ANALYSIS OF METHODS FOR FINE MAPPING QUANTITATIVE TRAIT LOCI USING LINKAGE DISEQUILIBRIUM

L. Grapes, R. L. Fernando and M. F. Rothschild

Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA

INTRODUCTION

Disequilibrium between unlinked loci will degrade rapidly under random mating (Crow, 1986). Thus, linkage disequilibrium (LD), which can occur between loci due to mutation, selection, population admixture, migration and finite population size, has been proposed to be a powerful tool for mapping QTL (Goddard, 1991). Based on the LD between a marker and a closely linked QTL locus, likelihood methods to estimate QTL position have been presented (Hill and Weir, 1994; Kaplan *et al*., 1995; Terwilliger, 1995). Given that an LD-creating event took place many generations ago in a region of the genome containing a QTL, only closely linked markers will be in disequilibrium with the QTL. In such regions, LD can be used to fine map QTL.

Recently, Meuwissen and Goddard (2000) described a method for fine mapping QTL using LD and marker haplotypes. Assuming that LD was created by mutation in a QTL allele several generations ago, their method estimates the covariance between marker haplotypes to identify regions that are identical by descent (IBD). In general, individuals with similar marker haplotypes flanking a QTL should contain QTLs that are IBD.

To study the advantage of using haplotype information for LD fine mapping, using a half-sib family structure, results from the IBD method were compared to those from a single marker TDT analysis (Meuwissen and Goddard, 2000). While the TDT method also uses LD information, it has the additional advantage of being unaffected by population admixture (Rabinowitz, 1997). This advantage, however, comes at a cost because only within-family information is used. In the IBD method, however, both within and between-family information are used.

Thus, in this study, we compared the IBD method with two regression methods that also use within and between-family information.

MATERIALS AND METHODS

We assumed that a previous linkage analysis study had mapped a QTL to a region of 4.5 cM, and within that region 10 bi-allelic markers were available.

In the founders, marker alleles were sampled from a Bernoulli distribution, and the QTL alleles were uniquely numbered. Subsequent generations were created by randomly mating parents from the previous generation to create the next. It was known which markers were maternally and paternally inherited so that marker haplotypes could be constructed.

Data were simulated by randomly selecting one QTL allele from the final generation with a frequency higher than 0.1. This allele was given an additive genetic value of 1, and the value of all other QTL alleles was set to 0. The phenotypic record for each individual was calculated by adding the QTL allele effects to an environmental effect sampled from $N(0, 1)$. A summary of the simulated population can be found in Table 1.

Table 1. Summary of the simulated population

Effective population size	100
Number of generations of random mating since QTL mutation occurred	100
Number of markers	10
Number of alleles per marker	2
Initial marker / QTL allele frequency	0.5 / 0.005
Distance (cM) between adjacent markers	0.5
Position of QTL	halfway between markers 5 and 6
Additive effect of one selected QTL allele	1
Residual standard deviation	1
Number of individuals (records) in final generation	100
Number of replicates used to calculate IBD matrices	100,000

Maximum-likelihood estimation. Modifying the method of Meuwissen and Goddard (2000), phenotypic data of the last generation for a single trait are modeled by

$$y = \boldsymbol{Z}a + e,$$

where y is a vector of phenotypic data, a is the vector of QTL effects, e is the vector of residuals, and $\boldsymbol{Z}$ is a known incidence matrix for a. For the purposes of simplifying our simulation, fixed effects are not incorporated into the population. The variance of the residuals is $\mathrm{Var}(e) = \sigma_e^2\mathbf{R}$, where $\mathbf{R}$ is an identity matrix. The variance of the QTL effects is $\mathrm{Var}(a) = \sigma_a^2\boldsymbol{H}_p$, where $\boldsymbol{H}_p$ is the covariance matrix for the QTL effects when the QTL is at position p.

By assuming multivariate normality, the residual loglikelihood of the data for our model is

$$L(\boldsymbol{H}_p, \sigma_a^2, \sigma_e^2) \propto -0.5[\ \ln(|\boldsymbol{V}|) + y'\boldsymbol{V}^{-1}y\ ],$$

which is similar to the loglikelihood equation used by Meuwissen and Goddard (2000) with fixed effects removed, and where $\boldsymbol{V}$ is the $\mathrm{Var}(y) = [\boldsymbol{Z}\,\boldsymbol{H}_p\,\boldsymbol{Z}'\sigma_a^2 + \boldsymbol{R}\sigma_e^2]$. All remaining steps were completed as described in Meuwissen and Goddard (2000) to estimate the QTL position. One hundred replications of this model were performed.

Single-locus (SL) regression model. For this model, a single locus is tested for association with the QTL. Phenotypic data for the last generation are modeled by

$$y = \boldsymbol{X}b + e,$$

where y is a vector of phenotypic data, b is a 2×1 vector of allelic effects (μ_0, μ_1) for a marker locus, and $\boldsymbol{X}$ is an incidence matrix for the allelic effects in b. The hypothesis $\mathbf{H}_0$: $\mu_0 = \mu_1$ vs. $\mathbf{H}_A$: $\mu_0 \neq \mu_1$ is tested for every marker locus. The locus with the smallest p-value is the center of a 0.5 cM region that estimates the QTL's position. One thousand replications of this model were performed.

Two-locus haplotype (HAP) regression model. In this model, a haplotype is constructed by selecting one marker on each side of the putative QTL. This creates marker brackets similar to those for the ML model. Phenotypic data for the last generation are modeled by

$$y = \boldsymbol{X}b + e,$$

where y is a vector of phenotypic data, b is a 4×1 vector of haplotype effects (μ_{00}, μ_{01}, μ_{10}, μ_{11}) for two marker loci, and $\boldsymbol{X}$ is an incidence matrix for the haplotype effects in b. The hypothesis $\mathbf{H}_0$: $\mu_{00} = \mu_{01} = \mu_{10} = \mu_{11}$ vs. $\mathbf{H}_A$: $\mu_{00} \neq \mu_{01} \neq \mu_{10} \neq \mu_{11}$ is tested for every marker bracket. The haplotype (marker bracket) with the smallest p-value is considered to contain the QTL. One thousand replications of this model were performed.

RESULTS AND DISCUSSION

Following Meuwissen and Goddard's (2000) comparison of haplotype-based vs. single-locus based tests, we allowed the estimated QTL position to be one marker bracket and up to two loci away from the QTL, respectively. The IBD method positioned the QTL in a 1.5 cM region in 77% of the replicates. The SL method estimated the QTL within a 2 cM region in 74% of the replicates and within 1 cM in 49% of the replicates. The HAP model placed the QTL within a 1.5 cM region in 52% of the replicates.

One advantage of the SL model is that knowledge of marker haplotypes and thus, parental genotypes, is not required to map the QTL. With the available resources, the last generation could be genotyped for additional markers. When 20 markers covering 4.5 cM (0.25 cM spacing) were tested for the SL model, more than 74% of the replicates positioned the QTL within a 1.5 cM region. See Figure 1 for complete results of all model simulations.

In addition, the two regression methods do not require any assumptions about the LD-creating event. The ML method assumes that mutation in a QTL caused LD to form. However, there are several causes of LD, including population admixture, which could affect the ML method. If the assumption about LD is not true, mapping results from the ML method may not be as accurate.

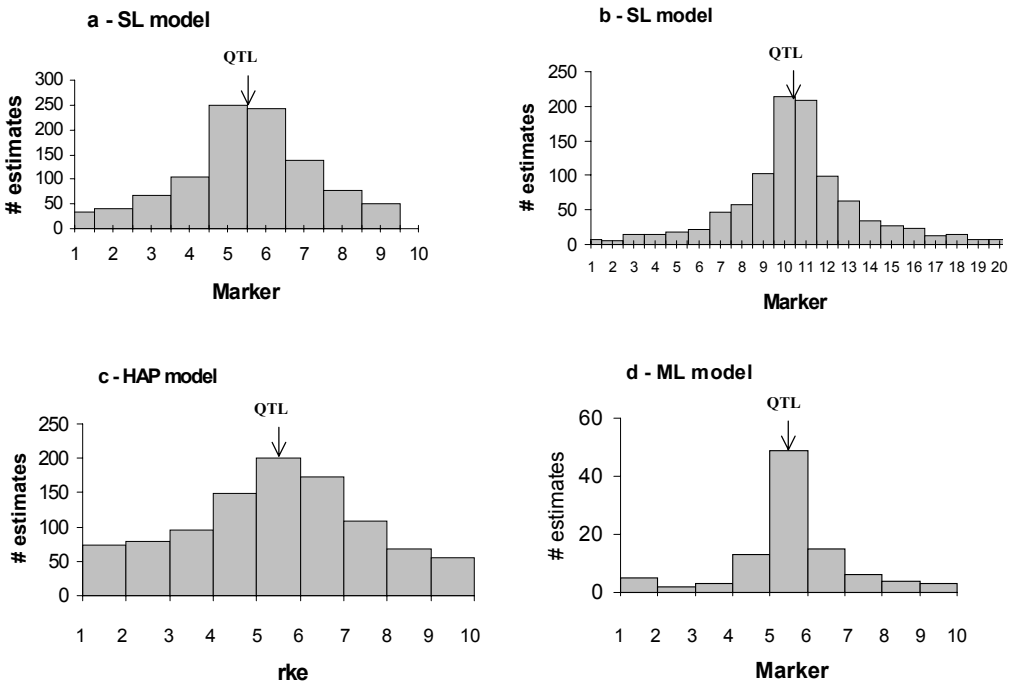

Figure 1. QTL position estimates for SL (a, b), HAP (c) and ML (d) models

CONCLUSION

A comparison between two regression models (SL and HAP) and a ML model revealed a moderate to large advantage of the ML method over the other models for fine mapping QTL. While the ML method is more complicated because assumptions about the creation of LD must be made and marker haplotypes must be known, the extra information does provide additional accuracy in positioning QTL.

REFERENCES

Crow J. A. (1986) "Basic Concepts in Population, Quantitative, and Evolutionary Genetics". W. H. Freeman and Company, New York.

Goddard, M. E. (1991) *Genet. Sel. Evol*. **23** (Suppl. 1) : 131s-134s.

Hill, W.G. and Weir, B.S. (1994) *Am. J. Hum. Genet*. **54** : 705-714.

Kaplan, N.L., Hill, W.G. and Weir, B.S. (1995) *Am. J. Hum. Genet*. **56** : 18-32.

Meuwissen, T.H.E. and Goddard, M.E. (2000) *Genetics* **155** : 421-430.

Rabinowitz, D. (1997) *Hum. Hered*. **47** : 342-350.

Terwilliger, J.D. (1995) *Am. J. Hum. Genet*. **56** : 777-787.