Kastamonu University
Journal of Engineering and Sciences
e-ISSN 2667-8209
http://dergipark.gov.tr/kastamonujes

QTL Analysis Methods

Özge Şahin*, [a], Orhan Kavuncu[b]

[a] *Faculty of Biology, Institute of Evolution and Biodiversity, Münster University, Muenster, Germany*
[b] *Department of Genetics and Bioengineering, Faculty of Engineering and Architecture, Kastamonu University, Kastamonu*

ARTICLE INFO

Received: May:13.2019
Reviewed: May:14.2019
Accepted: May:22.2019

Keywords:
Quantitative trait loci (QTL),
QTL mapping,
QTL detecting,
Marker loci,
Linkage.

Corresponding Author:
*E-mail: sahinoz@ankara.edu.tr

ABSTRACT

In recent years, quantitative characters have settled to a quite important place among genetic studies. The inheritance of these traits is usually achieved by multiple gene expression and by the additive and interactive effects of these genes. Therefore, the analysis methods used to understand the genetic basis of quantitative traits may be more complex than those of qualitative traits. The loci on the chromosome at which the genes controlling the genetic architecture of the quantitative traits are located are called the quantitative trait loci (QTL). There are different methods for detecting and analyzing the presence of these loci. In addition, it is possible to map the QTLs with the help of the molecular markers as a developing method thanks to today's technology. In this review, the methods of diagnosing QTL and mapping studies were analyzed.

Anahtar Kelimeler:
Kantitatif özellik lokusları (QTL),
QTL haritalama,
QTL belirleme,
Belirteç lokus,
Bağlantı.

ÖZ

Son yıllarda kantitatif özellikler genetik çalışmalar içinde oldukça önemli bir yer teşkil etmeye başlamıştır. Bu özelliklerin kalıtımı genellikle birden fazla gen açılımı ve bu genlerin eklemeli ve interaktif etkisi ile gerçekleşmektedir. Bu sebeple de kantitatif özelliklerin genetik temelini anlamada kullanılan analiz yöntemleri, kalitatif özelliklerinkinden daha karmaşık olabilmektedir. Kantitatif özelliklerin genetik mimarisini kontrol eden genlerin kromozom üzerinde bulunduğu lokuslar kantitatif özellik lokusları (QTL) olarak isimlendirilmektedir. Bu lokusların varlığını tespit etmek ve analiz etmek için farklı yöntemler mevcuttur. Ayrıca günümüz teknolojisi sayesinde gelişen bir yöntem olarak moleküler belirteçler yardımıyla QTL haritalamak mümkündür. Bu derlemede, QTL varlığını teşhis etmeye ve haritalama çalışmalarına yönelik analiz yöntemleri incelenmiştir.

1. Quantitative Traits and QTL

Quantitative traits have long been involved in genetic studies as a major field. The most striking part of the variation observed between populations or populations in both experimental studies and field research is the part resulting from quantitative characters. A quantitative trait is defined as traits that vary in quantitative terms and that are constantly phenotypically distributed by definition. These traits are usually controlled by multiple genes. Some of these genes have a large effect, some of which have a small effect. In addition, these traits are under the influence of environmental factors as well as genetic factors.

Genes that control the genetic change of quantitative traits are called quantitative trait loci (QTL) [20]. QTL may consist of small effective genes, large effective genes, or both. Variations in these traits are generally controlled by segregating of many loci. Therefore, quantitative traits are sometimes defined as polygenic traits. However, inheritance of these traits cannot be studied by methods developed to study the inheritance of traits controlled by only a few genes. Genes that are effective on quantitative traits are also inherited according to Mendel's rules. In terms of those which have a major effect on them, segregation can be studied methods based on Mendel's rules. However, polygenic

inheritance cannot be studied by methods based on classical Mendelian rules because of many genes having small effects, plus environment that alter their effects; hence it is not possible to detect the phenotypic segregation rates according to the genotypes [12]. Therefore, if there are large effective genes, determining their location on the genome will provide significant benefits in terms of applying breeding methods such as indirect selection. Thus, the idea that these large effective genes and/or small effective gene blocks, which may be linearly aligned on the genome, can be determined by their linkage to specific marker loci, led to the use of QTL definition from the 1990s [20], unlike the poligen definition. QTL is also referred to as large-effect gene sequences which are investigated to be related to one or more marker loci with small-effect genes.

When examining quantitative traits, it should be known that environmental differences play a significant role in the phenotypic variation of these traits. Most quantitative traits are controlled by multiple genes and environmental factors as mentioned above. Therefore, the change in the polygenic structure and the environment makes the studies on quantitative traits more difficult than studies on monogenic traits. These traits are genetically complex, although they are easy to measure as phenotypes and therefore the expression of complex traits are used by some researchers [11]. Because of its polygenicity and its sensitivity to environmental changes, it is necessary to use statistical tools to examine the genetic architecture of these traits in large populations.

In parallel with the developments in the field of molecular biology and biotechnology, analysis methods for QTL have been developed since the 1990s. Thanks to these developments and studies at the molecular level, the possibility of utilizing DNA sequences called genetic markers has emerged in determining the effect of QTLs and their settlements. The respective positions of the markers in the marker maps can be determined by the recombination events observed throughout the genome. The segregating marker samples may provide information on quantitative traits such as the number of linked QTLs and chromosomal locations and the effectivity of each QTL by virtue of phenotype and pedigree information. A description containing all QTL information is also referred to as the genetic structure of the quantitative trait of interest. The study about the genetic structures of quantitative traits using molecular markers means that QTL mapping is actually used by the relationship between marker loci and quantitative traits [24]. Therefore, using the possible relationship between the molecular markers and the quantitative trait interested, QTL is called QTL mapping to predict their position in the genome and predict QTL effects [20].

Identifying each gene separately and knowing the presence of QTL will enable many useful applications to be implemented. First, this may improve the effect of selective breeding for low-remodeling (ie, more affected by environmental differences) and for traits that can only be observed in single sex. Another important aspect is that transgenic technology can be applied to quantitative traits in the future. In addition, the identification of alleles that cause susceptibility to common multifactorial diseases such as heart disease or diabetes in the medical field can help in the development of some methods for prevention [2]. Considering the benefits that can be achieved through studies on the inheritance of quantitative traits, it can be said that in the future genetic researches on QTL will increase. For this reason, in this paper, methods for determining the presence and location of QTLs were reviewed.

2. QTL Mapping Methods

The experimental set-ups designed to predict the effect of QTL and its location on the genetic maps are based on two fairly recent mapping studies for single genes and methods of estimating the linkage disequilibrium between marker locus alleles and QTL alleles.

Required thing for QTL mapping is the variation in the quantitative traits of the linkage map of the polymorphic marker loci that covers the entire genome sufficiently or within the allel groups of marker alleles.

2.1 Marker loci

In recent years, the development of DNA marker technologies has contributed to the investigation of the genetic structures of living things and to the mapping of agriculturally important genes, including QTLs. In addition, markers serve as an important and powerful tool in the scientific world in many aspects such as analysis of evolutionary relations. Molecular markers may be genes on the DNA, or any genetic coding function, ie, DNA sequences that have no function or function as phenotypes.

An ideal marker should have some features. The first of these is the high level of polymorphicity. Accordingly, individuals or populations should have different alleles in marker loci. Another desirable feature is that the marker is abundant to cover the entire genome. It should also be neutral in relation to the quantitative trait of interest and

associated with the breeding success of the species. Finally, a marker should be codominant to help distinguish all possible genotypes in the locus. However, this last item does not always have to be fulfilled, because dominant or recessive markers can also be used successfully in the experimental desings.

The analyzes that can be carried out with molecular markers are listed as follows [5]: Selection by means of marker (Marker Assisted Selection, MAS), QTL analyzes, genetic mapping, gene isolation strategies, characterization of gene sources, phylogenetic analysis, identification of culture types and determination of genetic kinship, determination of parents. For example, the identification of subgroups of living species, such as mutant species, has been studied extensively with some techniques such as RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), SSR (Simple Sequence Repeat Polymorphism). In addition, in recent years, DNA Microarray technology with thousands of genes at the same time to learn the relationship with each other in the field of many areas of molecular biology and genetic research is being done extensively [25].

The identification of the mutation that affects the quantitative traits of interest is quite difficult, as there may be a large number of genes and many potential variants within each gene. Therefore, the first step here is to determine the locations of genes of interest through genome mapping [9]. Genetic markers have an important role in genome mapping. These markers are mapped by the help of their linkage analysis.

The presence of molecular markers and the chromosome maps made with their help enable the identification of elements such as large effective genes and small effective gene levels of the genetic structure that control quantitative traits. Due to the quantitative properties are affected by a large number of genes, it should be known where these genes are in the genome. By applying the method known as QTL analysis in a suitable population, the location of the relevant genes in the specific chromosomal region can be determined. In addition to this, the magnitude of the effects can be predicted and whether the gene effect is additive or dominant can be detected [5].

2.2 QTL mapping

It is possible to detect QTLs position and identify them by analyzing the expression of marker genes and the phenotypic value of interest in the population. QTL mapping is the analysis of the offspring by the selection and crossing of parents of different phenotypes in terms of one or more quantitative traits, and the detection of a possible linkage between known marker loci and QTL.

The term QTL was first proposed by Geldermann [3]. QTL detection was developed approximately 90 years ago after Sax's work [18]. The study of Sax on the relationship between seed size differences in seed bean (*Phaseolus vulgaris*) and the seed coating model and pigmentation has been one of the initial studies of the linkage between QTL and genes that affect a quantitative trait. Then Sax's experiments were repeated on *Drosophila melanogaster* by Mather [15]. Thoday [21] proposed the idea of using single locus markers to identify and map genetic regions that control variation in quantitative traits.

QTL mapping has quite simple principle. What is required is just two lines having different phenotypes of the quantitative traits studied and a linkage map for some polimorphic marker loci. The lines are crossed to get the F_1 offsipring and then it has to get the backcrossed and/or F_2 offspring. The data is then classified according to a marker locus (or some marker loci) and compared statistically if they are different means fort the quantitative trait. If there is a significant difference it is decided that the QTL and marker loci are linked [16, 17].

2.2.1 Single marker mapping

This method independently searches each of the markers without taking into account their position and sequence. Haley and Knott [4] and Martinez and Curnow [14], independently of each other, introduced the regression method analysis with the help of information on neighboring markers. Weller [22] and Weller *et al.* [23] developed mixed model analysis to find QTL by using single marker information in hybridization between inbreeding and outbreeding populations. A mixed model approach for QTL mapping in unrelated populations is also reported by Jansen *et al.* [7].

The single marker mapping shows which marker has a linkage with the quantitative trait of interest and thus indicates the presence of possible QTL. The H_0 hypothesis (control hypothesis) is that the mean value of the property of interest is independent of the marker. If the test statistic is greater than the significant value, this indicates that the QTL is linked to the marker. Hypothesis controls using t-test, ANOVA and simple linear regression approaches for

differences in phenotypic averages may be misleading in predicting the frequency of recombination between QTL or marker and QTL, which are predicted to be close to each other, although they are all similar.

QTL discovery with a single marker approach is a simple protocol in which statistical analysis software packages can be implemented and has the potential to identify many important markers. When considering these statistical results, you need to think carefully about two important issues. The first of these is the sample size. The higher the number of individuals studied, the more reliable the phenotypic mean and variance is estimated. Large sample sizes allow for the possibility to observe recombination events and more accurate estimation of parameters. Thus, the possibility of detecting QTL increases. The second important issue is the problems in multiple tests. They occur during the investigation of multiple markers by independent statistical tests. This may increase the level of statistical significance as determined by the investigator, which may lead to false-positive QTL.

2.2.2 Interval mapping

Interval mapping is an approach developed to determine which of the QTLs in the map can be in a previously prepared map of the marker locus. The most commonly used statistical method in this approach is the maximum likelihood method. Maximum likelihood (ML) methods are widely used in QTL mapping. Linear models use only marker averages, and use many information about the co-distribution of ML, marker and quantitative trait. Therefore, ML requires the use of statistical models with more intensive probability calculations. One approach can be seen in the interval mapping indicated by Eric Lander and David Botstein [10]. In this mapping, a predicted genetic map is used to find the location of QTL. Intervals defined for sequential markers are investigated and the probability of QTL being located in the interval investigated is tested using statistical methods. In the interval mapping defined by Lander and Botstein, single QTLs against each sequential marker in the genome are tested statistically. Test results are expressed as LOD (Logarithms of odd) scores. As the statistical values calculated from the samples taken from the default population with control hypotheses are generally very large, the logarithm of these values is taken to compare them easily [8]. This logarithm is called LOD scores.

Interval mapping investigates a systematic, linear model where sequential genetic markers are tested for the same H_0 hypothesis and the same probability form is used for each increment. In addition, the combined LOD scores represent a LOD profile versus the genetic map. The placement of the maximum LOD profile has the potential to display multiple or fictitious QTLs when using a single QTL model. Deciding the peaks indicating single QTLs is related to the determination of statistical significance results. Because this probability is often a function of the mixture of normal distributions when it is maximized under both control and alternative hypothesis, test statistics may be insufficient in observing standard statistical distributions. It is therefore difficult to specify a QTL in this case. Instead, the composite interval mapping method discussed below is proposed.

2.2.3 Composite interval mapping

Composite interval mapping method is defined by Zhao-Bang Zeng in 1993 [26]. In the same year Ritsert Jansen identified multiple QTL mapping and achieved the same result in reducing the number of models considered [6]. Both methods have the same idea that interval mapping includes additional markers as cofactors to aim at reducing variation of other QTLs in the genome. Both approaches have limitations. These can be studied in one-dimensional studies in comparison to genetic maps, and the research is difficult in cases where epistatic QTL effects are high. In addition, there is a risk of taking a large number of markers in the model as cofactors, and care should be taken to protect the information that is appropriate for the correct prediction of the QTL effect.

2.2.4 Multiple QTL mapping

The concept of planning for the placement of multiple QTLs is more powerful than single QTL, because these approaches have the potential to distinguish between QTLs that are connected and/or interacting. The effect of two or more QTL alleles is difficult to be predicted when the interaction occurs. It is not possible to see this effect in single QTL analysis. The goal is to investigate whether the effect of a new potential QTL with another QTL is in interaction and the new QTL segregates from the other independently. One of the many interesting situations may be the loss of the expression of the trait with a special combination of the multiple QTL's alleles. Another problem in the study for multiple QTL mapping is to think about each position in the genome at the same time. Therefore, the position of a QTL that behaves independently, QTL connected to another QTL or QTL that interacts with other QTL can be detected.

QTLs which are in the interaction are a field of special interest, because they can show specific regions in the genome. Otherwise, they may be related to quantitative traits using one-dimensional research.

QTLs in the interaction are open despite their multiple positions. However, it is difficult to consider that there are many potential QTLs and their interaction with numerous statistical models and difficult calculations. On the other hand, an approach can be presented as follows: First of all, all QTLs must be placed. Then a statistical model should be created for these QTLs and their interactions. Then, the model should be continued by removing the others from the model and remaining important interactions. Due to the computational density of a multidimensional research, it is impossible to investigate simultaneously and it is called semi-simultaneous research. Such approaches have the potential to work in many cases. However, it is limited to the QTL pool that passes the first QTL analysis, and it is unlikely that QTL epistatic effects are not properly established. Research through all potential models is a problem known as model selection and remains an active field of study in the field of theoretical statistics.

The importance of models developed for multiple QTLs is well understood for the linked QTL and plays an important role in the prediction and location of epistatic QTL. In the successful use of multiple QTLs, the limiting feature is not to write equations for the model, but not to define the best of many model subsets. Counting all possible QTL models that are considered suitable for the genetic structure such as linkage, epistasis in the experiment is a very difficult task. Accurate and fast simultaneous multidimensional researchs and their comparisons across most of the possible models allow decision on the most appropriate model for future research. One of the approaches to investigate the optimal multiple QTL genotype is to use genetic algorithms. Application of genetic algorithms to multiple QTL problems is one of many useful approaches. Because this application allows the sampling of QTL models corresponding to unequal QTL numbers and can be used in multidimensional investigation of a genome in conjunction with the QTL mapping methodology. In addition to this, distinguishing between the covariance caused by the linkage disequilibrium between QTLs and the variance from the epistatic (interactive) effect is waiting as an important problem.

Sen and Churchill proposed an approach for covariants, non-normal distribution traits, epistatic QTL and multiple simultaneous investigations in addition to the above mentioned challenges [19]. This approach divided the QTL problem into two different parts: the relationship between QTLs and the relationship between the quantitative trait and QTL. Separation of these two independent relationships intensely predicts unknown QTL genotypes. It then allowed different models to be searched and compared with the information obtained from QTL genotypes. The approach of dividing the problem into two parts is not new. In 1993, Jansen dealt with this issue. After estimating QTL genotypes, Sen and Churchill [19] described all possible QTL models using an approach that allows different models of various QTL numbers. The QTL genotypes calculated independently from QTL effects and locations reduced the epistasis and related QTL cases, because the state of the QTL genotype and the number of QTL are known before their effects and interactions are estimated.

3. Result and Discussion

There are two approaches to identifying a gene to be accepted as QTL in a specific genomic region. These are changings of positional cloning and QTL which is in the same region with the candidate locus.

Positional cloning requires the map position of the locus of interest at 0.3 cM. The reason of taking 0.3 cM is the average length that a vector research can carry [2]. This requires screening for high stability meiotic mapping in experimental organisms as well as polymorphic markers in random mating populations in the region where there is a strong linkage disequilibrium between the quantitative trait phenotype and QTL. Although everything seems clear in this respect, it is difficult to define the gene of interest and decide whether the polymorphisms are related to alleles or other effects. This method is suitable for loci identified by the broad effects of mutant alleles and is used to identify single loci that affect human diseases. From a medical point of view, the importance of loci with smaller effects will increase in the future. Therefore, it is important to develop this application in order to analyze QTLs at the single locus level.

The most common strategy used to go from the mapped region to the gene is the candidate gene approach. With this approach, many loci with known genetic function are identified and cloned from the unknown locus regions on the map. Therefore, the linkage between the molecular polymorphism and phenotype in each candidate locus in the region is investigated.

Exploring QTL in previous years was aimed at many scientific researches and was a targeted goal. Today, QTLs which determine many traits and the location of QTLs in multiple interactions are aimed to be found with the help of advanced statistical analysis. In addition to many statistical methods Bayesian approaches may alternatively be used in the estimation of some parameters involved in genome analyses [13]. More research is carried out by developing technologies and these researches bring together more information. Therefore, this information allows us to better understand central dogma, which includes replication, transcription and translation stages. Quantitative variation can be observed in each stage of central dogma [1]. When this variation information is combined with appropriate statistical significance, it can be seen how a genome works in a unique way. When one considers the ideas about emerging technologies and methodologies, it is accepted that a single technological progress or statistical method cannot be a solution to genomic problems. Instead, a combination of many techniques and analyzes will contribute to the solution.

4. References

[1] W. R. Doerge, "Mapping and analysis of quantitative trait loci in experimental populations," Nature, vol. 3, pp. 43-52, Jan. 2002. DOI: 10.1038/nrg703

[2] D. S. M. Falconer and T. F. C. Mackay, "Quantitative trait loci" in Introduction To Quantitative Genetics, 4th ed., Essex, England, 1996, pp. 356-377.

[3] H. Geldermann, "Investigations on inheritance of quantitative characters in animals by gene markers I. methods," Theoretical and Applied Genetics, vol. 46, no. 7, pp. 319-330, 1975.

[4] C.S. Haley and S.A. Knott, "A simple regression method for mapping quantitative trait loci in line crosses using flanking markers," Heredity, vol. 69, pp. 315-324**,** 1992.

[5] B. İşçi, "Asmada QTL (kantitatif karakter lokus) analizi," Anadolu (Journal of Aegean Agricultural Research Institute), vol. 18, no. 2, pp. 11-37, 2008.

[6] R. C. Jansen, "Interval mapping of multiple quantitative trait loci," Genetics, vol. 135, pp. 205-211, Sep. 1993.

[7] R. C. Jansen, D. L. Johnson and J. A. M. Van Arendonk, "A mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families," Genetics, vol. 148, pp. 391-399, Jan. 1998.

[8] O. Kavuncu, "İstatistik Teorisi ve Teorik Dağılımlar", Ankara, Turkey, 1995.

[9] H. Kozanoğlu and İ. Oğuz, "Importance of DNA markers in animal breeding and genetics," in II. International Symposium on Molecular Diagnosis and Applications, İzmir, Turkey, May 10-14, 2004, pp. 75-80.

[10] E. S. Lander and D. Botstein, **"**Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps," Genetics, vol. 121, pp. 185-199, Jan. 1989.

[11] E. S. Lander and N. J. Schork, "Genetic dissection of complex traits," Focus, vol. 4, no. 3, pp. 442–458, 2006.

[12] M. Lynch and B. Walsh, "Quantitative trait loci" in Genetics and Analysis of Quantitative Traits. Sunderland, USA, 1998, pp. 319-532.

[13] A.O. Maraghi, "A Bayesian generalized linear model approach to quantitative trait loci (QTL) detection of litter size in mice," Ph.D. dissertation, Dept. Animal Sci., Ankara Univ., Ankara, Turkey, 2011.

[14] O. Martinez and R. N. Curnow, "Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers," Theoretical and Applied Genetics, vol. 85, pp. 480-488**,** 1992.

[15] K. Mather, "Variation and selection of polygenic characters," Journal of Genetics, vol. 41, pp. 159-193, 1941.

[16] J. A. M. Arendonk and H. Bovenhuis, "Designs and Methods to Detect QTL for Production Traits Based on Mapped Genetic Markers," Poultry Genetics, Breeding and Biotechnology, W. M. Muir and S. E. Aggrey, Ed., Cambridge, USA, 2003, pp. 439-465.

[17] N. Yi and S. Xu, "Designs and Methods to Detect QTL for Production Traits Based on Random Genetic Models," Poultry Genetics, Breeding and Biotechnology, W. M. Muir and S. E. Aggrey, Ed., Cambridge, USA, 2003, pp. 465-497.

[18] K. Sax, "The association of size differences with seed co at pattern and pigmentation in *Phaseolus vulgaris*," Genetics, vol. 8, pp. 552-560, 1923.

[19] S. Sen and G. A. Churchill, "A statistical framework for quantitative trait mapping," Genetics, vol. 159, pp. 371-387, Sep. 2001.

[20] S. D. Tanksley, "Mapping polygenes," Annual Review of Genetics, vol. 27, pp. 205-233, 1993.

[21] J. M. Thoday, "Location of polygenes," Nature, vol. 191, pp. 368-370, Jul. 1961.

[22] J. I. Weller, "Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers," Biometrics, vol. 42, no. 3, pp. 627-640, Sep. 1986.

[23] J. I. Weller, Y. Kashi and M. Soller, "Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle," Journal of Dairy Science, vol. 73, pp. 2525–2537, 1990.

[24] S. Xu, "QTL analysis in plants," Methods in Molecular Biology: Quantitative Trait Loci: Methods and Protocols, vol. 195, pp. 283-310, 2002.

[25] H. E. Yurter, "Gen ekspresyon analizinde microarray teknolojisinin kullanımı," DEU Tıp Fakültesi Dergisi Özel Sayısı, pp. 41-48, 2001.

[26] Z. B. Zeng, "Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci," Genetics, vol. 90, pp. 10972-10976, Dec. 1993.