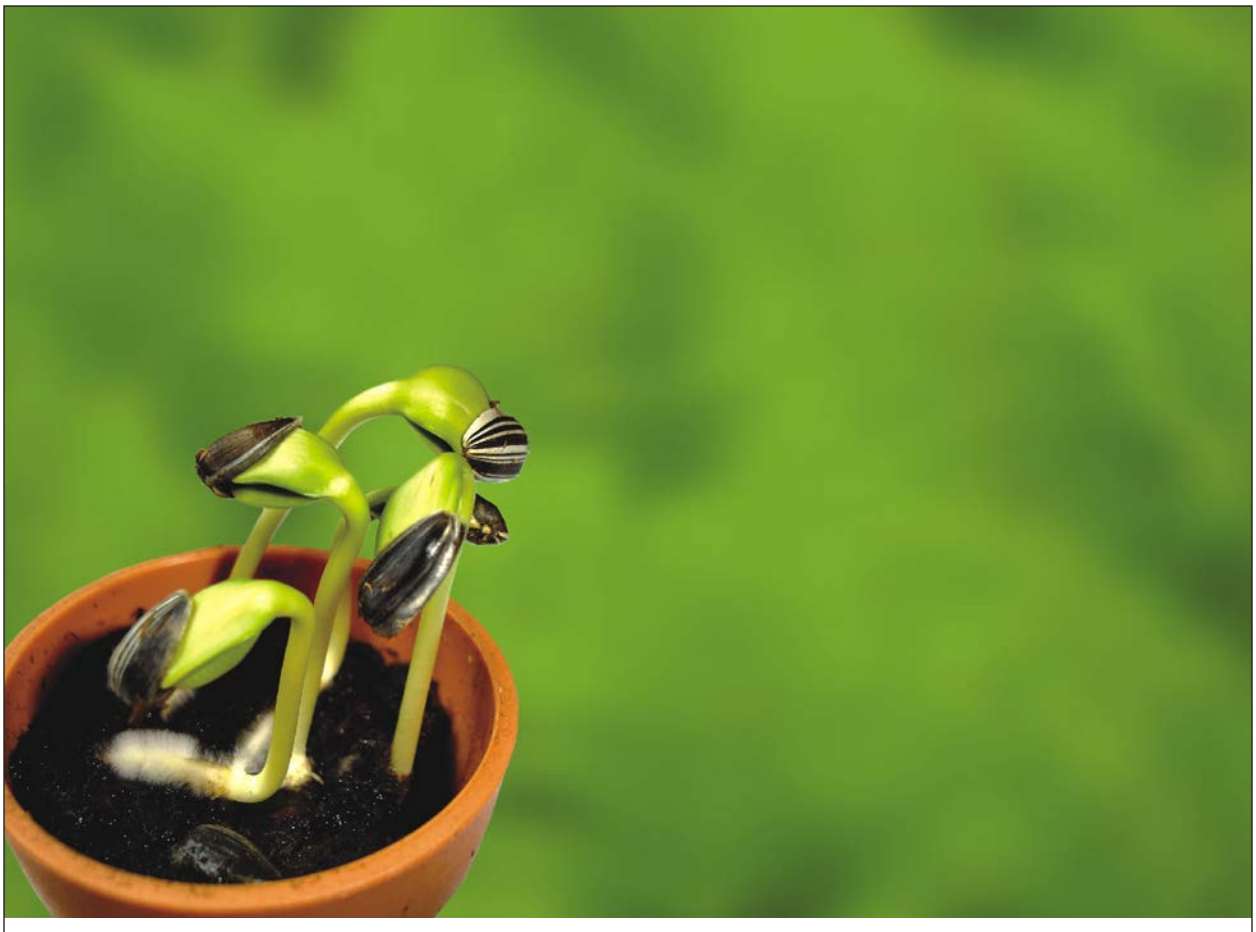

International Journal of Plant Breeding and Genetics

ISSN 1819-3595

Academic Journals Inc.

www.academicjournals.com

OPEN ACCESS International Journal of Plant Breeding and Genetics

ISSN 1819-3595
DOI: 10.3923/ijpbg.2016.58.68

Review Article
Advances in Quantitative Trait Loci, Mapping and Importance of Markers Assisted Selection in Plant Breeding Research

[1]Tewodros Mulualem and [2]Zelalem Bekeko

[1]Jimma Agricultural Research Center, P.O. Box 192, Jimma, Ethiopia
[2]School of Plant Sciences, Haramaya University, P.O. Box 138, Dire Dawa, Ethiopia

Abstract

Quantitative Trait Loci (QTL) mapping basically entails finding an association between a genetic marker and a measurable phenotype. Researchers work from the phenotype to the genotype, using statistical techniques to localize chromosomal regions that might contain genes contributing to the phenotypic variation in a quantitative trait of interest in a population. Most traits of interest in plant breeding show quantitative inheritance, which complicates the breeding process, since phenotypic performances partially reflects the genetic values of individuals. The genetic variation of a quantitative trait is controlled by the collective effects of Quantitative Trait Loci (QTLs), epistasis (interaction between QTLs) the environment and interaction between QTL and the environment. Linkage analysis and association mapping are the two most commonly used methods for QTL mapping and exploiting molecular markers in plant breeding involves finding a subset of markers associated with one or more QTLs that regulate the expression of complex traits. Currently, many QTL mapping identified QTLs that explained a significant proportion of the phenotypic variance and therefore, gave rise to an optimistic assessment of the prospects of Markers Assisted Selection (MAS). Therefore, the objective of this review is to provide an overview of current advances in QTL analysis such as mapping population development, population genetic structure of the mixture model, QTL mapping, factors affecting the power of QTL mapping and importance of employing MAS systems in crop improvement and marker-trait association analysis using different statistical methods employed in molecular plant breeding research activities.

Key words: Association mapping, markers assisted selection, molecular markers, QTL mapping, QTL analysis

Received: November 20, 2015 Accepted: January 30, 2016 Published: March 15, 2016

Citation: Tewodros Mulualem and Zelalem Bekeko, 2016. Advances in quantitative trait loci, mapping and importance of markers assisted selection in plant breeding research. Int. J. Plant Breed. Genet., 10: 58-68.

Corresponding Author: Tewodros Mulualem, Jimma Agricultural Research Center, P.O. Box 192, Jimma, Ethiopia

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Feeding ever-increasing population is the main challenge faced by the agricultural scientists and to meet this plant breeders and pathologists have to put continuous efforts to develop new crop varieties on fast track basis. The DNA based polymorphism, commonly known as DNA markers can be used for genetic improvement through selection for favorable traits such as disease resistance. Molecular markers are becoming an essential component in backcross breeding programs for tracking the resistance genes in gene pyramiding and physiological plant breeding.

Many agriculturally important traits such as yield, quality and some forms of disease resistance are controlled by many genes and are known as quantitative traits (Polygenic, multi-factorial or complex traits). A quantitative trait is a measurable trait that depends on the cumulative action of many genes and their interaction with the environment that can vary among individuals over a given range to produce a continuous distribution of phenotypes[1]. The genetic variation of a quantitative trait is assumed to be controlled by the collective effects of numerous genes, known as Quantitative Trait Loci (QTLs) and identification of QTLs based only on conventional phenotypic evaluation is not possible[2].

The general goals of QTL mapping in plant breeding are to (a) Increase our biological knowledge of the inheritance and genetic architecture of quantitative traits, both within a species and across related species and (b) Identify markers that can be used as indirect selection tools in breeding[3]. A major breakthrough in the characterization of quantitative traits that created opportunities to select for QTLs was initiated by the development of DNA (or molecular) markers in the 1980s. One of the main uses of DNA markers in agricultural research has been in the construction of linkage maps for diverse crop species. Linkage maps have been utilized for identifying chromosomal regions that contain genes controlling simple traits (controlled by a single gene) and quantitative traits using QTL[4]. The process of constructing linkage maps and conducting QTL analysis to identify genomic regions associated with traits is known as QTL mapping (Genetic, gene or genome mapping)[4,5]. The DNA markers that are tightly linked to agronomically important genes (called gene tagging) may be used as molecular tools for Marker-Assisted Selection (MAS) in plant breeding[6].

The MAS involves the presence/absence of a marker as a substitute for or to assist in phenotypic selection, in a way which may make it more efficient, effective, reliable and cost-effective compared to the more conventional plant breeding methodology. The use of DNA markers in plant breeding has opened a new realm in agriculture called molecular plant breeding[7]. Currently information regarding the application of QTL mapping in plant breeding programs in Ethiopia is very scanty. Therefore, the objective of this review is to assess the current advances in QTL analysis, mapping population development, QTL mapping and importance of employing MAS systems in plant breeding which can be used by plant breeders, plant pathologists and geneticists in any crop improvement program envisioned in the country.

DEVELOPMENT OF MAPPING POPULATION

Genes are situated on chromosomes in a linear order such that their position with respect to each other along the chromosome can be determined. Such a placement of genes on chromosomes constitutes gene mapping which is helpful for their cloning and transfer to desired places[8]. All the genes situated on a specific chromosome constitute a linkage group out of which any one of gene serves to identify the concerned chromosome.

Assignment of a new gene to appropriate linkage group is based up on its association in the transmission with either of the genes in a linkage group. The lack of independent transmission of two or more genes indicates their location on the same chromosome through the reverse may not always be true because genes situated on different chromosomes or the one sparsely (>50 cM) located even on the same chromosome can also show independent transmission. In such cases either some additional marker genes or specific cytogenetic stocks like aneuploids have to be used in the analysis. The true position of genes along the DNA in kbp constitutes the physical map where as a genetic map refers the relative position of genes determined from the frequency of recombinants of alleles of two genes. The genes situated on a chromosome recombine through the formation of chiasmata.

The construction of a linkage map requires a segregating plant population (i.e., a population derived from sexual reproduction). The parents selected for the mapping population will differ for one or more traits of interest. Population sizes used in preliminary genetic mapping studies generally range from 50-250 individuals[4] however, larger populations are required for high-resolution mapping. If the map will be used for QTL studies, then an important point to note is that the mapping population must be phenotypically before subsequent QTL mapping.

In self-pollinating species, mapping populations originate from parents that are both highly homozygous (inbred). In cross pollinating species, the situation is more complicated since most of these species do not tolerate inbreeding. Many cross pollinating plant species are also polyploid (contain several sets of chromosome pairs). Mapping populations used for mapping cross pollinating species may be derived from a cross between a heterozygous parent and a haploid or homozygous parent[9]. The F_1 generation mapping populations were successfully developed by pair crossing heterozygous parental plants that were distinctly different for important traits associated with plant persistence and seed yield[10].

Several different populations may be utilized for mapping within a given plant species, with each population type possessing advantages and disadvantages[5] (Fig. 1). The F_2 populations, derived from F_1 hybrids and backcross (BC) populations, derived by crossing the F_1 hybrid to one of the parents, are the simplest types of mapping populations developed for self pollinating species. Their main advantages are that they are easy to construct and require only a short time to produce. Inbreeding from individual F_2 plants allows the construction of Recombinant Inbred (RI) lines, which consist of a series of homozygous lines, each containing a unique combination of chromosomal segments from the original parents (Fig. 1).

The length of time needed for producing RI populations is the major disadvantage, because usually six to eight generations are required. Doubled Haploid (DH) populations may be produced by regenerating plants by the induction of chromosome doubling from pollen grains, however, the production of DH populations is only possible in species that are amenable to tissue culture (e.g., cereal species such as rice, barley and wheat). The major advantages of RI and DH populations are that they produce homozygous or true-breeding lines that can be multiplied and reproduced without genetic change occurring. This allows for the conduct of replicated trials across different locations and years (Fig. 1). Thus both RI and DH populations represent eternal resources for QTL mapping[5].

Types of mapping populations: The choice of appropriate mapping population is very critical for the success of any QTL mapping. Populations for QTL mapping can be broadly classified into two: experimental populations for linkage-based QTL mapping (e.g., inbred lines for autogamous or self pollinating species, half- or full- sib families for out-crossing or cross pollinating species) and natural or breeding populations for linkage disequilibrium-based association mapping (Fig. 1).

For association mapping, the populations can be classified into one of the following five groups[11] (1) Ideal sample with subtle population structure and familial relatedness, (2) Multi-family sample, (3) Sample with population structure, (4) Sample with both population structure and familial relationships and (5) Sample with severe population structure and familial relationships. Due to local adaptation, selection and breeding history in many plant species, many populations for association mapping would fall into category four[12]. Alternatively, populations for association mapping can be classified according to the source of materials as germplasm bank collections, synthetic populations and elite germplasm[13].

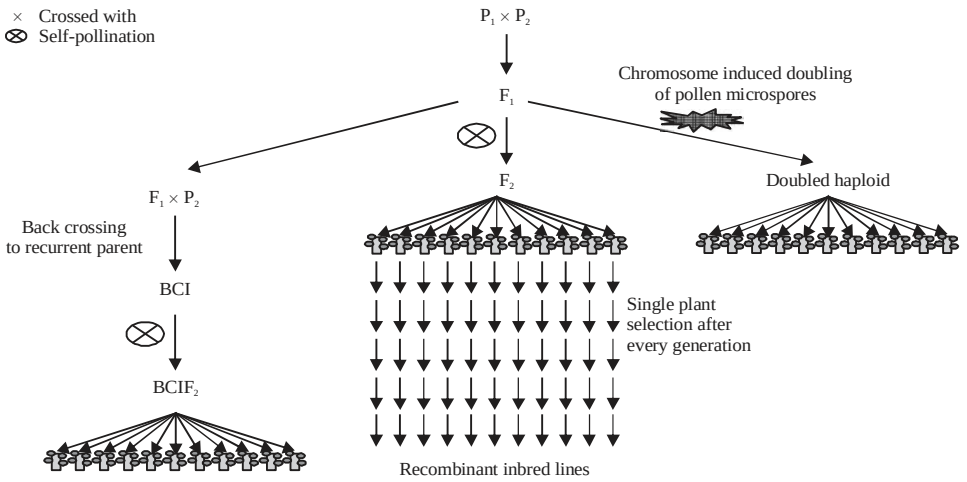

Fig. 1: Mapping populations for self-pollinating species, Source: Paterson[5]

Linkage-based QTL mapping depends on well defined populations developed by crossing two parents. In autogamous species, QTL mapping studies make use of F_2 or Fx derived families, backcross (BC), Recombinant Inbred Lines (RILs), Near Isogenic Lines (NILs) and Double Haploids (DH). These populations are developed by crossing two inbred parents with clear contrasting difference in phenotypic trait (s) of interest.

Each mapping population developed from inbred parents has its own advantages and disadvantages and vital to decide the appropriate population depending on our objectives of interest, trait complexity, available time and whether the molecular markers to be used for genotyping are dominant or co-dominant. Both F_2 and BC populations are the simplest types of mapping populations because they are easy to construct and require only a short time to produce. The F_2 is more powerful for detecting QTLs with additive effects and can also be used to estimate the degree of dominance for detected QTLs. When dominance is present, backcrosses give biased estimates of the effects because additive and dominant effects are completely confounded[14].

However, both F_2 and BC populations have three limitations. First, development of these populations require relatively few meioses such that even markers that are far from the QTLs remain strongly associated with it. Such long-distance associations hamper precise localization of the QTLs. Second, F_2 and backcross populations are temporary populations as they are highly heterozygous and cannot be propagated indefinitely through seeds (i.e., these populations can't be evaluated several times in different environmental conditions, years, locations, etc.). Finally, epistatic interactions could hardly be studied in both F_2 and backcross populations. The mapping population can be developed through different genetic structures models (Fig. 2).

POPULATION GENETIC STRUCTURE OF THE MIXTURE MODEL

The mixture proportions are actually the frequencies of QTL genotypes at a putative QTL in a population. For an entire mapping population initiated with two inbred lines, the frequencies of QTL genotypes can be predicted on the basis of the first Mendelian law. For example, the frequencies of QTL genotypes are 1/2 and 1/2 for the backcross or 1/4, 1/2 and 1/4 for the F_2. However, when marker information is associated with a putatively linked QTL, the frequencies of QTL genotypes given a particular marker genotype will not obey the Mendelian law but rather depend on the recombination fraction or linkage disequilibrium between the marker and QTL. The conditional probabilities of QTL genotypes upon the marker genotypes are the bridge associating the known marker information with unknown QTL information.

BACKCROSS/F_2

Two different inbred lines are crossed to generate the heterozygous F_1. When the F_1 is backcrossed to the original parents, or selfed or sibling-mated, different genes will be co-segregating, which leads to non-parental, recombinant types whose proportion depends on the degree of linkage

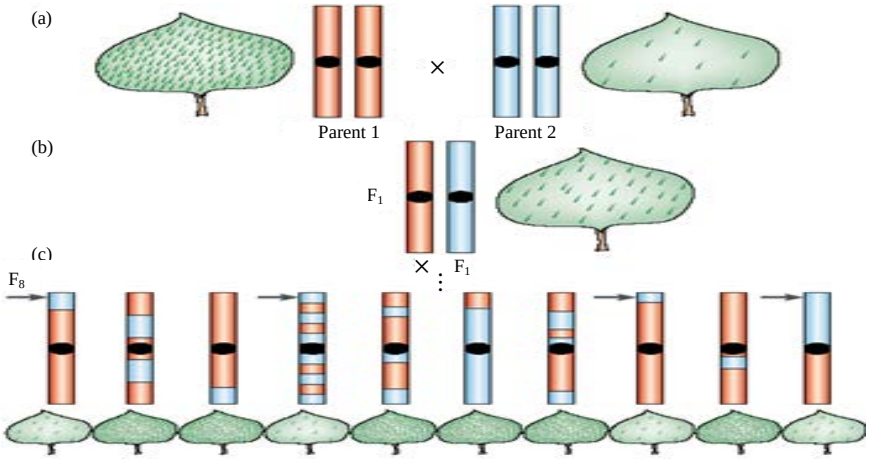

Fig. 2(a-c): Principles of mapping Quantitative Trait Loci (QTL)[13], (a) Inbred parents that differ in the density of trichomes (Parent 1: high trichome density, Parent 2: Low trichome density) are crossed to form an F_1 population with intermediate trichome density, (b) An F_1 individual is selfed to form a population of F_2 individuals and (c) Each F_2 is selfed for six additional generations ultimately forming a set of Recombinant Inbred Lines (RILs)

between different genes. By observing the number of recombinant types, easily estimate the linkage. However, the recombinants between a marker and QTL are not observable, because the QTL position is unknown and there is need to derive the conditional probabilities of QTL genotypes given known marker genotypes in terms of their recombination fractions. The advantage of the backcross or F_2 as a mapping population lies in the clear linkage phase between all genes, from which the parental origin of alleles can be precisely determined[15].

OUT BRED CROSSES

Many species, such as forest trees, cannot generate inbred lines because of their complicated biological features but they can be crossed to generate a segregating progeny population called an out bred cross. It is possible to use out cross progeny as an out bred population because crossover events occur during meioses. In fact, out bred lines as parents can be homozygous at some loci but are heterozygous at many loci and thus their controlled crosses can be backcross-like for some loci, F_2-like for other loci, or present new cross types. The QTL genotypes in out bred crosses can be inferred by their conditional probabilities given marker genotypes expressed as a function of the recombination fraction. A procedure is necessary to determine a correct linkage phase prior to the estimation of linkage.

RECOMBINANT INBRED LINES

Recombinant Inbred Lines (RILs) are powerful material for genetic mapping. They can be derived either by repeated selfing or by repeated brother-sister mating of the progeny from an F_1 cross between two inbred lines. The RILs can serve as a permanent mapping population for multiple uses because they are fixed and homozygous for two alternative alleles at all genes. Some lines are the same as parental (non-recombinant) types, whereas the others are recombinant types. The conditional probabilities of QTL genotypes given marker genotypes are derived in terms of the proportion of recombinant zygotes.

NATURAL POPULATIONS

For some species in which crosses are not possible, the co-segregation between the marker and QTL can be specified by Linkage Disequilibrium (LD). The LD represents nonrandom associations between different loci in a population and can be used to analyze an unstructured population. Linkage analysis has been widely used for genetic mapping by detecting the degree of LD between the marker and QTL. Unlike controlled crosses, the conditional probabilities of QTL genotypes given marker genotypes are expressed in terms of LD values[16].

QUANTITATIVE GENETIC STRUCTURE OF THE MIXTURE MODEL

It is generally assumed that a quantitative trait (y) of interest at a putative QTL is normally distributed with expected mean μj and variance σ^2. The genetic contributions of the QTL to the trait phenotype are reflected in the mean or variance of the normal distribution. For a given QTL genotype j, partition in to its expected mean (μj) into the different components: overall mean (μ), additive effect (a), dominance effect (d), additive × additive (iaa), additive × dominance (iad), dominance × additive (ida) and dominance × dominance (idd) epistatic effects. If it is assume that these effects are fixed, they can be directly estimated in the mixture model. As the variances of the fixed effects are viewed as zero, the variance contains only the residual variance σ^2 within a particular QTL genotype. The major task of a fixed-model-based mapping approach is to specify the genetic components of μj based on different genetic problems or mapping purposes.

ADDITIVE-DOMINANCE MODEL

In many cases, it is reasonable to assume that there is no non allelic interactions (epistasis) between different QTLs. Consider a QTL of three possible genotypes whose values and residual variances are defined as (Table 1).

In most situations, the residual variance is assumed to be identical among the three groups of QTL genotypes for the sake of computational simplicity. Statistical techniques are also available when the residual variances are genotype-specific (Table 1). When two or more QTLs are fit, more QTL genotypes will be included in the mixture model. For an F_2 population, the genotypic values, $\mu j_1 j_2$ (j1, j2 = 0, 1, 2) of nine QTL genotypes for two given QTLs, P and Q under the additive-dominance model can be defined as:

$$\begin{array}{cc} & \text{Q} \\ \text{P} & \begin{array}{ccc} 2 & 1 & 0 \end{array} \\ \mu j_1 J_2 = \begin{array}{c} 2 \\ 1 \\ 0 \end{array} & \left\{\begin{array}{ccc} \mu 22 = \mu + a1 + a2 & \mu 21 = \mu + a1 + d2 & \mu 20 = \mu + a1 - a2 \\ \mu 12 = \mu + d1 + a2 & \mu 11 = \mu + d1 + d2 & \mu 10 = \mu + d1 - a2 \\ \mu 02 = \mu - a1 + a2 & \mu 01 = \mu - a1 + d2 & \mu 00 = \mu - a1 - a2 \end{array}\right\} \end{array}$$

Table 1: QTL of three possible genotypes whose values and residual variances

Genotypes	QQ (2)	Qq (1)	qq (0)
Overall mean	μ	μ	μ
Effect	a	d	-a
Genotypic value	μ2 = μ+a	μ2 = μ+d	μ2 = μ-a
Residual	σ^2	σ^2	σ2

QLT: Quantitative trait loci

where, a1 and a2 are the additive effects and d1 and d2 are the dominance effects of these two QTLs, respectively.

ADDITIVE-DOMINANCE-EPISTASIS MODEL

If two or more QTLs interact to affect a quantitative trait, their epistatic effects should be modeled in the QTL genotypic values within the mixture model.

$$\begin{array}{cc} & \begin{array}{ccc} P & Q & \\ & 2 \quad\quad 1 \quad\quad 0 & \end{array} \\ \mu j_1 J_2 = & \begin{array}{c} 2 \\ 1 \\ 0 \end{array} \left\{ \begin{array}{ccc} \mu+a1+a2+i_{aa} & \mu+a1+d_2+i_{ad} & \mu+a1-a2+i_{aa} \\ \mu+d1+a2+ida & \mu+d1+d2+i_{dd} & \mu+d1-a2-+i_{ad} \\ \mu-a1+a2-i_{aa} & \mu-a1+d2-i_{ad} & \mu-a1-a2+iaa \end{array} \right\} \end{array}$$

where, i_{aa}, i_{ad}, i_{da} and i_{dd} are the additive×additive, additive×dominance, dominance×additive and dominance ×dominance epistatic effects, respectively.

MULTIPLICATIVE-EPISTATIC MODEL

The physiological basis of epistasis has been studied by modeling the relationship between genes and their products in many plant and animal experiments. The multiplicative interaction between a pair of loci may be an important form of epistasis for controlling complex traits, as anticipated. The multiplicative-epistasis model assumes that genotypes at a pair of loci have genotypic values equal to the product of genotypic values at the two different loci. For example, if genotypic values are μ2 for P genotype 2 and μ-2 for Q genotype 2, then the value of joint QTL genotype 22 is μ22 = μ2μ-2. Under the multiplicative-epistatic model, the genotypic values of two QTL can be modelled by:

$$\begin{array}{cc} & \begin{array}{ccc} P & Q & \\ & 2 \quad\quad 1 \quad\quad 0 & \end{array} \\ \mu j_1 J_2 = & \begin{array}{c} 2 \\ 1 \\ 0 \end{array} \left\{ \begin{array}{ccc} \mu_{22} = \mu_2 \mu'_2 & \mu 21 = \mu 2\ \mu' 1 & \mu 20 = \mu 2\ \mu' 0 \\ \mu 12 = \mu 1\ \mu' 2 & \mu 11 = \mu 1\ \mu' 1 & \mu 10 = \mu 1\ \mu' 0 \\ \mu 02 = \mu 0\ \mu' 2 & \mu 01 = \mu 0\ \mu' 1 & \mu 00 = \mu 0\ \mu' 0 \end{array} \right\} \end{array}$$

Although, multiplicative interactions between a pair of QTL are considered, two special cases, completely multiplicative action (both between and within loci) and pure additive action (without dominance) can also be manipulated by setting restrictions.

QTL MAPPING

Great majority of the economically important characters are inherited quantitatively and presumed to be controlled by a large number of genes called polygene whose exact number, mode of action and location is difficult to be ascertained through Mendelian analysis[17,18]. The association of this analysis can provide evidence for the genetic control of traits variation but is not very precise because the genetic effects associated with marker genotypes are confounded by the position of a functional QTL and its actual effect. If markers are so highly dense that they are generated at QTL positions, a simple marker-phenotype association analysis may be useful. The generation of such high-density maps is not possible for a majority of species in practice. Powerful analytical techniques are needed to separate the effects of a QTL from its location.

Unlike molecular markers, the genomic locations of QTLs are unknown and should be inferred on the basis of the association analysis of the phenotypes and markers. The role of statistical methods is in the identification, mapping and estimation of functional QTLs using location known, neutral markers. One of the most important statistical foundations for QTL mapping is laid out in the mixture different quantitative models, in which each observation is assumed to have arisen from one of unobservable QTL genotype groups, each group being suitably modeled by a density from some parametric family. This model provides a framework by which observations may be clustered together into genotype groups for discrimination.

PRINCIPLE OF QTL MAPPING

It is not difficult in populations of most crop plants to identify and map a good number of segregating markers (10-50) per chromosome. However, most of these markers would be in non-coding regions of the genome and might not affect the trait of interest directly; but, a few of these markers might be linked to genomic regions (QTLs) that do influence the trait of interest. Where such linkage occurs, the marker locus and the QTL will co-segregate. Therefore, the basic principle of determining whether a QTL is linked to a marker is to partition the mapping population into different genotypic classes based on genotypes at the marker locus and the apply correlative statistics to determine whether the

individuals of one genotype differ significantly with the individuals of other genotype with respect to the trait being measured.

Situations where genes fail to segregate independently are said to display linkage disequilibrium. The QTL analysis, thus, depends on linkage disequilibrium.

With natural populations, consistent association between QTL and marker genotype will not usually exist, except in a very rare situation where the marker is completely linked to the QTL. Therefore, QTL analysis is usually undertaken in segregating mapping populations, such as F_2-derived populations, Recombinant Inbred Lines (RILs), nearisogenic lines (NILs), doubled haploid lines (DHs) and backcross populations.

FACTORS AFFECTING THE POWER OF QTL MAPPING

The QTLs are statistically inferred from the data generated in an experiment. However, statistical influence does not always indicate biological significance due to multiple test problems associated with QTL mapping[19]. The following factors affect the power of QTL mapping:

- Number of genes controlling the target trait(s) and their genome positions
- Distribution of genetic effects and existence of genetic interactions
- Heritability of the trait
- Number of genes segregating in a mapping population
- Type and size of mapping population used
- Density and coverage of markers in the linkage map
- Statistical methodology employed and significance level used for QTL mapping
- The number of environments
- The phenotyping and genome coverage

The possible means to improve the power of QTL mapping are available[19,20]. Replicate progeny analysis, selective genotyping, sample pooling and sequential sampling are some of the suggested approaches for optimization of experimental designs, so as to enhance the power of QTL detection and estimation of QTL effects.

METHODS FOR QTL MAPPING

The basic objective in QTL mapping studies is to detect QTL while minimizing the occurrence of false positives (Type I errors, that is, declaring an association between a marker and QTL when in fact one does not exist). Tests for QTL/trait association are often performed by the following approaches:

Single marker approach: The single marker approach, sometimes referred to as the single factor analysis of variance (SF-ANOVA) or single point analysis, has been used extensively, especially with isozymes[21]. The SF-ANOVA is done for each marker locus independent of information from other loci. The F-tests provide evidence whether differences between marker locus genotype classes are significant or not. Although, computationally simple, this approach suffers from some major limitations: (1) The likelihood of QTL detection significantly decreases as the distance between the marker and QTL increases, (2) The method cannot determine whether the markers are associated with one or more QTLs and (3) The effects of QTL are likely to be underestimated because they are confounded with recombination frequencies.

Simple Interval Mapping (SIM): The SIM was first proposed by Lander and Botstein[17] and it takes full advantage of a linkage map. The method evaluates the target association between the trait values and the genotype of a hypothetical QTL (target QTL) at multiple analysis points between pair of adjacent marker loci (the target interval). Presence of a putative QTL is estimated if the log of odds ratio (LOD) exceeds a critical threshold. Lander and Botstein[17] developed formulae for calculating significance levels appropriate for interval mapping when the genome size, number of chromosomes, number of marker intervals and the overall false positive rate desired are given. The SIM has been the most widely approach as it can be easily accessed through statistical packages such as MAPMAKER/QTL. By using tightly linked markers for analysis, it is possible to compensate for recombination between markers and the QTL, thereby increasing the probability of statistically detecting the QTL and providing an unbiased estimate of QTL effect. However, when multiple QTLs are segregating in a cross (which is usually the case), SIM fails to take into account genetic variance caused by other QTLs. In such a case, SIM suffers essentially from the same shortcomings of single marker analysis.

Composite Interval Mapping (CIM): The combine interval mapping for a single QTL in a given interval with multiple regression analysis on marker associated with other QTL. It considers a marker interval plus a few other well-chosen single markers in each analysis, so that n-1 tests for interval-QTL

associations are performed on a chromosome with n markers. The advantages of CIM are: (1) Mapping of multiple QTLs can be accomplished by the search in one dimension, (2) By using linked markers as cofactors, the test is not affected by QTL outside the region, thereby increasing the precision of QTL mapping and (3) By eliminating much of the genetic variance by other QTL, the residual variance is reduced, thereby increasing the power of detection of QTL. The CIM is more powerful than SIM but is yet to be used extensively in QTL mapping.

Multiple Interval Mapping (MIM): Another recent and interesting development is Multiple Intervals Mapping (MIM). The MIM is the extension of interval mapping to multiple QTLs, just as multiple regressions extends analysis of variance. The MIM allows one to infer the location of QTLs to positions between markers, makes proper allowance for missing genotype data and can allow interactions between QTLs. Model selection is the principal problem in multiple QTL methods the chief concern is the formation of appropriate criteria for comparing models. The simplest multiple QTL method, multiple regression, should be used more widely, although like analysis of variance, it suffers in the presence of appreciable missing marker genotype data. A forward selection procedure using interval mapping (i.e., the calculation of conditional LOD curves) is appropriate in cases of QTLs that act additively and makes proper allowance for missing genotype data. The MIM is an improved method that, although computationally intensive, can in principle, map multiple QTLs and identify interactions between QTLs.

In determining whether a LOD score is sufficiently large for one to be confident of the presence of a QTL, consider the distribution of the LOD score under the null hypothesis of no segregating QTL. Adjustment must be made for the genome wide search for QTLs, so consider the distribution of the maximum LOD score genome wide. Permutation tests are valuable for determining significance landmarks for the LOD score, although computationally intensive, permutation tests allow for the observed phenotype distribution, marker density and pattern of missing genotype data. Once QTL are detected, the next step is to estimate the genotypic effect of the QTL and to localize the QTL to a precise genomic region. The interval mapping approach is superior to the ANOVA approach in terms of both localization and estimation of the effects of QTL. However, the reliability in terms of estimation of the QTL effect depends on the linkage between marker(s) and QTL, the number and type of progeny evaluated and the heritability of the trait. From multiple regression analysis, one can also obtain an R^2 value which gives the percentage of the total genetic variance explained by all of the markers. The R^2 value for the line is considered to be the amount of total genetic variation that is explained by the specific molecular marker.

Recently developed statistical packages also offer the means to analyze the QTL x environment interactions. Besides the QTL mapping methods described above, several other approaches are available, including Bayesian methods and the use of a genetic algorithm. The most commonly used statistical packages for QTL analysis are MAPMAKER/QTL, QTL Cartographer, PLABQTL, QTL Mapper and Qgene. Most of these statistical packages yield essentially similar QTL locations and gene effects on a given data set while there could be slight variation in the confidence intervals. Recent advances using nonparametric statistics or association-based approaches to identify QTLs and to calculate empirical critical (threshold) values for declaring significant QTLs will further refine QTL mapping.

INDICATION OF DNA MARKERS AND MARKER ASSISTED SELECTION

The fundamental basis of plant breeding is the selection of specific plants with desirable traits. Selection typically involves evaluating a breeding population for one or more traits in field trials (e.g., agronomic traits, disease resistance or stress tolerance) or with chemical tests (e.g., grain quality).

The goal of plant breeding is to assemble more desirable combinations of genes in new varieties. Standard breeding techniques for inbreeding cereal crops have been outlined[22]. In the commonly used pedigree breeding method, selecting desirable plants begins in early generations for traits of higher heritability. However, for traits of low heritability, selection is often postponed until the lines become more homozygous in later generations (F_5 or F_6). Selection of superior plants involves visual assessment for agronomic traits or resistance to stresses, as well as laboratory tests for quality or other traits. When the breeding lines become homozygous (F_5 or later) they can be harvested in bulk and evaluated in replicated field trials. The entire process involves considerable time (5-10 years for elite lines to be identified) and expense. The size and composition of a plant population is an important consideration for a breeding program.

The larger the number of genes segregating in a population, the larger the population size required in order to identify specific gene combinations. Typical breeding programmes usually grow hundreds or even thousands of populations and many thousands or millions of individual plants[23]. Given the extent and complexity of selection required in breeding programmes and the number

and size of populations, one can easily appreciate the usefulness of new tools that may assist breeders in plant selection. The scale of breeding programmes also underlines the challenges of incorporating a relatively expensive technology such as MAS.

MAIN TYPES OF DNA MARKERS USED IN MAS

There are five main considerations for the use of DNA markers in MAS: reliability; quantity and quality of DNA required, technical procedure for marker assay, level of polymorphism and cost[24].

Reliability: Markers should be tightly linked to target loci, preferably less than 5 cM genetic distance. The use of flanking markers or intragenic markers will greatly increase the reliability of the markers to predict phenotype.

DNA quantity and quality: Some marker techniques require large amounts and high quality of DNA, which may sometimes be difficult to obtain in practice and this adds to the cost of the procedures.

Technical procedure: The level of simplicity and the time required for the technique are critical considerations. High-throughput simple and quick methods are highly desirable.

Level of polymorphism: Ideally, the marker should be highly polymorphic in breeding material (i.e., it should discriminate between different genotypes), especially in core breeding material.

Cost: The marker assay must be cost-effective in order for MAS to be feasible.

The most widely used markers in major cereals crop are called Simple Sequence Repeats (SSRs) or microsatellites[25]. They are highly reliable (i.e., reproducible), co-dominant in inheritance, relatively simple and cheap to use and generally highly polymorphic.

The only disadvantages of SSRs are that they typically require polyacrylamide gel electrophoresis and generally give information only about a single locus per assay, although, multiplexing of several markers is possible[26]. These problems have been overcome in many cases by selecting SSR markers that have large enough size differences for detection in agarose gels, as well as multiplexing several markers in a single reaction. The SSR markers also require a substantial investment of time and money to develop and adequate numbers for high-density mapping are not available in some orphan crop species (Fig. 3). Sequence Tagged Site (STS), Sequence Characterized Amplified Region (SCAR) or Single Nucleotide Polymorphism (SNP) markers that are derived from specific DNA sequences of markers (e.g., restriction fragment length polymorphisms: RFLPs) that are linked to a gene or Quantitative Trait Locus (QTL) are also extremely useful for MAS[27,28].

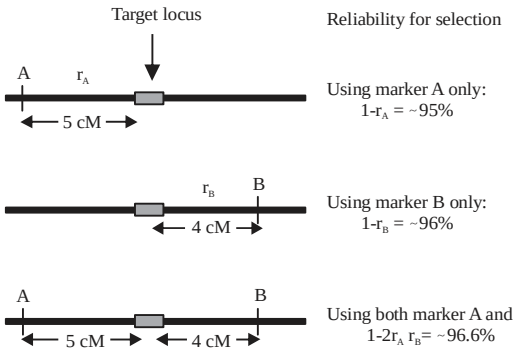

Fig. 3: Reliability of selection using single and flanking markers

QTL MAPPING AND MAS

The detection of genes or QTLs controlling traits is possible due to genetic linkage analysis, which is based on the principle of genetic recombination during meiosis[29]. This permits the construction of linkage maps composed of genetic markers for a specific population. Segregating populations such as F_2, F_3 or backcross (BC) populations are frequently used. However, populations that can be maintained and produced permanently, such as recombinant inbreds and doubled haploids, are preferable because they allow replicated and repeated experiments[30]. These types of populations may not be applicable to outbreeding cereals where inbreeding depression can cause non-random changes in gene frequency and loss of vigour of the lines. Using statistical methods such as single-marker analysis or interval mapping to detect associations between DNA markers and phenotypic data, genes or QTLs can be detected in relation to a linkage map[31]. The identification of QTLs using DNA markers was a major breakthrough in the characterization of quantitative traits[32].

IMPORTANCE OF QTL FOR MAS IN PLANT BREEDING

One of the most important aims of these marker analyses in plant breeding is to provide ordered hallmarks on

chromosomes with which one can map functional Quantitative Trait Loci (QTLs) determining complex phenotypic variation to particular genomic regions. The genome wide identification of QTLs, their locations and effects is one of fundamental importance for agricultural, evolutionary and biomedical genetics[33].

A variety of methods have been developed for QTL mapping. These methods can be classified as t-tests and analysis of variance, least-squares analysis (LS), maximum-likelihood analysis (ML) and Bayesian analysis. These methods differ in computational requirements, efficiency in terms of extracting information, flexibility with regard to handling different data structures and ability to map multiple QTLs. The simple LS method is efficient in terms of computational speed but cannot extract all information from the data and is restricted to specific mating designs. The technique of ML interval mapping is one of the most widely used methods for QTL analysis in controlled crosses or structured pedigrees. The interval mapping method has been extended to composite interval mapping and multiple interval mapping.

The t-test, analysis of variance and regression analysis of multiple markers and perform statistical tests based solely on single DNA marker information. For single-marker analysis, no genetic map is required and the calculations are based on phenotypic means and variances within each of the genotypic classes. Marker analysis can be extended to include all markers of the genome. Although, single marker analyses, confound the QTL effect and the QTL location, they provide preliminary results that facilitate the use of more advanced interval mapping to detect QTLs within a genomic interval bracketed by two linked markers. The ML interval mapping and its extension, composite interval mapping will be presented.

A QTL statistical model assumes that the QTL genotypes can be observed in a mapping population. This is not possible in practice. What we can do is to use observable markers to predict such unobservable QTLs through the linkage between markers and QTLs. Thus, by performing the association analysis between the markers and phenotypes, we can still infer the effect of a putative QTL on phenotypic variation. The use of a single-marker is limited for QTL identification since it cannot determine at which side of the marker, left or right the QTL is located. However, single marker analyses are useful for a preliminary test of the existence of a QTL, although they cannot estimate the QTL location. Two testing approaches for marker analysis based were introduced on the t and F test statistics.

CONCLUSION

Quantitative traits are controlled by polygenes and the regions within genomes that contain genes associated with a particular quantitative trait are known as Quantitative Trait Loci (QTLs). The identification of QTLs based on conventional evaluation is not possible. Therefore, the uses of DNA markers with construction of linkage maps for diverse crop species are vital to identifying chromosomal regions that contain genes controlling simple traits (controlled by a single gene) and quantitative traits using QTL analysis. Application of linkage maps and QTL analysis are important tool to identify genomic regions associated with traits. The DNA markers that are tightly linked to agronomically important genes may be used as molecular tools for Marker-Assisted Selection (MAS) in plant breeding. Despite optimism about continued yield improvement from conventional breeding, new technologies such as DNA marker technology will be needed to maximize the probability of success.

REFERENCES

1. Sham, P., J.S. Bader, I. Craig, M. O'Donovan and M. Owen, 2002. DNA pooling: A tool for large-scale association studies. Nat. Rev. Genet., 3: 862-871.
2. Lynch, M. and B. Walash, 1998. Genetic and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA., ISBN 0-87893-481-2, Pages: 980.
3. Bernardo, R., 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci., 48: 1649-1664.
4. Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, Y. Masohiro, C.R. Bhatia and T. Sasaki, 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed., 3: 87-103.
5. Paterson, A.H., 1996. Making Genetic Maps. In: Genome Mapping in Plants, Paterson, A.H. (Ed.). Academic Press, Austin, Texas, ISBN: 9780125465908, pp: 23-39.
6. Ribaut, J.M. and D. Hoisington, 1998. Marker-assisted selection: New tools and strategies. Trends Plant Sci., 30: 236-239.
7. Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol., 5: 94-100.
8. Semagn, K., A. Bjornstad, H. Skinnes, A.G. Maroy, Y. Tarkegne and M. William, 2006. Distribution of DArT, AFLP and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome, 49: 545-555.
9. Wu, R., C.X. Ma, I. Painter and Z.B. Zeng, 2002. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor. Populat. Biol., 61: 349-363.

10. Forster, J.W., E.S. Jones, R. Kolliker, M.C. Drayton and J.L. Dumsday *et al*., 2001. Development and Implementation of Molecular Markers for Forage Crop Improvement. In: Molecular Breeding of Forage Crops, Spangenberg, G. (Ed.). Springer, Netherlands, ISBN: 978-90-481-5676-4, pp: 101-133.
11. Yu, K., S.J. Park and V. Poysa, 2000. Marker-assisted selection of common beans for resistance to common bacterial blight: Efficacy and economics. Plant Breed., 119: 411-415.
12. Zhu, C., M. Gore, E.S. Buckler and J. Yu, 2008. Status and prospects of association mapping in plants. Plant Genome, 1: 5-20.
13. Breseghello, F. and M.E. Sorrells, 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics, 172: 1165-1177.
14. Carbonell, E.A., M.J. Asins, M. Baselga, E. Balansard and T. M. Gerig, 1993. Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. Theor. Applied Genet., 86: 411-416.
15. Hospital, F., 2005. Selection in backcross programmes. Philosophical Trans.: Biol. Sci., 360: 1503-1511.
16. Liu, S. and J.A. Anderson, 2003. Targeted molecular mapping of a major wheat QTL for *Fusarium* head blight resistance using wheat ESTs and synteny with rice. Genome, 46: 817-823.
17. Lander, E.S. and D. Botstein, 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 121: 185-199.
18. Paterson, A.H., S. Damon, J.D. Hewitt, D. Zamir and H.D. Rabinowitch *et al*., 1991. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations and environments. Genetics, 127: 181-197.
19. Liu, B.H., 2002. Statistical Genomics: Linkage, Mapping and QTL Analysis. 2nd Edn., CRC Press, Boca Raton, Florida.
20. Kearsey, M.J. and A.G.L. Farquhar, 1998. QTL analysis in plants; where are we now? Heredity, 80: 137-142.
21. Edwards, M.D., C.W. Stuber and J.F. Wendel, 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics, 116: 113-125.
22. Allard, R.W., 1999. Principles of Plant Breeding. 2nd Edn., John Wiley and Sons, New York, USA., ISBN-13: 9780471023098, Pages: 254.
23. Witcombe, J.R. and D.S. Virk, 2001. Number of crosses and population size for participatory and classical plant breeding. Euphytica, 122: 451-462.
24. Mohler, V. and C. Singrun, 2004. General Considerations: Marker-Assisted Selection. In: Molecular Marker Systems in Plant Breeding and Crop Improvement, Lorz, H. and G. Wenzel (Eds.). Springer, Berlin, Germany, ISBN: 978-3-540-20689-7, pp: 305-317.
25. Gupta, P.K. and P.K. Varshney, 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica, 113: 163-185.
26. Hospital, F. and A. Charcosset, 1997. Marker-assisted introgression of quantitative trait loci. Genetics, 147: 1469-1485.
27. Sanchez, A.C., D.S. Brar, N. Haung, Z. Li and G.S. Khush, 2000. Sequence tagged site marker-assisted selection for three bacterial blight resistance genes in rice. Crop Sci., 40: 792-797.
28. Sharp, P.J., S. Johnston, G. Brown, R.A. Mcintosh and M. Pallotta *et al*., 2001. Validation of molecular markers for wheat breeding. Austr. J. Agric. Res., 52: 1357-1366.
29. Tanksley, S.D., 1993. Mapping polygenes. Ann. Rev. Gene, 27: 205-233.
30. Hittalmani, S., A. Parco, T.V. Mew, R.S. Zeigler and N. Huang, 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor. Applied Genet., 100: 1121-1128.
31. Kearsey, M.J., 1998. The principles of QTL analysis (a minimal mathematics approach). J. Exp. Bot., 49: 1619-1623.
32. Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln and S.D. Tanksley, 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature, 335: 721-726.
33. Ribaut, J.M. and J. Betran, 1999. Single large-scale marker assisted selection (SLS-MAS). Mol. Breed., 5: 531-541.