVALUATION OF TRANSGENIC EVENTS**

Insertion of alleles through genetic transformation is a long, technically specialized, and expensive process. Genetic transformation protocols, in general, are dependent on tissue culture (callus induction and regeneration) response of a species. Transformation methods based on in planta avoid tissue culture. During the course of genetic transformation a foreign gene is introduced through particle bombardment or Agrobacterium method into a small plant tissue, which is competent to regenerate into a whole plant (Figure). Transgene refers to a set of nucleotide sequences that constitute a promoter, coding sequence of the gene, enhancer, and terminator. The promoter can be selected based on the selective spatial and temporal expression of the transgene. Levels of transgene expression are enhanced by inclusion of enhancer sequences. The desired gene is inserted randomly into a few cells of the target tissue; these transformed cells are selectively cultured to regenerate whole plants over time under controlled nourishment and culture. These transgenic plants (T0) are confirmed for the integration of the desired gene(s) through techniques like polymerase chain reaction (PCR) and Southern blotting. Expression of target character is verified through Western blotting, whereas quantification of the levels of trait expression is performed through ELISA (enzyme-linked immunosorbent assay) or related techniques. Integration of the transgene at a given locus or loci is called a transgenic event.

**Process of transgenic development**



**ADVANCEMENT OF GENERATION IN TRANSGENIC LINES**

Transgenic (T0) plants generated immediately after the process of genetic transformation are hemizygous for the gene of interest. Homozygous lines in crop species can be established by

1. Germination of T1 transgenic progeny seeds on selective media,
2. Evaluation for expression of transgenic trait (if diff erential expression between homozygotes and heterozygotes exists),
3. Basta paint tests, and
4. Real-time PCR (RT-PCR).

To distinguish between heterozygotes and homozygotes, RT-PCR can be used, but it is costly and requires specialized skill and equipment. Germination of seeds on media containing selection agents like antibiotics (hygromycin or kanamycin) or herbicides like Basta (glufosinate ammonium) can be used in a few cases to obtain homozygous lines. Instant diploidization of heterozygous lines is possible by regenerating plants through anther culture.

**LIMITATIONS FOR DIRECT ADOPTION OF TRANSGENIC PLANTS**

Transgenic plants cannot be brought to cultivation immediately after production because

1. Not all the cultivated or elite lines of a crop species are amenable to genetic transformation and
2. Pprolonged process of genetic transformation results in genetic drag of the initial variety. One of the challenges of crop improvement by genetic modification is the presence of unintended effects. Lack of understanding on such practical aspects becomes a significant barrier in transgenic breeding.

**FOREGROUND SELECTION**

Foreground selection refers to using markers that are tightly linked to the gene of interest in order to select for the target allele or gene. Background selection refers to using markers that are not tightly linked to the gene of interest in order to select against other DNA from the donor parent (i.e., to select for recurrent parent alleles at other loci than the target). Strategies for using markers for foreground selection The first step for foreground selection is to associate a molecular marker with the target trait by some genetic mapping method. In the best case, the marker itself would be the functional polymorphism – that is, the DNA change that causes the phenotypic difference between alleles. These markers can be called “functional markers”. Creating functional markers requires the gene of interest to be cloned. In the absence of a cloned gene, markers that are very tightly linked to the target gene are necessary to avoid recombination between the marker and the gene during backcrossing. A recombination, of course, would result in a situation where the breeder is selecting for a marker allele that is now linked to the undesirable trait allele. Assuming markers are available for the trait, then they can be applied at each backcross generation to select those plants carrying the desired allele (or gene, in the case of a novel transgene).

Markers are used as follows:

(1) BCxF1 seeds are planted and DNA is extracted from leaf tissue prior to flowering.

(2) The markers are screened on the DNA of each plant, typically requiring a separate reaction assay for each marker on each plant.

(3) Plants that have the desired marker allele – the one known to be associated with the desired trait allele – are selected for backcrossing to the recurrent parent.

A clear summary of an example of foreground selection in rice is provided. Their objective was to introgress three different bacterial leaf blight resistance alleles (each at a different chromosomal location) into elite “new plant type” rice breeding lines using marker-assisted selection (MAS). The donor parent was IRBB59, which carried all three resistance alleles, Xa21, xa13, and xa5, but was not a NPT line. The recurrent parents were IR65598-112, IR65600-42, and IR65600-96, all NPT lines.

**BACKGROUND SELECTION**

In background selection is designed to eliminate donor parent alleles other than the desired target gene. Of most importance is minimizing the size of the introgressed region, ideally limiting it to just the gene of interest. Also important, however, is the elimination of donor chromatin in regions of the genome unlinked to the target (i.e., non-carrier chromosomes and regions far from the target on the carrier chromosome). Strategies for using markers for background selection To remove unlinked donor genome regions, assaying 2-3 markers evenly spaced per 100 cM is the most efficient approach in early generations, but higher marker density may be useful in later backcross generations. A program including only a selection step on non-carrier chromosomes is more effective in later backcross generations (BC3 or BC4) than in early generations. Although this may be counter intuitive, it is based on the facts that in early generations, few recombinations have occurred to break up introgressed segments from the donor – thus, few markers are needed, and that most of the donor material will be removed in subsequent backcrosses. In later generations, the donor regions will be smaller, and hence more markers will be needed.

Selection against donor alleles in the region of the target – i.e., reducing the linkage block surrounding the target. The goal is to reduce the size of the donor DNA segment surrounding the target gene to as small as possible as soon as possible in the backcrossing program. To do this, recombinants between the target and closely linked markers need to be identified. Actually, individuals with a recombination on both sides of the target need to be identified (as shown in the figure below). Marker assisted selection is more efficient relative to traditional backcrossing for markers close to the target. If recombination is equally likely throughout a chromosome (which we know isn’t actually true), then the likelihood of a recombination between the target gene and a marker is proportional to the distance the marker is from the target. For markers far from the target, recombination is likely to occur early in the backcrossing process, and hence, the use of markers to select for these regions which are homozygous for the recurrent parent is not particularly necessary. However, as the marker gets closer to the target, the use of markers helps identify the rare recombinants. The problem is that few recombinations occur between the trait and markers close to the target, so that large populations need to be screened to find recombinants here. A double recombinant is even less frequent, necessitating a huge population size. Therefore, one possible approach is to identify a recombinant on one side of the target in one generation and on the other side in the next generation.

**MAS for gene introgression and pyramiding: MAS for specific traits with examples**

**Introduction**

Gene Introgression, also known as introgressive hybridization, in genetics is the movement of a gene (gene flow) from one species into the gene pool of another by the repeated backcrossing of an interspecific hybrid with one of its parent species. Purposeful introgression is a long-term process; it may take many hybrid generations before the backcrossing occurs. Introgression differs from simple hybridization. Introgression results in a complex mixture of parental genes, while simple hybridization results in a more uniform mixture, which in the first generation will be an even mix of two parental species.

Steps of Gene Introgression

• Interspecific crossing with wild source containing the desired gene

• Back cross procedure and removal of disturbing linkage drag

• Crosses with breeding parent for variety selection

• Resistant variety with linkage drag

**GENE PYRAMIDING**

**Introduction**

Watson and Singh (1953) first introduced the concept called gene pyramiding. Gene pyramiding is defined as a method aimed at assembling multiple desirable genes from multiple parents into a single genotype. The end product of a gene pyramiding program is a genotype with all of the target genes. Gene pyramiding is a breeding method aimed at assembling multiple genes with known effects on target traits. It is mainly used in improving existing elite cultivars for a few unsatisfactory traits, for which genes with large positive effects are identified. Traditionally, the identification of the sources of useful gene is very slow and breeder‟s capability to trace the presence or absence of the target genes is limited. This limits the number of genes to be incorporated into elite cultivars at any times. The development of modern molecular and genomics technology has not only accelerated the discovery of favorable gene but also widened the sources of useful genes.

Objectives of gene pyramiding

1) Enhancing trait performance by combining two or more complementary genes,

2) Remedying deficits by introgressing genes from other sources,

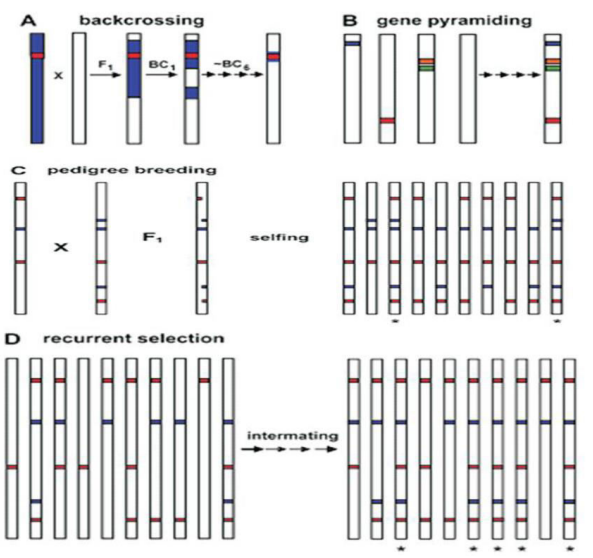
3) Increasing the durability.

Types of gene pyramiding

1. Conventional technique: Serial gene pyramiding: Genes are deployed in same plant one after other (Figure ). \* Pedigree breeding \* Backcross breeding \* Recurrent selection

**Molecular technique Simultaneous gene pyramiding**:

Genes are deployed at a time in a single plant. \* Marker assisted selection \* Transgenic method



**Figure-Conventional techniques of Plant Breeding**

**Disadvantages of conventional methods**

1. Gene pyramiding is mainly used to improve qualitative traits such as disease and insect resistance. This is associated with the fact that the presence of target trait genes must be confirmed by phenotyping mostly at the individual level and that individual phenotypic performance is a good indicator of the genotype only if genes have a major effect on phenotypic performance and the error of phenotyping is minimal.

2. In addition to the reliability of phenotyping at individual level other factors influencing the success of gene pyramiding are the inheritance model of the genes for the target traits, linkage and/or pleiotropism between the target trait and other traits.

3. For instance, allelic genes cannot be combined in the same genotype. The effect conferred by a recessive gene cannot be evaluated on heterozygous individuals and progeny testing is required.

4. If the target gene is tightly linked to genes with large negative effects on other traits, these undesirable genes may be transferred together with the target gene into the recipient line and result in reduced performance of other traits (linkage drag).

Therefore, any improvement in the knowledge of the trait genetics (inheritance, genetic relationship, etc.) and techniques for inferring for genotypes. Marker assisted gene pyramiding Marker-assisted selection (MAS) is a method of rapidly incorporating valuable traits into new cultivars. Molecular markers, or DNA tags, that have been shown to be linked to traits of interest are particularly useful for incorporating genes that are highly affected by the environment, genes for resistance to diseases and pests, and to accumulate multiple genes for resistance to specific diseases and pests within the same cultivar – a process called gene pyramiding. One of the first wheat cultivars to be developed using MAS was the soft winter wheat cultivar “Madsen”, released in 1986 by the USDA-Agricultural Research Service (ARS) and Washington State University. “Madsen” was developed using the isozyme marker from the endopeptidase protein, EpD1b, to incorporate a gene for resistance to eyespot (Tapesia yallunde) (Allan et al., 1989). Since 1990, detailed molecular maps of wheat have been constructed that include more than 3,000 molecular markers and several important traits have been associated with DNA markers. Additional markers can be developed from the 8,000 expressed sequence tags (ESTs) that have been mapped in wheat [Maps and references are available online (USDA-ARS 2005)].

**Advantages gene pyramiding**

1. Widely used for combining multiple disease resistance genes for specific races of a pathogen 2. Pyramiding is extremely difficult to achieve using conventional methods

3. Consider - phenotyping a single plant for multiple forms of seedling resistance – almost impossible

4. Important to develop 'durable' disease resistance against different races

5. Main used to improve existing elite cultivar

6. Eliminates extensive phenotyping

7. Control linkage drag

8. Reduces breeding duration

**Main factors affecting gene pyramiding**

1. Characteristics of the target traits/genes

2. Reproductive characteristics

3. A breeder's capability to identify the 'desired' genotypes

4. Operating capital Success stories

Table 1 below shows the successfully pyramided genes with their traits of some important crop plants, rice, wheat, cotton, pea, barley, broccoli, soyabean and chickpea.

