Indian Journal of Agriculture and Allied Sciences
A Refereed Research Journal

ISSN 2395-1109
Volume: 1, No.: 4, Year: 2015

Received: 10.10.2015, Accepted: 25.11.2015

QTL MAPPING AND ITS APPLICATIONS IN PLANT BREEDING

Dan Singh Jakhar, Abhishek Singh, Mukh Ram and Rajesh Singh
Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi–221 005, UP, India, E-mail; dansingh410@gmail.com, Corresponding Author: Dan Singh Jakhar

Abstract: *A QTL is defined as "a region of the genome that is associated with an effect on a quantitative trait". Conceptually, a QTL can be a single gene, or it may be a cluster of linked genes that affect the trait. The procedure for finding and locating QTLs on a particular chromosome is called QTL mapping. QTL mapping is based on the principle of detecting an association between phenotype and the genotype of markers expecting that genes and markers segregate via chromosome recombination. QTL mapping studies have been reported in most crop plants for diverse traits including yield, quality, disease and insect resistance, abiotic stress tolerance, and environmental adaptation. Here, a brief overview of the principle of QTL mapping, salient requirements for QTL mapping, common statistical tools and techniques employed in QTL analysis and major problems and factors limiting QTL mapping for crop improvement shall be provided. Hence identification of putative QTL locations and DNA markers linked to QTLs has opened up opportunities for isolation and molecular characterization of QTLs via map-based cloning.*
Keywords: *QTL Mapping, Quantitative trait, Mapping Population, Markers*

Introduction: Quantitative characters have been a major area of study in genetics for over a century, as they are a common feature of natural variation in populations of all eukaryotes, including crop plants. For most of the period up to 1980, the study of quantitative traits has involved statistical techniques based on means, variances and co-variances of relatives. These studied provided a conceptual base for partitioning the total phenotypic variance into genetic and environmental variances, and further analyzing the genetic variance in terms of additive, dominance and epistatic effects. From this information, it became feasible to estimate the heritability of the trait and predict the response of the trait to selection. It was also possible to estimate the minimum number of genes that controlled the trait of interest. However, little was known about what these genes were, where they are located, and how they controlled the trait(s), apart from the fact that for any given trait, there were several such genes segregating in a Mendelian fashion in any given population, and in most cases their effects were approximately additive [1]. These genes were termed 'polygenes' [2]. Sax's (1923) experiment with beans demonstrated that the effect of an individual locus affecting a quantitative trait could be isolated though a series of crosses resulting in randomization of the genetic background with respect to all genes not linked to the genetic markers under observation. Even though all of the markers used by Sax were morphological seed markers with complete dominance, he was able to show a significant effect on seed weight associated with some of his markers. Despite this demonstration, there were extremely few successful detections of marker-QTL linkage in crop plants during 1930-80s, and of these, even fewer were repeated. The major limitation was the lack of availability of adequate polymorphic markers. 'QTL' (Quantitative Trait Loci), a term first coined [3].

Principle of QTL Mapping: Mapping of QTL is based on a systematic search for linkage disequilibrium between marker loci and QTL, In simple terms, QTL analysis is based on the principle of detecting an association between phenotype and the genotype of markers expecting that genes and markers segregate via chromosome recombination.

Application of QTL Mapping: The two general goals of QTL mapping in plants are

1. To know the inheritance pattern of a quantitative trait.

2. To identify markers that can be used as indirect selection tools in breeding.

Basic Requirements for QTL Detection

1. To develop appropriate mapping population
2. Development of a saturated linkage map based on molecular markers
   a. Use of appropriate marker for genotyping mapping population
   b. Use of appropriate number of markers for genotyping mapping population
3. Precise phenotypic data of target traits in mapping population
4. To identify molecular markers linked to the trait(s) of interest using statistical programs

Mapping Population: A population used for QTL mapping is known as mapping population. It may be derived from crossing between 2 or more than 2 parents or may be mixture of diverse germplasm.

Requirements for Development of Mapping Population

1. Target trait needs to be polymorphic between parents.
2. Trait should be heritable.

Factors Affecting Development of Mapping Population

1. Diversity of Parents

- Optimum, excess may cause sterility.
- Too diverse will cause less transgressive segregants.
- Inter-varietal will be better.

2. Size of Population: The size of the mapping population for QTL analysis depends on several factors including the type of mapping population employed for analysis, genetic nature of the target trait, objectives of the experiment, and the resources availed for handing a sizable mapping population on terms of phenotyping and genotyping, while analysis of a large number of individuals (500 or more) would enable detection of even QTLs having small effects on the target trait. The basic purpose of QTL mapping would be largely served if one can detect the QTLs with major effects. This would require, the general, a mapping population of a size of 200-300 individuals.

Types of Mapping Populations

1. RILs / SSD Lines

- Develop through selfing or sibmating.
- Complete homozygous lines but small proportion of heterozygosity may be present.
- Product of multiple meiotic divisions.
- Many different recombinants are present.
- Replication and multi-location trails are possible.
- 1:1 with both Dominant & Codominant markers

Merits	Demerits
Permanent Population	Requires more season to develop
Product of multiple meiotic divisions.	Difficult to develop in crops showing high inbreeding depression.
Best for mapping.	Can be estimated only additive components.
Replicated over years, locations are possible	

2. F2 Population

- Simplest to develop
- Outcome of single meiotic division,
- 1:1, (with dominant marker)
- 1:2:1 (with codominant marker)
- Mortal Population
- Replication and multi-location trails can‟t be conducted hence precise phenotyping is not possible

Merits	Demerits
Easy to develop	Mortal population
Requires less time to develop	Linkage is based on one cycle of meiotic division.
For preliminary mapping it is best	G × E cannot be estimated
Can estimate both additive and dominant components.	

3. NILs are Ideal for

- Construction of high-resolution mapping,
- Gene expression profiling,
- Studying phenotypic effect of a QTL (gene tagging).

Merits	Demerits
Permanent Population	Requires more seasons to develop.
Replication and multi-location trail may be conducted.	Directly useful only for molecular tagging of the gene concerned, but not for linkage mapping
Best population for gene tagging.	Linkage drag is a potential problem in constructing NILs, which has to be taken care of.
	Can estimate only additive components.

4. Backcross Mapping Population

- Mortal Population
- F1 is crossed with either parent.
- F1× P1 (Dominant parent) resulted No Ratio (all the plants will be of same type) in case of dominant but codominant marker will give 1:1.
- F1 × P2 (Recessive parent) resulted 1:1 progeny in both the dominant and codominant markers.

Merits	Demerits
Easy to develop	Mortal population
Requires less time to develop	Linkage is based on one cycle of meiotic division.
It is best for mapping quality traits not for quantitative trait.	G × E cannot be estimated

5. Double Haploids

- Anther culture using pollen grains of F1
- Completely homozygous.
- Product of one meiotic division anther culture.
- Comparable to RILs.
- 1:1 in both the cases.

Merits	Demerits
Takes two seasons for development.	Male side gametes are accounted only.
Completely homozygous, no residual heterozygosity	Development of DH is costly, needs equipped lab and skilled personnel.
Permanent population,	
Replicated & Multilocation testing is possible	DH is product of single meiotic product.
Useful for mapping qualitative as well as quantitative traits	Production of DHs is only possible for species with a well established protocol for haploid production.
	Can estimate only additive components.
	Somaclonal variation may arise during development of DH

Generating a Reasonably Saturated Linkage Map: By screening the mapping population using polymorphic molecular markers (popularly called as 'genotyping'), we can analyze the segregation patterns for each of the markers. The segregation patterns are usually in consonance with the type of mapping population used. The genotypic data is then analyzed using a statistical package such as MAPMAKER [4] or JOINMAP [5], for construction of a linkage map of the molecular markers analyzed in the study. Mapping means placing the markers in order, indicating the relative genetic distances between them, and assigning them to their linkage groups on the basis of recombination values from all pairwise combinations between the markers.

To perform a whole-genome QTL scan, it is desirable to have a saturated marker map. In such a map, markers are available for each chromosome from one end to the other, and adjacent markers are spaced sufficiently close that recombination events only rarely occur between them. For practical purposes, this is generally considered to be less than 10 recombinations per 100 meioses, or a map distance of less than 10 centi-Morgans (cM). In the model plant Arabidopsis thaliana, which has a particularly small genome, this requires as few as 50 markers. Several-fold more markers are needed for plant genomes such as wheat and maize. In crops like maize, a broad 'rule- of- the-thumb' is to cover each of the chromosomal locations with at least one or two polymorphic molecular markers.

Phenotyping of Mapping Population and Sample Size: The target quantitative traits have to be measured as precisely as possible and limited amounts of missing data can be tolerated. The power to resolve the QTL location is limited first by sample size, and then by genetic marker coverage of the genome. Often, the number of individuals in a sample might appear to be large, but missing data or skewed allele frequencies in the population cause the effective sample size to diminish, thus sacrificing statistical power. Sometimes, it may be necessary to sacrifice population size in favour of data quality, and this trade-off means that only major QTL can be detected. Data is typically pooled over locations and replications to obtain a single quantitative trait value for the line. It is also preferable to measure the target trait(s) in experiments conducted in multiple locations to have a better understanding of the QTL x Environment interaction, if any.

Statistical 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:

1. 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 [6,7]. SF-ANOVA is

done for each marker locus independent of information from other loci. 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: (i) the likelihood of QTL detection significantly decreases as the distance between the marker and QTL increases; (ii) the method cannot determine whether the markers are associated with one or more QTLs; (iii) the effects of QTL are likely to be underestimated because they are confounded with recombination frequencies.

2. Simple Interval Mapping (SIM): SIM was first proposed [8] 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. 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 [8]. 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 by other $QTL_{s.}$

3. Composite Interval Mapping (CIM): CIM [9] and MQM (multiple-QTL model or marker-QTL-marker analysis) developed [10] 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 as follows: (i) mapping of multiple QTLs can be accomplished by the search in one dimension; (ii) by using linked markers as cofactors, the test is not affected by QTL outside the region, thereby increasing the precision of QTL mapping; and (iii) By eliminating much of the genetic variance by other QTL, the residual variance is reduced, thereby increasing the power of detection of QTL. CIM is more powerful than SIM, but is yet to be used extensively in QTL mapping.

4. Multiple Interval Mapping (MIM): MIM uses multiple marker intervals simultaneously to fit various putative QTL directly into the multiple putative QTL model for mapping QTL, developed MIM. MIM tends to be more powerful than SMA and CIM [11]. MIM leads to more accurate QTL position and QTL effect estimates [12]. MIM is appropriate for the identification and estimation of genetic architecture parameters, including the number, genomic positions effects and interactions of significant QTL and their contribution to the genetic variance.

Major Problems and Factors Limiting QTL Analysis

- The major problem associated with QTL analysis is that the individual QTL effects are small, heritability for most trials is generally less than 50% so that the heritability associated with individual QTL is a small fraction of this [13].
- The efficiency of QTL may also be reduced when the environment and the interaction of the environment and genetic background (G × E), affect largely the final phenotypic trait.
- The accuracy of the analysis may be influenced by the experimental design (including the type of segregating population), its size, the number and contributions of each QTL, evaluation of the trait, trait etc.
- We do not have information on
- Are these QTL coding for specific enzymes involved in a particular pathway, do they act as regulators of gene expression?
- Are they non-coding regions that have some influence in the expression of the trait of interest?

Conclusion: Hence identification of putative QTL locations and DNA markers linked to QTLs has opened up opportunities for isolation and molecular characterization of QTLs via map-based cloning.

References

1. Kearsey, M.J. and Pooni, H.S. (1996). The Genetical Analysis of Quantitative Traits. Chapman and Hall, London.
2. Mather K. (1949). Biometrical genetics. Dover publications, New York, pp. ix, 158.
3. Gelderman, H. (1975). Investigation on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theoretical and Applied Genetics* 46, 300–19

4. Lander, E. S., and Botstein, D. (1967). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 238, 1567-1570.
5. Stam, P. (1993). Construction of integrated genetic linkage map by means of a computer package: Joinmap. *The Plant J*., 5: 739-744.
6. Tanksley SD, Medina-Filho H, Rick CM (1982) Use'ofnaturally occurring enzyme variation to detect and n1ap genes control ling quantitative traits in an interspecific backcross of tomato. *Heredity* 49: 11-25
7. Edwards, M.D., Stuber, C.W. and Wendel, J.F.C. (1987). Molecular marker facilitated investigations of quantitative trait loci in maize. *Genetics*, 116: 113-125.
8. Lander, E.S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121: 185-199.
9. Zeng, Z.B. (1994). Precision mapping of quantitative trait loci. *Genetics*, 136: 1457-1468.
10. Jansen, R.C. and Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, 136: 1447-1455.
11. Kao, C., Z. Zeng, and R. Teasdale (1999). Multiple interval mapping for quantitative trait loci. *Genetics 152*, 1203-1216.
12. Mayer, R. E. (2005a). Cognitive theory of multimedia learning. In R.E. Mayer (Ed.).

The Cambridge Handbook of Multimedia Learning.New York: Cambridge University Press.

13. Kearsey, M.J. (1998). The principles of QTL analysis (a minimal mathematics approach). *J. Exp. Bot.*, 49: 1619-1623.