Theor Appl Genet (1991) 81:333–338

THEORETICAL AND APPLIED GENETICS

© Springer-Verlag 1991

Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny

S. J. Knapp

Department of Crop Science, Oregon State University, Corvallis, OR 97331, USA

Received November 15, 1989; Accepted August 22, 1990
Communicated by A. R. Hallauer

Summary. To maximize parameter estimation efficiency and statistical power and to estimate epistasis, the parameters of multiple quantitative trait loci (QTLs) must be simultaneously estimated. If multiple QTL affect a trait, then estimates of means of QTL genotypes from individual locus models are statistically biased. In this paper, I describe methods for estimating means of QTL genotypes and recombination frequencies between marker and quantitative trait loci using multilocus backcross, doubled haploid, recombinant inbred, and testcross progeny models. Expected values of marker genotype means were defined using no double or multiple crossover frequencies and flanking markers for linked and unlinked quantitative trait loci. The expected values for a particular model comprise a system of nonlinear equations that can be solved using an interative algorithm, e.g., the Gauss-Newton algorithm. The solutions are maximum likelihood estimates when the errors are normally distributed. A linear model for estimating the parameters of unlinked quantitative trait loci was found by transforming the nonlinear model. Recombination frequency estimators were defined using this linear model. Certain means of linked QTLs are less efficiently estimated than means of unlinked QTLs.

Key words: Allozymes – Restriction fragment length polymorphism – Linkage maps – Nonlinear models

Introduction

Restriction fragment length polymorphism (RFLP) linkage maps make it technically feasible to map quantitative trait loci (QTLs) using the phenotypes of markers as independent variable values (Lander and Botstein 1989; Knapp et al. 1990). Several methods have been described for mapping QTLs using backcross (BC), doubled haploid (DH), recombinant inbred (RI), F_2, or F_3 progeny (Weller 1986; Jensen 1989; Lander and Botstein 1989; Simpson 1989; Knapp et al. 1990). These methods use models that describe the cosegregation of a QTL and linked flanking codominant marker loci, hereafter called flanking marker models (Lander and Botstein 1989; Jensen 1989; Knapp et al. 1990), or the cosegregation of a quantitative trait locus and a linked codominant marker locus, hereafter called individual marker models (Weller 1986; Simpson 1989).

Methods appropriate for either model are inadequate when multiple loci affect a quantitative trait. Statistical power is compromised, estimates of means of QTL genotypes are biased, and epistasis cannot be estimated. In this paper, I describe methods for simultaneously estimating the parameters of linked or unlinked QTLs using flanking marker models for BC, DH, or RI. These methods apply to certain testcross progeny as well (Knapp et al. 1990). The expected values for these progeny types are equivalent except for notation and linear transformation differences (Knapp et al. 1990). Doubled haploid progeny models are used throughout this paper to describe models that are applicable to the entire group.

Multiple independent quantitative trait loci

A genetic model for two unlinked QTLs

The flanking marker model for two unlinked quantitative trait loci describes the cosegregation of six loci. Let A and B denote linked codominant molecular marker loci, C and D denote linked codominant molecular marker loci that segregate independently of A and B, and Q and Q* denote quantitative trait loci between A and B and C and

D, respectively. In addition, let $r_1, r_2, r_{12}, r_3, r_4$, and r_{34} denote recombination frequencies between A and Q, B and Q, A and B, C and Q*, D and Q*, and C and D, respectively.

Suppose DH lines are derived by self-fertilizing doubled haploids derived from the F_1 between inbred lines P_1 and P_2. Alleles in P_1 and P_2 are assigned subscripts of 1 and 2, respectively. The genotypes of P_1 and P_2 are $A_1A_1Q_1Q_1B_1B_1C_1C_1Q_1^*Q_1^*D_1D_1$ and $A_2A_2Q_2Q_2B_2B_2C_2C_2Q_2^*Q_2^*D_2D_2$, respectively. QQ* genotypes in the doubled haploid population are $Q_1Q_1Q_1^*Q_1^*$, $Q_1Q_1Q_2^*Q_2^*$, $Q_2Q_2Q_1^*Q_1^*$, and $Q_2Q_2Q_2^*Q_2^*$. Let μ_{1111}, μ_{1122}, μ_{2211}, and μ_{2222} denote means of QTL genotypes $Q_1Q_1Q_1^*Q_1^*$, $Q_1Q_1Q_2^*Q_2^*$, $Q_2Q_2Q_1^*Q_1^*$, and $Q_2Q_2Q_2^*Q_2^*$, respectively.

Expected values of marker genotype means were derived for two unlinked loci using expected frequencies of quantitative trait locus gentoypes within ABCD marker gentoypic classes (Table 1). These frequencies were derived by taking products of probabilities of Q locus genotypes within AB marker genotypic classes and Q* locus genotypes within CD marker genotypic classes, using no double crossover recombination frequencies ($r_{12}=r_1+r_2$ and $r_{34}=r_3+r_4$) (Table 1). I substituted $r_{12}-r_1$ for r_2 and $r_{34}-r_3$ for r_4 to simplify the model (Table 1); thus, expected values were defined as functions of $\mu_{1111}, \mu_{1122}, \mu_{2211}, \mu_{2222}, \varrho_1=r_1/r_{12}$, and $\varrho_2=r_3/r_{34}$.

Progeny within parental marker classes for unlinked QTLs have known QTL genotypes when there are no double crossovers within the A and B or C and D chromosome segments. This is well known and extends to k independent loci. Because double crossovers arise in practice, estimators defined for these models may be genetically biased. The bias, which is usually negligible, is a function of the double crossover frequency. Bias decreases as the distance between flanking markers and the coefficient of coincidence decreases, assuming positive interference.

Estimates of means of QTL genotypes can be used to estimate the intralocus effects of Q and Q* and the interlocus effect between Q and Q*. The additive effects of Q and Q* are estimated using the contrasts

$$\alpha=\hat{\mu}_{1111}+\hat{\mu}_{1122}-\hat{\mu}_{2211}-\hat{\mu}_{2222}$$

and

$$\alpha^*=\hat{\mu}_{1111}-\hat{\mu}_{1122}+\hat{\mu}_{2211}-\hat{\mu}_{2222},$$

respectively, whereas the additive-by-additive epistatic effect between Q and Q* is estimated using the contrast

$$\alpha\alpha^*=\hat{\mu}_{1111}-\hat{\mu}_{1122}-\hat{\mu}_{2211}+\hat{\mu}_{2222}.$$

Statistical models for two unlinked QTLs

The doubled haploid equations (Table 1) for two independent QTLs comprise a system of nonlinear equations

Table 1. Marker genotypes and expected values of marker genotype means for doubled haploid progeny for two unlinked quantitative trait loci. Expected values are for no double crossovers in the AQB or CQ*D chromosome segments where A and B are linked loci, C and D are linked loci that are independent of A and B, and Q and Q* are quantitative trait loci between A and B and C and D, respectively

Marker genotype	Expected value of marker genotype mean[a]
$A_1A_1B_1B_1C_1C_1D_1D_1$	$\theta_1=\mu_{1111}$
$A_1A_1B_1B_1C_1C_1D_2D_2$	$\theta_2=(1-\varrho_2)\,\mu_{1111}+\varrho_2\,\mu_{1122}$
$A_1A_1B_1B_1C_2C_2D_1D_1$	$\theta_3=\varrho_2\,\mu_{1111}+(1-\varrho_2)\,\mu_{1122}$
$A_1A_1B_1B_1C_2C_2D_2D_2$	$\theta_4=\mu_{1122}$
$A_1A_1B_2B_2C_1C_1D_1D_1$	$\theta_5=(1-\varrho_1)\,\mu_{1111}+\varrho_1\,\mu_{2211}$
$A_1A_1B_2B_2C_1C_1D_2D_2$	$\theta_6=(1-\varrho_1)(1-\varrho_2)\,\mu_{1111}+(1-\varrho_1)\varrho_2\,\mu_{1122}+\varrho_1(1-\varrho_2)\,\mu_{2211}+\varrho_1\varrho_2\,\mu_{2222}$
$A_1A_1B_2B_2C_2C_2D_1D_1$	$\theta_7=(1-\varrho_1)\varrho_2\,\mu_{1111}+(1-\varrho_1)(1-\varrho_2)\,\mu_{1122}+\varrho_1\varrho_2\,\mu_{2211}+\varrho_1(1-\varrho_2)\,\mu_{2222}$
$A_1A_1B_2B_2C_2C_2D_2D_2$	$\theta_8=(1-\varrho_1)\,\mu_{1122}+\varrho_1\,\mu_{2222}$
$A_2A_2B_1B_1C_1C_1D_1D_1$	$\theta_9=\varrho_1\,\mu_{1111}+(1-\varrho_1)\,\mu_{2211}$
$A_2A_2B_1B_1C_1C_1D_2D_2$	$\theta_{10}=\varrho_1(1-\varrho_2)\,\mu_{1111}+\varrho_1\varrho_2\,\mu_{1122}+(1-\varrho_1)(1-\varrho_2)\,\mu_{2211}+(1-\varrho_1)\varrho_2\,\mu_{1122}$
$A_2A_2B_1B_1C_2C_2D_1D_1$	$\theta_{11}=\varrho_1\varrho_2\,\mu_{1111}+\varrho_1(1-\varrho_2)\,\mu_{1122}+(1-\varrho_1)\varrho_2\,\mu_{2211}+(1-\varrho_1)(1-\varrho_2)\,\mu_{2222}$
$A_2A_2B_1B_1C_2C_2D_2D_2$	$\theta_{12}=\varrho_1\,\mu_{1122}+(1-\varrho_1)\,\mu_{2222}$
$A_2A_2B_2B_2C_1C_1D_1D_1$	$\theta_{13}=\mu_{2211}$
$A_2A_2B_2B_2C_1C_1D_2D_2$	$\theta_{14}=(1-\varrho_2)\,\mu_{2211}+\varrho_2\,\mu_{2222}$
$A_2A_2B_2B_2C_2C_2D_1D_1$	$\theta_{15}=\varrho_2\,\mu_{2211}+(1-\varrho_2)\,\mu_{2222}$
$A_2A_2B_2B_2C_2C_2D_2D_2$	$\theta_{16}=\mu_{2222}$

[a] $\varrho_1=r_1/r_{12}$ and $\varrho_2=r_3/r_{34}$. $\mu_{1111}, \mu_{1122}, \mu_{2211}$, and μ_{2222} are means of QTL genotypes $Q_1Q_1Q_1^*Q_1^*$, $Q_1Q_1Q_2^*Q_2^*$, $Q_2Q_2Q_1^*Q_1^*$, and $Q_2Q_2Q_2^*Q_2^*$, respectively

that can be solved for $\theta=[\mu_{1111}\ \mu_{1122}\ \mu_{2211}\ \mu_{2222}\ r_1\ r_3]'$. I found equalities between expected values for certain marker classes and used them to simplify the system of equations. For example,

$$\theta_1+\theta_4=\theta_2+\theta_3=\mu_{1111}+\mu_{1122}$$
$$=(1-\varrho_2)\,\mu_{1111}+\varrho_2\,\mu_{1122}+\varrho_2\,\mu_{1111}+(1-\varrho_2)\,\mu_{1122};$$

thus, $\theta_3=\theta_1+\theta_4-\theta_2$ where $\varrho_2=r_3/r_{34}$. To simplify the model, I substituted $\theta_1+\theta_4-\theta_2$ for θ_3. Likewise, I substituted the right-hand sides of

$$\theta_9=\theta_{13}+\theta_{16}-\theta_5,$$

$$\theta_{12}=\theta_1+\theta_4-\theta_8,$$

and

$$\theta_{15}=\theta_{13}+\theta_{16}-\theta_{14}$$

for the left-hand sides (Table 1). These simplifications led to the nonlinear model

$$\begin{aligned}y_i=&\mu_{1111}(x_1+x_3+x_{12})\\&+[(1-\varrho_2)\mu_{1111}+\varrho_2\mu_{1122}](x_2-x_3)\\&+\mu_{1122}(x_3+x_4+x_{12})\\&+[(1-\varrho_1)\mu_{1111}+\varrho_1\mu_{2211}](x_5-x_9)\\&+[(1-\varrho_1)\mu_{1122}+\varrho_1\mu_{2222}](x_8-x_{12})\\&+\mu_{2211}(x_9+x_{13}+x_{15})\\&+[(1-\varrho_2)\mu_{1111}+\varrho_2\mu_{1122}](x_{14}-x_{15})\\&+\mu_{2222}(x_9+x_{15}+x_{16})\\&+[(1-\varrho_1)(1-\varrho_2)\mu_{1111}+(1-\varrho_1)\varrho_2\mu_{1122}\\&\quad+\varrho_1(1-\varrho_2)\mu_{2211}+\varrho_1\varrho_2\mu_{2222}]x_6\\&+[(1-\varrho_1)\varrho_2\mu_{1111}+(1-\varrho_1)(1-\varrho_2)\mu_{1122}\\&\quad+\varrho_1\varrho_2\mu_{2211}+\varrho_1(1-\varrho_2)\mu_{2222}]x_7\\&+[\varrho_1(1-\varrho_2)\mu_{1111}+\varrho_1\varrho_2\mu_{1122}\\&\quad+(1-\varrho_1)(1-\varrho_2)\mu_{2211}+(1-\varrho_1)\varrho_2\mu_{2222}]x_{10}\\&+[\varrho_1\varrho_2\mu_{1111}+\varrho_1(1-\varrho_2)\mu_{1122}\\&\quad+(1-\varrho_1)\varrho_2\mu_{2211}\\&\quad+(1-\varrho_1)(1-\varrho_2)\mu_{2222}]x_{11}+e_i,\end{aligned}\tag{1}$$

where y_i is the i^{th} observation of the dependent variable (quantitative trait), e_i is the experimental error of the i^{th} observation, $i=1, 2, \ldots, n$, n is the number of doubled haploid lines, x_k is a dummy variable indexing the k^{th} marker genotypic class, and $k=1, 2, \ldots, 16$. $x_k=1$ if an observation belongs to the k^{th} marker genotypic class, otherwise $x_k=0$.

To facilitate estimating the parameters of model 1, $\hat{r}_{12}-r_1$ and $\hat{r}_{34}-r_3$ can be substituted for r_2 and r_4, respectively, where $\hat{r}_{12}$ and $\hat{r}_{34}$ are maximum likelihood estimates (Allard 1956) of recombination frequencies between marker loci. Estimates within the parameter space are assured by using these constraints and bounding estimation using $0.0\leq r_1\leq\hat{r}_{12}$ and $0.0\leq r_3\leq\hat{r}_{34}$. We have described programs for estimating the parameters of model 1 which use these constraints and bounds (Knapp and Bridges 1990). These estimates are maximum likelihood estimates when the e_i are independently and normally distributed (Gallant 1987).

A linear model can be defined by expressing the parameters of model 1 as a function of marker genotype means instead of the underlying QTL parameters. This leads to

$$\begin{aligned}y_i=&\theta_1(x_1+x_3+x_{12})+\theta_2(x_2-x_3)+\theta_4(x_3+x_4+x_{12})\\&+\theta_5(x_5-x_9)+\theta_6x_6+\theta_7x_7+\theta_8(x_8-x_{12})\\&+\theta_{10}x_{10}+\theta_{11}x_{11}+\theta_{13}(x_9+x_{13}+x_{15})\\&+\theta_{14}(x_{14}-x_{15})+\theta_{16}(x_9+x_{15}+x_{16})+e_i\\=&\mu_{1111}(x_1+x_3+x_{12})+\theta_2(x_2-x_3)+\mu_{1122}(x_3+x_4+x_{12})\\&+\theta_5(x_5-x_9)+\theta_6x_6+\theta_7x_7+\theta_8(x_8-x_{12})\\&+\theta_{10}x_{10}+\theta_{11}x_{11}+\mu_{2211}(x_9+x_{13}+x_{15})\\&+\theta_{14}(x_{14}-x_{15})+\mu_{2222}(x_9+x_{15}+x_{16})+e_i.\end{aligned}\tag{2}$$

This linear model can be used to estimate recombination frequencies, in addition to QTL means, but the nonlinear model is superior for this purpose. I found estimators of r_1 and r_3 that are functions of the parameters of model 2 and r_{12} and r_{34}. Solving

$$\begin{aligned}\theta_5+\theta_8=&(1-\varrho_1)\mu_{1111}+(1-\varrho_1)\mu_{1122}\\&+\varrho_1\mu_{2211}+\varrho_1\mu_{2222}\end{aligned}$$

for r_1, I found

$$r_1=\frac{r_{12}[\theta_5+\theta_8-(\mu_{1111}+\mu_{1122})]}{\mu_{2211}+\mu_{2222}-(\mu_{1111}+\mu_{1122})}.\tag{3}$$

Likewise, solving

$$\begin{aligned}\theta_2+\theta_{14}=&(1-\varrho_2)\mu_{1111}+\varrho_2\mu_{1122}\\&+(1-\varrho_2)\mu_{2211}+\varrho_2\mu_{2222}\end{aligned}$$

for r_3, I found

$$r_3=\frac{r_{34}[\theta_2+\theta_{14}-(\mu_{1111}+\mu_{2211})]}{\mu_{1122}+\mu_{2222}-(\mu_{1111}+\mu_{2211})}.\tag{4}$$

r_1 and r_3 can be estimated by substituting $\hat{r}_{12}$ and $\hat{r}_{34}$ and estimates of μ_{1111}, μ_{1122}, μ_{2211}, μ_{2222}, θ_5, θ_8, θ_2, and θ_{14} from model 2.

Extension to additional unliked loci are straightforward. The number of marker genotypic classes and equations is 4^k where k is the number of quantitative trait loci. The practical limits of k are quickly reached because the number of equations increases exponentially; nevertheless, the parameters of the nonlinear models can be estimated when there are missing marker phenotypic classes, as long as there is information about every parameter. Recombination frequency estimation efficiency, however, may suffer if the number of observations for certain equations is sparse.

QTL means can be efficiently estimated using multilocus models, but the sample sizes of the multilocus recombinant classes may be insufficient to get meaningful estimates of recombination frequencies.

Model 1 can be used to simultaneously estimate the parameters of two independent QTLs. This model can be used to estimate epistatic effects and to get statistically unbiased estimates of QTL genotype means. Estimates of QTL means are biased when loci affecting a quantitative trait are excluded from the model. Bias arises when observations for loci excluded from the model are unbalanced, i.e., when genotypes excluded from the model are un-

equally sampled, and may inflate or deflate estimates of means for genotypes included in the model. The size of the bias is a function of sample size, the extent of the unbalance, and the size of effects of QTLs excluded from the model.

Multiple linked quantitative trait loci

Genetic models for two or three linked QTLs

The model for two linked QTLs describes the cosegregation of five loci (Table 2). Let A, B, and C denote linked codominant molecular marker loci and Q and Q* denote quantitative trait loci between A and B and B and C, respectively, where r_1, r_2, r_{12}, r_3, r_4, and r_{34} are recombination frequencies between A and Q, B and Q, A and B, B and Q*, C and Q*, and B and C, respectively. This model was derived using no multiple crossover frequencies and definitions equivalent to those for unlinked QTLs (Table 2). There are several different kinds of multiple crossovers, but there are no observations for determining certain kinds of multiple crossovers. The genotypes of the parental lines P_1 and P_2 are $A_1A_1Q_1Q_1B_1B_1Q_1^*Q_1^*C_1C_1$ and $A_2A_2Q_2Q_2B_2B_2Q_2^*Q_2^*C_2C_2$, respectively. Hypothesized QTL genotypes in a doubled haploid population are $Q_1Q_1Q_1^*Q_1^*$, $Q_1Q_1Q_2^*Q_2^*$, $Q_2Q_2Q_1^*Q_1^*$, and $Q_2Q_2Q_2^*Q_2^*$. Let μ_{1111}, μ_{1122}, μ_{2211}, and μ_{2222} denote means of QTL genotypes $Q_1Q_1Q_1^*Q_1^*$, $Q_1Q_1Q_2^*Q_2^*$, $Q_2Q_2Q_1^*Q_1^*$, and $Q_2Q_2Q_2^*Q_2^*$, respectively.

The model for three linked QTLs describes the cosegregation of seven loci (Table 3). Let A, B, C, and D denote linked codominant molecular marker loci and Q, Q* and Q′ denote quantitative trait loci between A and B, B and C, and C and D, respectively, where r_1, r_2, r_{12}, r_3, r_4, r_{34}, r_5, r_6, and r_{56} are recombination frequencies between A and Q, B and Q, A and B, B and Q*, C and Q*, B and C, C and Q′, D and Q′, and C and D, respectively. The three-locus model was derived using no multiple crossover frequencies and definitions equivalent to those for two linked QTLs (Table 3). The genotypes of the parents, P_1 and P_2, are $A_1A_1Q_1Q_1B_1B_1Q_1^*Q_1^*C_1C_1Q_1'Q_1'D_1D_1$ and $A_2A_2Q_2Q_2B_2B_2Q_2^*Q_2^*C_2C_2Q_2'Q_2'D_2D_2$, respectively. Hypothesized QQ*Q′ genotypes are $Q_1Q_1Q_1^*Q_1^*Q_1'Q_1'$, $Q_1Q_1Q_2^*Q_2^*Q_2'Q_2'$, $Q_2Q_2Q_1^*Q_1^*Q_1'Q_1'$, $Q_1Q_1Q_1^*Q_1^*Q_2'Q_2'$, $Q_2Q_2Q_2^*Q_2^*Q_1'Q_1'$, and $Q_2Q_2Q_2^*Q_2^*Q_2'Q_2'$. Let μ_{111111}, μ_{112222}, μ_{221111}, μ_{111122}, μ_{222211}, and μ_{222222} denote means of QTL genotypes $Q_1Q_1Q_1^*Q_1^*Q_1'Q_1'$, $Q_1Q_1Q_2^*Q_2^*Q_2'Q_2'$, $Q_2Q_2Q_1^*Q_1^*Q_1'Q_1'$, $Q_1Q_1Q_1^*Q_1^*Q_2'Q_2'$, $Q_2Q_2Q_2^*Q_2^*Q_1'Q_1'$, and $Q_2Q_2Q_2^*Q_2^*Q_2'Q_2'$, respectively.

Statistical models for two or three linked QTLs

Models for linked loci (Tables 2 and 4) are quite different from models for unlinked loci (Table 1). Unlike flanking marker models for unlinked loci (Table 1), the means of certain genotypes for linked QTLs are not well determined; specifically, the means of recombinant QTL genotypes are not well determined, whereas the means of fully nonrecombinant classes are well determined. The differ-

Table 2. Marker genotypes and expected values of marker genotype means for doubled haploid progeny for two linked quantitative trait loci. Expected values are for no multiple crossovers in the AQBQ*C chromosome segments where A, B and C are linked codominant molecular marker loci and Q and Q* are quantitative trait loci between A and B and B and C, respectively

Marker genotype	Expected value of marker genotype mean[a]
$A_1A_1B_1B_1C_1C_1$	$\theta_1=\mu_{1111}$
$A_1A_1B_2B_2C_2C_2$	$\theta_2=(1-\varrho_1)\,\mu_{1122}+\varrho_1\,\mu_{2222}$
$A_2A_2B_1B_1C_1C_1$	$\theta_3=\varrho_1\,\mu_{1111}+(1-\varrho_1)\,\mu_{2211}$
$A_1A_1B_1B_1C_2C_2$	$\theta_4=(1-\varrho_2)\,\mu_{1111}+\varrho_2\,\mu_{1122}$
$A_2A_2B_2B_2C_1C_1$	$\theta_5=\varrho_2\,\mu_{2211}+(1-\varrho_2)\,\mu_{2222}$
$A_2A_2B_2B_2C_2C_2$	$\theta_6=\mu_{2222}$

[a] $\varrho_1=r_1/r_{12}$ and $\varrho_2=r_3/r_{34}$, where $r_{12}=r_1+r_2$, $r_{34}=r_3+r_4$, and r_1, r_2, r_{12}, r_3, r_4, and r_{34} are recombination frequencies between A and Q, B and Q, A and B, B and Q*, C and Q*, and B and C, respectively. μ_{1111}, μ_{1122}, μ_{2211}, and μ_{2222} are means of QTL genotypes $Q_1Q_1Q_1^*Q_1^*$, $Q_1Q_1Q_2^*Q_2^*$, $Q_2Q_2Q_1^*Q_1^*$, and $Q_2Q_2Q_2^*Q_2^*$, respectively

Table 3. Marker genotypes and expected values of marker genotype means for doubled haploid progeny for three linked quantitative trait loci. Expected values are for no multiple crossovers in the AQBQ*CQ′D chromosome segment where A, B, C and D are linked codominant molecular marker loci and Q, Q*, and Q′ are quantitative trait loci between A and B, B and C, and C and D, respectively

Marker genotype	Expected value of marker genotype mean[a]
$A_1A_1B_1B_1C_1C_1D_1D_1$	$\theta_1=\mu_{111111}$
$A_1A_1B_2B_2C_2C_2D_2D_2$	$\theta_2=(1-\varrho_1)\,\mu_{112222}+\varrho_1\,\mu_{222222}$
$A_2A_2B_1B_1C_1C_1D_1D_1$	$\theta_3=\varrho_1\,\mu_{111111}+(1-\varrho_1)\,\mu_{221111}$
$A_1A_1B_1B_1C_2C_2D_2D_2$	$\theta_4=(1-\varrho_2)\,\mu_{111122}+\varrho_2\,\mu_{112222}$
$A_2A_2B_2B_2C_1C_1D_1D_1$	$\theta_5=\varrho_2\,\mu_{221111}+(1-\varrho_2)\,\mu_{222211}$
$A_1A_1B_1B_1C_1C_1D_2D_2$	$\theta_6=(1-\varrho_3)\,\mu_{111111}+\varrho_3\,\mu_{111122}$
$A_2A_2B_2B_2C_2C_2D_1D_1$	$\theta_7=\varrho_3\,\mu_{222211}+(1-\varrho_3)\,\mu_{222222}$
$A_2A_2B_2B_2C_2C_2D_2D_2$	$\theta_8=\mu_{222222}$

[a] $\varrho_1=r_1/r_{12}$, $\varrho_2=r_3/r_{34}$, and $\varrho_3=r_5/r_{56}$, where $r_{12}=r_1+r_2$, $r_{34}=r_3+r_4$, and $r_{56}=r_5+r_6$ and r_1, r_2, r_{12}, r_3, r_4, r_{34}, r_5, r_6, and r_{56} are recombination frequencies between A and Q, B and Q, A and B, B and Q*, C and Q*, B and C, C and Q′, and C and D, respectively. μ_{111111}, μ_{112222}, μ_{221111}, μ_{111122}, μ_{22211}, and μ_{222222} are means of QTL genotypes $Q_1Q_1Q_1^*Q_1^*Q_1'Q_1'$, $Q_1Q_1Q_2^*Q_2^*Q_2'Q_2'$, $Q_2Q_2Q_1^*Q_1^*Q_1'Q_1'$, $Q_1Q_1Q_1^*Q_1^*Q_2'Q_2'$, $Q_2Q_2Q_2^*Q_2^*Q_1'Q_1'$, and $Q_2Q_2Q_2^*Q_2^*Q_2'Q_2'$, respectively

ence is equivalent to the difference between individual and flanking marker models (Lander and Botstein 1989; Knapp et al. 1990). Because of this, models for multiple linked QTLs cannot be transformed to get linear models for estimating the means of QTL genotypes (Tables 2 and 3).

The genetic models for linked QTLs are systems of nonlinear equations which, like model 1, can be solved using an iterative estimation method (Galland 1987). A nonlinear model for two linked QTLs is

$$\begin{aligned} y_i = {} & \mu_{1111} x_1 + [(1-\varrho_1)\mu_{1122} + \varrho_1 \mu_{2222}] x_2 \\ & + [\varrho_1 \mu_{1111} + (1-\varrho_1)\mu_{2211}] x_3 \\ & + [(1-\varrho_2)\mu_{1111} + \varrho_2 \mu_{1122}] x_3 \\ & + [(1-\varrho_2)\mu_{2222} + \varrho_2 \mu_{2211}] x_5 + \mu_{2222} x_6 + e_i, \end{aligned} \quad (5)$$

where x_k indexes the k^{th} marker phenotypic class (Table 2). Likewise, a nonlinear model for three linked QTLs is

$$\begin{aligned} y_i = {} & \mu_{111111} x_1 + [(1-\varrho_1)\mu_{112222} + \varrho_1 \mu_{222222}] x_2 \\ & + [\varrho_1 \mu_{111111} + (1-\varrho_1)\mu_{221111}] x_3 \\ & + [(1-\varrho_2)\mu_{111122} + \varrho_2 \mu_{112222}] x_4 \\ & + [(1-\varrho_2)\mu_{222211} + \varrho_2 \mu_{221111}] x_5 \\ & + [(1-\varrho_3)\mu_{111111} + \varrho_3 \mu_{111122}] x_6 \\ & + [(1-\varrho_3)\mu_{222222} + \varrho_3 \mu_{222211}] x_7 + \mu_{222222} x_8 + e_i, \end{aligned} \quad (6)$$

where x_k indexes marker phenotypic classes (Table 3). We have described programs for estimating the parameters of models 5 and 6 (Knapp and Bridges 1990) using nonlinear least squares (Gallant 1987). Derivatives needed for this purpose are listed in the Appendix.

The estimation of the parameters of 5 and 6 is straightforward; however, the structures of these models are quite different from those flanking marker models for unlinked loci and lead to estimation problems. The means of the $Q_1Q_1Q_2^*Q_2^*$ and $Q_2Q_2Q_1^*Q_1^*$ genotypes of 5, for example, are less well determined than the means of the $Q_1Q_1Q_1^*Q_1^*$ and $Q_2Q_2Q_2^*Q_2^*$ genotypes, because there are no marker classes that are strictly functions of specific recombinant QTL genoptypes (Tables 2 and 4). To illustrate this, I simulated a population of 250 doubled haploid lines segregating for two linked QTLs flanked by marker loci. The parametric values used to generate genotypic and phenotypic observations for marker and quantitative trait loci were $r_1 = 0.05$, $r_2 = 0.10$, $r_3 = 0.05$, $r_4 = 0.07$, $\mu_{1111} = 50.0$, $\mu_{1122} = 55.0$, $\mu_{2211} = 52.0$, $\mu_{2222} = 57.0$, and $\sigma^2 = 4.0$, where σ^2 is the error variance.

The variances of $\hat{\mu}_{1122}$ and $\hat{\mu}_{2211}$ were several times greater than those of $\hat{\mu}_{1111}$ and $\hat{\mu}_{2222}$ (Table 4). In addition, estimates of μ_{1122} and μ_{2211} were significantly different from those estimated using known QTL genotypes (known because the observations were simulated), whereas estimates of μ_{1122} and μ_{2211} were not different from those estimated using known QTL genotypes (Table 4).

Table 4. Estimates of the parameters of two linked QTLs in a doubled haploid population simulated using $r_1 = 0.05$, $r_2 = 0.10$, $r_3 = 0.05$, $r_4 = 0.07$, $\mu_{1111} = 50.0$, $\mu_{1122} = 55.0$, $\mu_{2211} = 52.0$, $\mu_{2222} = 57.0$, $\sigma^2 = 4.0$, and $n = 250$, where n is the number of doubled haploid lines. The estimated variance, coefficient of determination, and recombination frequencies between marker loci A and B and C and D were $\hat{\sigma}^2 = 4.09$, $R^2 = 0.75$, and $\hat{r}_{12} = 0.116$ and $\hat{r}_{34} = 0.136$, respectively

Parameter	Point estimate		Model 5 standard error	Model 5 interval estimate[c]
	Known QTL genotype[a]	Model 5[b]		
μ_{1111}	49.79	49.75	0.24	49.29, 50.21
μ_{1122}	55.37	53.30	3.85	45.71, 60.89
μ_{2211}	52.04	55.44	2.04	51.42, 59.46
μ_{2222}	57.17	57.15	0.23	56.70, 57.61
r_1	0.044	0.107	0.029	0.050, 0.017
r_3	0.068	0.120	0.133	−0.142, 0.382

[a] The QTL genotypes for these data were known because they were simulated. These estimates were determined using the known QTL genotypes

[b] Maximum likelihood estimates determined using nonlinear least squares

[c] The confidence coefficient was $1-\alpha = 0.95$

The statistical limitations of models for linked loci, which are equivalent to those for individual marker models, make it difficult to resolve ambiguities between adjacent flanking marker segments where significant effects are found for the individual segments. Differences found for a given segment may be due to effects of QTL within an adjacent segment or to a QTL within the segment. Either way, multilocus models must be used to test these hypotheses but, as the models and example illustrate, the means of certain QTL genotypes means are not efficiently estimated. This makes it difficult to differentiate between alternative hypotheses. Furthermore, these models do not address the problem of multiple QTLs within a particular segment.

Acknowledgements. This work was supported by a grant from the Medical Research Foundation of Oregon. Oregon Agricultural Experiment Station Technical Paper No. 9027.

Appendix

First-order partial derivatives are listed for nonlinear models 1, 5, and 6. Derivatives of 1, the model for two unlinked QTLs, are

$$\partial/\partial\mu_{1111} = x_1 + x_3 + x_{12} + (1-\varrho_2)(x_2 - x_3) + (1-\varrho_1)(x_5 - x_9),$$

$$\partial/\partial\mu_{1122} = \varrho_2(x_2 - x_3) + x_3 + x_4 + x_{12} + (1-\varrho_1)(x_8 - x_{12}),$$

$$\partial/\partial\mu_{2211} = \varrho_1(x_5 - x_9) + x_9 + x_{13} + x_{15} + (1-\varrho_2)(x_{14} - x_{15}),$$

$$\partial/\partial\mu_{2222} = \varrho_1(x_8 - x_{12}) + \varrho_2(x_{14} - x_{15}) + x_9 + x_{15} + x_{16},$$

$$\partial/\partial r_1 = [(\mu_{2211} - \mu_{1111})/r_{12}](x_5 - x_9) + [(\mu_{2222} - \mu_{1122})/r_{12}](x_8 - x_{12}),$$

and

$$\partial/\partial r_3 = [(\mu_{1122}-\mu_{1111})/r_{34}](x_2-x_3) + [(\mu_{2222}-\mu_{2211})/r_{34}](x_{14}-x_{15}).$$

Derivatives of 5, the model for two linked QTLs, are

$$\partial/\partial\mu_{1111} = x_1+\varrho_1 x_4+(1-\varrho_2)\,x_5,$$

$$\partial/\partial\mu_{1122} = (1-\varrho_1)\,x_3+\varrho_2 x_5,$$

$$\partial/\partial\mu_{2211} = (1-\varrho_1)\,x_4+\varrho_2 x_6,$$

$$\partial/\partial\mu_{2222} = x_2+\varrho_1 x_3+(1-\varrho_2)\,x_6,$$

$$\partial/\partial r_1 = [(\mu_{2222}-\mu_{1122})/r_{12}]\,x_3+[(\mu_{1111}-\mu_{2211})/r_{12}]\,x_4,$$

and

$$\partial/\partial r_3 = [(\mu_{2211}-\mu_{1111})/r_{34}]\,x_5+[(\mu_{2211}-\mu_{2222})/r_{34}]\,x_6.$$

Derivatives of 6, the model for three linked QTLs, are

$$\partial/\partial\mu_{111111} = x_1+\varrho_1 x_4+(1-\varrho_3)\,x_7,$$

$$\partial/\partial\mu_{112222} = (1-\varrho_1)\,x_3+\varrho_2 x_5,$$

$$\partial/\partial\mu_{221111} = (1-\varrho_1)\,x_4+\varrho_2 x_6,$$

$$\partial/\partial\mu_{111122} = (1-\varrho_2)\,x_5+\varrho_3 x_7,$$

$$\partial/\partial\mu_{222211} = (1-\varrho_2)\,x_6+\varrho_3 x_8,$$

$$\partial/\partial\mu_{222222} = x_2+\varrho_1 x_3+(1-\varrho_3)\,x_8,$$

$$\partial/\partial r_1 = [(\mu_{222222}-\mu_{112222})/r_{12}]\,x_3+[(\mu_{111111}-\mu_{221111})/r_{12}]\,x_4,$$

$$\partial/\partial r_3 = [(\mu_{112222}-\mu_{111122})/r_{34}]\,x_5+[(\mu_{221111}-\mu_{222211})/r_{34}]\,x_6,$$

and

$$\partial/\partial r_5 = [(\mu_{111122}-\mu_{111111})/r_{56}]\,x_7+[(\mu_{222211}-\mu_{222222})/r_{56}]\,x_8.$$

References

Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235–278.

Gallant RA (1987) Nonlinear statistical models. Wiley and Sons, New York

Jensen J (1989) Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. Theor Appl Genet 78:613–618

Knapp SJ, Bridges WC (1990) Programs for mapping quantitative trait loci using flanking molecular markers and nonlinear models. J Hered 81:234–235

Knapp SJ, Bridges WC, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage maps. Theor Appl Genet 79:583–592

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Simpson SP (1989) Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor Appl Genet 77:815–819

Weller JI (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627–640