

Maximum likelihood estimation of linkage between a marker gene and a quantitative locus

Z. W. Luo and
M. J. Kearsey

Department of Genetics, University of Birmingham,
P.O. Box 363, Birmingham B15 2TT, U.K.

A maximum likelihood approach is developed for estimating the recombination fraction in a segregating population (F_2), between a marker gene and a locus affecting a quantitative trait as well as estimating the means and variances of the three genotypes of the quantitative trait. The experimental results from computer simulations show that even with experimental sizes of 500, estimates of the parameters can be obtained by aid of the codominant marker gene as long as the heritability of the quantitative trait in question is not less than 0.10. However at low heritabilities the variances of estimates are very large.

INTRODUCTION

The location of polygenes has always been an important aim of biometrical genetics since it could lead to a better understanding of the structure of the polygenic system as well as its possible manipulation both by classical breeding and genetical engineering. However the phenotype of a quantitative character is a consequence of both genetic and environmental sources and, in a polygenic system, an individual gene substitution may well contribute a relatively small effect to the phenotype of the character governed by the system (Mather and Jinks, 1971).

In general these properties make it almost impossible for the experimenter to identify individual genotypes among the offspring of a cross from the quantitative character studied and they therefore prevent the direct use of the traditional methods of locating major genes. Nevertheless, the chromosomal inheritance of polygenes suggests the possibility of detecting and estimating the genetic linkage between the genes using an appropriate experimental design.

One reliable method of demonstrating the presence of a gene for a quantitative trait on a particular chromosome is to determine a linkage relationship between that trait and a distinguishable Mendelian locus (so called marker locus) and the work of Penrose (1938, 1946) is considered as being pioneering in this field. Breese and Mather (1957, 1960)

investigated the distribution of polygenic activity affecting chaeta number and viability characters along chromosome III of *Drosophila melanogaster* by using appropriate mutant loci on the same chromosome as genetic markers. In a series of fundamental studies, Thoday and his colleagues showed how major mutants can be used to identify, map and evaluate individual loci affecting quantitative traits (Thoday, 1961, 1979; Spickett and Thoday, 1966; Thoday *et al.*, 1964). Since then this method has been applied in basic studies of biometrical or quantitative genetics of various organisms such as in *Drosophila* (Williams, 1968, 1977, 1978 and 1980; Davies and Workman, 1971; Davies, 1971; Dominguez and Rubio, 1986; Shrimpton and Robertson, 1988*a, b*), in wheat (Patterson *et al.*, 1968; Law, 1966), in maize (Edwards *et al.*, 1987) and in tomato (Tanksley and Rick, 1980; Weller, 1987; Weller *et al.*, 1988). However others have identified major flaws in this approach (McMillan and Robertson, 1974).

In theoretical studies, Hill (1975) demonstrated a statistical test for the presence of linkage between a marker gene and quantitative trait locus (QTL) in natural families. By using the intercrossing and backcrossing experimental design from two inbred lines, Jayakar (1970) described a statistical procedure for the detection and estimation of linkage between a marker gene and a QTL by using full-sib data. Lander and Botstein (1989) recently developed a new method of locating the polygenes

underlying quantitative traits using RFLP linkage maps. Given a detailed knowledge of RFLP positions throughout the genome, the probability of a putative QTL being located in a given interval can be estimated using the method of log-odds (LOD). This method of interval mapping, first described by Lander and Green (1987), allows efficient detection of QTLs while limiting the overall occurrence of false positives. It also allows accurate estimation of the phenotypic effects of QTLs and their location to specific regions. By using this systematic QTL mapping approach, Paterson *et al.* (1988) resolved six quantitative traits of tomato into Mendelian factors.

Weller (1986) attempted to estimate the recombination fraction between a marker locus and a QTL by applying maximum likelihood techniques to the analysis of the F_2 generation of a cross between two inbred lines. In his study the moment method suggested by Zhuchenko *et al.* (1979) was used to reduce the number of parameters to be estimated. Although the likelihood function based on an F_2 family involved just three unknown parameters it was still difficult to obtain maximum likelihood estimates of these parameters analytically or numerically due to the complexity of the likelihood function of the mixed distribution. Because of the need to search the likelihood surface for each parameter separately and ignorance of the internal mathematical relationships among the parameters to be estimated, the method used by Weller (1986) could not guarantee that the estimates obtained were in fact the maximum likelihood estimates. In the present paper we describe the use of a likelihood function involving only one unknown parameter, the recombination fraction between the marker and QTL. The estimate of the parameter which maximizes the likelihood function will, therefore, be the maximum likelihood estimate. In turn the maximum likelihood estimates of the basic parameters of the quantitative trait are also obtained. Finally, the results from simulation experiments are used to support the theory.

THEORETICAL APPROACH

The breeding programme

Consider a breeding programme starting with two inbred lines, one of which is homozygous for the alleles M_1 and Q_1 of the locus of the genetic marker and the QTL respectively, and the other is homozygous for the alleles M_2 and Q_2 . Thus the parental genotypes are $M_1M_1Q_1Q_1$ and $M_2M_2Q_2Q_2$. The

marker alleles are assumed to be codominant, therefore the three genotypes at this locus are distinguishable. The recombination fraction between the two loci is denoted by r . The F_1 is selfed or sibmated to produce an F_2 family and the means and variances of the quantitative trait among the three marker genotypes of the F_2 generation are as shown in table 1.

Table 1 Basic statistics of the three marker genotypes in the F_2

| Statistics of quant. trait | Marker genotypes | | |
|----------------------------|------------------|----------------|----------------|
| | M_1M_1 | M_1M_2 | M_2M_2 |
| Means | \bar{X}_{11} | \bar{X}_{12} | \bar{X}_{22} |
| Variance | S_{11}^2 | S_{12}^2 | S_{22}^2 |
| Sample size | n_1 | n_2 | n_3 |

Development of the analytical method

If the linkage between the marker and QTL is incomplete, the individuals within each of the three marker genotypes are a mixture of three quantitative trait genotypes, Q_1Q_1 , Q_1Q_2 and Q_2Q_2 . Let μ_{11} , μ_{12} and μ_{22} stand for the means and σ_{11}^2 , σ_{12}^2 and σ_{22}^2 stand for the variances of the three genotypes for the quantitative trait. These variances represent both environmental and genetical variation at other loci affecting the quantitative trait. The means and variances of the marker groups can thus be partitioned into the following:

$$\bar{X}_{11} = (1-r)^2\mu_{11} + 2r(1-r)\mu_{12} + r^2\mu_{22} \quad (1)$$

$$\bar{X}_{12} = r(1-r)\mu_{11} + [1-2r(1-r)]\mu_{12} + r(1-r)\mu_{22} \quad (2)$$

$$\bar{X}_{22} = r^2\mu_{11} + 2r(1-r)\mu_{12} + (1-r)^2\mu_{22} \quad (3)$$

and

$$S_{11}^2 = (1-r)^2[\sigma_{11}^2 + (\mu_{11} - \bar{X}_{11})^2] + 2r(1-r)[\sigma_{12}^2 + (\mu_{12} - \bar{X}_{11})^2] + r^2[\sigma_{22}^2 + (\mu_{22} - \bar{X}_{11})^2] \quad (4)$$

$$S_{12}^2 = r(1-r)[\sigma_{11}^2 + (\mu_{11} - \bar{X}_{12})^2] + [1-2r(1-r)][\sigma_{12}^2 + (\mu_{12} - \bar{X}_{12})^2] + r(1-r)[\sigma_{22}^2 + (\mu_{22} - \bar{X}_{12})^2] \quad (5)$$

$$S_{22}^2 = r^2[\sigma_{11}^2 + (\mu_{11} - \bar{X}_{22})^2] + 2r(1-r)[\sigma_{12}^2 + (\mu_{12} - \bar{X}_{22})^2] + (1-r)^2[\sigma_{22}^2 + (\mu_{22} - \bar{X}_{22})^2]. \quad (6)$$

If it is further assumed that the observed value of an individual with the quantitative trait genotype Q_iQ_j is a random variable which is normally distributed with mean μ_{ij} and variance σ_{ij}^2 , then the distribution of each of the three marker groups will be a mixture of three normal distributions. Therefore the density function of the combined distribution can be written as a linear combination of the three component normal distributions weighted according to their proportions (Day, 1969; Tan and Chang, 1972). *i.e.*,

$$f_{M_{11}}(x) = (1-r)^2 f_{11}(x) + 2r(1-r)f_{12}(x) + r^2 f_{22}(x) \tag{7}$$

$$f_{M_{12}}(x) = r(1-r)f_{11}(x) + [1-2r(1-r)]f_{12}(x) + r(1-r)f_{22}(x) \tag{8}$$

$$f_{M_{22}}(x) = r^2 f_{11}(x) + 2r(1-r)f_{12}(x) + (1-r)^2 f_{22}(x). \tag{9}$$

Where

$$f_{11}(x) = \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp\left[-\frac{(x-\mu_{11})^2}{2\sigma_{11}^2}\right]$$

$$f_{12}(x) = \frac{1}{\sqrt{2\pi\sigma_{12}}} \exp\left[-\frac{(x-\mu_{12})^2}{2\sigma_{12}^2}\right]$$

$$f_{22}(x) = \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp\left[-\frac{(x-\mu_{22})^2}{2\sigma_{22}^2}\right].$$

Therefore the likelihood function for the entire F_2 population can be written as

$$L = \left[\prod_{i=1}^{n_1} f_{M_{11}}(x_i) \right] \left[\prod_{j=1}^{n_2} f_{M_{12}}(x_j) \right] \left[\prod_{k=1}^{n_3} f_{M_{22}}(x_k) \right] \tag{10}$$

The logarithm function of (10) will be

$$L(r) = \sum_{i=1}^{n_1} \ln f_{M_{11}}(x_i) + \sum_{j=1}^{n_2} \ln f_{M_{12}}(x_j) + \sum_{k=1}^{n_3} \ln f_{M_{22}}(x_k) \ln \frac{1}{\sqrt{2\pi}} (n_1 + n_2 + n_3) + \sum_{i=1}^{n_1} \ln \left\{ \frac{(1-r)^2}{\sigma_{11}} \exp\left[-\frac{(x_i-\mu_{11})^2}{2\sigma_{11}^2}\right] + \frac{2r(1-r)}{\sigma_{12}} \exp\left[-\frac{(x_i-\mu_{12})^2}{2\sigma_{12}^2}\right] + \frac{r^2}{\sigma_{22}} \exp\left[-\frac{(x_i-\mu_{22})^2}{2\sigma_{22}^2}\right] \right\} + \sum_{j=1}^{n_2} \ln \left\{ \frac{r(1-r)}{\sigma_{11}} \exp\left[-\frac{(x_j-\mu_{11})^2}{2\sigma_{11}^2}\right] \right.$$

$$\left. + \frac{[1-2r(1-r)]}{\sigma_{12}} \exp\left[-\frac{(x_j-\mu_{12})^2}{2\sigma_{12}^2}\right] + \frac{r(1-r)}{\sigma_{22}} \exp\left[-\frac{(x_j-\mu_{22})^2}{2\sigma_{22}^2}\right] \right\} + \sum_{k=1}^{n_3} \ln \left\{ \frac{r^2}{\sigma_{11}} \exp\left[-\frac{(x_k-\mu_{11})^2}{2\sigma_{11}^2}\right] + \frac{2r(1-r)}{\sigma_{12}} \exp\left[-\frac{(x_k-\mu_{12})^2}{2\sigma_{12}^2}\right] + \frac{(1-r)^2}{\sigma_{22}} \exp\left[-\frac{(x_k-\mu_{22})^2}{2\sigma_{22}^2}\right] \right\} \tag{11}$$

Equation (11) includes the other six unknown parameters besides r . However simultaneous equations (1)-(3) and (4)-(6), respectively, consist of three independent linear equations which involve three independent unknowns provided that r is given. Thus unique solutions exist for both of the groups of simultaneous equations.

The solutions for equations (1)-(3) are

$$\mu_{22} = \frac{1}{1-4r+4r^2} \{ \bar{X}_{11} - 2r(1-r)(\bar{X}_{12} + \bar{X}_{22}) - \frac{1}{1-2r} (\bar{X}_{11} - \bar{X}_{22})[(1-r)^2 - 2r^2(1-r)] \} \tag{12}$$

$$\mu_{11} = \frac{1}{1-2r} (\bar{X}_{11} - \bar{X}_{22}) + \mu_{22} \tag{13}$$

$$\mu_{12} = \bar{X}_{12} + \bar{X}_{22} - [r\mu_{11} + (1-r)\mu_{22}]. \tag{14}$$

If we define

$$d_{11} = (\mu_{11} - \bar{X}_{11})^2$$

$$d_{12} = (\mu_{12} - \bar{X}_{11})^2$$

$$d_{13} = (\mu_{22} - \bar{X}_{11})^2$$

$$d_{21} = (\mu_{11} - \bar{X}_{12})^2$$

$$d_{22} = (\mu_{12} - \bar{X}_{12})^2$$

$$d_{23} = (\mu_{22} - \bar{X}_{12})^2$$

$$d_{31} = (\mu_{11} - \bar{X}_{22})^2$$

$$d_{32} = (\mu_{12} - \bar{X}_{22})^2$$

$$d_{33} = (\mu_{22} - \bar{X}_{22})^2$$

$$e_1 = S_{11}^2 - [(1-r)^2 d_{11} + 2r(1-r)d_{12} + r^2 d_{13}] \tag{15}$$

$$e_2 = S_{12}^2 - \{ r(1-r)d_{21} + [1-2r(1-r)]d_{22} + r(1-r)d_{23} \} \tag{16}$$

$$e_3 = S_{22}^2 - [r^2 d_{31} + 2r(1-r)d_{32} + (1-r)^2 d_{33}]. \tag{17}$$

Then the solutions of equations (4)–(6) are

$$\sigma_{22}^2 = \frac{1}{1-4r+4r^2} \{e_1 - 2r(1-r)(e_2 + e_3) - \frac{1}{1-2r} (e_1 - e_3)[(1-r)^2 - 2r^2(1-r)]\} \quad (18)$$

$$\sigma_{11}^2 = \frac{1}{1-2r} (e_1 - e_3) + \sigma_{22}^2 \quad (19)$$

$$\sigma_{12}^2 = e_1 + e_3 - [r\sigma_{11}^2 + (1-r)\sigma_{22}^2]. \quad (20)$$

When the estimates (12)–(14) and (18)–(20) are incorporated into (11), the logarithm of the likelihood function (11) will only involve the recombination fraction r . By maximizing equation (11) about r , the true maximum likelihood estimate of the recombination fraction can be obtained. Since the means and variances of the quantitative trait are monotonic functions about r , and r is the maximum likelihood estimate, then the means and variances are also maximum likelihood estimates. This follows from 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, then a maximum likelihood estimate of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Furthermore, the three genotypes of the QTL in the F_2 population, Q_1Q_1 , Q_1Q_2 and Q_2Q_2 , are the parental and F_1 genotypes respectively. These genotypes should have the same environmental variance, *i.e.*, the theoretical values of σ_{11}^2 , σ_{12}^2 and σ_{22}^2 should be statistically homogeneous;

i.e.,

$$\sigma^2 = \frac{n_1}{N} e_1 + \frac{n_2}{N} e_2 + \frac{n_3}{N} e_3$$

where

$$N = n_1 + n_2 + n_3.$$

Description of the algorithm

From the previous section, equations (11), (12)–(14) and (18)–(20) give the likelihood function about one unknown parameter r , the recombination fraction between the marker and the QTL. In theory it is possible to find the solution of the following equation

$$\frac{dL(r)}{dr} = 0$$

The root of the above equation gives the maximum likelihood estimate of r , but this involves solving an algebraic equation with high exponent. Fortunately, for only one unknown variable, it is very easy to find the maximum value point of the equation (11) by an iterative numerical method. As the defined boundary of r is between 0.0 to 0.5, a computer programme can be designed to search the surface of the maximum likelihood function in this interval through the trace defined by (11) for any given precision.

Efficiency of the maximum likelihood estimate

Cavalli-Sforza and Bodmer (1971) recommended that Fisher's information statistic be used to give the variance of the maximum likelihood estimate for one parameter. In the present case equation (11) is the logarithm of the likelihood function $L(r)$ of the recombination fraction r and it can be shown quite generally that

$$\frac{1}{E(-(d^2L/d\theta^2))}$$

is approximately the variance of the maximum likelihood estimate of r . Just as the means and variances are functions of the recombination fraction r , so also are the distribution parameters in the likelihood function. This makes the second differential function involved in the above variance estimator too complicated to be expressed briefly. Therefore it is necessary to obtain the means and variances of the estimates by computer simulation. Because of the time involved only 20 replicates of each situation were simulated. If the means and the expected value of the QTL-marker recombination fraction are denoted by \bar{r} and r^* respectively, then

$$\frac{|\bar{r} - r^*|}{s_r} \quad (21)$$

will be distributed as Student's t distribution with 19 degrees of freedom and provides a significance test between the maximum likelihood estimate and its theoretical value.

For a given theoretical recombination fraction between the marker and QTL, the realised recombination fraction r_e , is an observation of its theoretical value r^* . The difference between them will be an indicator of the adequacy of the simulation programme used in the present study. Therefore

$$\frac{(r_e - r^*)^2}{V(r_e)} = \chi_1^2.$$

Computer simulation of F₂ population

A Fortran-77 computer programme was designed to simulate the genetic behaviour of linkage between the marker locus and the QTL. The programme allows F₂ progeny from any two defined inbred lines to be generated by combining gametes generated from the defined parents by a "random walk" procedure (Crosby, 1973). Although any number of markers and quantitative trait loci are easily simulated by this programme, in the present study only two loci are considered, one the genetic marker the other the quantitative trait locus. The simulation programme also allows varying degrees of linkage to be accommodated between the two loci as well as various dominance ratios and heritabilities of the quantitative trait to be considered.

RESULTS

All possible combinations of the three recombination fractions (0.15, 0.25, 0.35), two heritabilities (0.1, 0.5) and three dominance ratios (0.0, 0.5, 1.0) for the QTL were simulated. Table 2 shows the maximum likelihood estimates (M.L.E.) of the recombination fraction between the marker and QTL and the standard error given a sample size of 500. The maximum likelihood estimates of means and environmental variances of the QTL and their corresponding standard errors are given in tables 3 and 4.

The simulation results consistently show that there is no significant difference between r_e and r^* over all levels of the recombination fractions *i.e.*, 0.15, 0.25 and 0.35 ($\chi^2_{0.15} = 0.13$, $\chi^2_{0.25} = 0.08$ and $\chi^2_{0.35} = 0.09$), which indicates that the simulation data on which the linkage analysis is performed are reliable.

When the QTL has a low heritability ($h_n^2 = 0.1$), the estimates of the means (Table 3) for the increas-

ing QTL homozygote (μ_{11}) and heterozygote (μ_{12}) and the environmental variation of the quantitative trait (Table 4) are regularly biased from their actual values for all combinations of the different recombination fractions and dominance ratios. The mean estimates for the decreasing homozygote (μ_{22}), which consistently have the largest standard error, rarely differ significantly from their true values. This is because the means of the homozygote (μ_{22}) do not change with increasing dominance ratio from 0.0 to 1.0 unlike the means of the homozygote (μ_{11}) and heterozygote (μ_{12}). The differences between the means at low heritability are smaller than at high heritability. This makes it more difficult to separate the means of the component distributions from that of the mixed distribution. For a given heritability, codominance yields the maximum differences among the means and thus the means of the three genotypes at the QTL will be more easily distinguished.

On increasing the heritability ($h_n^2 = 0.5$), the estimates of the means and the nonheritable variance of the QTL are less biased and their standard errors decline. Most of the means and the nonheritable variances of the QTL have estimates nonsignificantly different from their expected values.

Nearly consistent estimates of the recombination fraction are obtained at both heritability levels. It has been shown from table 2 that the estimate of the recombination fraction depends more on the dominance ratio than on the heritability of the QTL. When alleles at the QTL show codominance or partial dominance the method presented in the paper can lead to very good estimates of r . Moreover the standard errors of the estimates also decrease dramatically with increasing heritability. On the other hand, it is clear that the accuracy of the recombination fraction estimate is also dependent on the true value of r , the tighter the QTL-marker linkage, the better the estimate of the recombination fraction. At a theoretical

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

| h_n^2 | dr | True Recombination Fraction | | |
|---------|------|-----------------------------|-------------------|-------------------|
| | | 0.15 | 0.25 | 0.35 |
| 0.1 | 0.0 | 0.1498 ± 0.0177 | 0.2690 ± 0.0114 | 0.3381 ± 0.0101 |
| 0.1 | 0.5 | 0.1312 ± 0.0125 | 0.2625 ± 0.0110 | 0.3277 ± 0.0095* |
| 0.1 | 1.0 | 0.1661 ± 0.0125 | 0.2751 ± 0.0117* | 0.3072 ± 0.0101** |
| 0.5 | 0.0 | 0.1498 ± 0.0098 | 0.2348 ± 0.0072 | 0.3311 ± 0.0076** |
| 0.5 | 0.5 | 0.1523 ± 0.0054 | 0.2463 ± 0.0085 | 0.3372 ± 0.0095 |
| 0.5 | 1.0 | 0.1377 ± 0.0069 | 0.2253 ± 0.0078** | 0.3281 ± 0.0092* |

Table 3 The maximum likelihood estimates and expected values of the means of the 3 QTL genotypes for different genetic situations, where h_n^2 , dr EVE, and r are narrow heritability, dominance ratio, expected mean of the quantitative trait and true recombination fraction between the genetic marker and QTL respectively

| dr | r | $h_n^2 = 0.1$ | | | $h_n^2 = 0.5$ | | |
|------|------|-----------------|-----------------|----------------|-----------------|-----------------|-----------------|
| | | μ_{11} | μ_{12} | μ_{22} | μ_{11} | μ_{12} | μ_{22} |
| 0.0 | 0.15 | 109.30 ± 0.31 | 105.07 ± 0.15** | 99.63 ± 0.32 | 119.46 ± 0.53 | 109.29 ± 0.68 | 99.57 ± 0.25 |
| | 0.25 | 109.62 ± 0.40 | 103.72 ± 0.23** | 99.99 ± 0.40 | 119.43 ± 0.46 | 110.12 ± 0.20 | 100.90 ± 0.27** |
| | 0.35 | 109.00 ± 0.77 | 105.42 ± 0.63 | 98.69 ± 0.85 | 119.03 ± 1.03 | 110.34 ± 0.31 | 101.24 ± 0.39** |
| EVE | | 108.94 | 104.47 | 100.00 | 120.00 | 110.00 | 100.00 |
| | 0.15 | 108.46 ± 0.42 | 105.31 ± 0.36** | 99.88 ± 0.46 | 119.33 ± 0.30 | 113.80 ± 0.23 | 100.29 ± 0.17 |
| 0.5 | 0.25 | 109.64 ± 0.58** | 104.56 ± 0.57* | 99.64 ± 0.65 | 119.32 ± 0.37 | 110.71 ± 0.15** | 100.33 ± 0.23 |
| | 0.35 | 105.90 ± 0.76** | 107.61 ± 0.68* | 100.03 ± 0.65 | 118.61 ± 0.67 | 112.08 ± 0.61** | 100.74 ± 0.37* |
| | | 108.44 | 106.33 | 100.00 | 118.86 | 114.14 | 100.00 |
| EVE. | | 107.74 ± 0.25* | 106.58 ± 0.32* | 99.13 ± 0.45 | 118.93 ± 0.18* | 116.36 ± 0.17 | 100.22 ± 0.18 |
| | 0.25 | 108.53 ± 0.55* | 107.00 ± 0.33 | 98.00 ± 0.47** | 119.05 ± 0.49** | 115.69 ± 0.36* | 100.41 ± 0.24 |
| 1.0 | 0.35 | 106.63 ± 0.76 | 106.63 ± 0.68 | 98.63 ± 0.65 | 116.37 ± 0.75 | 115.87 ± 0.36 | 100.46 ± 0.30 |
| | | 107.30 | 107.30 | 100.00 | 116.32 | 116.32 | 100.00 |
| | EVE. | | | | | | |

recombination fraction of 0.35, the estimates obtained are always below their true values irrespective of heritability and dominance ratio and the bias becomes larger as heritability decreases.

DISCUSSION

The method suggested in the paper has concentrated on estimating QTL-marker linkage using a technique based on maximum likelihood. The results in table 2 indicate that the maximum likelihood estimates of the recombination fraction agree well with their theoretical values under the different genetical situations simulated; even though some estimates are significantly different from their actual values, the bias is never more than 15 per cent. In the linkage analysis, estimates of the recombination fraction r are always restricted to the range $0.0 \leq r \leq 0.5$, and this results in a downward bias, i.e., $r < e(r)$ (Ott, 1985). Nonetheless, as an asymptotically unbiased estimate, the

maximum likelihood estimate of the QTL-marker recombination fraction is expected to tend towards its true value with increasing sample size. However, it can be seen from table 2 that the standard deviations of r based on 20 replicate simulations (= standard errors given $\times \sqrt{20}$) given a sample size of 500, can be large particularly when h^2 is low, e.g., at the heritability of 0.1 and the recombination fraction of 0.15, the standard deviation of the estimate of true r is $0.0177 \times \sqrt{20} = 0.0792$. On the other hand the standard deviation of the μ_{ij} s (table 3) are sufficiently small to distinguish the three means even at low heritability.

The likelihood function used in the paper involved only one parameter, the QTL-marker recombination fraction r . For a given set of experimental data, the basic statistics of the marker groups were determined. Using these statistics, the means and the non-heritable variances of the quantitative trait were derived using those values of r which maximized the likelihood function (equation (11)) by following the formulae (12)–(14) and (18)–(20). In this way the likelihood function always has its peak at a value of r above that for which the variance estimate becomes negative; unlike the method of Weller (1986), the negative variance can thus be avoided.

It has been shown from table 4 that the non-heritable variance at the QTL is consistently underestimated with low heritability ($h_n^2 = 0.1$), while the biases have nearly vanished when the heritability reaches 0.5. This shows that the accuracy of the estimated variance is also dependent on the heritability. The standard errors of the variance estimates consistently increase with r , as do the standard errors of the mean (table 3). The formulae

Table 4 The maximum likelihood estimates of the environmental variances associated with the QTL, where h_n^2 , dr and EVE are narrow heritability, dominance ratio and expected variance of the quantitative trait

| h_n^2 | dr | EVE. | True Recombination Fractions | | |
|---------|-----|-------|------------------------------|----------------|----------------|
| | | | 0.15 | 0.25 | 0.35 |
| 0.1 | 0.0 | 90.00 | 86.19 ± 1.14** | 83.50 ± 1.39** | 75.87 ± 3.50** |
| 0.1 | 0.5 | 90.00 | 87.34 ± 1.22* | 78.45 ± 1.89** | 78.75 ± 2.79** |
| 0.1 | 1.0 | 90.00 | 83.43 ± 1.69** | 77.28 ± 2.23** | 78.75 ± 2.67** |
| 0.5 | 0.0 | 50.00 | 47.12 ± 1.45* | 52.51 ± 2.04 | 57.74 ± 2.84** |
| 0.5 | 0.5 | 50.00 | 49.10 ± 1.07 | 47.77 ± 2.57 | 53.69 ± 3.80 |
| 0.5 | 1.0 | 50.00 | 48.41 ± 1.18 | 47.43 ± 2.03 | 51.42 ± 2.24 |

for the solutions of these estimates (12)–(14) and (18)–(20) have indicated that these estimates decrease monotonically with decreasing r when the other parameters are fixed and these solutions lose their meaning if and only if $r = 0.5$. This may explain the dependence of the standard errors of the means and variances on the values of r .

Table 3 shows that there is a good agreement between observed and expected values of the means of the quantitative trait and no obvious bias exists between them over the different situations simulated. However the standard deviations of these estimates reveal their poor precision.

So far we have simply considered a quantitative trait governed by a single locus and the variation remaining is thus only ascribed to environmental variance. More realistically, the character under study would be controlled by many genes which may be distributed on a number of chromosomes. For a trait with heritability h^2 governed by n loci, the heritability of any one locus will be h^2/n . As the number of loci increases, the heritability of any one locus will become very small. Weller (1986) pointed out that his proposed method of QTL-marker linkage analysis was useful only for QTL's with effects greater than 1.0 phenotypic standard deviation. With the same assumption about the marker locus as that made by Weller, the present study also obtained satisfactory results for QTL's with $h_n^2 = 0.1$, i.e., the QTL has an effect less than 1.0 phenotypic standard deviation. However, when the heritability of the QTL decreased to 0.01, the method recommended in this article did not yield meaningful results either. Since recombination fraction is defined between two loci, when there are many loci controlling the trait studied it is clearly impossible to summarize the linkage relationship between the quantitative trait and the given marker locus in terms of a single recombination fraction. But if, among the quantitative trait loci, there is only one QTL linked with the given marker, which has a heritability not lower than that discussed in this paper, the method proposed in this study is still appropriate, although the variance will now include both the environmental variation and the genetic variance of the other segregating loci, i.e. the heritability refers to that locus only. It can be concluded that the analytical procedure described here is at least effective for quantitative traits controlled by a single linked locus with heritability not less than 0.1 but it is ineffective for traits with the same heritability equally shared by several loci.

Soller and Brody (1976) showed that it was possible to detect a codominant QTL responsible

for 1 phenotypic standard deviation in an F_2 population of 1000 individuals (i.e., $\sigma_p = Q_1Q_1 - Q_2Q_2 = 2d$). Weller (1986) attempted to estimate the recombination fraction between a codominant marker locus and a codominant QTL with effects greater than 1.0 phenotypic standard deviation in the F_2 population of 2000 individuals. In this paper, a possibly more realistic sample size of 500 was used, a quarter of that studied by Weller.

In theory the method assumed that the distributions of the phenotypes among the marker groups were normal. This assumption is probably wrong since incomplete linkage between the QTL and the marker will result in skewness in the F_2 marker groups. This may be another reason why the basic statistics of the QTL are biased.

REFERENCES

- BREESE, E. L. AND MATHER, K. 1957. The organisation of polygenic activity within a chromosome in *Drosophila*: I. Hair characters. *Heredity*, 11, 373–395.
- BREESE, E. L. AND MATHER, K. 1960. The organisation of polygenic activity within a chromosome in *Drosophila*: II. Viability. *Heredity*, 14, 375–399.
- CAVALLI-SFORZA, L. L. AND BODMER, W. F. 1971. *The Genetics of Human Populations*. W. H. Freeman and Company, San Francisco.
- CROSBY, J. L. 1973. *Computer Simulation in Genetics*. John Wiley & Sons, London.
- DAY, N. E. 1969. Estimating the components of a mixture of normal distributions. *Biometrika*, 56, 463–474.
- DAVIES, R. W. AND WORKMAN, P. L. 1971. The genetic relationship of two quantitative characters in *Drosophila melanogaster*, I. Responses to selection and whole chromosome analysis. *Genetics*, 69, 353–361.
- DAVIES, R. W. 1971. The genetic relationship of two quantitative characters in *Drosophila melanogaster*, II. Location of the effects. *Genetics*, 69, 363–375.
- DOMINGUEZ, A. AND RUBIO, J. 1986. Analysis of genetic factors affecting egg production in *Drosophila melanogaster*. *Heredity*, 57, 305–313.
- EDWARDS, M. D., STUBER, C. W. AND 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.
- HILL, A. P. 1975. Quantitative linkage: A statistical procedure for its detection and estimation. *Ann. Human Genet.*, 38, 439–450.
- 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.
- LAW, C. N. 1966. The location of genetic factors affecting a quantitative character in wheat. *Genetics*, 53, 487–498.
- LANDER, E. S. AND BOTSTEIN, D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121, 185–199.
- LANDER, E. S. AND GREEN, P. 1987. Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA.*, 84, 2363–67.
- MCMILLAN, I. AND ROBERTSON, A. 1974. The power of methods for the detection of major genes affecting quantitative characters. *Heredity*, 32, 349–356.

- MATHER, K. AND JINKS, J. L. 1971. *Biometrical Genetics*. Chapman and Hall, London.
- MOOD, A. M., GRAYBILL, F. A. AND BOES, D. C. 1974. *Introduction to the Theory of Statistics*. McGraw-Hill Book Company, New York.
- OTT, J. 1985. *Analysis of Human Genetics Linkage*. The Johns Hopkins University Press. Baltimore and London.
- PATERSON, A. H., LANDER, E. S., HEWITT, J. D., PETERSON, S., LINCOLN, S. E. AND TANKSLEY, S. D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. *Nature*, 335, 721-726.
- PATTERSON, F. L., SCHAFER, J. F. AND CALDWELL, R. M. 1968. Effect of selected linkage blocks on yield and yield components in wheat. In Finley, K. W. and Shepherd K. W. (eds) *3rd Int. Wheat Genet. Symp., Canberra, 5-9 August 1968*. Aust Acad Sci, Canberra.
- PENROSE, L. S. 1938. Genetic linkage in graded human character. *Ann. Eugen.*, 8, 233-237.
- PENROSE, L. S. 1946. A further note on the sib-pair linkage method. *Ann. Eugen.*, 13, 25-29.
- SHRIMPTON, A. E. AND ROBERTSON, A. 1988a. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. I. Allocation of third chromosome sternopleural bristle effects to chromosome sections. *Genetics*, 118, 437-443.
- SHRIMPTON, A. E. AND ROBERTSON, A. 1988b. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics*, 118, 437-443.
- SPICKETT, S. G. AND THODAY, M. J. 1966. Regular responses to selection. 3 Interaction between located polygenes. *Genet. Res. Camb.*, 7, 96-121.
- 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.
- 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. *Journal of the American Statistical Association*, 67, 702-708.
- TANKSLEY, S. D. AND RICK, C. M. 1980. Isozymic gene linkage map of the tomato: Application in genetics and breeding. *Theor. Appl. Genet.*, 57, 161-170.
- THODAY, J. M. 1961. Location of polygenes. *Nature*, 191, 368-370.
- THODAY, J. M., GIBSON, J. B. AND SPICKETT, S. G. 1964. Regular responses to selection. II. Recombination and accelerated response. *Genet. Res.*, 13, 205-218.
- THODAY, J. M. 1979. Polygene mapping: Uses and limitations. In Thompson, J. N. and Thoday, J. M. (eds) *Quantitative Genetic Variation*. pp. 219-233.
- WILLIAM, C. 1968. Correlations between egg production and chromosomal regions in *Drosophila melanogaster*. *Can. J. Gen. Cytol.*, 10, 951-960.
- WILLIAM, C. 1977. Correlations between chromosome segments and fitness in *Drosophila melanogaster*. I. The X chromosome and egg production. *Genetics*, 85, 721-732.
- WILLIAM, C. 1978. Correlation between chromosomal segments and fitness in *Drosophila melanogaster*. II. The X chromosome and egg viability. *Genetics*, 92, 595-601.
- WILLIAM, C. 1980. Correlation between chromosome segments and fitness in *Drosophila melanogaster*. IV. Fecundity, viability and the third chromosome. *Can. J. Genet. Cytol.*, 22, 405-410.
- 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 tomato with the aid of genetic markers using approximate maximum likelihood methods. *Heredity*, 59, 413-421.
- WELLER, J. I., SOLLER, M. AND BRODY, T. 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato by means of genetic markers. *Genetics*, 118, 329-339.
- ZHUCHENKO, A. A., KOROL, A. B. AND ANDRYUSHCHENKO, V. K. 1979. Linkage between loci of quantitative characters and marker loci. *Genetika*, 14, 771-778.