

Chapter 10

A method for detecting multiple interacting QTL

Brian Kinghorn

University of New England

Introduction..... 92
The Gametic Relationship Matrix approach..... 92
Detecting multiple QTL..... 94
Detecting multiple interacting QTL..... 95
 Model for fitting interacting QTL..... 95
 Some results 96
References..... 98

Introduction

In the last lecture you found out about regression and maximum likelihood methods for detecting QTL. The extension to cater for multiple interacting QTL is best illustrated on the basis of analysis by regression.

The Gametic Relationship Matrix approach

In the fullest implementation of this approach, we first set up a symmetrical matrix that contains a row and column for each gametic haplotype (2 per animal, one from each parent) in the population of animals that we have. Such a matrix is specific to the chromosomal region of current interest. Each element in this matrix is then the probability of identity-by-descent for the representations of this region (one representation per gamete). Here are simple examples of this “Gametic Relationship Matrix” (GRM). Notice that without marker information we must resort to simple segregation probabilities – however, marker information allows us to be more ‘surgical’ in allocating identity-by-descent probabilities:

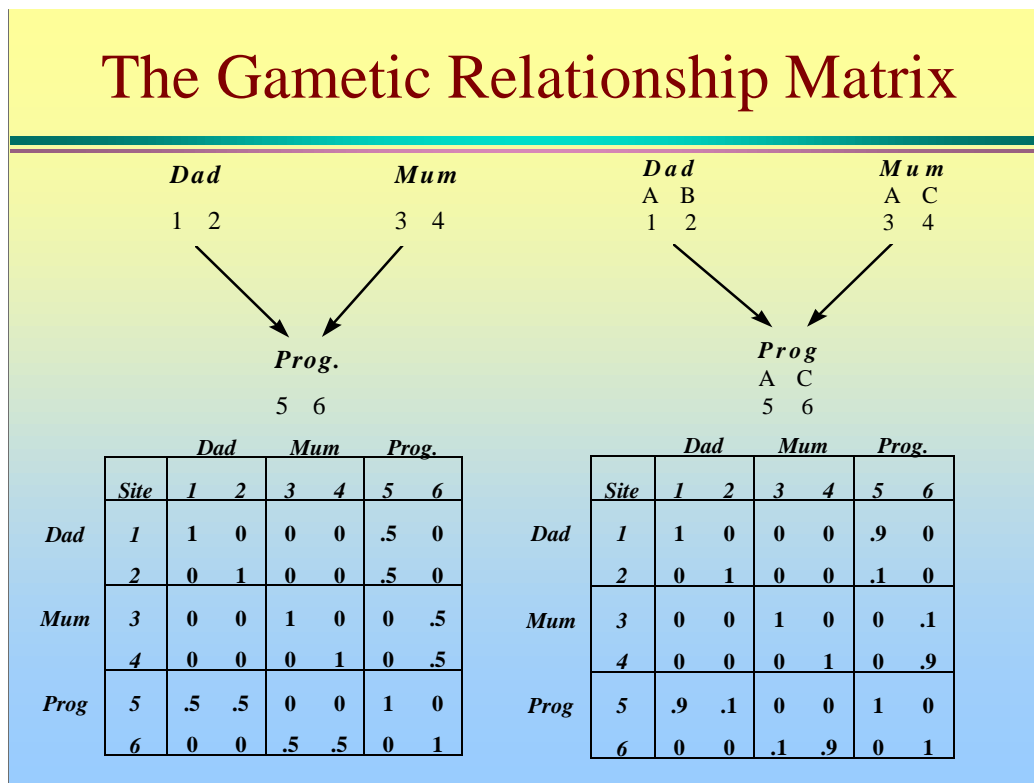
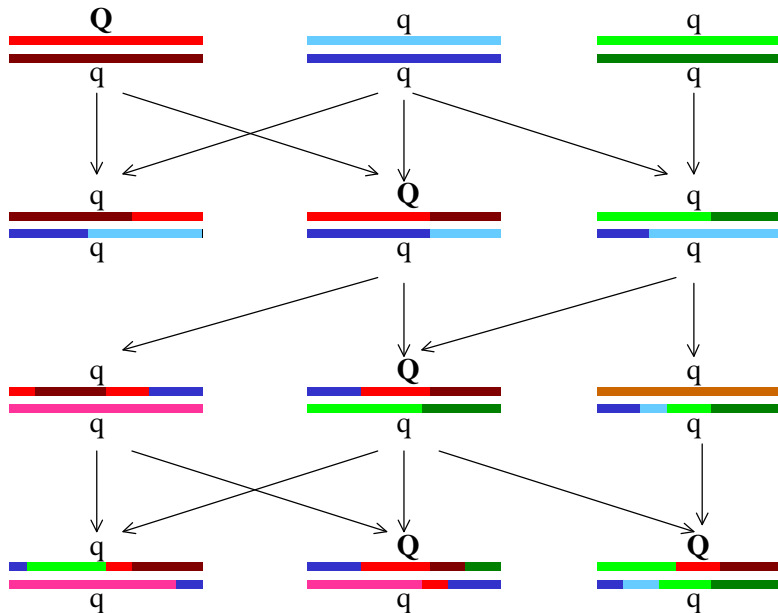


Figure. (From Kinghorn and Clarke (1997). Gametic relationship matrices (GRM) for a QTL are of dimension 6 sites x 6 sites for the simple 3-animal pedigree shown. Elements of the GRM are probability of identity by descent of the alleles at the prevailing pair of sites. In the GRM to the left, no marker information is available, and, for example, probability of identity by descent between sites 4 and 6 is 0.5, as site 6 (maternal) could have inherited from sites 3 or 4 with equal probability. In the GRM to the right, a marker with alleles A, B and C is available, and for example, probability of identity by descent between sites 4 and 6 is 1, *for the marker locus*. If the QTL is linked with a recombination fraction of 0.1, then the probability of identity by descent between sites 4 and 6 is 0.9, *for the QTL*, with a 0.1 probability (in the event of recombination) for sites 3 and 6. Special attention is required where there is ambiguity of marker allele inheritance (Wang et al., 1995).

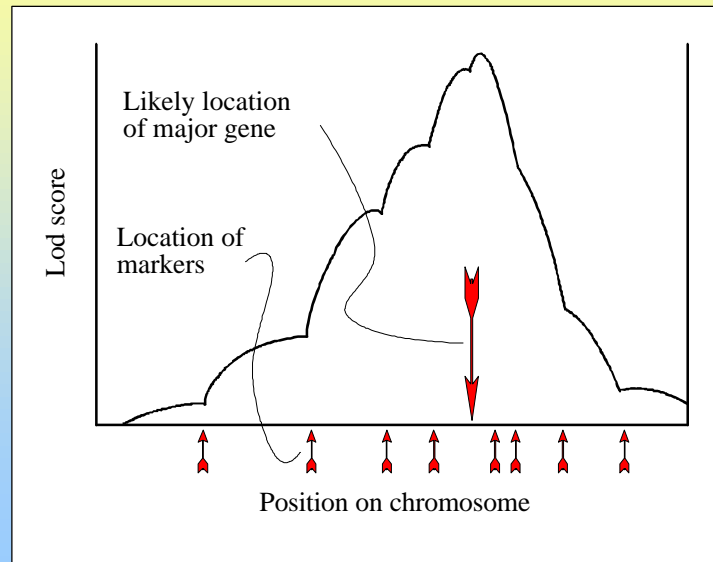
You can visualise regions of identity-by-descent in the following diagram (it looks better in colour!):



With a good data set, the GRM gives us a lot of information for mapping. In the diagram above, the top-left founder animal has QTL allele Q in its paternally inherited region of haplotype (coloured red). For all its descendants, the GRM gives us probabilities that they have inherited the same bit of DNA, holding that Q allele. We can then simply *regress* their phenotypes on these probabilities to get an estimate of the effect of Q on phenotype. Quite simple really!

The strategy is to construct a GRM (or a subset of it) for each location in the genome, and test the goodness of fit of the resulting regression. We end up with something like this for each chromosome:

QTL detection with markers



Here the goodness of fit is a LOD score, described elsewhere.

Detecting multiple QTL

This all works well. However, if there is more than one QTL of significant effect, we can do better. One approach is *Composite Interval Mapping*. Once we are happy about the most likely position and effect of a QTL, we fix that in the analysis – we correct all the animal phenotypes for the most likely impact of that QTL on their performance – and then repeat the process to look for another QTL.

This has two problems:

- The estimated position of the first QTL can be influenced by the second QTL, and vice-versa. This is especially dangerous for linked QTL. A method to simultaneously locate the two QTL is preferable.
- Life is complex – and that means that genes (or gene products) interact with each other to produce the organisms that we all are. The value of a particular gene variant will differ between genetic backgrounds. In some cases it will be the weak link to achieving high merit, and in others it will not. This means that we should ideally look for interacting sets of genes. Otherwise we could miss some important genes – and opportunities to exploit them.

Detecting multiple interacting QTL

We can nominate two separate positions in the genome as candidate locations for two QTL. We can then construct a GRM for each position, and carry out a 2-locus regression, as outlined below, fitting interaction effects between the two loci, as well as additive and dominance effects within each locus.

How can we find the best fitting two positions? The following paper demonstrates an approach that works efficiently, using a genetic algorithm:

Carlborg, O., Andersson, L. and Kinghorn, B.P. 2000. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics*. In Press

The genetic algorithm (GA) works by “breeding” the best solution to the prevailing mathematical problem. In this case, the “DNA” that the GA uses is simply the candidate positions for the two (or more) QTL. Each of these is a candidate solution to the problem of QTL locations. Each candidate solution competes to become a “parent” in the next generation. They compete on a criterion that is simply the goodness of fit of these positions to the phenotypic data and pedigree on hand.

The successful “parent” solutions then combine in some way – exchanging information, and mutate to some extent, to generate a new generation of candidate solutions.

If this is difficult to understand, it is because you are a geneticist, and not an engineer. We geneticists get confused at first because the thing we want to optimise is all to do with genetics and life becomes confusing!! If we were optimising the design of a supersonic jet, then there would only be one set of genetic parameters to think about.

Model for fitting interacting QTL

Here is a simple one-locus model of genetic effects, similar to that found in all texts in this area. II , Ii and ii are the genotype values for combinations of the two alleles I and i , m is a general mean, A_i is the additive affect and D_i the dominance effect at locus i .

$$\begin{pmatrix} II \\ Ii \\ ii \end{pmatrix} = \begin{pmatrix} m + A_i \\ m + D_i \\ m - A_i \end{pmatrix}$$

We can now expand this to cater for effects at two loci. The classical statistical approach (eg. Jana 1971) is typified as follows:

$$\begin{pmatrix} II JJ & II Jj & II jj \\ Ii JJ & Ii Jj & Ii jj \\ ii JJ & ii Jj & ii jj \end{pmatrix} = \begin{pmatrix} \bar{i} + A_i + A_j + AA_{ij} & m + A_i + D_j + AD_{ij} & m + A_i - A_j - AA_{ij} \\ m + D_i + A_j + AD_{ji} & m + D_i + D_j + DD_{ij} & m + D_i - A_j - AD_{ji} \\ m - A_i + A_j - AA_{ij} & m - A_i + D_j - AD_{ij} & m - A_i - A_j + AA_{ij} \end{pmatrix}$$

The number of parameters to handle has increased from three (m , A_i and D_i) to nine (m , A_i , A_j , D_i , D_j , plus interaction terms AA_{ij} , AD_{ij} , AD_{ji} , and DD_{ij}). Notice that each locus here has two alleles.

More detail is here extracted from Carlborg et al. (2000):

“

The objective function used was the residual sum of squared errors from a weighted least squares approach to QTL mapping. The method is the extension of the method of Jansen (1992) to the two-loci linear model $G = m + A1 + A2 + D1 + D2 + AA12 + AD12 + AD21 + DD22$ as indicated by the author. The parameters of the model will be explained below. Markers have not been used as cofactors and successive iterations in the EM algorithm have been removed to increase the computational efficiency during the evaluation procedure. The modifications needed to the single QTL mapping procedure described by Jansen and Stam (1994) when implementing the two QTL model included duplication of each individual nine times (instead of three times i.e. once for every possible two-QTL genotype) and the use of an expanded design matrix (X). The design matrix for the two-locus linear model has been described by Jana (1971). The weight for each observation was taken to be the product of the conditional probabilities of the single QTL-genotypes given the markers (Haley and Knott 1992) at each of the two fitted QTL. The estimates of the model parameters can be found as:

$$\beta = (X^T W X)^{-1} X^T W Y$$

$$\sigma^2 = (1/N)(Y - X \beta)^T W (Y - X \beta)$$

where Y is the complete data vector, X is the design matrix for the complete data, W is the diagonal matrix of weights, β is the vector of the regression parameters, σ^2 is the normal variance and N is the number of individuals (Jansen and Stam 1994).

The residual sums of squared errors can then be calculated as:

$$SSE = (Y - X \beta)^T W (Y - X \beta)$$

The method described above can easily be extended to take account of background QTL in the analysis. Two extra ga-genes are added to the genetic algorithm and two extra columns are added to the X matrix for each background QTL. The extra ga-genes represent the chromosomal location for the QTL and the columns in the design matrix are to contain the QTL indicator variables a and d (Haley and Knott 1992), for a QTL at the location given by the ga-genes. The rest of the evaluation procedure is the same as before. We have evaluated the increase in computational demand for a simultaneous search for more than two QTL using this method, but have not investigated any other properties.

“

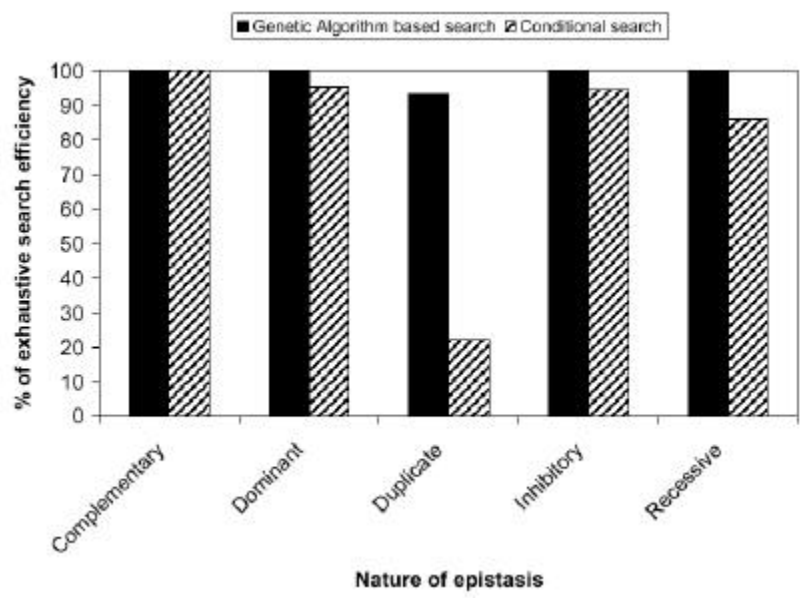
Some results

There are two advantages in this approach:

- The genetic algorithm gives a fast search, saving much computer time. It increases the computational demand by a factor of 3 to 5 when compared to the conditional search (Carlborg et al. 2000). The improvement in computational efficiency of the GA as compared to an exhaustive enumerative search (looking at all pairs of locations in a genome size of 2,000 cM using a resolution of 1 cM) was by a factor 133 for two QTL. An expansion of the search to additional dimensions by also searching for background QTL simultaneously leads to further computational

advantages for the GA based search. Improvements are in the order of 65,000 for three QTL and 1.7×10^7 for four simultaneously fitted QTL.

- As Carlborg et al. (2000) report, the results from the simulation study with 18 QTL (Figure below) showed that the genetic algorithm based search had higher relative efficiency to detect the simulated pair of epistatically interacting QTL than the conditional search (ie. composite mapping approach, as described above) for all epistatic models tested. The genetic algorithm had a relative efficiency of 100% for all epistatic models except for the duplicate. The conditional search had between 86 and 96% relative efficiency for the dominant, recessive and inhibitory epistatic models and 100% relative efficiency for the complementary model. The difference in relative efficiency for the search methods was very large for the duplicate epistatic model, where the conditional search only had a relative efficiency of 21%, while the genetic algorithm based search had a relative efficiency of 93% (this could grow to 100% with better tuning of the GA parameters). In the simulation where two interacting QTL explained all genetic variation, both methods had a relative efficiency of 100%.



As Carlborg et al. note: “The genetic algorithm is a general tool to search large parameter spaces and could be of use in many other areas in QTL mapping. In this study we have used a genetic algorithm in the search for two interacting QTL in a cross between inbred lines, but the method can also be used for analyses of crosses between outbred lines and in searches for more than two QTL. For analyses of outbred lines, the genetic algorithm could also be used when testing for QTL segregation within the founder lines. This would be implemented by using a genetic algorithm to group the haplotypes from the founders in allelic groups and in this way obtain the most likely allelic constitution for the founders and other individuals in the pedigree. This results in greater detection power because of more extreme

probabilities of identity-by-descent of chromosomal regions between phenotyped individuals and each founder.“

Acknowledgement: Thanks to Örjan Carlborg*, Leif Andersson* and Brian Kinghorn for permission to use direct quotation.

References

- Carlborg, O., Andersson, L. and Kinghorn, B.P. 2000. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics*. In Press
- Haley, C.S., and S.A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using ²anking markers. *Heredity* 69: 315 -324.
- Jana, S., 1971 Simulation of quantitative characters from qualitatively acting genes. I. Nonallelic gene interactions involving two or three loci. *Theor. Appl. Genet.* 41: 216-226.
- Jansen, R. C., 1992 A general mixture model for mapping quantitative trait loci by using molecular markers. *Theor. Appl. Genet.* 85: 252-260.
- Jansen, R. C., 1993 Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205-211.
- Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447-1455.
- Kinghorn, B.P. and Clarke B. E. 1997. Genetic evaluation at individual QTL. *Animal Biotechnology*, 8:63-68.
- Wang, T., Fernando, R.L., van der Beek, S., Grossman, M. and van Arendonk, J.A.M. 1995. Covariance between relatives for a marked quantitative trait locus. *Gen. Sel. Evol.* 27:251-274.