Review Article

Marker assisted plant breeding; principle and practices

Abstract:

Marker assisted plant breeding achieved various plant breeding objectives in a cost effective and time consuming manner through molecular marker techniques or statistical and bioinformatics tools. Molecular markers indicate the differences in the nucleotide sequence of different organs or species. Apart from application of molecular markers in linkage map construction, they are suitable for assessing genetic variations within cultivars and germplasm, in genome selection and in fingerprinting studies. The most engaging fact of molecular markers is in Marker assisted selection (MAS). In comparison with traditional breeding, molecular markers have efficiency to increase the effectiveness of breeding programmes. In this study, the applications of molecular markers in plant breeding studies are described.

Keywords: Plant breeding, Molecular markers, Marker assisted selection, linkage map and Quantitative trait loci (QTLs).

Introduction:

The main objective of plant breeder in agriculture is to improve the existing cultivars which are lack in one or more traits. The improvement can be made by crossing these cultivars with lines that possess the desired trait. In conventional breeding programme, whole genomes are crossed followed by selection of superior recombinants which is laborious and time consuming process because it involves several cross, several generation, careful phenotypic selection and the linkage drag (tight linkage of the undesired loci with the desired loci). These limitations may make it difficult to achieve desired objectives. Advancement in DNA marker technology make it possible to overcome many of the problems face during conventional breeding due to development of several types of molecular markers and molecular breeding strategies.

Molecular markers are used in many steps of a plant breeding program, e.g. germplasm characterization, parental selection for crossing, F_2 hybrid confirmation test, in pedigree and evolution studies, seed purity test, cultivar protection, construction of linkage maps, mapping of genes or quantitative trait loci (QTLs) and in breeding strategies establishment. These markers are mainly independent of developmental stages and environmental conditions which are useful to map traits governed by major genes as well as for QTLs. Marker assisted selection (MAS) involves the indirect selection of desired trait

through genetically associated molecular markers. MAS are generally applied in following situations: when direct phenotypic selection is less efficient, time consuming or expensive, low heritability of the traits, requirement of specific biological or environmental conditions for gene expression and QTL (for multiple traits or several genes) for the same trait are simultaneous or cumulative under selection (pyramiding).

The MAS in plant breeding open up new era to develop improved cultivars or fix trangenes within the cultivars through backcross marker assisted method [1-5]. In each backcross generation, transgenic individuals are selected on the basis of molecular marker or markers linked to the transgene, while other polymorphic markers of donor and recurrent parents are used to eliminate linkage drag or to retrieve recurrent parent genome. MAS can be used in any plant breeding method to transfer single gene which is linked with reliable marker (e.g.backcross marker assisted method) or in indirect selection which is more advantageous than the direct selection of the trait. Most of the agronomic important traits are under quantitative genetic control, having low heritability. So MAS strategies (e.g.backcross marker assisted method) is used to introgress favorable alleles at quantitative trait loci in corn [6], common bean [7], and in rice [8-10]. For selection of such trait, accurate statistical analysis or well established field experiment strategies are required to overcome part of environment effects. Additionally, molecular markers linked to QTL are mainly used to increase the genetic gain. The magnitude of increased genetic gain is depends on positions and effect of QTLs, stability of QTLs across multiple environments and across relevant breeding germplasm [2].

Molecular markers in plant breeding:

Genetic markers represent genetic differences between species and individuals by act as ‘sign’ or ‘flags’ to close proximity to genes. Such markers do not affect the functioning of genes because they are present near or linked to the genes controlling the traits. Sometimes markers present over the gene of interest, those markers are known as ‘perfect markers’. Genetic markers are of three types: Morphological markers which are phenotypic characters or traits, Biochemical markers are allelic variant of enzymes called as isozymes and DNA (molecular) markers represents variation at DNA levels [11]. Morphological and biochemical markers are not frequently used because they are limited in number and influenced by environmental and developmental stages [12]. These limitations of morphological and biochemical markers can be overcome by molecular markers which are mainly polymorphic and abundantly distributed throughout the genome.

1. Construction of linkage map:

Linkage map is a ‘road map’ of the chromosome which depict the order of genetic markers and relative distance between them in terms of recombination frequencies. Linkage maps helps in the identification of genes or QTLs associated with traits of interest on chromosome. Markers which are tightly or closely linked to gene of interest will be transmitted together from parents to progeny as compared to markers that are located further apart. Lower recombination frequency between two markers indicates that they are closer situated on a

chromosome. The recombination frequencies are converted into map units called centi Morgan (cM) using mapping functions. Two most commonly used mapping functions are Kosambi (indicates that the recombination events influence the occurrence of adjacent recombination events) and Haldane (assumes that no interference between crossover events) [13, 14]. Construction of Linkage maps are divided into three types: Production of mapping population, identification of polymorphism and linkage analysis of markers.

a) Mapping populations:

Mapping population is required for linkage map construction. Parents used in mapping populations should be different for one or more traits of interest. Size of mapping populations should be varying from 50-250 individuals but larger populations size is required for high resolution mapping. If map will used for QTL analysis than the populations must be phenotypically evaluated. In self pollination, mapping population is originated from two homozygous parents whereas, in case of cross pollination, mapping population is mainly derived from a cross between a heterozygous parent and a haploid or homozygous parent [11]. Different types of populations are utilized for mapping within plant species, with each having advantages and disadvantages.

b) Identification of polymorphism

The second step in construction of linkage map is to identify polymorphic marker which provide sufficient difference between parents. It is critical to identify sufficient amount of polymorphism for linkage map construction. In general, cross pollinated species revealed higher DNA polymorphism as compared to self pollinated species. So, sometimes in self pollination, parents are selected on the basis of the level of genetic diversity between species. Only polymorphic markers are screened around the entire population along with parents. This process is known as 'genotyping'.

c) Linkage analysis of markers:

The final step of linkage maps construction is the data analysis using computer software (Mapmaker/EXP, MapManager QTX and JionMap) for each marker on each individual. The linkage between markers is usually calculated in terms of logarithm of odds (LOD) value or LOD score [15]. The LOD value of >3 is used for linkage map construction because it shows that linkage is 1000 times more likely (i.e. 1000:1) than no linkage. The major difficulty in obtaining equal number of linkage groups and chromosomes is the non-random distribution of markers and unequal recombination frequency over the chromosome.

2. Linkage mapping of molecular markers and oligogenes:

i) Analysis of QTLs:

The region of genome which contains genes associated with quantitative traits is known as quantitative trait loci (QTLs). QTL analysis is worked on the principle of detecting association between phenotypic and genotypic data of markers. The marker which is closely associated with QTLs are inherited together, shows lower chances of occurrence of recombination. When two markers are linked on each side of QTLs, these markers are known as 'flanking' markers. Flanking markers based selection is more reliable than that of single markers because of the lower recombination. The QTLs are mainly detected by three methods: single-marker analysis, simple interval mapping and composite interval mapping [16,17].

Single-marker analysis is the simplest method for QTLs detection with single markers. The statistical approaches used are analysis of variance (ANOVA), t-tests and linear regression. An advantage of single marker analysis is that it does not require a complete linkage map and can be performed with simple statistical programs. Along with its advantage, the major disadvantage is that this method is unable to detect further a same QTL as it is detected previously. MapManager QTX and QGene are mostly used computer programmes for single marker analysis [18, 19].

Simple interval mapping (SIM) analyses intervals between adjacent pairs of linked markers on linkage maps [20]. This method is statistically more powerful than that of single point analysis because linked markers compensates for recombination between the markers and the QTL. Many researcher conducted SIM using Map Maker/QTL [21] and QGene [18]. Recently popular method for QTLs mapping is the composite interval mapping (CIM). The interval mapping is combined with linear regression along with statistical models to detect QTLs with CIM [22-25]. CIM have advantage that it is effective and precise method for QTL analysis, especially when linked QTLs are involved, as compared to single-point analysis and interval mapping. Researchers have used QTL cartographer [26, 27], PLABQTL [28] and MapManager QTX [19] to perform CIM.

ii) Advanced backcross-QTL analysis (AB-QTL):

Advanced backcross QTL analysis is method of combining QTL with varietal development. This method involves the identification of valuable QTLs from unadapted germplasm (e.g wild species, land races) and transferring them into established elite lines [29]. In first generation, BC1 population undergoes negative selection to reduce deleterious donor allele whereas BC2 and BC3 populations are evaluated for traits of interest and genotyped using molecular markers. In this way the QTLs are transferred to adapted genetic background. This method has been applied in various crops, e.g Tomato (*Lycopersicon esculentum* L.) [29-31], rice (*Oryza sativa*) [32-34], maize (Zea mays) [35, 36], Wheat (*Triticum aestivum*) [37] and barley (*Hordeum vulgare* L.) [38].

iii) **Association mapping:**

Association mapping or linkage disequilibrium mapping is the method of QTL identification by utilizing historic linkage disequilibrium to link phenotype (observable characteristic) to genotype (the genetic constitution of the organism). In association mapping genetic markers lies within the candidate genes and association mapping based on linkage disequilibrium between the candidate gene markers and causal polymorphisms caused in the gene. Association mapping is also known as LD mapping. LD is the nonrandom association of alleles at different loci (frequency of occurrence of genes in population than it would be expected). Association mapping is the powerful tool for identification of natural variations in genes. The main advantage of association mapping is that it is applied within current existing populations instead of creating new population. Association mapping has been applied in number of crops e.g flowering time in maize [39], growth habit and bolting in sea beet (*Beta vulgaris* sp. maritima) [40], kernel composition in maize [41] and kernel size in wheat [42], and flowering time and pathogen resistant genes in Arabidopsis [43].

3. Gene tagging:

The QTL analysis and linkage map construction is a laborious and expensive process. Therefore there is an alternative method for QTL detection which saves both time and money, especially useful in those cases where resources are limited. The two methods are bulked segregant analysis (BSA) and selective genotyping which identify markers that tag QTLs. Mapping populations are required in both the cases. BSA method helps in the identification of markers present in specific chromosomal regions [44]. In this case DNA from two contrasting bulks of 10-20 individuals (e.g. resistant vs. susceptible for a particular disease) were combined. Markers are used to screen across the two bulks. The marker which shows polymorphism across the bulks may represent as markers that are linked to a QTL or gene of interest. Once polymorphic markers are identified then entire populations genotyped with these markers and localized linkage map may be developed. This method usually used to tag genes controlling simple traits but in some cases markers linked to major QTLs are also identified [11]. In selective genotyping (also known as 'trait-based marker analyses' or 'distribution extreme analysis') only those individuals that have phenotypic extremes or trails of the trait are selected for analysis [20, 45-46]. Linkage map are construction or QTLs are detected only using those individuals which shows extreme phenotypes for trait of interest. Then subsample of population is being selected for genotyping which reduces mapping cost. Selecting genotyping is performed in those cases when growing or phenotyping of individuals are easier or cheaper in mapping population as compared to genotyping using DNA marker assays. The major disadvantage associated with this method is that only one tait can be tested at a time and in some cases it is not efficient in determining the effects of QTLs.

MARKER ASSISTED PLANT BREEDING APPLICATIONS:

Marker assisted selection is the combination of traditional genetics and molecular biology. In MAS, phenotype is selected on the basis of genotype of the markers. The markers used in preliminary mapping studies are suitable for marker-assisted selection without further testing and development. Markers that are not used in previous MAS studies may not be reliable for predicting phenotype, and will be useless. The genes which control traits of interest such as color, quantity, or disease resistance etc are selected through MAS. The major steps included in MAS are: high resolution mapping, validation of markers and marker conversion, if required. In high resolution mapping, greater number of markers and larger population size is used to found tightly-linked marker. Then the effectiveness of these markers is validated on independent population or on different genetic background to determine the target phenotype.

1. FINGERPRINTING AND GENE CLONNING

a) DNA Fingerprinting

With the advent of DNA fingerprinting, that is the profiling of the DNA which is unique and different for all individuals like the fingerprints, in humans by sir Alec Jeffrey a decade ago, soon it was adopted for other organisms as well like fungai, plants etc for the betterment of mankind. Plant DNA fingerprinting has evolved beginning from the RFLP- in conjuction with southern hybridisation based fingerprinting to mostly employed PCR-based

fingerprinting approaches for single or multi locus profiling and nowdays the developing next generation sequencing strategies being exploited for generating high throughput plant DNA fingerprinting. Though NGS invention will definitely support fingerprinting approach either in terms of marker development or to adopt Genotype-by-Sequencing but there's no second thought about the fact that the wide applications of plant DNA fingerprinting still relies on PCR-based molecular markers to a greater extent.

Shortly after the invention of PCR (1980s), many approaches came into lime light for generating plant fingerprints. The strategy adopted was such that PCR fragments from the plant genomic DNA were produced using short oliegonucleotides primers along with arbitrary sequences. Further after electrophoretic separation and visualisation of bands using autoradiography, multilocational banding patterns were observed. The markers employed for this purpose were starting from RAPD, AFLP, ISSR to newly developed and less frequently used markers like SRAP (sequence related amplified polymorphism) where the junction which is polymorphic in between the exons and flanking introns are amplified, TRAP (Target region amplification polymorphism), SAMPL (Selective amplification of polymorphic microsatellite loci) a combination of AFLP and microsatellite based primers, DAMD (Direct amplification of minisatellite DNA) that use primers specific to minisatellites only, RGAP (Resistant gene analog polymorphism) that makes use of primers binding to only conserved domains of resistant gene [47,48].

Likewise PCR based single locus markers gained importance due to their high reproducibility, being codominant in nature, increased polymorphism rate and high accuracy ratio as compared to other fingerprinting strategies. Here two microsatellite flanking primers are used to amplify the target region using PCR and bands are separated electrophoretically using polyacrylamide or capilliary separation and bands are visualised using autoradiography or fluorography. EST-SSRs and SNPs have shown high applicability nowdays where ESTs are generated either from cDNA cloning or from existing databases of ESTs available. Single nucleotide polymorphism (SNPs) on the other hand are nowdays most widely employed authentic single locus markers. Apart from these, SCAR (Sequence characterised amplified region) and CAPs (cleaved amplified polymorphic sequences) have also been developed [49].

Plant DNA fingerprinting possess many advantages in the crop improvement programmes such as marker assisted breeding, identification of genotypes, genetic similarities and variations, population structure and gene flow, introgression, polyploidy, association and linkage mapping studies etc. And soon the high throughput sequencing strategies will possess a strong hold in the fingerprinting field and in alliance with both the strategies new qualitative and quantitative outcomes can be generated.

b) Gene cloning

Gene cloning refers to the phenomena of construction of a recombinant molecule first and then transforming into a suitable host for its multiplication and increasing copy number or say producing clones. Basically the two major components of gene cloning process are

- The DNA fragment to be cloned

- The plasmid/vector in which the fragment is to be ligated

The DNA fragment containing the desired gene of interest along with operons, regulatory elements etc are isolated from the genomic DNA either using restriction enzymes or amplifying by PCR and similarly side by side plasmid which is the extra chromosomal circular DNA in bacterial cells is also linearised using restriction enzymes. Further the desired gene is ligated into the plasmid by physical joining with the help of ligase enzyme that creates phosphodiester bonds between them and a new recombinant molecule is constructed. The recombinant molecule is then transformed into suitable host say E.coli where it replicates along with the host organism (fig 1).

Plasmid vectors enable the rapid multiplication of the desired gene along with substituting the required control elements required for further transcription and translation of the cloned gene. Gene cloning is not considered successful until its presence is checked which can be done by plating the bacterial host on to the selective media and further if the colonies observed then confirmation is done using colony PCR and sequencing.

The recombinant DNA technology and gene cloning has enabled the researchers and plant breeders to exchange and inculcate desired genes conferring traits of interest into target crops and overcome the barriers of traditional breeding approaches. The cloned gene when successfully express in the host may possess the ability of producing such trait which otherwise would be lacking like disease resistance, enhanced vitamin content etc. This rDNA technology has enabled the researchers to inculcate and characterize all genes for crop improvement irrespective of those found in a specie or its close relative only. Also the technique is quite specific, only desired traits are inserted unlike wide crosses made by breeders where undesired genes can also transfer. Gene cloning is widely employed in case of woody trees and vines where it takes number of years to acquire a phenotypic trait by traditional breeding whereas genetic manipulation eliminates the need of full growth of the crop rather with molecular techniques, researchers can analyse the genetic makeup of whole plant at an early stage [50].

2. GENOMIC SELECTION

Although phenotyping and pedigree analysis have been successful in providing information regarding traditional crop improvement strategies still they are inefficient in providing precise values. Due to the favourable results depicted by gene tagging and QTL mapping, MAS (Marker-assisted selection) have been extensively used in crop improvement programmes since 1990. But the strategy face many limitations as single gene effect is not solely responsible for governing all the economically important traits hence DNA markers can only identify small number of genetic variations, also single genes posess minute effects and therefore high amount of data is required for compilation of the effects. Complication is further worsen if QTL is traced by a haplotype of marker. The modified strategy of MAS to overcome all these problems and named it as Genomic selection [51]. Basically genomic selection is a type of Marker-assisted selection only where the complete genome is traced with the help of genetic markers where all the quantative trait loci (QTL) are in linkage

disequilibrium with atleast one of the genetic marker so that the estimated effects per QTL is small. The estimation in the target population is based according to the distinct but related clusters namely training and breeding populations. The resultant breeding values confirm the selection estimated by the breeding population. Instead of estimating novel genes governing a particular trait, GS estimates highest genomic estimated breeding value (GEBV) which means, using statistical model(s) to estimate the genotyped population to inturn predict the breeding values of future phenotypes in the candidate specie (fig 2).

To enhance the authenticity of statistical analysis, phenotyping can be considered as a key source. Also GS along with gene pyramiding, and Marker assisted recurrent selection (MARS) can serve as integrated platform for diversifying corp improvement programmes.

HOW TO ESTIMATE GEBV...

1. Inferring genotype of the plant at each QTL using genetic markers
2. Estimating the consequence of each QTL on the trait
3. Acquiring the GEBV value by summing up the total effects of QTL for selection candidates.

Prediction models used to estimate GEBV ...

1. Prediction using genome-wide dense marker maps
2. LASSO-related penalyzed regression method
3. Whole genome regression model
4. Kernel Hilbert spaces regression method
5. Random forest method
6. Elastic net algorithm
7. Bayesian regression model etc

Advantage over other breeding strategies:

Genome-wide selection in crop protection is an emerging area as compared to animals where its significance is increasing day by day. Genomic selection needs to be implemented in plants as an integrated breeding programme to enhance the efficiency of other conventional breeding strategies.

Breeding strategies like MAS which deals with the markers associated with only qualitative traits and association genetics on the other hand focus on LD mapping rather than evaluating the functional alleles. A successful variety is a combination of valuable traits governed by minor genes which might get skipped in MAS and association mapping and hence GS comes into play to overcome these drawbacks by covering the whole genome with the help of markers and estimating GEBVs to identify phenotype of target trait. Phenotypic prediction accuracy of genomic selection models is much higher than any other model say multiple linear regression, where GS models outperformed than MLR by approximately 47% in case of biparental wheat experiment [52]. The annual genetic gains per annum given by GS is much higher than MAS and phenotypic selection as depicted in case of wheat and maize [53]. The only hurdles in the path of exercising genomic selection for crop improvement is

that the quantity of genomic markers required is high along with the genotyping cost. However to counteract both these hurdles, next generation sequencing techniques and single nucleotide polymorphism (SNP) have been inculcated that has enhanced the rate of implementation of genomic selection for crop improvement strategies.

3. HIGH THROUGHPUT SNP GENOTYPING

Because of their high density and even distribution among the genomes of various crop plants, SNPs have been reported as the most applicable genetic markers as of now. They are readily used for association mapping studies and analysing genes for disease loci which require highthroughput SNP genotyping. There are many strategies being employed nowdays for SNP genotyping as it deals with large number of SNP markers and efficient technology (fig 3).

The efficient genotyping strategy involves:

- Amplification of target fragment
- Discrimination of alleles
- Product identification with allele specificity

Target amplification is mostly done using Polymerase chain reaction (PCR) where specific amplicons are applied to achieve accurate amplification as using primers that amplify multiple loci can lead to serious genotyping errors. For designing primers for SNP genotyping, pseudogenes, conserved sequences and repetitive sequences should be neglected as they may hinder the amplification step. After amplication, a purification step is carried out to remove excess of dNTPs, PCR primer leftovers so as to initiate the extention step. For purification, shrimp alkaline phosphatise (SAP) or E.coli exonuclease can be used.

Allelic discrimination is a crucial step where this discrimination strength emerge from DNA polymerase and DNA ligase and also the thermodynamics of matched and mismatched DNA duplex are exploited for the same [54]. This step is important for its accurate and specific product outcome. Protocols employed nowdays like Molecular Beacons, the TaqMan assay, FRET-DOL assay etc are single step assays combining PCR amplification and allele discrimination.

Various techniques being exploited for the identification of product with allele specificity are Mass Spectrometry which identifies products on the basis of their molecular weight, DNA sequencing approaches that detect product on the basis of fluorescence and size. Also other methods such as FRET (fluorescence resonance energy transfer), FP (Florescence polarisation), absorbance, luminescence etc uses direct correlation of the indexes with the product. Detection can be divided into two groups: Homogeneous based detection and Solid-phase mediated detection [55].

Homogeneous based detection are more efficient for automated scoring because no purification and clean up steps are required but also face the limitation of reduced multiplexing whereas, solid-phase detection methods employ purification strategies using

mass spectrometry, illumina color beads, Zip Code technology, Orchid SNP IT technology etc. Most of the detection protocols employs simultaneous genotyping assays by using 96 or 384 well plates or microarray and above all most detection systems have automatic genotyping scoring depending upon the cost and throughput [56].

The cost and throughput structures rely on their applicability in different fields like in clinical and diagnostic field where SNP marker requirement is low but sample size is high and methods like TaqMan assay, Molecular beacon and Invader technology fit efficiently in this field and where requirement of SNP markers is high, microarray method is employed [57]. Applications like gene mapping and linkage disequilibrium for different traits require both high amount of SNP markers and large sample size and for this most cost effective and flexible methods should be taken into account. The most popular methods employed for SNP genotyping as of now are FP-TDI, taqMan assay and pyrosequencing. Other methods such as invader's technology and fluorescence detection are efficient in terms of accuracy and success rate.

4. PHYLOGENETIC RELATIONSHIPS AND GENETIC DIVERSITY

Molecular markers play a crucial role in extracting genetic profiles from the tested germplasm along with the clarification of the evolutionary relationships among different groups, communities, genera, tribe etc. This evolutionary relationship studies helps to elucidate the various ancestral information, relatedness among different species and further exploiting this fact for various genetic variation studies [58]. Phylogenetic relationships are depicted in form of a evolutionary tree where it highlights the relatedness among different species and genera, their migration pattern, infection prevalence etc in a graphical manner.

The basic elements of a phylogenetic tree which comprise of an edge connected to nodes which can be internal or external [59]. External nodes dipicts operational taxonomic units (OTU) which are the molecular sequences from which tree was hypothesized. The last common ancestors (LCA) is depicted by the internal nodes, also the phylogenetic tree could be a result of single gene or multiple gene. Root depicts the common ancestor of all the taxa and in absence of common one, it can be placed in middle of the tree.

A molecular phylogenetic relationship can be predicted using evolutionary information generated by using biomolecular sequence alignment of amino acids, DNA, RNA or single nucleotide polymorphism (SNP), morphological data etc. Statistical methods that are applied for the generation of a pylogenetic tree are : maximum parsimony, maximum likelihood, UPGMA (Unweighted pair group method with arithmetic mean), transform distance method, neighbour joining method etc. Various softwares now dealing with the phlogenetic tree assessment are Paup, PAML, PHYLIP, Pfam, TREEfam, PANTHER etc [60].

Plant genetic diversity

Genetic diversity is an important aspect for the crop improvement which is being exploited since ages for supplementing the growing food demands. Genetic diversity in plant genetic resources have enabled to inculcate desired traits as per choice of farmers as well as plant

breeders. Basically plant genetic diversity deals with the genetic changes in a specie for its adaptation to various biotic and abiotic changes and inherit that ability of coping with the environmental changes in further generations. These genetically diverse species can be preserved in the form of plant genetic resources in biorepositories, gene banks, DNA libraries etc for longer duration.

Assessment of genetic diversity

Various techniques being employed for the assessment of genetic diversity are:

Morphological evaluation

Biochemical evaluation (allozyme)

Molecular marker based evaluation

Morphological and biochemical based evaluation were being employed in the pregenomic era where morphological parameters dealt with visual characteristics for the diverse traits and evaluating them. On the other hand biochemical evaluation deals with allelic variants called isozymes that are detected using electrophoresis and staining.

The third and most applicable evaluation system is by employing molecular markers. These molecular markers help in assessing genetic diversity such that they can detect all sort of mutagenic actions like deletions, additions, invertions etc and unlike biochemical markers, these can be both dominant as well as codominant. Advantages of using these genetic markers are that they are capable of assessing cultivar purity, parental selection, assessing genetic variability and cultivar identity along with marker assisted backcrossing and gene pyramiding. The various genetic markers being employed for the genetic variability assessment are RAPD, RFLP, AFLP,SSRs, ISSRs, EST-SSRs, SNP etc.

Genetic diversity analysis

There are various statistical approaches that can be employed to assess the genetic diversity which are based on following concepts:

- Measuring polymorphism
- Similarity coefficient
- Shannon's information index
- Allele frequency based approaches
- Heterozygosity
- F statistics
- Effective population size etc

Nowadays, various analytical programs are available for estimating the genetic diversity using computational tools like : NTSYSpc, Arlequin, DARWIN, Power Marker, DnaSP, MEGA, STRUCTURE, fastSTRUCTURE, fineSTRUCTURE, POPGENE etc. many of these sftwares perform similar tasks only differing in modes of input and output data, user interface and platform, solely depending on the user's choice.

With the advent of molecular marker assessment, new data can be characterized with accuracy and speed with low cost and high output and hence germplasm rich in qualitative and quantitative traits can be further analysed and stored for longer durations in the repositories.

BIOINFORMATICS AND DATABASES FOR GENOMIC RESEARCH

For the development of statistical tools and programs along with computer softwares for efficient storage, accumulation and visualisation of the biological samples, Bioinformatics comes into play. Bioinformatics emerged in early 1980s when it came into consideration that personal computers could be used as storage and evaluation device for the biological samples as well. And with passing time bioinformatics tools started showing up in sites such as European Molecular Biology Laboratory (EMBL), The National Centre for Biotecnology Information (NCBI), the DNA Databank of Japan (DDBJ) etc which are the international conventions which supplements the need of researchers all around the globe and are also progressing day by day.

For sequence search and similarity approaches, analytical tools like BLAST and CLUSTAL have shown their applicability since 1990s. In addition to that, for retrieving biological information based on their sequence information, several other databases like AutoSNP, SNP_2CAPS, TASSEL, STRUCTURE are also being used nowdays. Bioinformatics also include some databases which are specifically designed for the purpose of storage of vast genetic information and the efficient ones are GenBank, Phytozome, SwissProt, UniProt, the EMBL nucleotide database etc. All these databases are in open access to te users and are readily available to them. KEGG (Kyoto Encyclopedia of Genes and Genomes) on the other hand aims at providing information regarding the metabolic pathways and the gene interaction among different organisms [61-63].

With the advancement in science and technology, many new approaches are being employed and with the advent of NGS techniques, genome sequencing is no big task today. And bioinformatics play a crucial role in the storage and analysis of the data and comeup with valuable information to be exploited for crop improvement. Many new databases have been created for the analysis and functional annotation of plant genomes such as Blast2GO which provide information regarding functional regions within DNA sequence, another database i.e. SSR Locator that enables user to identify the appropriate targets for primers to bind to the genomic DNA and ensure that they are unique in nature. It also plays a significant role in primer designing and contains a PCR simulator that helps in hypothetically analysing the comparison of amplified products among different crops. Likewise, few databases are mentioned in table 1 which are widely employed in case of crop plants.

Conclusion:

Molecular markers play an important role in plant breeding to increase crop production or productivity. However, plant breeders are facing major problems in crop yield production due to global warming, new biotypes of diseases and insects or with abiotic stress. The progress made in genomics is creating the path to identify new genes for resistance to biotic and abiotic stress using DNA markers. Integration of those desired genes from unadapted cultivars to elite cultivars will help in the development of improved crop varieties suitable for different agro-climate conditions.

References:

1. Micallef MC, Austin S, Bingham ET. Improvement of transgenic alfalfa by backcrossing. In Vitro Cellular & Developmental Biology.1995; 31: 187-192.
2. Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiology. 2008; 147: 969-977.
3. Liu YB, Wei W, Ma KP, Darmency H. Backcrosses to Brassica napus of hybrids between *B. juncea* and *B. napus* as a source of herbicide-resistant volunteer-like feral populations. Plant Science. 2010; 179: 459-465.
4. Roy S, Banerjee A, Tarafdar J, Senapati BK. Detection of probable marker-free transgene-positive rice plants resistant to rice tungro disease from backcross progenies of transgenic Pusa Basmati 1. Journal of Genetics. 2012; 91: 213-218.
5. Roy S, Banerjee A, Tarafdar J, Senapati BK, Dasgupta I. Transfer of transgenes for resistance to rice tungro disease into high-yielding rice cultivars through genebased marker-assisted selection. Journal of Agricultural Science. 2011; 150: 610-618.
6. Bouchez A, Hospital F, Causse M, Gallais A, Charcosset A. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics. 2002; 162: 1945-1959.
7. Miklas PN. Marker-assisted backcrossing QTL for partial resistance to sclerotinia white mold in dry bean. Crop Science. 2007; 47: 935-942.
8. Siangliw JL, Jongdee B, Pantuwan G, Toojinda T. Developing KDML105 backcross introgression lines using marker-assisted selection for QTLs associated with drought tolerance in rice. ScienceAsia. 2007; 33: 207-214.
9. Iftekharuddaula KM, Newaz MA, Salam MA, Ahmed HU, Mahbub MAA, Septiningsih EM, Collard BCY, Sanchez DL, Pamplona AM, Mackill DJ. Rapid and high-precision marker assisted backcrossing to introgress the *SUB1* QTL into BR11, the rainfed lowland rice mega variety of Bangladesh. Euphytica. 2011; 178: 83-97.
10. Cuc LM, Huyen LTN, Hien PTM, Hang VTT, Dam NQ, Mui PT, Quang VD, Ismail AM, Ham LH. Application of marker assisted backcrossing to introgress the submergence tolerance QTL SUB1 into the Vietnam elite rice variety-AS996. American Journal of Plant Sciences. 2012;3: 528-536.
11. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica. 2005; 142: 169–196.
12. Winter P, Kahl G. Molecular marker technologies for plant improvement. World Journal of Microbiology & Biotechnology. 1995; 11: 438–448.
13. Kearsey M, Pooni H. The genetical analysis of quantitative traits. Chapman & Hall, London. . 1996
14. Hartl D, Jones E. Genetics: Analysis of Genes and Genomes, Jones and Bartlett Publishers, Sudbury, MA. 2001.
15. Risch N. Genetic linkage: Interpreting LOD scores. Science. 1992; 255: 803–804.

16. Tanksley SD. Mapping polygenes. Annu Rev Genet. 1993; 27: 205– 233.
17. Liu B. Statistical Genomics: Linkage, Mapping and QTL Analysis CRC Press, Boca Raton. 1998.
18. Nelson JC. Qgene—software for marker-based genomic analysis and breeding. Mol Breed. 1997; 3: 239–245.
19. Manly KF, Cudmore H, Robert J, Meer JM. Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome. 2001; 12: 930–932.
20. Lander E, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989; 121: 185–199.
21. Lincoln S, Daly M, Lander E. Mapping genes controlling quantitative traits using MAPMAKER/QTL. Version 1.1. Whitehead Institute for Biomedical Research Technical Report, 2nd Edn. 1993.
22. Jansen R. Interval mapping of multiple quantitative trait loci. Genetics. 1993; 135: 205–211.
23. Jansen R, Stam P. High resolution of quantitative traits into multiple loci via interval mapping. Genetics. 1994; 136: 1447–1455.
24. Zeng ZB. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci USA. 1993; 90: 10972–10976.
25. Zeng ZB. Precision mapping of quantitative trait loci. Genetics. 1994; 136: 1457–1468.
26. Basten CJ, Weir BS, Zeng ZB. 1994. Zmap-a QTL cartographer. In: J.S.G.C. Smith, B.J. Benkel,W.F. Chesnais, J.P. Gibson, B.W. Kennedy & E.B. Burnside (Eds.), Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software, Guelph, Ontario, Canada. Published by the Organizing Committee, 5th World Congress on Genetics Applied to Livestock Production.
27. Basten C, Weir B, Zeng ZB. QTL Cartographer. Department of Statistics, North Carolina State University, Raleigh, NC. 2001.
28. Utz H, Melchinger A. PLABQTL: A program for composite interval mapping of QTL. J Quant Trait Loci 2(1). 1996.
29. Tanksley SD, Nelson JC. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet. 1996; 92: 191–203.
30. Bernacchi, D., T. Beck-Bunn, Y. Eshed, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, S.D. Tanksley, 1998. Advanced backcross QTL analysis in tomato. I: Identification of QTLs for traits of agronomic importance from Lycopersicon hirsutum. Theor Appl Genet. 97:381-397
31. Bernacchi, D., T. Beck-Bunn, D. Emmatty, Y. Eshed, S. Inai, J. Lopez, V. Petiard, H. Sayama, J. Uhlig, D. Zamir, S.D. Tanksley. 1998. Advanced backcross QTL analysis in tomato. II: Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from Lycopersicon hirsutum and L. pimpinellifolium. Theor Appl Genet. 97:170-180.
32. Xiao, J., J. Li, S. Grandillo, S. Ahn, L. Yuan, S.D. Tanksley, and S.R. McCouch. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, Oryza rufipogon. Genetics. 150:899-909.

33. Moncada, P., C.P. Martínez, J. Borrero, M. Chatel, J. H. Gauch, E. Guimaraes, J. Tohme, and S.R. McCouch. 2001. Quantitative trait loci for yield and yield components in an Oryza sativa × O. rufipogon BC2F2 population evaluated in an upland environment. Theor Appl Genet. 102:41-52

34. Chai L, Zhang J, Pan X B, Zhang F, Zheng T Q, Zhao X Q, Wang W S, Jauhar A, Xu J L and Ll Z K. 2014. Advanced Backcross QTL Analysis for the Whole Plant Growth Duration Salt Tolerance in Rice (*Oryza sativa* L.). Journal of Integrative Agriculture. 13: 1609-1620.

35. Ho, J.C., S.R. McCouch, and M.E. Smith. 2002. Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. Theor Appl Genet. 105:440-448

36. Mano Y, Omori F, Kindiger B, Takahashi H. A linkage map of maize × teosinte Zea luxurians and identification of QTLs controlling root aerenchyma formation. Mol. Breed. 2008; DOI 10.1007/s11032-007-9132-8 (in press).

37. Narasimhamoorthy B, Gill BS, Fritz AK, Nelson JC, Brown-Guedira GL. Advanced backcross QTL analysis of a hard winter wheat x synthetic wheat population. Theor Appl Genet. 2006; 112(5):787-96.

38. Yun SJ, Gyenis L, Bossolini E, Hayes PM, Matus I, Smith KP, Steffenson BJ, Tuberosa R, Muehlbauer GJ. Validation of quantitative trait loci for multiple disease resistance in barley using advanced backcross lines developed with a wild barley. Crop Science. 2006; 46: 1179–1186.

39. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 2001; 28: 286–289. 10.1038/90135

40. Hansen M, Kraft T, Ganestam S, Sall T, Nilsson NO. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. Genet Res. 2001; 77(1): 61-66.

41. Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell. 2004; 16: 2719–2733. 10.1105/tpc.104.025700.

42. Breseghello F, Sorrells ME. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci. 2006; 46: 1323–1330. 10.2135/cropsci2005.09-0305

43. Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M. Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes. PLoS Genetics. 2005; 0531-0539

44. Michelmore R, Paran I, Kesseli R.. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA. 1991; 88: 9828–9832.

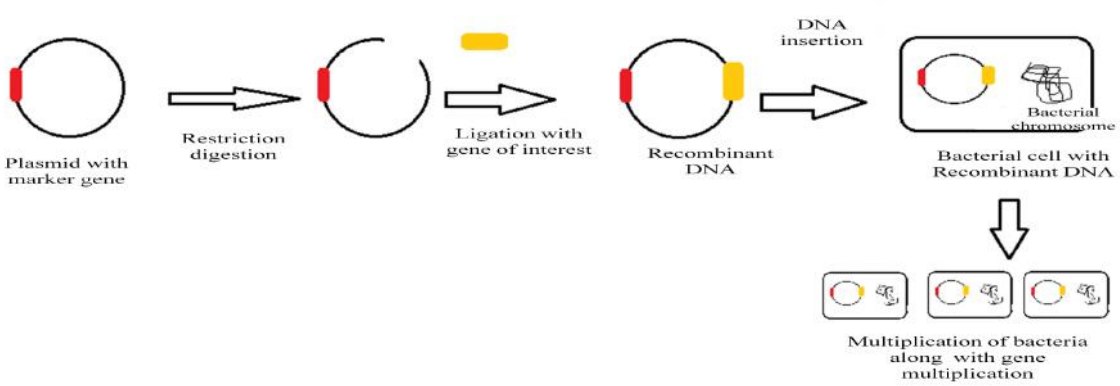
45. Foolad M, Jones R. Mapping salt-tolerance genes in tomato (*Lycopersicon esculentum*) using trait-based marker analysis. Theor Appl Genet. 1993; 87: 184–192.

46. Zhang LP, Lin GY, Liu ND, Foolad MR. Mapping QTLs conferring early blight (*Alternaria solani*) resistance in a *Lycopersicon esculentum* × *L. hirsutum* cross by selective genotyping. Mol Breed. 2003; 12: 3–19.
47. Li G, Quiros CF. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theor Appl Genet. 2001; 103:455–461.
48. Li G, McVetty PBE, Quiros CF. SRAP molecular marker technology in plant science. In Plant Breeding from Laboratories to Fields. Croatia: InTech. 2013.
49. Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett. 2006; 9:615–629.
50. Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F and Petit RJ. Current trends in microsatellite genotyping. Mol Ecol Resources.2011; 11:591–611.
51. Meuwissen TH, Goddard ME. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics. 2001; 155: 421–430
52. Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME. 2011. Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. Crop Science. 2011; 51: 2597-2606
53. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME. 2010. Plant breeding with Genomic selection: Gain per unit time and cost. Crop Science 50*:* 1681-1690
54. Tyagi S, Bratu DP, Kramer FR.1998. Multicolor molecular beacons for allele discrimination. Nat Biotechnol.1998; 16: 49–53.
55. Livak KJ. Allelic discrimination using fluorogenic probes and the 50 nuclease assay. Genet Anal.1999; 14: 143–149.
56. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat. 2001; 17: 296–304.
57. Chen X, Sullivan PF. Single Nucleotide Polymorphism Genotyping: biochemistry, protocol, cost and throughput. The Pharmacogenomics Journal. 2003; 3:77–96
58. Baldauf SL. Phylogeny for the faint of heart: a tutorial. Trends in Genetics. 2003; 19: 345-351
59. Hall BG. Phylogenetic Trees Made Easy: A How-To Manual, 2nd ed. Sinauer Associates, Inc.: Sunderland, MA. 2004.
60. Govindraj M, Vetriventhan M, Srinivasan M. Importance of Genetic Diversity Assessment in Crop Plants and its Recent Advances: An Overview Of Its Analytic Perspectives. Genetics Research International. 2015; http://dx.doi.org/10.1155/2015/431487
61. Arend D, Lange M, Chen J, Colmsee C, Flemming S, Hecht D, Scholz U. EDAL a framework to store, share and publish research data. BMC Bioinf. 2014; 15: 214
62. Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KFX. Plant genome sequencing – applications for crop improvement. Curr. Opin. Biotechnol. 2014; 26: 31–37.
63. Skuse GR, Du C. Bioinformatics Tools For Plant Genomics. International Journal of Plant Genomics 1-2. 2008.

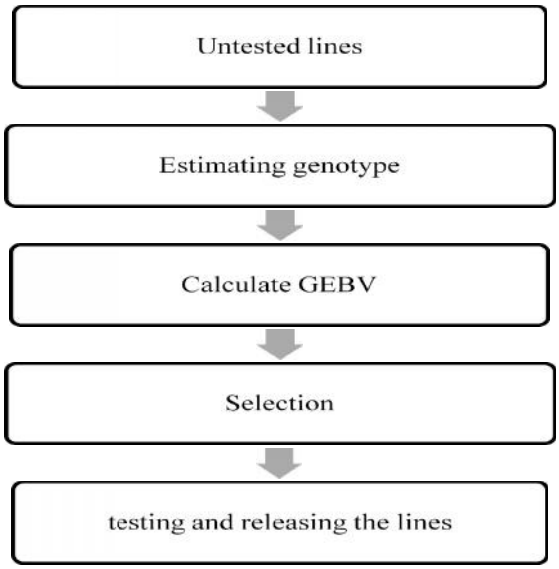
651 **Table 1: List of databases used in genomic research**

Sr No	DATABASES	URL	
1	European molecular biology laboratory (EMBL)	http://www.ebi.ac.uk/embl.html	Collection of all nucleotide and protein sequences
2	GenBank	http://www.ncbi.nlm.nih.gov/web/genbank/	Collection of all nucleotide and protein sequences
3	DNA databank of Japan (DDBJ)	http://www.ddbj.nig.ac,jp	Collection of all nucleotide and protein sequences
4	Genome sequence database	http://www.ncgr.org/gsdb	Collection of all nucleotide and protein sequences
5	Unigene	http://www.ncbi.nlm.nih.gov/unigene/	Gene oriented clusters
6	Clusters of orthologus groups (COG)	http://www.ncbi.nlm.nih.gov/COG/	Phylogenetic protein classification
7	ASDB	http://cbcg.nersc.org/asdb	Expression patterns of alternative spliced genes
8	EID	http://mcb.harvard.edu/gilbert/EID/	Intron containing genes
9	PLANTCARE	http://sphinx.rug.ac.be.8080/PLANTCARE	Cis-acting regulatory elements of plants
10	PLACE	http://www.dna.affrc.go.jp/htdocs/PLACE/	Cis-acting regulatory elements of plants
11	Arabidopsis Database (AtDB)	http://genome-www.stanford.edu/Arabidopsis	*Arabidopsis thaliana* genome
12	Crop Net	http://syntny.nott.ac.uk	Genome mapping in crop plants
13	INE	http://www.dna.affrc.go.jp:82/giot/INE.html	physical map and sequence data of rice
14	Mendel database	http://jiio6.jic.bbsrc.ac.uk/	EST and STS database sequence of plants
15	ZmDB	http://zmdb.iastate.edu/	Genome database of maize
16	Kyoto encyclopedia of genes and genomes (KEGG)	http://www.genome.ad.jp/Kegg	Metabolic and regulatory pathways
17	PROSITE	http://www.expasy.ch/prosite/	Protein patterns and profiles
18	CATH	http://www.biochem.ucl.ac.uk/bsm/cath/	Protein domain structure classification
19	SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/	Structural protein relationships

652



653 Fig 1: diagrammatic presentation of gene cloning



654 Fig 2: Genome selection strategy

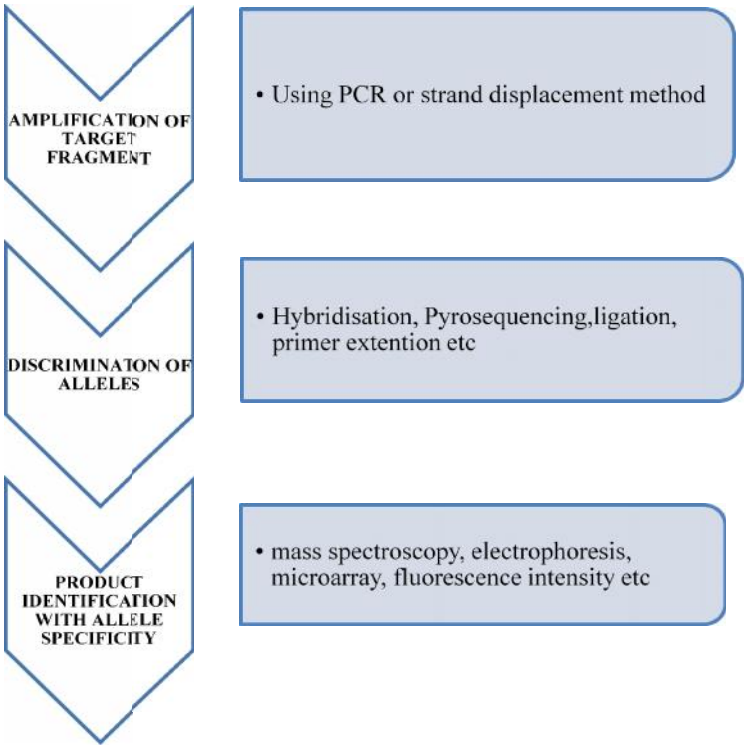


Fig 3: Efficient genotypic stratergies

655