Genetica (2005) 123: 25–37

QTL mapping and the genetic basis of adaptation: recent developments

Zhao-Bang Zeng
Departments of Statistics and Genetics & Bioinformatics Research Center, North Carolina State University, Raleigh, NC 27695-7566, USA (E-mail: zeng@statgen.ncsu.edu)

Received 28 July 2003 Accepted 1 May 2004

Key words: genetic architecture, genetic basis of adaptation, genetic correlation, genotype by environment interaction, microarrays, QTL mapping, quantitative trait loci

Abstract

Quantitative trait loci (QTL) mapping has been used in a number of evolutionary studies to study the genetic basis of adaptation by mapping individual QTL that explain the differences between differentiated populations and also estimating their effects and interaction in the mapping population. This analysis can provide clues about the evolutionary history of populations and causes of the population differentiation. QTL mapping analysis methods and associated computer programs provide us tools for such an inference on the genetic basis and architecture of quantitative trait variation in a mapping population. Current methods have the capability to separate and localize multiple QTL and estimate their effects and interaction on a quantitative trait. More recent methods have been targeted to provide a comprehensive inference on the overall genetic architecture of multiple traits in a number of environments. This development is important for evolutionary studies on the genetic basis of multiple trait variation, genotype by environment interaction, host–parasite interaction, and also microarray gene expression QTL analysis.

Abbreviations: CIM – composite interval mapping; EM – expectation and maximization algorithm; IM – interval mapping; MIM – multiple interval mapping; QTL – quantitative trait loci.

Introduction

Quantitative trait loci (QTL) mapping is a genome-wide inference of the relationship between genotype at various genomic locations and phenotype for a set of quantitative traits in terms of the number, genomic positions, effects, interaction and pleiotropy of QTL and also QTL by environment interaction. The primary purpose of QTL mapping is to localize chromosomal regions that significantly affect the variation of quantitative traits in a population. This localization is important for the ultimate identification of responsible genes and also for our understanding of the genetic basis of quantitative trait variation.

Applied to natural populations, most QTL mapping experiments are designed to study the genetic basis of phenotypic differences between different natural populations or between different species (Mackay, 2001; Mauricio, 2001). Starting from two differentiated populations, a cross is usually made between the populations to create a hybrid, and then either backcross the hybrid to the parental population(s) to create backcross population(s) or intercross among hybrids (if possible) to create a F_2 population. Recombinant inbred lines can also be created from the cross and are popular for QTL mapping study. QTL mapping analysis is performed in these segregating populations to locate QTL that are responsible for the difference between the parental populations which could be due to adaptation. QTL mapping analysis in these populations can help us to understand a number of issues that are associated with the genetic basis of adaptation. It can estimate how many QTL that have different alleles between populations and contribute

significantly to the population difference. It can estimate where they are located in the genome; what their effects are; how they interact; and how QTL interact with the environment. All these are critically important for the study of the genetic basis of adaptation.

QTL analysis certainly has many limitations. The number of QTL is likely to be downwardly biased estimated due to linkage and limited sample size. There is also likely a bias in the estimation of QTL effect distribution as only QTL with relatively large effects are likely to be detected and some QTL effects may represent the joint effects of multiple closely linked genes. Analysis of epistasis may only detect a part of gene interactions and there could be many other hidden interactions between detected and undetected QTL. Certainly, there is a big gap between QTL that are mapped with a confidence interval in many cM units and genes that are responsible for the variation. Mauricio (2001) discussed some caveats in using these methods for interpreting the genetic basis of adaptation for evolutionary biology studies.

In this article, I review some statistical methods used for QTL mapping analysis, particularly the methods used to map multiple QTL simultaneously for studying QTL epistasis and for estimating the overall genetic architecture of quantitative trait variation. I will use two QTL mapping experiments to illustrate the use of these methods and interpretation of the mapping analysis. One experiment is the study of genetic basis of a morphological shape difference between two *Drosophila* species due to adaptation (Zeng et al., 2000). The other experiment is the study of genetic basis of long-term selection response on wing size of *Drosophila melanogaster* (Weber et al., 1999, 2001). I also describe a method to study details of genetic correlation between multiple traits and to test QTL by environment interaction. In the end, I discuss the connection of this multiple trait QTL analysis with microarray gene expression data and outline an approach in using this method for the construction of genetic effect network between QTL, gene expressions and quantitative trait phenotypes.

Statistical framework

Statistical analysis of QTL mapping works with two data sets. One is the molecular marker data set that provides information of segregation of a genome at various marker positions in a population, and the other is the quantitative trait data set that provides information of segregation and effects of QTL. The connection between the two data sets is QTL. The variation of trait values in a population is partially due to the segregation of QTL alleles, and QTL are linked to some molecular markers. It is this linkage that provides information to localize QTL in a genome.

Let Y denote the trait data and X denote the marker data. In a joint analysis of marker and trait data, we study the joint probability of Y and X

$$\begin{aligned} P(Y,X) &= P(Y|X)P(X) \\ &= \sum_{Q} P(Y|Q,X)P(X) \\ &= \sum_{Q} P(Y|Q)P(Q|X)P(X) \end{aligned} \tag{1}$$

This joint probability can be split into two parts. One is $P(X)$ which can be modeled as a function of marker linkage order ω, linkage phases ϕ and recombination frequencies γ between markers. This analysis is the marker linkage analysis and $P(X|\gamma,\phi,\omega)$ is the likelihood of marker data.

The other part is $P(Y|X)$ which represents the QTL analysis, analyzing the conditional probability of trait values Y given marker genotypes X through QTL genotypes Q. $P(Q|X)$ is a function of QTL positions λ in relation to markers, and involves the segregation analysis of QTL given marker genotypes. $P(Y|Q)$ is a link function between QTL genotypes Q and trait phenotypes Y, and can be modeled as a function of QTL effect parameters θ, such as additive, dominance and epistatic effects of QTL and any other parameters that link QTL genotypes to trait phenotypes. Together, λ and θ represent the genetic architecture parameters of QTL. In this form, we generally represent $P(Y|X)$ as

$$P(Y|X,\lambda,\theta) = \sum_{Q} P(Y|Q,\theta)P(Q|X,\lambda) \tag{2}$$

which is the likelihood of trait data given marker data and is the main focus of this article.

Another statistical approach that has been used for QTL analysis is Bayesian posterior inference. In Bayesian statistics, model parameters are regarded as random variables, and we are concerned with the inference of posterior probability of