Heredity **66** (1991) 117–124

Received 17 May 1990

Maximum likelihood estimation of linkage between a marker gene and a quantitative trait locus. II. Application to backcross and doubled haploid populations

Z. W. LUO* & M. J. KEARSEY
School of Biological Sciences, University of Birmingham, Birmingham B15 2TT, UK

The algorithm for estimating both the recombination fraction between a marker gene and a locus affecting a quantitative trait, and also the means and variances of the QTL genotypes, is extended to backcross and doubled haploid populations. The simulation experiments show that estimates of these parameters can be obtained with acceptable accuracy and results are compared with those obtained using F_2 populations studied previously (Luo & Kearsey, 1989).

Keywords: doubled haploid populations, linkage, marker gene, quantitative trait.

Introduction

Since the publication of the paper by Botstein *et al.* (1980), we have witnessed growing interest in the use of molecular genetic markers to locate unknown genes, particularly those genetic factors associated with quantitative variation. These approaches are all developed from a variety of methods examined by biometric geneticists in the last 40 years.

Three aspects of the use of markers have attracted the interest of biometricians and geneticists. Firstly, the detection of polygenes (or QTLs) with associated developments of statistical approaches appropriate to various breeding experiments (Thoday, 1961; Jayakar, 1970; Hill, 1975; Lander & Botstein, 1989 and Luo, 1989); secondly, measurement of the statistical power of different experimental designs for polygene detection (McMillan & Robertson, 1974; Soller & Brody, 1976; Soller & Genizi, 1978; Luo & Kearsey, 1989); finally, estimation of linkage between a given marker gene and a putative QTL. Jayakar (1970) developed formulae to estimate marker-QTL recombination fractions in two different designs for use in studying natural populations, but they were unrealistic in practice through lack of appropriate estimates of environmental variance. Weller (1986) applied maximum likelihood techniques to the analysis of the F_2 generation of a cross between two inbred lines in order to estimate not only the recombination fraction between a marker locus and a QTL but also the nature and size of the effect. Because the numerical algorithm employed by Weller to search the likelihood surface was based on an unreasonable assumption, i.e. the parameters to be estimated in the likelihood function were independent of each other, his algorithm could not confirm that the estimates obtained were, in fact, the maximum likelihood estimates. In an attempt to overcome this problem we described, in a previous paper (Luo & Kearsey, 1989), a maximum likelihood approach for estimating the recombination fraction in a segregating population (F_2) between a marker gene and a QTL as well as estimating the means and variances of the three genotypes of the QTL. In this paper, we extend the algorithm to two other experimental designs, i.e. backcrosses and doubled haploids.

*Present address: Department of Zoology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.

Theoretical approach

We consider a breeding programme starting with two inbred lines, one of which is homozygous for the alleles M_1 and Q_1 for the marker and QTL respectively, while the other is homozygous for the alleles M_2 and Q_2. The marker alleles are assumed to be co-dominant, and the

recombination fraction between the two loci is denoted by r. The F_1 is either crossed to the parent with genotype $M_2M_2Q_2Q_2$ to produce a backcross generation with family size of n_b or used to generate n_d, doubled haploid families (e.g. by anther culture). For simplicity, we further assume that all the families have a constant size of n. The means and variances of the quantitative trait among the marker genotypes are as shown in Table 1.

Development of the analytical method

Since the linkage between marker and QTL is incomplete, the individuals within each of the two possible marker genotypes are a mixture of two quantitative trait genotypes, i.e. Q_1Q_2 and Q_2Q_2 for the backcross design but Q_1Q_1 and Q_2Q_2 for the doubled haploids. Let μ_{11}, μ_{12} and μ_{22} represent the means and σ^2_{11}, σ^2_{12} and σ^2_{22} represent the variances of the three genotypes for the quantitiative trait. The variances include not only environmental variation but also genetic variation at other loci affecting the quantitative trait. Therefore, the means of the marker classes can be partitioned into the following:

for the backcross,

$$X_{12} = (1-r)\mu_{12} + r\mu_{22} \qquad (1a)$$
$$X_{22} = r\mu_{12} + (1-r)\mu_{22}$$

and for the doubled haploids,

$$X_{11} = (1-r)\mu_{11} + r\mu_{22} \qquad (1b)$$
$$X_{22} = r\mu_{11} + (1-r)\mu_{22}$$

while the variances of the marker groups can be partitioned as

$$S^2_{12} = (1-r)[(\mu_{12} - X_{12})^2 + \sigma^2_{12}] + r[(\mu_{22} - X_{12})^2 + \sigma^2_{22}]$$
$$S^2_{22} = r[(\mu_{12} - X_{22})^2 + \sigma^2_{12}] + (1-r)[(\mu_{22} - X_{22})^2 + \sigma^2_{22}] \qquad (2a)$$

Table 1 Basic statistics of the marker genotypes in (a) the backcross and (b) the doubled haploid

Statistics of the quantitative trait	M_1M_1	M_1M_2	M_2M_2
Means	X_{11}	X_{12}	X_{22}
Variances	S^2_{11}	S^2_{12}	S^2_{22}
Sample size			
(a)	0	n_1	n_2
(b)	n_1	0	n_2

for the backcross, and as

$$S^2_{11} = (1-r)[(\mu_{11} - X_{11})^2 + \sigma^2_{11}] + r[(\mu_{22} - X_{11})^2 + \sigma^2_{22}]$$
$$S^2_{22} = r[(\mu_{11} - X_{22})^2 + \sigma^2_{11}] + (1-r)[(\mu_{22} - X_{22})^2 + \sigma^2_{22}] \qquad (2b)$$

for the doubled haploids.

The composite distribution densities of the relevant marker classes can be written as (Hasselblad, 1966; Day, 1969)

$$f_{M_{12}}(x) = (1-r)f_{12}(x) + rf_{22}(x)$$
$$f_{M_{22}}(x) = rf_{12}(x) + (1-r)f_{22}(x) \qquad (3a)$$

for the backcross design; and as

$$f_{M_{11}}(x) = (1-r)f_{11}(x) + rf_{22}(x)$$
$$f_{M_{22}}(x) = rf_{11}(x) + (1-r)f_{22}(x) \qquad (3b)$$

for the doubled haploids, where

$$f_{ij}(x) = \frac{1}{\sqrt{2\pi\sigma^2_{ij}}} \exp\left[-\frac{(x-\mu_{ij})^2}{\sigma^2_{ij}}\right] \qquad (4)$$

and

$i,j = 1,2.$

Therefore the likelihood function can be written as

$$L_B(\mu_{12}, \mu_{22}, \sigma^2_{12}, \sigma^2_{22}, r) = \left[\prod_{i=1}^{n_1} f_{M_{12}}(x_i)\right]\left[\prod_{j=1}^{n_2} f_{M_{22}}(x_j)\right] \qquad (5a)$$

for the entire backcross generation, and as

$$L_D(\mu_{11}, \mu_{22}, \sigma^2_{11}, \sigma^2_{22}, r) = \left[\prod_{i=1}^{n_1} f_{M_{11}}(x_i)\right]\left[\prod_{j=1}^{n_2} f_{M_{22}}(x_j)\right] \qquad (5b)$$

for the entire doubled haploid population.

The logarithm functions of (5) will be, respectively,

$$\begin{aligned} L_B(r) = \sum_{i=1}^{n_1} \ln f_{M_{12}}(x_i) + \sum_{j=1}^{n_2} \ln f_{M_{22}}(x_j) &= \ln \frac{1}{\sqrt{2\pi}}(n_1 + n_2) \\ &+ \sum_{i=1}^{n_1} \ln\left\{\frac{1-r}{\sigma_{12}} \exp\left[-\frac{(x_i-\mu_{12})^2}{2\sigma^2_{12}}\right]\right. \\ &\left. + \frac{r}{\sigma_{22}} \exp\left[-\frac{(x_i-\mu_{22})^2}{2\sigma^2_{22}}\right]\right\} \\ &+ \sum_{j=1}^{n_2} \ln\left\{\frac{r}{\sigma_{12}} \exp\left[-\frac{(x_j-\mu_{12})^2}{2\sigma^2_{12}}\right]\right. \\ &\left. + \frac{1-r}{\sigma_{22}} \exp\left[-\frac{(x_j-\mu_{22})^2}{2\sigma^2_{22}}\right]\right\} \end{aligned} \qquad (6a)$$

and

$$L_D(r)=\sum_{i=1}^{n_1}\ln f_{M_{11}}(x_i)+\sum_{j=1}^{n_2}\ln f_{M_{22}}(x_j)=\ln\frac{1}{\sqrt{2\pi}}(n_1+n_2)$$
$$+\sum_{i=1}^{n_1}\ln\left\{\frac{1-r}{\sigma_{11}}\exp\left[-\frac{(x_i-\mu_{11})^2}{2\sigma_{11}^2}\right]\right.$$
$$\left.+\frac{r}{\sigma_{22}}\exp\left[-\frac{(x_i-\mu_{22})^2}{2\sigma_{22}^2}\right]\right\}$$
$$+\sum_{j=1}^{n_2}\ln\left\{\frac{r}{\sigma_{11}}\exp\left[-\frac{(x_j-\mu_{11})^2}{2\sigma_{11}^2}\right]\right.$$
$$\left.+\frac{1-r}{\sigma_{22}}\exp\left[-\frac{(x_j-\mu_{22})^2}{2\sigma_{22}^2}\right]\right\}. \qquad (6b)$$

Equation 6(a and b) include the other four unknown parameters besides r. However, following the method used to analyse the F_2 generation, equations (1a) and (2a) provide the following solutions to the unknowns in (6a)

$$\mu_{22}=\frac{1}{1-2r}[(1-r)X_{M_{12}}-rX_{M_{22}}]$$
$$\mu_{12}=\{X_{M_{12}}+X_{M_{22}}\}-\mu_{22} \qquad (7a)$$

and

$$\sigma_{22}^2=\frac{1}{1-2r}[(1-r)e_2-re_1]$$
$$\sigma_{12}^2=e_1+e_2-\sigma_{22}^2 \qquad (7b)$$

where

$$e_1=S_{12}^2-[(1-r)(\mu_{12}-X_{12})^2+r(\mu_{22}-X_{12})^2]$$
$$e_2=S_{22}^2-[r(\mu_{12}-X_{22})^2+(1-r)(\mu_{22}-X_{22})^2]$$

while equations (1b) and (2b) yield the corresponding solutions to the means and variances of the QTL genotypes in the doubled haploid design (6b)

$$\mu_{22}=\frac{1}{1-2r}[(1-r)X_{M_{22}}-rX_{M_{11}}]$$
$$\mu_{11}=\{X_{M_{11}}+X_{M_{22}}\}-\mu_{22} \qquad (8a)$$

and

$$\sigma_{22}^2=\frac{1}{1-2r}[(1-r)e_2-re_1]$$
$$\sigma_{11}^2=e_1+e_2-\sigma_{22}^2 \qquad (8b)$$

where

$$e_1=S_{11}^2-[(1-r)(\mu_{11}-X_{11})^2+r(\mu_{22}-X_{11})^2]$$
$$e_2=S_{22}^2-[r(\mu_{11}-X_{22})^2+(1-r)(\mu_{22}-X_{22})^2].$$

When these estimates of the means (μ_{ij}) and variances (σ_{ij}^2) obtained from equations (7) or from (8) are incorporated into the log likelihood functions (6a) and (6b) for these two designs, respectively, then these functions will involve only one unknown parameter, the recombination fraction r. Searching these functions numerically for their maximum values with respect to r, will yield the maximum likelihood estimates of r for each design. Since the means and variances of the quantitative trait have been expressed as the monotonic functions about r, according to the invariant property of the maximum likelihood estimator (Mood *et al.*, 1974), i.e. if $\hat{\theta}$ is the maximum likelihood estimate of θ in the distribution density $f(x_j;\theta)$ and $\tau(\cdot)$ is a transformation of the parameter space Θ, then a maximum likelihood estimate of $\tau(\theta)$ is $\tau(\hat{\theta})$. Thus formulae (7) and (8) directly give maximum likelihood estimates of the means and variances of the quantitative trait in the two different designs.

Furthermore, the two QTL genotypes, i.e. Q_1Q_2 and Q_2Q_2, in the backcross generation and the two corresponding genotypes, i.e. Q_1Q_2 and Q_2Q_2, in the doubled haploid population, should have the same environmental variances. We can therefore use the following

$$\sigma_b^2=\frac{n_1}{N}\sigma_{12}^2+\frac{n_2}{N}\sigma_{22}^2$$

and

$$\sigma_d^2=\frac{n_1}{N}\sigma_{11}^2+\frac{n_2}{N}\sigma_{22}^2$$

to approximate the environmental variances for the two populations respectively, where N represents the total experimental size of backcross design but represents the number of families for the doubled haploids.

Description of the algorithm

In the last section, it has been demonstrated that the likelihood functions [$L_B(r)$ or $L_D(r)$] involve only one unknown parameter r. This greatly simplifies the algorithm to search the likelihood surface for the relevant maximum likelihood estimates. In our previous paper (Luo & Kearsey, 1989), an iterative algorithm was described to obtain the estimates from F_2 populations. The same numerical algorithm will be employed here.

Simulation of the two designs and data analysis

Two FORTRAN-77 computer programs were designed both to simulate the genetic behaviour of linkage between a marker locus and a QTL and to work out the required maximum likelihood estimates folowing the algorithms described in the previous discussion. The principle of the simulation of genetic behaviour of the marker-QTL linkage is described in detail elsewhere (Luo, 1989; Luo & Kearsey, 1989).

All possible combinations of three recombination fractions between the marker gene and the QTL (r=0.15, 0.25, 0.35), two narrow heritabilities (h_n^2=0.1, 0.5) and three dominance ratios (dr=0.0, 0.5, 1.0) for the QTL (these genetic parameters are defined by using an F_2 population as a standard) were simulated for both designs with fixed size of 500 while each of the 500 doubled haploid families has a size of 5. These are the same parameter combinations used for the F_2 generation by Luo & Kearsey (1989).

Results

The maximum likelihood estimates of the recombination fraction between the marker locus and the QTL and the corresponding standard errors, given a sample size of 500, are recorded in Table 2 for the doubled haploid population and in Table 3 for the backcross generation. The maximum likelihood estimates of the means of the two possible genotypes at the QTL can be found, together with their standard errors, in Tables 4 and 5 for the doubled haploids and the backcross generation, respectively. The maximum likelihood estimates of the environmental variances and their standard errors for the two designs are given in Table 6.

It is clear from Table 2 that there is no significant difference between the maximum likelihood estimates of the recombination fraction obtained from the data of means of doubled haploid families and their true values. The standard errors of these estimates consistently decrease as the heritability of the QTL increases, but neither the dominance ratio of the QTL nor the true values of the estimated parameter has any obvious influence on the estimates or their standard errors.

Table 3 shows that the recombination fractions can also be estimated adequately by use of the backcross data. However, when the heritability is low, these fractions are slightly underestimated at its true value of 0.15 but are slightly overestimated when the true value is equal to or higher than 0.25. The standard errors of

Table 2 The maximum likelihood estimates of the recombination fraction between the marker and QTL, where h_n^2 and dr represent narrow heritability and dominance ratio of the QTL respectively (doubled haploid families)

		True recombination fraction		
h_n^2	dr	0.15	0.25	0.35
0.1	0.0	0.1800 ± 0.0458	0.2856 ± 0.0374	0.3564 ± 0.0447
0.1	0.5	0.1218 ± 0.0316	0.2880 ± 0.0469	0.3742 ± 0.0648
0.1	1.0	0.1731 ± 0.0490	0.2849 ± 0.0469	0.3821 ± 0.0400
0.5	0.0	0.1550 ± 0.0200	0.2666 ± 0.0332	0.3603 ± 0.0316
0.5	0.5	0.1470 ± 0.0224	0.2581 ± 0.0374	0.3559 ± 0.0300
0.5	1.0	0.1561 ± 0.0245	0.2427 ± 0.0265	0.3750 ± 0.0283

Table 3 The maximum likelihood estimates of the recombination fraction between the marker and QTL, where h_n^2 and dr represent narrow heritability and dominance ratio of the QTL respectively (backcross generation)

		True recombination fraction		
h_n^2	dr	0.15	0.25	0.35
0.1	0.0	0.1287 ± 0.0700	0.2742 ± 0.0557	0.3807 ± 0.0500
0.1	0.5	0.1105 ± 0.0721	0.2796 ± 0.0529	0.3720 ± 0.0633
0.1	1.0	0.0854 ± 0.0656	0.2817 ± 0.0510	0.3720 ± 0.0633
0.5	0.0	0.1644 ± 0.0361	0.2588 ± 0.0480	0.3690 ± 0.0400
0.5	0.5	0.1554 ± 0.0361	0.2627 ± 0.0353	0.5392 ± 0.0300
0.5	1.0	0.1575 ± 0.0332	0.2554 ± 0.0374	0.3428 ± 0.0283

Table 4 The maximum likelihood estimates and expected values of the means of the two QTL genotypes for different genetic situations, where h_n^2, dr, μ and r are narrow heritability, dominance ratio, expected means of the QTL genotypes and true recombination fraction between the genetic marker and QTL respectively (doubled haploid families)

		$h_n^2=0.1$		$h_n^2=0.5$	
dr	r	$\hat{\mu}_{11}$	$\hat{\mu}_{22}$	$\hat{\mu}_{11}$	$\hat{\mu}_{22}$
0.0	0.15	109.63 ± 1.1091	99.31 ± 1.0488	119.87 ± 0.6633	100.08 ± 0.5568
	0.25	109.63 ± 0.8775	98.68 ± 1.0392	115.85 ± 5.3768	100.17 ± 0.9434
	0.35	109.23 ± 1.0100	99.70 ± 1.0247	116.45 ± 5.2900	99.97 ± 1.3266
μ		108.94	100.00	120.00	100.00
0.5	0.15	109.13 ± 2.5436	99.91 ± 1.0198	118.94 ± 0.5745	100.09 ± 0.4583
	0.25	109.76 ± 1.2042	99.20 ± 1.1533	117.19 ± 3.3601	100.05 ± 0.5658
	0.35	109.16 ± 1.4071	99.58 ± 0.8485	116.08 ± 5.0665	100.23 ± 0.9434
μ		108.44	100.00	118.86	100.00
1.0	0.15	110.39 ± 2.1111**	99.60 ± 0.7000	117.01 ± 0.6083	100.33 ± 0.5000
	0.25	109.98 ± 1.0440**	99.16 ± 1.0583	118.64 ± 3.2655	100.25 ± 0.3873
	0.35	109.14 ± 1.1091	99.61 ± 0.8889	119.03 ± 2.8160	99.75 ± 0.9185
μ		107.30	100.00	116.32	100.00

Table 5 The maximum likelihood estimates and expected values of the means of the two QTL genotypes for different genetic situations, where h_n^2, dr, μ and r are narrow heritability, dominance ratio, expected means of the QTL genotypes and true recombination fraction between the genetic marker and QTL respectively (backcross generation)

		$h_n^2=0.1$		$h_n^2=0.5$	
dr	r	$\hat{\mu}_{12}$	$\hat{\mu}_{22}$	$\hat{\mu}_{12}$	$\hat{\mu}_{22}$
0.0	0.15	102.80 ± 0.9434**	99.52 ± 0.9274	109.91 ± 0.5657	99.89 ± 0.5657
	0.25	101.06 ± 0.9899**	99.45 ± 1.6462	110.41 ± 0.5292	99.65 ± 0.3000
	0.35	106.61 ± 2.3706	98.77 ± 2.3706	110.14 ± 0.7280	99.77 ± 0.7746
μ		104.47	100.00	110.00	100.00
0.5	0.15	106.02 ± 1.0149	100.17 ± 0.7348	114.27 ± 0.5385	100.06 ± 0.6481
	0.25	107.40 ± 1.1446	98.96 ± 1.1402	113.26 ± 1.6817	100.01 ± 0.8718
	0.35	107.64 ± 1.2000	98.75 ± 1.3000	113.84 ± 0.9381	100.34 ± 0.9747
μ		106.33	100.00	114.14	100.00
1.0	0.15	112.52 ± 3.4380	100.68 ± 0.6083	111.24 ± 0.5385	99.93 ± 0.4243
	0.25	105.81 ± 1.4213	99.02 ± 1.2961	111.09 ± 0.6557	99.97 ± 0.6403
	0.35	106.17 ± 1.6026	99.14 ± 1.3528	110.73 ± 1.3038	100.41 ± 0.7211
μ		107.30	100.00	116.32	100.00

the estimates obtained from this design are mostly higher than those obtained from doubled haploid family means, suggesting that using the doubled haploid family means can yield a more accurate estimate of the recombination fraction for a fixed sample size. Of course each family mean of the doubled haploids with respect to the quantitative trait was based on five individuals.

The phenotypic means of the two genotypes of the QTL are denoted by μ_{11} and μ_{22} in the doubled haploid population and by μ_{12} and μ_{22} in the backcross generation. The maximum likelihood estimates of these means are listed in Tables 4 and 5 for the two designs. It is clear that these estimates do not deviate significantly from their expected values when the quantitative trait has a high heritability ($h_n^2=0.5$). Moreover, these means can also be estimated with a moderate heritability ($h_n^2=0.1$). The effect of the level of dominance of the increasing allele at the QTL on the accuracy of these estimates differs in the two designs. As can be seen from Tables 4 and 5, increasing dominance makes the expected values of μ_{11} and μ_{22} more alike but μ_{11} and μ_{12} more unlike. The effect of dominance on the doubled haploids is an artefact of the model since, with

Table 6 The maximum likelihood estimates of the environmental variances associated with the QTL, where h_n^2, *dr* EVE are narrow heritability, dominance ratio and expected variance of the quantitative trait for doubled haploids and backcrosses

			True recombination fraction		
h_n^2	*dr*	EVE	0.15	0.25	0.35
Doubled haploids					
0.1	0.0	90.00	79.73 ± 9.0500	75.19 ± 12.2593	83.44 ± 9.2952
0.1	0.5	90.00	87.87 ± 8.7178	72.72 ± 10.2387	81.00 ± 10.7624
0.1	1.0	90.00	77.83 ± 11.3309	78.44 ± 8.7304	79.52 ± 12.3778
0.5	0.0	50.00	49.20 ± 4.2661	65.20 ± 10.2401	61.64 ± 10.6300
0.5	0.5	50.00	49.25 ± 3.6810	54.83 ± 4.6750	60.82 ± 11.3978
0.5	1.0	50.00	47.95 ± 3.3985	53.42 ± 7.7188	49.10 ± 6.1237
Backcrosses					
0.1	0.0	90.00	79.79 ± 5.8915	82.79 ± 5.7879	73.82 ± 14.1421
0.1	0.5	90.00	85.04 ± 11.4873	75.20 ± 9.9600	72.92 ± 9.7783
0.1	1.0	90.00	72.53 ± 4.3336**	80.20 ± 7.9869	76.08 ± 10.9316
0.5	0.0	50.00	46.70 ± 3.9370	44.44 ± 5.3413	46.48 ± 6.6588
0.5	0.5	50.00	48.37 ± 3.6986	52.19 ± 7.5050	49.88 ± 8.6742
0.5	1.0	50.00	46.03 ± 3.4395	48.82 ± 4.3070	50.54 ± 7.6551

fixed phenotypic variance and heritability, increasing dominance effectively reduces additivity. As a result, with dominance the means are estimated less well in the doubled haploids but better with the backcrosses. The reverse is true with no dominance. The standard errors of the estimates also decline with increasing heritability for both designs.

The maximum likelihood estimates of the environmental variance, σ^2, do not differ significantly from their true values in the doubled haploid families and only one of the 18 situations simulated in the backcross design.

In terms of their abilities to provide accurate estimates of marker-QTL linkage, as well as the phenotypic means and variances of the QTL genotypes, neither of the two designs is consistantly superior to the other over all possible genetic backgrounds of the QTL. The above results do, however, suggest that at low heritabilities the doubled haploid families would be more powerful than the backcross design when the dominance ratio is low.

Discussion

This paper has concentrated on the use of maximum likelihood techniques to estimate the marker-QTL recombination fraction and the relevant genetic and environmental effects of the QTL. The results obtained from the simulation experiments with 20 replications of each of 18 different combinations of genetic parameters indicate that recombination fractions could be well estimated using these designs. In fact, none of the estimates differed significantly from their expected values for all combinations of genetic parameters considered with experimental size of 500. However, some significantly biased estimates were observed in the F_2 generation design for both the heritabilities simulated (Luo & Kearsey, 1989).

Accurate estimates of the means and variances of the two QTL genotypes were regularly found from both the designs and the frequency of significantly biased estimates was again lower than that in the F_2 experiments. The maximum likelihood estimates of the environmental variances associated with the QTL were rarely observed to be different from their actual values in the simulated backcross and doubled haploid experiments, while biased estimates for this parameter were commonly observed in the simulated F_2.

The estimation problem, discussed here and in our previous work, is statistically equivalent to resolving a single mixed normal distribution into a few component distributions with common variance. As we have noted in the previous discussion, if the putative QTL is linked with the co-dominant marker gene, the phenotypic distribution of the quantitative trait in each marker class will be an unequal component mixture of two component normal distributions for the backcross and doubled haploids, but of three component normal

distributions for the F_2 population. The power of any algorithm leading to a solution to the relevant parameters of a composite distribution depends on the absolute difference between the means of the component distribution for a fixed common variance. For a given heritability, it is obvious that the absolute difference between the means of the homozygous and heterozygous genotypes (e.g. two of the three component distributions) at the QTL in the F_2 generation is much less than that between the two means of the two possible genotypes (e.g. two component distributions) at the same QTL in either the backcross or the doubled haploid family designs. This would explain why the latter two designs always exhibit higher efficiency in producing the relevant estimates than the F_2 generation design. In the present studies, the heritability of the quantitative trait plays an important role in determining the mean difference and variance of the component distributions, i.e. the higher the heritability the larger the difference between the means, the smaller the variance and therefore the more efficient the algorithm. Furthermore, it is interesting to note that the marker-QTL linkage estimates obtained from the F_2 generation regularly have smaller standard errors than those yielded by the other two designs; i.e. the F_2 data supply more information about linkage for a fixed experimental size (Mather, 1936; 1938).

So far, there have been three basic approaches developed respectively by Hasselblad (1966), Bhattacharya (1967) and Cohen (1967) to carry out dissection of mixed distributions. As a specific genetic application to the problem of separating the mixed distributions, the algorithms developed in the present studies have used the genetic characteristics to provide useful information. This has effectively simplified the estimation procedure and in turn increased the efficiency of the algorithm. Firstly, the proportions of the component distributions were completely determined by the recombination fraction between the marker gene and the linked QTL. Secondly, if the phenotypic means and variances of the recombinant genotypes, estimated from the experimental sample, were used as unbiased estimates of the corresponding population parameters of the composite distributions, then the means and variances of the relevant genotypes at the QTL were uniquely determined by the recombination fraction and these estimates. This results in the maximum likelihood function involving only one unknown parameter, i.e. the recombination fraction. The maximum likelihood estimates of the remaining parameters, including the means and environmental variances of the relevant QTL genotypes, could easily be obtained directly from their functional relationships to the means and variances of the marker genotypes, thus avoiding use of the complicated iterative algorithm suggested by previous authors. In fact, the present algorithms obviously show higher efficiency than the general methods. Hasselblad (1966) claimed that it was extremely difficult to dissect three sub-distributions by use of his maximum likelihood algorithm, which was considered by Tan & Chang (1972) to be the best of the three approaches mentioned before, when the means were separated by less than 2 standard deviation units even with an experimental size of 1,000. However, using the algorithms developed in this and our previous paper (Luo & Kearsey, 1989), the component distributions were regularly well-estimated even though the means were separated by much less than one standard deviation, for example $h_n^2 = 0.1$ with an experimental size of 500.

In the previous discussion, we have assumed normality of the distributions of the phenotypes among the marker groups. This assumption is at odds with the fact that incomplete linkage between the marker and the QTL will result in skewness in these marker groups. This assumption may be a source of bias in the estimates of the basic parameters.

References

BHATTACHARYA, G. G. 1967. A simple method of resolution of a distribution into Gaussian components. *Biometrics*, **23**, 115–135.

BOTSTEIN, D., WHITE, R. L., SKOLNICK, M. AND DAVIES, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Hum. Genet.*, **32**, 314–331.

COHEN, A. C. 1967. Estimation in mixtures of two normal distributions. *Technometrics*, **9**, 15–28.

DAY, N. E. 1969. Estimating the components of a mixture of normal distributions. *Biometrika*, **56**, 463–474.

HASSELBLAD, F. 1966. Mixture of normal distributions. *Technometrics*, **8**, 231–250.

HILL, A. P. 1975. Quantitative linkage: a statistical procedure for its detection and estimation. *Ann. Hum. Genet. London*, **38**, 439–449.

JAYAKAR, S. D. 1970. On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics*, **26**, 451–464.

LANDER, E. S. AND BOTSTEIN, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.

LUO, Z. W. 1989. Ph.D. Thesis. University of Birmingham.

LUO, Z. W. AND KEARSEY, M. J. 1989. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity*, **63**, 401–408.

MATHER, K. 1936. Types of linkage data and their value. *Ann. Eugenics*, **7**, 251–264.

MATHER, K. 1938. *The Measurement of Linkage in Heredity*. Methuen and Co. Ltd, London.

MCMILLAN, I. AND ROBERTSON, A. 1974. The power of methods for the detection of major genes affecting quantitative characters. *Heredity*, **32**, 349–356.

MOOD, A. M., GRAYBILL, F. A. AND BOES, D. C. 1974. *Introduction to the Theory of Statistics*. McGraw-Hill Book Company, New York.

SOLLER, M. AND 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. *Theor. Appl. Genet.*, **47**, 35–39.

SOLLER, M. AND GENIZI, A. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics*, **34**, 47–55.

TAN, W. Y. AND CHANG, W. C. 1972. Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of two normal densities. *J. Am. Stat. Assoc.*, **67**, 702–708.

THODAY, J. M. 1961. Location of polygenes. *Nature*, **191**, 368–370.

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.