Genet. Res., Camb. (1996), **67**, *pp.* 43–54 *With 2 text-figures*

Models for mapping quantitative trait loci (QTL) in progeny of non-inbred parents and their behaviour in presence of distorted segregation ratios

R. SCHÄFER-PREGL*, F. SALAMINI AND C. GEBHARDT
Max-Planck-Institut für Züchtungsforschung, Carl-v.-Linné-Weg 10, 50829 Köln, Germany

(*Received 5 April 1995 and in revised form 26 June 1995*)

Summary

In plants, models for mapping quantitative trait loci (QTL) based on flanking markers have been mainly developed for progenies of inbred lines. We propose two flanking marker models for QTL mapping in F_1 progenies of non-inbred parents. The first is based on the segregation of four different scorable alleles at a marker locus (the four-allele model) and the second (the common-allele model) on one scorable allele per marker locus segregating in both parents. These models are suitable for the majority of the allelic configurations which may occur in crosses between heterozygous parents. For both cases, when four scorable or one common-allele per marker locus segregate, additional algorithms were developed to estimate the recombination frequency between two marker loci. Tests carried out with simulated populations of various sizes indicate that the models provide a good estimate of QTL genotypic means and of recombination frequencies between flanking markers and between the marker loci and the QTL. The estimates of QTL genotypic means have a higher precision than the estimates of recombination frequencies. The four-allele model shows a higher ability to detect QTLs than the common-allele model. If segregation ratios are distorted, the power of both models and the precision of the estimates of recombination frequencies are reduced, whereas the accuracy of estimates of QTL genotype means is not affected by distorted segregation ratios. The power of the common-allele model is substantially reduced if QTL genotypic means depend on additive allelic interactions, whereas the four-allele model is less affected by the non-additive behaviour of QTL alleles.

1. Introduction

The quantitative variation which can be observed in populations of living organisms has a genetic component depending on small effects of numerous genes distributed over the whole genome (Nilson-Ehle, 1909; East, 1916). The loci influencing quantitative inherited traits are called quantitative trait loci (QTLs, Geldermann, 1975). The linkage between major genes and QTLs was first analysed considering a limited number of morphological markers (Sax, 1923; Rasmusson, 1933). Molecular markers such as RFLP (restriction fragment length polymorphisms), RAPDs (random amplified polymorphic DNA), microsatellites and markers based on AFLP technique (Vos *et al.* in press), have been recently developed, which are, in principle, unlimited in number and allow the construction of highly saturated linkage maps (Phillips & Vasil, 1994). In crop plants, RFLP markers map at distances between 1 and 20 Centimorgans and can be efficiently used for QTL mapping.

The linkage of a QTL to a single marker locus has been investigated using linear models (Soller & Brody, 1976; Soller *et al.* 1979; Edwards *et al.* 1987, 1992; Osborn *et al.* 1987; Stuber *et al.* 1987, 1992; Tanksley & Hewitt, 1988; Martin *et al.* 1989). These models tend to underestimate the effect of QTLs since the contrast between the mean values of a trait in two subpopulations with alternate marker alleles decreases when distance from a putative QTL to a marker locus increases (Lander & Botstein, 1986, 1989; Knapp, 1989). Nonlinear single marker models which estimate the effect of a QTL and the recombination frequency between a marker and a QTL have been published. In these models, however, the precision of relevant parameter estimates decreases when the recombination frequency increases (Weller, 1986).

Flanking marker models describe the location of putative QTLs within an interval defined by two

* To whom correspondence should be addressed.

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

linked marker loci. These models utilize more genetic information and, as a consequence, the estimate of the recombination frequency between marker loci and the QTL is more efficient (Weller, 1987; Lander & Botstein, 1989; Knapp & Bridges, 1990; Knapp *et al.* 1990; Jansen, 1992; Knapp, 1994). Lander & Botstein (1989) proposed a linear model where values of QTL genotypes were estimated by treating them as missing values. Jansen (1992) introduced a 'general mixture model' for several flanking marker configurations. Knapp *et al.* (1990) have elaborated nonlinear models for several types of segregating populations and used linear least square methods to estimate unknown QTL parameters.

Methods and models published so far for QTL analysis in plants are mainly designed for populations descending from inbred lines. For several reasons, however, inbred lines cannot be routinely obtained in some crop plants. A few cases were reported where QTLs have been allocated to chromosomes by adapting models developed for progenies of inbred lines (Leonards-Schippers *et al.* 1994). Haley *et al.* (1994) described a linear least square method to be used in crosses between outbred lines, where, for a series of fixed putative QTLs, phenotypic values are regressed onto additive and dominance coefficients for each individual.

In this paper, we describe flanking marker-based models for QTL analysis designed for F_1 populations derived from crosses among heterozygous parents. In addition, we propose models for the estimation of the recombination frequency R between two marker loci. To test the accuracy of the parameter estimated by the models, QTL mapping experiments have been simulated and the results presented.

2. Methods and results

(i) *Flanking marker model based on four sorable alleles per marker locus (the four-allele model)*

In the F_1 population derived from crossing of non-inbred parents, up to four different alleles can exist at a marker locus. Let A_1, A_2, A_3, and A_4 denote the alleles at the marker locus A, and B_1, B_2, B_3, and B_4 those at the locus B. Q_1, Q_2, Q_3, and Q_4 denote four different alleles at a putative QT locus lying in the interval between loci A and B. When the parents P1 and P2 of a cross have the allelic configurations $A_1A_2Q_1Q_2B_1B_2$ and $A_3A_4Q_3Q_4B_3B_4$, respectively, the four non-recombinant genotypes $A_1A_3Q_1Q_4B_1B_3$, $A_1A_4Q_1Q_4B_1B_4$, $A_2A_3Q_2Q_3B_2B_3$ and $A_2A_4Q_2Q_4B_2B_4$ are expected in equal frequency in their F_1 progeny (Fig. 1). In such a cross, however, and without consideration of double crossover events, the recombination between the two marker loci and the QTL generates up to 36 different genotypes. Given the fact that both the existence and the position of the QTL are unknown, only 16 different marker genotypes are scorable in the F_1 progeny (Table 1).

Let *r1* and *r2* denote the recombination frequency between A and QTL and B and the QTL, respectively, while q13, q14, q23 and q24 correspond to the average value of the QTL genotypes Q_1Q_3, Q_1Q_4, Q_2Q_3 and Q_2Q_4, respectively. When the alleles at the two marker loci are linked in coupling (A_1, B_1 and A_2, B_2 for P1 and A_3, B_3 and A_4, B_4 for P2), then

$$
\begin{aligned}
Y = {} & m_1 * q13 + m_2 * q14 + m_3 * q23 + m_4 * q24 + m_5 * \frac{q13 * r2 + q14 * r1}{r1 + r2} + m_6 * \frac{q14 * r2 + q13 * r1}{r1 + r2} \\
& + m_7 * \frac{q13 * r2 + q23 * r1}{r1 + r2} + m_8 * \frac{q23 * r2 + q13 * r1}{r1 + r2} + m_9 * \frac{q14 * r2 + q24 * r1}{r1 + r2} + m_{10} * \frac{q24 * r2 + q14 * r1}{r1 + r2} \\
& + m_{11} * \frac{q23 * r2 + q24 * r1}{r1 + r2} + m_{12} * \frac{q24 * r2 + q23 * r1}{r1 + r2} + m_{13} * \frac{q13 * r2^2 + q24 * r1^2 + q14 * r1 * r2 + q13 * r1 * r2}{r1^2 + r2^2 + 2 * r1 * r2} \\
& + m_{14} * \frac{q24 * r2^2 + q13 * r1^2 + q14 * r1 * r2 + q23 * r1 * r2}{r1^2 + r2^2 + 2 * r1 * r2} \\
& + m_{15} * \frac{q14 * r2^2 + q23 * r1^2 + q13 * r1 * r2 + q24 * r1 * r2}{r1^2 + r2^2 + 2 * r1 * r2} \\
& + m_{16} * \frac{q23 * r2^2 + q14 * r1^2 + q13 * r1 * r2 + q24 * r1 * r2}{r1^2 + r2^2 + 2 * r1 * r2} \\
& + e, \qquad (1)
\end{aligned}
$$

where Y represents the value expected for the trait mean of the marker classes m_1 to m_{16}, while e is the experimental error. m_1 to m_{16} are dummy variables given to the marker classes (a class groups all

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

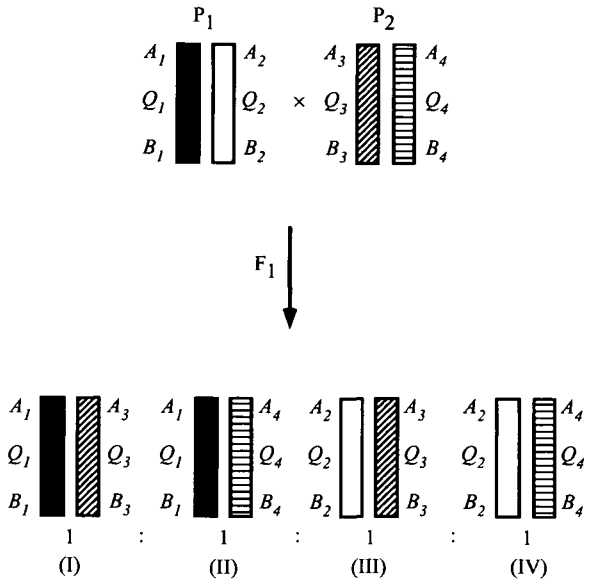

Fig. 1. Inheritance of a QTL flanked by marker loci A and B in the F_1 generation derived from crosses of heterozygous plants. Only the four non-recombinant genotypes are shown.

individuals with identical marker genotypes). If, for example, the marker genotype is $A_1A_3B_1B_3$, then the variable $m_1 = 1$, m_2 to m_{16} are 0 and Y represents the expected value of the marker genotype $A_1A_3B_1B_3$ ($Y = q13$), and so on. Y is a nonlinear function of the unknown parameters $r1$, $r2$, $q13$, $q14$, $q23$ and $q24$, which can be estimated using least square methods (see below). The QTL parameters which are estimated using this model are those listed in the first column of Table 3.

(ii) *Flanking marker model based on one scorable allele per marker locus common to both parents (common-allele model)*

Flanking marker models for QTL analysis have been developed for F_2 populations derived from inbred lines (Lander & Botstein, 1989; Knapp, 1990). In these models, two marker alleles, A_1 and A_2, are scored at the locus A, and two others, B_1 and B_2, at the linked locus B. Given an F_1 with the allelic configuration

$$\frac{A_1Q_1B_1}{A_2Q_2B_2} \times \frac{A_1Q_1B_1}{A_2Q_2B_2},$$

nine different marker genotypes will be present in F_2. Including the three QTL genotypes Q_1Q_1, Q_1Q_2 and Q_2Q_2, 27 different genotypes should be considered by the model. When only one marker allele can be scored per locus (in the case of, for example, genetic maps based on RAPD markers), the marker will segregate in an F_2 with a 3:1 ratio (presence *v.* absence). In crosses among heterozygous parents a proportion of the segregating marker alleles are common to the heterozygous parents P1 and P2. In such an F_1, if one of the two alleles is scored as null (= absence of the RFLP band; Gebhardt *et al.* 1991, 1994), the markers will segregate according to a 3:1 ratio. We will consider a cross where two flanking marker loci are flanking a QTL. This configuration can be written as

$$\frac{A_1Q_1B_1}{A_0Q_0B_0}(\mathrm{P1}) \times \frac{A_1Q_1B_1}{A_0Q_0B_0}(\mathrm{P2}),$$

where A_1 and B_1 indicate the scorable, common marker alleles and A_0 and B_0 the non-scorable ones. The marker loci have to be linked in coupling, since

Table 1. *Marker classes, marker genotypes and possible QTL genotypes in an F_1 population where four alleles segregate at both marker loci A and B*

Marker class	Marker genotype	QTL genotypes
m_1	$A_1A_3B_1B_3$	Q_1Q_3
m_2	$A_1A_4B_1B_4$	Q_1Q_4
m_3	$A_2A_3B_2B_3$	Q_2Q_3
m_4	$A_2A_4B_2B_4$	Q_2Q_4
m_5	$A_1A_3B_1B_4$	Q_1Q_3 or Q_1Q_4
m_6	$A_1A_4B_1B_3$	Q_1Q_3 or Q_1Q_4
m_7	$A_1A_3B_2B_3$	Q_1Q_3 or Q_2Q_3
m_8	$A_2A_3B_1B_3$	Q_1Q_3 or Q_2Q_3
m_9	$A_1A_4B_2B_4$	Q_1Q_4 or Q_2Q_4
m_{10}	$A_2A_4B_1B_4$	Q_1Q_4 or Q_2Q_4
m_{11}	$A_2A_3B_2B_4$	Q_2Q_3 or Q_2Q_4
m_{12}	$A_2A_4B_2B_3$	Q_2Q_3 or Q_2Q_4
m_{13}	$A_1A_3B_2B_4$	Q_1Q_3 or Q_1Q_4 or Q_2Q_3 or Q_2Q_4
m_{14}	$A_2A_4B_1B_3$	Q_1Q_3 or Q_1Q_4 or Q_2Q_3 or Q_2Q_4
m_{15}	$A_1A_4B_2B_3$	Q_1Q_3 or Q_1Q_4 or Q_2Q_3 or Q_2Q_4
m_{16}	$A_2A_3B_1B_4$	Q_1Q_3 or Q_1Q_4 or Q_2Q_3 or Q_2Q_4

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

the repulsion type and coupling/repulsion type of linkage cannot be distinguished based on distribution of marker phenotypes (Ritter *et al.* 1990). In the F_1 population only four different marker phenotypes will be distinguished (Table 2).

Let *r1* and *r2* again denote the recombination frequency between the locus *A* and the QTL and locus *B* and the QTL, respectively, while *q11*, *q10* and *q00* correspond to the means of the QTL genotypes Q_1Q_1, Q_1Q_0, and Q_0Q_0, respectively. Setting $R = r1 + r2$, where *R* is the recombination frequency between Locus *A* and locus *B*, and assuming that the alleles at the two marker loci are linked in coupling, then

$$\begin{aligned} NN = {} & (m_1+m_2+m_3+m_4)*(\tfrac{1}{2}-\tfrac{1}{2}*R)^2*NT \\ & +(m_5+m_6+m_7+m_8+m_9+m_{10}+m_{11}+m_{12}) \\ & *\tfrac{1}{4}*(1-R)*R*NT+(m_{13}+m_{14}+m_{15}+m_{16}) \\ & *\tfrac{1}{4}*R^2*NT. \end{aligned} \tag{4}$$

$$\begin{aligned} Y = {} & m_1*\left\{\left[(\tfrac{1}{2}-\tfrac{1}{2}R)^2*q11+\tfrac{1}{2}*(1-R)*(q11*r2+q10*r1)+\tfrac{1}{2}*(1-R)*(q11*r1+q10*r2)+\tfrac{1}{2}*(1-R)^2*q10\right.\right. \\ & \left.\left.+\tfrac{1}{2}*\frac{R^2*(q11*r1*r2+q00*r1*r2+q10*r1^2+q10*r2^2)}{2*r1*r2+r1^2+r2^2}\right]\Big/[3*(\tfrac{1}{2}-\tfrac{1}{2}*R)^2+(1-R)*R+\tfrac{1}{2}*R^2]\right\} \\ & +m_2*\frac{\tfrac{1}{4}*\dfrac{R^2*(q11*r2^2+2*q10*r1*r2+q00*r1^2)}{2*r1*r2+r1^2+r2^2}+\tfrac{1}{2}*(1-R)*(q10*r2+q00*r1)}{\tfrac{1}{4}*R^2+\tfrac{1}{2}*(1-R)*R} \\ & +m_3*\frac{\tfrac{1}{2}*(1-R)*(q10*r1+q00*r2)+\tfrac{1}{4}*\dfrac{R^2*(q11*r1^2+2*q10*r1*r2+q00*r2^2)}{2*r1*r2+r1^2+r2^2}}{\tfrac{1}{4}*R^2+\tfrac{1}{2}*(1-R)*R}+m_4*q00+e, \end{aligned} \tag{2}$$

where *Y* represents the expected value for the trait mean of the marker classes m_1 to m_4, and *e* is the experimental error. m_1 to m_4 are dummy variables given to the marker classes as described for the previous model. *Y* is a nonlinear function of the unknown parameters *r1*, *r2*, *q11*, *q10* and *q00*, which can be estimated using least square methods (see later).

(iii) *Estimation of the recombination frequency between the marker loci A and B*

In the models described above, r_1 and r_2 are not known. The relationship between r_1 and r_2 can be expressed as $RHO = r1/(r1+r2) = r1/R$ (Knapp *et al.* 1990, see Appendix I and II), thus, an estimation of *R* is required to calculate the values for r_1 and r_2.

An appropriate model for estimating *R* between marker loci with four scorable alleles per locus is:

The equivalent equation for estimating the recombination frequency *R* between two marker loci with one common allele each is:

$$\begin{aligned} NN = {} & m_1*(3*(\tfrac{1}{2}-\tfrac{1}{2}*R)^2+2*(\tfrac{1}{2}-\tfrac{1}{2}*R)*R)*NT \\ & +(m_2+m_3)*(\tfrac{1}{4}*R^2+(\tfrac{1}{2}-\tfrac{1}{2}*R)*R)*NT \\ & +m_4*\tfrac{1}{4}*R^2*NT, \end{aligned} \tag{5}$$

where *NN* represents the expected number of marker genotypes in the marker classes m_1 to m_4, and *NT* is the population size. Thus, *NN* is a nonlinear function of the recombination frequency *R*, which can be estimated as described below.

(iv) *Simulations, parameter estimation and test statistics*

SAS programs were developed using the SAS-functions RANNOR and RANUNI (SAS Institute

$$\begin{aligned} NN = {} & m_1*(\tfrac{1}{2}-\tfrac{1}{2}*R)^2*NT+m_2*(\tfrac{1}{2}-\tfrac{1}{2}*R)^2*NT+m_3*(\tfrac{1}{2}-\tfrac{1}{2}*R)^2*NT+m_4*(\tfrac{1}{2}-\tfrac{1}{2})*R^2*NT \\ & +m_5*\tfrac{1}{4}*(1-R)*R*NT+m_6*\tfrac{1}{4}*(1-R)*R*NT+m_7*\tfrac{1}{4}*(1-R)*R*NT+m_8*\tfrac{1}{4}*(1-R)*R*NT \\ & +m_9*\tfrac{1}{4}*(1-R)*R*NT+m_{10}*\tfrac{1}{4}*(1-R)*R*NT+m_{11}*\tfrac{1}{4}*(1-R)*R*NT+m_{12}*\tfrac{1}{4}*(1-R)*R*NT \\ & +m_{13}*\tfrac{1}{4}R^2*NT+m_{14}*\tfrac{1}{4}R^2*NT+m_{15}*\tfrac{1}{4}R^2*NT+m_{16}*\tfrac{1}{4}R^2*NT, \end{aligned} \tag{3}$$

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

Table 2. *Marker classes, marker phenotypes, marker genotypes and possible QTL genotypes in an F_1 population where one allele common to both parents segregates at both marker loci A and B*

Marker class	Marker phenotype	Marker genotype	QTL genotypes
m_1	α_1, β_1	$A_1A_1B_1B_1$	Q_1Q_1
		$A_1A_1B_1B_0$	Q_1Q_1 or Q_1Q_0
		$A_1A_0B_1B_1$	Q_1Q_1 or Q_1Q_0
		$A_1A_0B_1B_0$	Q_1Q_1 or Q_1Q_0 or Q_0Q_0
m_2	α_0, β_1	$A_0A_0B_1B_1$	Q_1Q_1 or Q_1Q_0 or Q_0Q_0
		$A_0A_0B_1B_0$	Q_1Q_0 or Q_0Q_0
m_3	α_1, β_0	$A_1A_1B_0B_0$	Q_1Q_1 or Q_1Q_0 or Q_0Q_0
		$A_1A_0B_0B_0$	Q_1Q_0 or Q_0Q_0
m_4	α_0, β_0	$A_0A_0B_0B_0$	Q_0Q_0

Inc., 1990) which assemble virtual populations with the marker allele configurations as described in the previous sections and with given values of the parameters which are usually estimated by the two models. Double crossover events were not considered. Populations with a size of 100, 200 and 1000 F_1 individuals were created 200 times, each time with different SEED-values for the RANNOR- and RANUNI-functions. For each of these replications new normal distributions were generated as follows: with a variance of 2; with fixed values for the parameters *r1* and *r2*; and with mean values for QTL genotypes specified as parameter for the RANNOR-function. Similar test simulations were performed with 1000 datasets, enlarging the size of the F_1 from 35 to 1035 in steps of 1, to analyse the relationship between the population size and the average *P* value (see also later and Table 3), where the *P* value is the probability of the non-existence of a QTL. The lower the value of *p*, the more likely is the existence of a QTL. In all cases, the largest variation among QTL genotypic values was approximately 1 standard deviation.

In simulations performed with fixed population sizes, the average values for QTL genotypes were chosen assuming codominant inheritance and no epistasis. In simulations with increasing population sizes additional average values for QTL genotpes were set according to non-additive allelic interactions. Distorted segregations ratios were simulated by reducing the frequency of one non-recombinant gamete from 0·5 to 0·2, while adjusting the sum of all gamete frequencies to 1. After having created the populations according to the chosen frequencies of gametes, Chi-square tests were performed to determine the significance of the deviation of the observed segregation ratio from the ratios 1:1:1:1 and 3:1.

The methods of parameter estimation and test statistics described below for the analysis of a putative QTL located in an interval between two segregating markers have been first utilized and adopted for QTL analysis in F_2 and BC populations by Knapp *et al.* (1990).

To estimate the parameters of the models, iterative techniques of the NLIN procedure (SAS Institute, Inc., 1989) were used. As the four-allele and the common-allele models described above are nonlinear regression functions with unknown parameters, NLIN treats them as a series of linear regressions at each iteration step and evaluates the residual error sum of squares. The vector of the unknown parameters is changed at each iteration step, until the error sum of squares is minimized. The Gauss–Newton method was chosen to compute parameter changes. To use this algorithm, derivatives of the models had to be specified (SAS Institute Inc., 1989, see Appendix).

Parameters were estimated for the full model assuming that a QTL exists between a pair of markers and for the reduced model which assumes that no QTL is located between the same pair of markers. The Likelihood ratio *L* for testing the hypothesis of no QTL is

$$L = \frac{(SSE_R - SSE_F)/(df_R - df_F)}{(SSE_F/df_F)}, \quad (6)$$

where SSE_F and SSE_R are the error sum of squares derived from the NLIN procedure for the full model and the reduced model, respectively, and df_F and df_R are the degrees of freedom for the full model and the reduced model, respectively (Gallant, 1987; Knapp & Bridges, 1990). df_F and df_R are given by $N-P_F$ and $N-P_R$, where N is the number of valid observations and P_F and P_R are the number of the estimated parameters for the full model and the reduced model, respectively.

L is distributed as an *F* random variable (Gallant, 1987; Knapp & Bridges, 1990), such that if $L > F_{\alpha, df_N, df_F, 0}$ where $F_{\alpha, df_N, df_F, 0}$ is a value from the *F*-distribution chosen for a desired α-threshold, and df_N is $df_F - df_R$, the hypothesis of no QTL can be rejected. The *P* value obtained for a Likelihood ratio is given

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

Table 3. *Simulated data, parameter estimates and their variances for a QTL flanked by two marker loci with four alleles per locus (the four-allele model), assuming an F_1 population size of (a) 100 and (b) 200 individuals, and normal and distorted segregation ratios*

Parameter...		Estimated mean[b]		Variance[b]	
Segregation...	Simulated (mean)[a]	Normal	Distorted[c]	Normal	Distorted[c]
		(*a*) 100 Individuals			
R (%)	19·00	17·78	15·18	10·81	28·35
RHO[d] (%)	36·84	41·62	37·94	984·74	895·88
r1 (%)	7·00	7·53	5·59	35·56	25·47
r2 (%)	12·00	10·25	9·58	33·18	37·23
q13	50·60	50·59	50·59	0·11	0·22
q14	51·30	51·30	51·30	0·11	0·18
q23	51·30	51·32	51·32	0·10	0·08
q24	52·00	52·07	52·05	0·10	0·09
L	—	3·72	3·46	3·34	3·60
P (%)	—	5·60	8·23	2·45	227·99
Chi[e] (%)	—	77·82	16·93	532·85	397·24
		(*b*) 200 Individuals			
R (%)	19·00	18·51	15·82	4·73	13·05
RHO[d] (%)	36·84	34·25	38·43	580·41	580·38
r1 (%)	7·00	6·36	6·07	21·08	17·72
r2 (%)	12·00	12·37	9·75	22·14	20·96
q13	50·60	50·60	50·57	0·05	0·08
q14	51·30	51·33	51·34	0·06	0·08
q23	51·30	51·29	51·32	0·05	0·03
q24	52·00	52·00	52·00	0·05	0·04
L	—	5·97	5·45	6·23	4·94
P (%)	—	0·54	1·63	2·45	40·14
Chi[e] (%)	—	66·57	1·34	796·31	23·72

[a] Data sets were created 200 times with new random parameters (seed values) each time.
[b] Means and variances were derived from parameter estimations performed for each of the 200 created datasets.
[c] Distorted segregation ratios were simulated by reducing the frequency of non-recombinant gametes from 0·5 to 0·2.
[d] $RHO = r1/R$.
[e] Chi-square values obtained from χ^2-tests.

by $1 - \text{PROBF}(L, df_N, df_F, 0)$, where PROBF is a SAS-function which returns the probability that an observation of an F distribution is less than or equal to *L* (SAS Institute Inc., 1990). If the *P* value was smaller than the threshold $\alpha = 0{\cdot}05$, the hypothesis that a QTL exists within the marker bracket was accepted.

(v) *Precision of estimates and sensitivity of models in the presence of distorted segregation ratios and non-additive allelic effects*

Means of simulated and estimated parameters for F_1 populations with fixed size ranging between 100 and 1000 individuals are presented in Tables 3 and 4. Separate simulations were performed for populations with normal and distorted segregation ratios. The variances of the values estimated for 200 replications were calculated to measure the precision of the estimates. Variance values ranging between 0 and 1 for parameters *R*, *RHO*, *r1*, *r2*, and for *P* and *Chi* were transformed into percentage values to allow a better comparison of variances. The estimations of QTL genotypic mean provided by both models, as indicated by the very small variances in Tables 3 and 4, were highly accurate, irrespective of the occurrence of distorted segregations and of population sizes. For both models, the estimates of *RHO*, which represents the relationship between *r1* and *r2*, had a high degree of variance. The means of the estimations for *RHO* differed slightly from the values given in the simulations. The accuracy of the estimate of this parameter was better for the four-allele model than for the common-allele model (Tables 3*b*, 4*a*). For both models, the variance of the estimates and their differences from the simulated values decreased as the population size increased (compare *a* and *b* in Tables 3 and 4).

The models used for estimating the recombination frequency *R* between the two marker loci slightly underestimated this parameter when the segregation ratios were not distorted. In the presence of distorted segregation ratios, however, the estimates deviated substantially from the expected values, even when the variances of the estimated *R* means were relatively small compared to the variances of *RHO*. In the simulation with distorted segregation ratios, the estimates of *R* obtained with the common-allele model

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

Table 4. *Simulated data, parameter estimates and their variances for a QTL flanked by two marker loci both with the same alleles in the two parents (the common-allele model), assuming an F_1 population size of (a) 100 and (b) 200 individuals, and normal and distorted segregation ratios*

Parameter...		Estimated mean[b]		Variance[b]	
Segregation...	Simulated (mean)[a]	Normal	Distorted[c]	Normal	Distorted[c]
		(*a*) 200 Individuals			
R (%)	19·00	17·50	35·67	37·04	39·72
RHO[d] (%)	36·84	40·61	40·36	996·50	787·27
r1 (%)	7·00	7·10	14·52	42·69	113·83
r2 (%)	12·00	10·40	21·15	47·98	111·99
q11	50·60	50·52	50·67	1·45	1·60
q10	51·30	51·35	51·37	0·44	0·51
q00	52·00	52·01	52·00	0·05	0·04
L	—	5·52	5·15	7·85	5·79
P (%)	—	2·41	3·36	27·04	97·14
Chi[e] (%)	—	50·57	5·01	832·80	149·75
		(*b*) 1000 Individuals			
R (%)	19·00	17·15	35·01	6·00	6·60
RHO[d] (%)	36·84	35·76	38·34	299·60	184·67
r1 (%)	7·00	6·12	13·44	9·45	24·00
r2 (%)	12·00	11·03	21·57	11·26	24·56
q11	50·60	50·61	50·59	0·25	0·22
q10	51·30	51·29	51·39	0·07	0·07
q00	52·00	52·00	51·99	0·01	0·01
L	—	23·36	23·54	33·84	33·80
P (%)	—	0·00	0·00	0·00	0·00
Chi[e] (%)	—	50·30	0·00	800·12	0·00

[a] Data sets were created 200 times with new random parameters (seed values) each time.
[b] Means and variances were derived from parameter estimations performed for each of the 200 created datasets.
[c] Distorted segregation ratios were simulated by reducing the frequency of non-recombinant gametes from 0·5 to 0·2.
[d] $RHO = r1/R$.
[e] Chi-square values obtained from χ^2-tests.

differed more from the expected values than those from the four-allele model (compare Tables 3 and 4). For both models, the precision of the estimates increased with the population size.

Since *r1* and *r2* were calculated from *R* and *RHO*, the precision of the estimates of these parameters was influenced by the same variables as found for *R*.

In the four-allele model, with an average *P* value of 0·54%, 96·5% of all *P* values derived from simulations with 200 F1 individuals were significant at the 5% level ($\alpha = 0{\cdot}05$), while only 86% of them were significant in the common-allele model. With a population size of 1000 individuals the mean value of *P* for both models resulted less than 10^{-5} (Table 4*b*, data for the four-allele model not shown). As shown in Fig. 2, the power of both models increases as the population size increases. On average, the values of *P* obtained from the four-allele model were always lower than the corresponding values computed for the common-allele model (compare Fig. 2*a* with 2*b* and Table 3*b* with 4*a*). In the presence of distorted segregations the power of both models decreased, as indicated by higher *P*-values for *D1* compared to those for *D0* (Fig. 2).

Using QTL genotypic means reflecting non-additive allelic interactions, the power and precision of the common-allele model decreased significantly, whereas the estimation of relevant parameters by the four-allele model was less affected by a non-additive behaviour of QTL alleles (compare Fig. 2*a* with 2*c* and 2*b* with 2*d*).

The usage of the models for QTL analysis performed on experimental data has shown that the precision of estimates and the *P* values obtained from the analysis were not influenced by deviations of phenotypic trait values from normal distribution (unpublished results, data not shown).

3. Discussion

Numerical tests show that the models described in this paper provide adequate estimates of QTL genotypic values and recombination frequencies between a QTL and flanking markers in an F_1 population derived from crosses between heterozygous parents. Using these models on the same experimental data to which models developed for progenies of inbred lines have been applied previously (Leonards-Schippers *et al.* 1994), the overall significance levels for QTLs and their chromosomal locations were comparable. Other

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

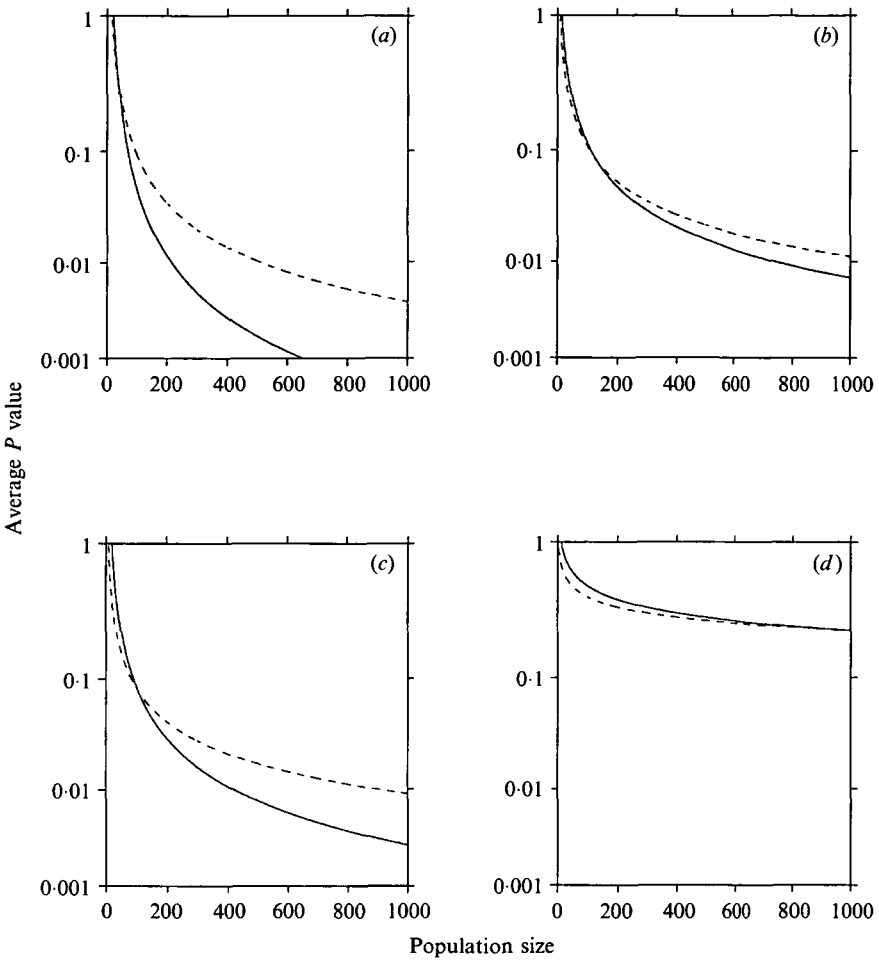

Fig. 2. Relationship between population size and the average significance level P derived from the analysis of two series (*D0*, —— and *D1*, – – –) of 1000 simulated data with ascending population sizes and a maximum QTL mean difference of one standard deviation for (*a*) the four-allele model with additive allelic interaction, (*b*) the common-allele model with additive allelic interaction, (*c*) the four-allele model with non-additive allelic interaction and (*d*) the common-allele model with non-additive allele interaction. Data sets of series *D0* were created with normal segregation ratios, and data sets of series *D1* with distorted segregation, by reducing the frequencies of non-recombinant gametes from 0·5 to 0·2, while adjusting the sum of all gamete frequencies to 1.

statistical methods, like the analysis of variance at a single marker locus, also detected, with approximations, the same QTLs. It is known that the mere detection of QTLs is limited by marker density, population size and differences between alternative QTL alleles rather than by the method of QTL analysis (Lander & Botstein, 1989). The best estimates of QTL effects and QTL location may be derived from models developed for particular types of crosses. The way we perform QTL analysis is only one of several possible existing approaches. Beside the fact that nonlinear models provide good estimates, the kind of QTL analysis chosen is relatively robust against deviations from assumptions underlying the models: in addition to the low sensitivity to distorted segregation ratios (except for models providing the estimations of R), the estimates and the significance levels are not affected by non-normal distributions of phenotypic trait values.

The 'four-allele model' was designed for a situation where four putative QTL alleles are simultaneously segregating in a progeny, together with four different alleles at each of the flanking marker loci. The model, however, does not necessarily require the presence of four different QTL alleles. When, for example, only two different QTL alleles are present in the parents, e.g. $A_1A_2Q_1Q_2B_1B_2 \times A_3A_4Q_1Q_2B_3B_4$, the estimates of '*q14*' (= q12) and '*q23*' (= q21) provided by the four-allele model will be equal, whereas 'q13' (= q11) and 'q24' (= q22) will be different. The estimates of q13, q14, q23 and q24 provide, therefore, information on the degree of homo- and heterozygosity of the QTL alleles studied.

In addition, it is only required to score two different marker alleles (= two restriction fragments), one for each parent, per marker locus. If, for example, A_1 stands for the presence of fragment x in parent P_1, the A_2 stands for the absence of the same fragment. The same is true for alleles A_3 and A_4 of parent P_2. The four genotypic classes in the F_1 for each locus are then described with $xy, x0, 0y, 00$ (A_1 = fragment x, A_3 = fragment y). When two linked markers are scored in

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

this way, all 16 marker classes can be distinguished in the F1 generation.

Furthermore, it is possible to use different markers each segregating for only one parent: if the marker fragments are derived from two different probes, which are known to approximately the same position, they can be treated as allelic fragments although they do not have to be so in the molecular sense.

As the four-allele model does not make assumptions with respect to the parental QTL alleles, it is suitable for a wide range of QTL and marker allelic configurations occurring in crosses among non-inbred plants. Thus, using a map with a sufficient density of markers and the four-allele model only, it is possible to perform a QTL analysis covering the whole genome (a paper with results is in preparation). Only regions distal to markers at the end of a linkage group may contain QTLs lying outside of an interval which is suitable for QTL analysis. In addition, this model can be used, for example, in four-way crosses $((A \times B) \times (C \times D))$ between inbred lines.

In the four-allele model marker symbols are assigned considering them to be linked in coupling. The actual phase (coupling or repulsion) of marker alleles in specific crosses can be readily defined by using, for example, the algorithms described by Ritter *et al.* (1990), and the assignment of allelic symbols can be properly done electronically before computing.

The QTL analysis of common fragments with the common-allele model has to be performed independently from the four-allele model and provides additional estimates for the questionable parameters, possibly even for the same interval(s). The utilization of the common-allele model is limited to configurations in which the marker alleles are linked in coupling. The other two possibilities, repulsion and coupling/repulsion, as described in Ritter *et al.* (1990), cannot be distinguished, and therefore an appropriate reassignment of marker alleles is not possible. In addition, the common-allele model can be used for QTL analysis in organisms for which only genetic maps based on dominant RAPD markers are available.

The two models presented in this paper provide a good estimate for QTL genotypic means, irrespective of the population size and segregation distortion. The simulation tests have shown, however, that the estimates of the recombination frequencies between the QTL and the flanking markers have lower precision than the ones recorded for the QTL genotypic means. This reflects both the relatively high variance of *RHO*, resulting from the relationship between *r1* and *r2*, and the underestimation of *R* characteristic of the models written for estimating this parameter. In addition, if segregation ratios are highly distorted, the two models for estimating *R*, in particular the model for common-alleles, do not provide precise estimates of this parameter. Because of their sensitivity to deviation from normal segregation ratios, these models should be used with caution when a χ^2 test reveals distorted segregation ratios. Particularly in the case of parents with common-alleles, maximum likelihood estimation is then to be preferred (e.g. Ritter *et al.* 1990) in order to provide better estimates for *R*. The precision of estimates of *RHO* provided by the two QTL models discussed in this paper is, however, hardly influenced by the occurrence of distorted segregation ratios and is enhanced by using larger populations.

As expected, the four-allele model shows a better estimation efficiency for *RHO* and provides more power than the common-allele model. Using a population size of 200 F_1 individuals with the four-allele model, QTL genotypic differences of one standard deviation were significant in approximately 95% of cases analysed, as compared to 86·5% for the common-allele model. In addition, the latter model is more susceptible to deviations due to the codominant behaviour of alleles. This results from the loss of information generated by the phenotypic dominance of the allele represented by the electrophoretic fragment over the null allele; '11', '01' and '10' marker genotypes cannot, in fact, be distinguished (Table 2). To optimize the power of the common-allele model, a population size of more than 300 individuals should be used. Additionally, an analysis of variance with specified linear contrasts should be performed to detect putative allelic interactions.

The problems we have encountered while estimating recombination frequencies are characteristic for non-linear models (Knapp, 1990), but can exist even with a higher degree of probability when using linear mixture models (Jansen, 1992).

The authors thank S. J. Knapp and P. Stam for their helpful suggestions and critical reading of the manuscript. This work was supported by grant 1.10 of the Bundesministerium für Forschung und Technologie (BMFT). Those who have an interest in using the models described in this paper can contact R. Schäfer-Pregl.

References

East, E. M. (1916). Studies on size inheritance in Nicotinia. *Genetics* **1**, 164–176.

Edwards, M. D., Hellentjaris, T., Wright, S. & Stuber, C. W. (1992). Molecular-marker-facilitated investigations of quantitative trait loci in maize. 4. Analysis based on genome saturation with isozyme and restriction fragment length polymorphism markers. *Theoretical and Applied Genetics* **83**, 765–774.

Edwards, M. D., Stuber, C. W. & Wendel, J. F. (1987). Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* **116**, 113–125.

Gallant, R. A. (1987). *Nonlinear Statistical Models*. New York: John Wiley & Sons.

Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschabel, U., Kaufmann, H., Thompson, R. D., Bonierbale, M. W., Ganal, M. W., Tanksley, S. D. & Salamini, F. (1991). RFLP maps of potato and their alignment with the homologous tomato genome. *Theoretical and Applied Genetics* **83**, 49–57.

Gebhardt, C., Ritter, E. & Salamini, F. (1994). RFLP map of the potato. In *DNA-based Markers in Plants* (ed. R. L. Phillips and I. K. Vasil), pp. 271–285. The Netherlands: Kluwer Academic Publishers.

Geldermann, H. (1975). Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theoretical and Applied Genetics* **46**, 319–330.

Haley, C. S., Knott, S. A. & Elsen, J.-M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207.

Jansen, R. C. (1992). A general mixture model for mapping quantitative trait loci by using molecular markers. *Theoretical and Applied Genetics* **85**, 252–260.

Knapp, S. J. (1989). Quasi mendelian analyses of quantitative traits using molecular marker linkage maps: an overview of parameter estimation methods. In *Proc 12th Eucarpia Congr., Goettingen, FRG* (ed. G. Roebellen), pp. 51–67.

Knapp, S. J. (1994). Mapping quantitative trait loci. In *DNA-based Markers in Plants* (ed. R. L. Phillips and I. K. Vasil), pp. 271–285. The Netherlands: Kluwer Academic Publishers.

Knapp, S. J. & Bridges, W. C. (1990). Programs for mapping quantitative trait loci using flanking molecular markers and nonlinear models. *Journal of Heredity* **81**, 234–235.

Knapp, S. J., Bridges, W. C. & Birkes, D. (1990). Mapping quantitative trait loci using molecular marker linkage maps. *Theoretical and Applied Genetics* **79**, 583–592.

Lander, E. S. & Botstein, D. (1986). Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proceedings of the National Academy of Sciences, USA* **83**, 7353–7357.

Lander, E. S. & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Leonards-Schippers, C., Gieffers, W., Schäfer-Pregl, R., Ritter, E., Knapp, S. J., Salamini, F. & Gebhardt, C. (1994). Quantitative resistance to Phytophthora infestans in potato: a case study for QTL mapping in allogamous plant species. *Genetics* **137**, 67–77.

Martin, B., Niehuis, J., King, G. & Schaefer, A. (1989). Restriction fragment length polymorphisms associated with water use efficiency in tomato. *Science* **243**, 1725–1728.

Nilsson-Ehle, H. (1909). Kreuzungsuntersuchungen an Hafer und Weizen. *Lund* 1909.

Osborn, T. C., Alexander, D. C. & Forbes, J. F. (1987). Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. *Theoretical and Applied Genetics* **73**, 350–356.

Phillips, R. L. & Vasil, I. K. (eds.) (1994). *DNA-based Markers in Plants*, pp. 271–285. The Netherlands: Kluwer Academic Publishers.

Rasmusson, J. M. (1933). A contribution to the theory of quantitative character inheritance. *Hereditas* **18**, 245–261.

Ritter, E., Gebhardt, C. & Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous plants. *Genetics* **125**, 645–654.

SAS Institute Inc. (1989). *SAS/STAT Users Guide, Version 6*, 4th edn, vol. 2, 1042 pp. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1990). *SAS Language: Reference, Version 6*, 4th edn, vol. 2, 846 pp. Cary, NC: SAS Institute Inc.

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**, 552–560.

Soller, M. & Brody, T. (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**, 35–39.

Soller, M., Brody, T. & Genizi, A. (1979). The expected distribution of marker-linked quantitative effects in crosses between inbred lines. *Heredity* **43**, 170–190.

Stuber, C. W., Edwards, M. D. & Wendel, J. F. (1987). Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Science* **27**, 639–648.

Stuber, C. W., Lincoln, S. E., Wolff, D. W., Helentjaris, T. & Lander, E. S. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* **132**, 823–839.

Tanksley, S. D. & Hewitt, J. (1988). Use of molecular markers in breeding for soluble solids content in tomato – a re-examination. *Theoretical and Applied Genetics* **75**, 811–823.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Hornes, M., Freijters, A., Pot, J., Peleman, J., Kuiper, M. & Zabeau, M. AFLP: a new technique for DNA fingerprinting (in press).

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

Weller, J. I. (1987). Mapping and analysis of quantitative trait loci in Lycopersicon (tomato) with the aid of genetic markers using approximate maximum likelihood methods. *Heredity* **59**, 413–421.

Appendix: Alternative notation of the models and their derivatives required by common statistical software

I(*a*) *Four-allele model* (*equation 1*)

MODEL Y = x1 * Q13 + x2 * Q14 + x3 * Q23 + x4 * Q24 + x5 * (Q13 * (1 − RHO) + Q14 * RHO) + x6 * (Q14 * (1 − RHO) + Q13 * RHO) + x7 * (Q13 * (1 − RHO) + Q23 * RHO) + x8 * (Q23 * (1 − RHO) + Q13 * RHO) + x9 * (Q14 * (1 − RHO) + Q24 * RHO) + x10 * (Q24 * (1 − RHO) + Q14 * RHO) + x11 * (Q23 * (1 − RHO) + Q24 * RHO) + x12 * (Q24 * (1 − RHO) + Q23 * RHO) + x13 * (Q13 * (1 − RHO) * * 2 + Q24 * RHO * * 2 + Q14 * RHO * (1 − RHO) + q23 * RHO * (1 − RHO)) + x14 * (Q24 * (1 − RHO) * * 2 + Q13 * RHO * * 2 + Q14 * RHO * (1 − RHO) + q23 * RHO * (1 − RHO)) + x15 * (Q14 * (1 − RHO) * * 2 + Q23 * RHO * * 2 + Q13 * RHO * (1 − RHO) + q24 * RHO * (1 − RHO)) + x16 * (Q23 * (1 − RHO) * * 2 + Q14 * RHO * * 2 + Q13 * RHO * (1 − RHO) + q24 * RHO * (1 − RHO)).

I(*b*) *Derivatives of the four-allele model*

DER.Q13 = x1 + x5 * (1 − RHO) + x6 * RHO + x7 * (1 − RHO) + x8 * RHO + x13 * (1 − RHO) * * 2 + x14 * RHO * * 2 + x15 * RHO * (1 − RHO) + x16 * RHO * (1 − RHO).

DER.Q14 = x2 + x5 * RHO + x6 * (1 − RHO) + x9 * (1 − RHO) + x10 * RHO + x13 * RHO * (1 − RHO) + x14 * RHO * (1 − RHO) + x15 * (1 − RHO) * * 2 + x16 * RHO * * 2.

DER.Q23 = x3 + x7 * RHO + x8 * (1 − RHO) + x11 * (1 − RHO) + x12 * RHO + x13 * RHO * (1 − RHO) + x14 * RHO * (1 − RHO) + x15 * RHO * * 2 + x16 * (1 − RHO) * * 2.

DER.Q24 = x4 + x9 * RHO + x10 * (1 − RHO) + x11 * RHO + x12 * (1 − RHO) + x13 * RHO * * 2 + x14 * (1 − RHO) * * 2 + x15 * RHO * (1 − RHO) + x16 * RHO * (1 − RHO).

DER.RHO = x5 * (−Q13 + Q14) + x6 * (−Q14 + Q13) + x7 * (−Q13 + Q23) + x8 * (−Q23 + Q13) + x9 * (−Q14 + Q24) + x10 * (−Q24 + Q14) + x11 * (−Q23 + Q24) + x12 * (−Q24 + Q23) + x13 * (Q13 * (−2 * (1 − RHO)) + Q24 * 2 * RHO − Q14 * RHO + Q14 * (1 − RHO) − q23 * RHO + Q23 * (1 − RHO)) + x14 * (Q24 * (−2 * (1 − RHO)) + Q13 * 2 * RHO − Q14 * RHO + Q14 * (1 − RHO) − q23 * RHO + Q23 * (1 − RHO)) + x15 * (Q14 * (−2 * (1 − RHO)) + Q23 * 2 * RHO − Q13 * RHO + Q13 * (1 − RHO) − Q24 * RHO + Q24 * (1 − RHO)) + x16 * (Q23 * (−2 * (1 − RHO)) + Q14 * 2 * RHO − Q13 * RHO + Q13 * (1 − RHO) − Q24 * RHO + Q24 * (1 − RHO)).

II(*a*) *Common-allele model* (*equation 2*)

Model Y = x11 * ((1/2 − 1/2 * R) * * 2 * Q11 + 1/2 * (1 − R) * R * (Q11 * (1 − RHO) + Q12 * RHO) + 1/2 * (1 − R) * R * (Q11 * RHO + Q12 * (1 − RHO)) + 1/2 * (1 − R) * * 2 * Q12 + 1/2 * R * * 2 * (Q11 * RHO * (1 − RHO) + Q22 * RHO * (1 − RHO) + Q12 * RHO * * 2 + Q12 * (1 − RHO) * * 2)) / (3 * (1/2 − 1/2 * R) * * 2 + (1 − R) * R + 1/2 * R * * 2) + x10 * (1/4 * R * * 2 * (Q11 * (1 − RHO) * * 2 + 2 * Q12 * RHO * (1 − RHO) + Q22 * RHO * * 2) + 1/2 * (1 − R) * R * (Q12 * (1 − RHO) + Q22 * RHO)) / (1/4 * R * * 2 + 1/2 * (1 − R) * R) + x01 * (1/2 * (1 − R) * R * (Q12 * RHO + Q22 * (1 − RHO)) + 1/4 * R * * 2 * (Q11 * RHO * * 2 + 2 * Q12 * RHO * (1 − RHO) + Q22 * (1 − RHO) * * 2)) / (1/4 * R * * 2 + 1/2 * (1 − R) * R) + x00 * Q22.

II(*b*) *Derivatives of the common-allele model*

DER.Q11 = (x11 * ((1/2 − 1/2 * R) * * 2 + 1/2 * (1 − R) * R * (1 − RHO) + 1/2 * (1 − R) * R * RHO + 1/2 * R * * 2 * RHO * (1 − RHO))) / (3 * (1/2 − 1/2 * R) * * 2 + (1 − R) * R + 1/2 * R * * 2) + 1/4 * ((x10 * R * * 2 * (1 − RHO) * * 2) / (1/4 * R * * 2 + 1/2 * (1 − R) * R)) + 1/4 * ((x01 * R * * 2 * RHO * * 2) / (1/4 * R * * 2 + 1/2 * (1 − R * R)).

DER.Q12 = (x11 * (1/2 * (1 − R) * R * RHO + 1/2 * (1 − R) * R * (1 − RHO) + 1/2 * (1 − R) * * 2 + 1/2 * R * * 2 * (RHO * * 2 + (1 − RHO * * 2))) / (3 * (1/2 − 1/2 * R) * * 2 + (1 − R) * R + 1/2 * R * * 2) + (x10 * (1/2 * R * * 2 * RHO * (1 − RHO) + 1/2 * (1 − R) * R * (1 − RHO))) / (1/4 * R * * 2 + 1/2 * (1 − R) * R) + (x01 * (1/2 * (1 − R) * R * RHO + 1/2 * R * * 2 * RHO * (1 − RHO))) / (1/4 * R * * 2 + 1/2 * (1 − R) * R).

DER.Q22 = 1/2 * ((x11 * R * * 2 * RHO * (1 − RHO)) / (3 * (1/2 − 1/2 * R) * * 2 + (1 − R) * R + 1/2 * R * * 2)) + (x10 * (1/4 * R * * 2 * RHO * * 2 + 1/2 * (1 − R) * R * RHO)) / (1/4 * R * * 2 + 1/2 * (1 − R) * R) + (x01 * (1/2 * (1 − R) * R * (1 − RHO) + 1/4 * R * * 2 * (1 − RHO) * * 2)) / (1/4 * R * * 2 + 1/2 * (1 − R) * R) + x00.

DER.RHO = x11 * (1/2 * (1 − R) * R * (−Q11 + Q12) + 1/2 * (1 − R) * R * (Q11 − Q12) + 1/2 * R * * 2 * (Q11 * (1 − RHO) − Q11 * RHO + Q22 * (1 − RHO) − Q22 * RHO + 2 * Q12 * RHO − 2 * Q12 * (1 − RHO))) / (3 * (1/2 − 1/2 * R) * * 2 + (1 − R) * R + 1/2 * R * * 2) + x10 * (1/4 * R * * 2 * (−2 * Q11 * (1 − RHO) + 2 * Q12 * (1 − RHO) − 2 * Q12 * RHO + 2 * Q22 * RHO) + 1/2 * (1 − R) * R * (−Q12 + Q22)) / (1/4 * R * * 2 + 1/2 * (1 − R) * R) + x01 * (1/2 * (1 − R) * R * (Q12 − Q22) + 1/4 * R * * 2 * (2 * Q11 * RHO + 2 * Q12 * (1 − RHO) − 2 * Q12 * RHO − 2 * Q22 * (1 − RHO))) / (1/4 * R * * 2 + 1/2 * (1 − R) * R).

III(*a*) *Model for estimating the recombination frequency R between two marker loci with four different alleles* (*equation 3*)

Model NN = x1 * ((1 − R)/2) * * 2 * NT + x2 * ((1 − R)/2) * * 2 * NT + x3 * ((1 − R)/2) * * 2 * NT + x4 * ((1 − R)/2) * * 2 * NT + x5 * ((1 − R)/2) * (R/2) * NT + x6 * ((1 − R)/2) * (R/2) * NT + x7 * ((1 − R)/2) * (R/2) * NT + x8 * ((1 − R)/2) * (R/2) * NT + x9 * ((1 − R)/2) * (R/2) * NT + x10 * ((1 − R)/2) * (R/2) * NT + x11 * ((1 − R)/2) * (R/2) * NT + x12 * ((1 − R)/2) * (R/2) * NT + x13 * (R * * 2/4) * NT + x14 * (R * * 2/4) * NT + x15 * (R * * 2/4) * NT + x16 * (R * * 2/4) * NT.

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press

III(*b*) *Derivatives*

DER.R = −x1 * (1/2−1/2 * R) * NT−x2 * (1/2−1/2 * R) * NT−x3 * (1/2−1/2 * R) * NT−x4 * (1/2−1/2 * R) * NT−1/4 * x5 * R * NT+1/2 * x5 * (1/2−1/2 * R) * NT−1/4 * x6 * R * NT+1/2 * x6 * (1/2−1/2 * R) * NT−1/4 * x7 * R * NT+1/2 * x7 * (1/2−1/2 * R) * NT−1/4 * x8 * R * NT+1/2 * x8 * (1/2−1/2 * R) * NT−1/4 * x9 * R * NT+1/2 * x9 * (1/2−1/2 * R) * NT−1/4 * x10 * R * NT+1/2 * x10 * (1/2−1/2 * R) * NT−1/4 * x11 * R * NT+1/2 * x11 * (1/2−1/2 * R) * NT−1/4 * x12 * R * NT+1/2 * x12 * (1/2−1/2 * R) * NT+1/2 * x13 * R * NT+1/2 * x14 * R * NT+1/2 * x15 * R * NT+1/2 * x16 * R * NT.

IV(*a*) *Model for estimating the recombination frequency R between two marker loci with common alleles* (*equation 4*)

Model NN = x11 * (3 * ((1−R)/2) * * 2+4 * ((1−R)/2) * (R/2)) * NT+x10 * ((R/2) * * 2+2 * ((1−R)/2) * (R/2)) * NT+x01 * ((R/2) * * 2+2 * ((1−R)/2) * (R/2)) * NT+x00 * (R/2) * * 2 * NT.

IV(*b*) *Derivatives*

DER.R = x11 * (−1/2−1/2 * R) * NT+x10 * (1/2−1/2 * R) * NT+x01 * (1/2−1/2 * R) * NT+1/2 * x00 * R * NT.

https://doi.org/10.1017/S0016672300033462 Published online by Cambridge University Press