A Population Genetics Model of Marker-Assisted Selection

Z. W. Luo,* R. Thompson† and J. A. Woolliams†

** Institute of Genetics, Fudan University, Shanghai, P.R. China and † Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, United Kingdom*

Manuscript received January 15, 1997
Accepted for publication March 27, 1997

ABSTRACT

A deterministic two-loci model was developed to predict genetic response to marker-assisted selection (MAS) in one generation and in multiple generations. Formulas were derived to relate linkage disequilibrium in a population to the proportion of additive genetic variance used by MAS, and in turn to an extra improvement in genetic response over phenotypic selection. Predictions of the response were compared to those predicted by using an infinite-loci model and the factors affecting efficiency of MAS were examined. Theoretical analyses of the present study revealed the nonlinearity between the selection intensity and genetic response in MAS. In addition to the heritability of the trait and the proportion of the marker-associated genetic variance, the frequencies of the selectively favorable alleles at the two loci, one marker and one quantitative trait locus, were found to play an important role in determining both the short- and long-term efficiencies of MAS. The evolution of linkage disequilibrium and thus the genetic response over several generations were predicted theoretically and examined by simulation. MAS dissipated the disequilibrium more quickly than drift alone. In some cases studied, the rate of dissipation was as large as that to be expected in the circumstance where the true recombination fraction was increased by three times and selection was absent.

SEVERAL statistical models have been developed to investigate the efficiency of marker-assisted selection (MAS) for the improvement of a quantitative trait. In an attempt to analyze the effects of 10 blood loci on the milk production traits in a dairy cattle population, NEIMANN-SORENSEN and ROBERTSON (1961) first developed a statistical method to assess the genetic effects associated with known marker loci. A fundamental measure established for such assessment in their article is the additive genetic variance explained by the marker loci. The study assumed a direct influence of the marker gene on the trait. This was also assumed by SMITH (1967) who investigated the potential of MAS in a range of livestock selection schemes. The study showed that the extra improvement in genetic response from MAS relative to the genetic response from phenotypic selection (*i.e.,* termed here the relative efficiency of MAS) of the selection schemes increased as the proportion of additive genetic variance explained by the marker increases and the heritability of the trait under selection decreases.

LANDE and THOMPSON (1990) have recently established a general framework for the relative efficiency of MAS schemes over the corresponding phenotypic selection schemes for a single generation. In this study markers were not assumed to have a direct effect on the trait. Their formulas again indicated that the relative efficiency of MAS under various circumstances was determined by the proportion of the additive genetic variance in the trait explained by the markers and the heritability of the trait. These results were mainly derived from the infinite-loci model using linear statistical models commonly used in the classical selection index theory (HAZEL 1943; FALCONER 1989) with large sample size assumptions. The infinite-loci model in LANDE and THOMPSON (1990) considered the quantitative traits that are additively influenced by numerous genes. Computer simulation studies by ZHANG and SMITH (1992) and GIMELFARB and LANDE (1994) confirmed that use of MAS could be more efficient in improving selection response than the corresponding phenotypic selection schemes and investigated factors affecting the efficiency of MAS. These studies provided a more realistic assessment for practice of MAS and showed that LANDE and THOMPSON had overestimated the potential.

It has been clear from these studies that MAS could be useful in improving selection response for a quantitative trait with very low heritability provided that the marker loci explain a substantial proportion of additive genetic variance of the quantitative trait under selection. Furthermore, when a quantitative trait has a low heritability, the effect of each underlying individual quantitative trait locus (QTL) must be very small in relation to the phenotypic standard deviation and even a selectively advantageous allele at the QTL could be lost due to random genetic drift in populations with a finite size. The use of MAS may thus be advantageous in reducing the random loss and increasing the fixation probability of the selectively favored allele. Neverthe-

Corresponding author: Zewei Luo, Institute of Genetics, Fudan University, 220 Handan Rd., Shanghai 200433, P.R. China.
E-mail: zwluo@ms.fudan.sh.cn

TABLE 1

Selection model of a quantitative trait locus by use of a genetic marker

Marker genotypes: Genotypes at QTL:	*MM*			*Mm*			*mm*		
	AA	*Aa*	*aa*	*AA*	*Aa*	*aa*	*AA*	*Aa*	*aa*
Frequencies (f_{ij})	$p_t^2 Q_t^2$	$2p_t^2 Q_t \times (1 - Q_t)$	$p_t^2(1 - Q_t)^2$	$2p_t(1 - p_t) Q_t R_t$	$2p_t(1 - p_t) \times (Q_t + R_t - 2Q_t R_t)$	$2p_t(1 - p_t) \times (1 - Q_t) \times (1 - R_t)$	$(1 - p_t)^2 R_t^2$	$2(1 - p_t)^2 \times R_t(1 - R_t)$	$(1 - p_t)^2 \times (1 - R_t)^2$
Fitness	1	$1 - h_2 s_2$	$1 - s_2$	$1 - h_1 s_1$	$1 - h_1 s_1 - h_2 s_2$	$1 - h_1 s_1 - s_2$	$1 - s_1$	$1 - s_1 - h_2 s_2$	$1 - s_1 - s_2$

s_1 (or s_2) and h_1 (or h_2), selection coefficient and degree of dominance in the selection coefficient at the marker locus (or at the QTL), respectively; p_t and Q_t (or R_t), frequency of marker allele M and proportion of allele A at the QTL among chromosomes carrying the marker allele M (or m).

less, it remains unclear how the proportion of the additive genetic variance in the trait explained by the marker loci evolves with successive generations of selection. This will determine the overall effectiveness of such uses of markers.

In this article, a population genetic model of MAS is developed for a single marker and QTL and the factors involved in the model are examined for their effect on the efficiency of MAS in the short and long term, with particular emphasis on change of linkage disequilibrium between the marker and QTL using MAS.

MODEL

For simplicity, a random-mating diploid population is assumed with an effective size N_e and nonoverlapping generations. Two autosomal loci are assumed: one additively affects a quantitative trait (QTL) while the other is a codominant marker that has no direct effect on the trait. The two loci may be linked with a recombination frequency of $r(0 \leq r \leq 0.5)$, which is the same for both sexes. The two alleles are denoted by M and m at the marker locus and by A and a at the QTL. The phenotype of the trait (Z) is assumed to have a residual variance σ_e^2 and to have an initial phenotypic variance of 1. The residual variance may be considered as an overall variation comprising genetic variation caused from segregation of QTLs that are in linkage equilibrium with the marker as well as a component of environmental variation. The phenotype was assumed to be normally distributed although this was only necessary where stated. The QTL is assumed to affect the trait additively and the difference in genotypic value between the two homozygotes at the QTL is $2d$. The frequencies of M and A at generation t before selection are denoted by p_t and q_t, respectively.

The population genetic model for selection on both the marker and QTL is shown in Table 1. The selective advantages of the two loci were assumed to combine additively as described in KIMURA and OHTA (1971) and the fitness of an individual in the population before selection was assumed to be entirely determined by its genotype at the two loci. In Table 1, s_1 and h_1 are selection coefficient and degree of dominance in the selection coefficient at the marker locus while s_2 and h_2 are the corresponding selection parameters at the QTL; Q_t (or R_t) is the frequency of allele A at the QTL among chromosomes carrying M (or m); f_{ij} represents frequency of the ith genotype at the marker locus ($i = 1$, 2, 3 referring to genotype MM, Mm and mm, respectively) and the jth genotype at the QTL ($j = 1$, 2, 3 corresponding to genotype AA, Aa and aa).

Using the model of two-loci selection discussed by EWENS (1979), the increments in the frequencies due to selection at generation t are shown, with p_t and $q_t > 0$:

$$\Delta p_t = \frac{\begin{array}{r} p_t(1 - p_t)\{[1 - p_t - (1 - 2p_t)h_1]s_1 \\ + [1 - q_t - (1 - 2q_t)h_2](Q_t - R_t)s_2\} \end{array}}{\begin{array}{r} 1 - (1 - p_t)[1 - (1 - 2h_1)p_t]s_1 \\ - (1 - q_t)[1 - (1 - 2h_2)q_t]s_2 \end{array}} \quad (1.1)$$

$$\Delta Q_t = \frac{\begin{array}{r} Q_t(1 - Q_t)[1 - q_t - (1 - 2q_t)h_2]s_2 \\ - r(1 - h_1 s_1 - h_2 s_2)(1 - p_t)(Q_t - R_t) \end{array}}{\begin{array}{r} 1 - (1 - p_t)h_1 s_1 - \{(1 - q_t)(1 - Q_t) \\ + [q_t + Q_t(1 - 2q_t)]h_2\}s_2 \end{array}} \quad (1.2)$$

$$\Delta q_t = \frac{\begin{array}{r} q_t(1 - q_t)[1 - q_t - (1 - 2q_t)h_2]s_2 \\ + p_t(1 - p_t)[1 - p_t - (1 - 2p_t)h_1] \\ \times (Q_t - R_t)s_1 \end{array}}{\begin{array}{r} 1 - (1 - p_t)[1 - (1 - 2h_1)p_t]s_1 \\ - (1 - q_t)[1 - (1 - 2h_2)q_t]s_2 \end{array}} \quad (1.3)$$

$$\Delta R_t = \frac{\begin{array}{r} R_t(1 - R_t)[1 - q_t - (1 - 2q_t)h_2]s_2 \\ + r(1 - h_1 s_1 - h_2 s_2)p_t(Q_t - R_t) \end{array}}{\begin{array}{r} 1 - [1 - (1 - h_1)p_t]s_1 - \{(1 - R_t) \\ \times (1 - q_t) + [q_t + R_t(1 - 2q_t)]h_2\}s_2 \end{array}} \quad (1.4)$$

The term $D_t = p_t(1 - p_t)(Q_t - R_t)$ represents linkage disequilibrium between the marker locus and the QTL at generation t. D_t is the covariance at generation t between the variates X_M and X_A; where $X_M = 1$ or 0 depending upon whether M is present or absent on a randomly chosen chromosome and X_A is similarly defined.

It is clear from Equation 1.3 that change in the allelic frequency at the QTL can be partitioned into two parts: the first due to direct selection at the QTL with a selec-

tion intensity described by h_2 and s_2, and the second due to the effect of selection at the marker locus determined by h_1 and s_1 mediated by $(Q_t - R_t)$. A sufficient condition for the second term to be nonzero is that the marker and the QTL are in linkage disequilibrium. The system of Equations 1 has only 3 d.f. since a dependency exists in that $q_t = p_t Q_t + (1 - p_t) R_t$.

ANALYTICAL METHODS

In the present study with the simplified model of QTL and marker locus, response to MAS and truncation selection for the quantitative trait can be measured by the increase in the frequency of the allele A at the QTL, since $E(\Delta Z) = 2d\Delta q$. A deterministic model is developed to model the dynamics of the linkage disequilibrium under recombination and selection, which is necessary for evaluating the response to continuous MAS. Second, a framework is given for the comparison of the two-loci model with the methods of LANDE and THOMPSON (1990) based on the infinite-loci model. Finally, a simulation model was set up to examine the validity of approximations made in the theoretical analyses.

Deterministic model: Equations 1.1–1.4 are used to provide the joint evolution of p, q, Q and R over time for an infinite population. Analytical solutions to this group of differential equations are difficult because of their nonlinearity, but numerical solutions can be obtained iteratively using the initial allelic frequencies at the marker (p_0) and the QTL (q_0), and Q_0 since $R_0 = (q_0 - p_0 Q_0)(1 - p_0)^{-1}$, and from these $D_t (t = 0, 1, 2 \cdots)$ can be evaluated in a recursive fashion. It should be noted that D_0 will be only a function of recombination rate between the two loci and population size if the disequilibrium is created entirely from hybridization between two inbred lines as showed in the APPENDIX.

It was found (HILL and ROBERTSON 1968; OHTA and KIMURA 1969) that when effect of selection is ignored, the dynamics of the linkage disequilibrium between two loci can be described by the recurrence equation $D_t = (1 - r)[1 - (2N_e)^{-1}] D_{t-1}$. The equations given by (1.1) – (1.4) already account for the effect of recombination assuming selection in an infinite population, but the linkage disequilibrium was modified by a factor $[1 - (2N_e)^{-1}]$ as a partial adjustment for finite population size. The effective population size N_e was approximated by $4N_m N_f (N_m + N_f)^{-1}$, which is appropriate when there is no selection on a population with N_m male and N_f female parents with family sizes Poisson distributed. The use of the approximation in this context is simply as a first-order correction. With the population sizes used in this study (see later) the approximation is examined in the following numerical analyses and works well.

Rate of gain using the deterministic model applied to index selection: In the present model, strategies of selection on both marker locus and QTL are entirely determined by the selection parameters s_1, s_2, h_1 and h_2. This section will focus on determination of these parameters from which genetic gain from optimizing improvement in the population mean as in LANDE and THOMPSON (1990) may be expected.

The approach adopted by LANDE and THOMPSON, and its natural extension over multiple generations, was to use methods appropriate for an infinite-loci model and to construct a selection index that was a linear combination of the phenotype and marker score (S). The marker score was defined analogously to LANDE and THOMPSON (1990) as the sum of the additive effects associated with the marker loci: in the present context the scores are $S_1 = 2Dd/p$, $S_2 = Dd(1 - 2p)/[p(1 - p)]$ and $S_3 = -2Dd/(1 - p)$ for marker genotypes $i = 1$, 2 and 3 (these can be deduced from the APPENDIX). The index is of the form $I = b_S S + b_Z Z$.

LANDE and THOMPSON (1990) showed that the selection index with $b_S b_Z^{-1} = (1/h^2 - 1)(1 - \rho^2)^{-1}$ would optimize the rate of improvement in the mean phenotype of the selected trait, where ρ^2 is the proportion of the additive genetic variation explained by the marker and h^2 is the heritability of the QTL in the present context. The APPENDIX shows that $\rho^2 = D^2[p(1 - p) q(1 - q)]^{-1}$. This term ρ^2 is the square of the correlation of the variates X_M and X_A defined earlier. Therefore in this context,

$$b_S b_Z^{-1} = (1/h^2 - 1) \times \{1 - D^2[p(1 - p) q(1 - q)]^{-1}\}^{-1}. \quad (2)$$

When h^2 is small, this ratio is very large with the index emphasizing the marker information. Here the genetic variation arises only from the QTL; thus $h^2 = 2q(1 - q) d^2$. The index was scaled to give $b_Z = 1$, so the expression in Equation 2 is the value of b_S used in the index.

The prediction of the genetic gain in terms of the deterministic model of Table 1 does require estimation of s_1, s_2, h_1 and h_2. With a population of infinite size and a proportion π selected according to the index, the proportion of selected individuals with the genotypes ij at the two loci is $\pi_{ij} = f_{ij} \pi^{-1} \Phi(\xi_{ij} - T)$, where Φ is the standardized normal distribution function and T is a truncation point, ξ_{ij} is proportional to the expected value of the index for subclass ij and is given by $\xi_{ij} = \sigma_e^2 [b_S b_Z^{-1} S_i + G_j]$ where G_j is $2(1 - q) d$, $(1 - 2q) d$ and $-2qd$ for $j = 1$, 2 and 3, respectively. The truncation point T can be found numerically to satisfy $\pi = \sum_{i=1}^{3} \sum_{j=1}^{3} f_{ij} \Phi(\xi_{ij} - T)$. Then the expected frequency of the allele A in the next generation is $q_1 = \sum_{i=1}^{3} (\pi_{i1} + {}^1/_2 \pi_{i2})$ and $\Delta q = q_1 - q_0$. The method of HILL (1969) was extended to model selection from the nine genotypes in order to examine the accuracy of this procedure in finite populations of the sizes considered here, and a very good agreement was obtained.

This result, however, used a fully parameterized model with fitness coefficients derived for each individual genotype and thus required 8 d.f. Table 1 on the other hand is a model determined by only 4 d.f. (*i.e.*, the four selection parameters). The fit of this latter model and the resulting accuracy were examined using weighted least squares. The relative fitness of each group ($\pi_{ij} f_{ij}^{-1}$) were regressed upon the frequency of M and the frequency of A within the group (2 d.f.), which described the additive effects (*i.e.*, $h_1 = h_2 = 0.5$), and upon a further two variables describing dominance deviations at each locus (2 d.f.). The f_{ij} were used as weights for the nine observations in the regression analysis, and the multiple regression was constrained so that the mean population fitness was 1. The algorithm and the corresponding numerical subroutine to fit a regression model using weighted least squares can be found elsewhere (DRAPER and SMITH 1966, p. 77ff; GENSTAT 5 release 3, Chapter 8).

From these analyses the additive, dominance and residual components of variance for fitness were obtained and the parameters s_1, s_2, h_1 and h_2 were obtained from the regression coefficients. The regression coefficients were rescaled so that the {*MM*, *AA*} genotype had a fitness of 1 as in Table 1. The rate of gain was then predicted from the estimates of s and h for each locus using Equation 1.3.

Predicting gain assuming an infinite-loci model: LANDE and THOMPSON predict that the ratio of the rate of gain using index weights given by Equation 2 (denoted by ΔZ_{MAS}) and phenotypic selection (denoted by ΔZ_T) is given by

$$\Delta Z_{MAS} \Delta Z_T^{-1} = \sqrt{\rho^2 h^{-2} + (1 - \rho^2)^2 (1 - h^2 \rho^2)^{-1}},$$

where ρ^2 is the proportion of additive genetic variation explained by the markers. With the single QTL model but following LANDE and THOMPSON (1990),

$$\Delta q_T = q(1 - q)\, id \qquad (3)$$

$$\Delta q_{MAS} = \Delta q_T \sqrt{\rho^2 h^{-2} + (1 - \rho^2)^2 (1 - h^2 \rho^2)^{-1}}, \qquad (4)$$

where Δq_T is the estimate obtained from the infinite-loci model in which the rate of gain in phenotypic selection is predicted by ih^2, where i is the selection intensity. The implication of Equation 4 is that the rate of progress with MAS and its relative efficiency depends solely on h^2, ρ^2 and the intensity of selection.

SIMULATION MODELS

Index selection over one generation: For 11 schemes with parameters given in Table 2, base populations were simulated using Monte-Carlo methods. These schemes represented changes in census number, selection intensity, marker and QTL allele frequencies and standardized gene effects (d). Truncation selection was carried out for one generation with two selection procedures:

TABLE 2

Parameters defining the 11 schemes simulated for one generation of MAS

Schemes	N	π	p	q	r	D	d
1	100	0.1	0.5	0.5	0.1	0.2	1/4
2	200	0.2	0.5	0.5	0.1	0.2	1/4
3	200	0.1	0.5	0.5	0.1	0.2	1/4
4	200	0.1	0.5	0.5	0.1	0.2	1/8
5	200	0.1	0.5	0.5	0.1	0.1	1/4
6	200	0.1	0.5	0.5	0.0	0.0	1/4
7	200	0.1	0.5	0.5	0.5	0.2	1/4
8	200	0.1	0.3	0.3	0.1	0.2	1/4
9	200	0.1	0.7	0.7	0.1	0.2	1/4
10	200	0.1	0.3	0.5	0.1	0.1	1/4
11	200	0.1	0.5	0.3	0.1	0.1	1/4

N, the number of individual per sex before selection; π is the selection proportion for each sex; p, q and r the gene frequencies of the M allele and the A allele and the recombination fraction between the marker and QTL, respectively; D, the linkage disequilibrium; and d, the standardized additive effect of A.

(i) on phenotype alone (*i.e.*, mass selection, $b_S = 0$ in the equation of index I), and (ii) on the index constructed with both phenotypic and marker information using Equation 2 with the assumed values of p, q, D and d. Genetic gain was assessed by Δq in the selected parents, which determines the genetic gain in the next generation assuming equal contribution of parents to the unselected offspring. Each scheme was simulated 200 times with each selection procedure. One exception was made to using Equation 2 in the indices and that was in scheme 6, in which the linkage disequilibrium was zero but the suboptimal index, $I = S + Z$ was used.

Evaluation of disequilibrium over multiple generations: Further populations were simulated to examine the evolution of the linkage disequilibrium over several generations of selection. The various combinations of gene frequencies, recombination rates and selection coefficients were examined. The selection coefficients were assumed constant over all generations.

For a given set of simulated parameters, the base population was generated by random union of the haploid genotypes at the marker locus and QTL whose frequency distribution was described in Table 1. Individual phenotype for the quantitative trait was generated from its corresponding genotypic value at the QTL plus a random number that was sampled from a normal distribution with mean zero and the residual variance that was the difference of phenotypic variance (assumed to be 1) and the genetic variance due to the QTL. The populations were of census size of $2N$ in each generation before selection and were propagated using random mating of equal numbers of selected male and female parents (n of each sex, where $n = \pi N$ with selection proportion π), each pair producing full-sib

families with N/n offspring of each sex. Selection of parents was mimic in accordance to the procedure as follows: Let w_{xi} represent the fitness of the ith individual with sex x ($=m$ or f referring to male or female, respectively) at generation $t-1$ defined as in Table 1 and $w'_{xi} = \sum_{k=1}^{i} w_{xk} / \sum_{k=1}^{n} w_{xk}$. The kth individual was selected if $w'_{x(k-1)} < y \leq w'_{xk}$, where y is a realization of a random variate uniformly distributed over [0, 1]. This was repeated until the required numbers of parents had been selected. The "random walk" procedure for generating a progeny genotype from two defined parental genotypes accounting for linkage between the loci involved has been described elsewhere (LUO and KEARSEY 1989).

A characteristic of the simulations of one generation of MAS using indices was that the selectively favored allele at the marker locus was very quick to reach fixation. To avoid the quick fixation of the favorable marker allele and to investigate how the persistency of the linkage disequilibrium contributes to genetic gain of continuous MAS, selection coefficients were assigned to be less than those arising from the single-generation simulations and, unless otherwise stated in the text, (s_1, s_2) = (0.3, 0.1) or (0.6, 0.2) with $h_1 = h_2 = 0.5$. Two or five hundred replicates were carried out for each model depending upon magnitude of standard errors of means of the simulation replicates.

NUMERICAL RESULTS

Goodness of fit of the two-loci model: Table 3 shows the proportions of the total genotypic variance for fitness that is attributable to additive, dominance and residual components of genetic variance. The size of the residual component is a measure of the goodness of fit of the model in Table 1, and in these schemes this term always accounted for 3% or less of the total variance.

TABLE 3

The fitted values for the selection coefficients and the additive and the dominance variations in the fitness defining the 11 different MAS schemes using linear indices combining phenotype and marker score

Scheme	s_1	h_1	s_2	h_2	V_A	V_D
1	0.887	1.01	0.047	0.97	0.67	0.33
2	0.954	1.01	0.047	0.97	0.67	0.33
3	0.877	1.01	0.114	0.99	0.67	0.33
4	0.941	1.01	0.059	0.99	0.67	0.33
5	0.904	1.02	0.112	0.86	0.65	0.33
6	0.863	0.81	0.227	0.56	0.82	0.15
7	0.887	1.01	0.114	0.99	0.67	0.33
8	0.985	0.98	0.015	0.98	0.50	0.50
9	0.913	1.01	0.095	0.93	1.00	0.00
10	0.991	0.98	0.012	0.87	0.51	0.49
11	0.805	1.03	0.203	0.81	0.65	0.33

s_1, h_1, s_2 and h_2, selection coefficients; V_A, additive variation; and V_D, dominance variation.

This is indicative of a good fit using only additive and dominance in fitness.

The selection coefficients obtained (see Table 3) indicated that the favorable marker tended to be recessive in that only *MM* homozygotes had a significant selective advantage. The proportional contribution by the dominance component tended to vary inversely with the marker gene frequency, with little effect of variation in the other parameters. In all of the schemes the selection coefficient on the marker locus was much greater than that on the QTL. It should be remembered that the fit depends on the genotypic frequencies, for example, in scheme 9 the fitness predicted by the model for {*MMaa*} is close to 1 but its observed fitness was zero; however this genotype comprised only 0.0001 of the population.

Genetic gain to one generation of MAS: Table 4 summarizes the simulated values of Δq following one generation of MAS and their expected responses using Equations 1.3 and 4. It shows that the estimates from the model with only four selection coefficients accurately predicted the changes in Δq and Δp (data not shown) found in the simulations. This is consistent with the good fit found in the previous section. However the predicted response based on the infinite-loci model is usually higher than the observed response by simulation with a maximum overprediction of 96% (scheme 9), and underprediction also occurred with scheme 8 by 18%. Schemes 8 and 9 are of particular interest since the proportions of genetic variance explained by the marker and heritability were identical in the two schemes. However, the simulated genetic progress of scheme 8 was more than twice that of scheme 9. This comparison shows that in finite-loci models the effectiveness of MAS is not only a function of h^2 and ρ^2 and that major discrepancies from this presumption can occur.

In the indices for these schemes the marker score was the major component of the index, indicating that the value attached to the disequilibrium and, as an example, 53% of genetic response of the scheme with $D = 0.2$ was obtained if the value of D was halved (*cf.* schemes 3 *vs.* 5). Comparison of schemes 1 and 3 showed that for these populations differences in census size made little impact on the progress obtained. Scheme 6 was included to examine the consequence of incorporating marker information while the loci were in linkage equilibrium ($D = 0.0$); when $b_S = b_Z$, Δq was reduced by 26% in comparison to the optimum response ($b_S = 0$, $b_Z > 0$). This would have been predicted from the results of SALES and HILL (1967).

The use of MAS with two loci was effective in maintaining progress with reduced selection intensity, and this was reflected in the predictions from Equation 1.3, but the estimate based on the infinite-loci model predicted a proportional reduction of 20% (*cf.* schemes 2 and 3).

Frequencies of the selectively favorable alleles at the

TABLE 4

The expected change in gene frequency of the favorable allele at the QTL following one generation of index selection

	Phenotypic selection		Marker assisted selection		
Scheme	Simulated	Predicted Equation (3)	Simulated	Predicted Equation (1.3)	Predicted Equation (4)
1	0.101	0.110	0.408	0.420	0.498
2	0.089	0.087	0.410	0.405	0.397
3	0.112	0.110	0.412	0.420	0.498
4	0.054	0.055	0.410	0.413	0.498
5	0.112	0.110	0.234	0.250	0.265
6	0.112	0.110	0.083	0.090	0.110
7	0.112	0.110	0.414	0.420	0.498
8	0.094	0.092	0.619	0.617	0.542
9	0.094	0.092	0.276	0.293	0.542
10	0.112	0.110	0.322	0.326	0.285
11	0.094	0.092	0.260	0.261	0.259

Δq, change in gene frequency. The number of replicates in the simulation were 200 for each scheme of MAS and phenotypic selection. Standard errors of the simulated means averaged over 200 replicates ranged from 0.006 to 0.010. Simulations for schemes that are genetically equivalent for phenotypic selection have been pooled before presentation.

marker and QTL had a substantial effect on genetic response to MAS. Comparison of response for schemes 10 and 11 shows that the frequencies p and q were not interchangeable in determining the response. In these two schemes the efficiencies of MAS relative to phenotypic selection predicted by LANDE and THOMPSON were 2.59 and 2.81, respectively. The observed ranking of efficiencies was, in fact, reversed with ratios 2.88 and 2.78.

Evolution of linkage disequilibrium: Sustaining the contribution of a marker locus to genetic response will depend upon the maintenance of linked disequilibrium between the marker locus and the QTL under selection, recombination and genetic drift. Figure 1 shows the predicted and the simulated dynamics of linkage disequilibrium for a range of initial conditions. The recursive use of the deterministic model (Equations 1.1–1.4) with the adjustment for the finite population size with the initial conditions predicted the dynamics of D averaged over simulation replicates very well. Linkage disequilibrium decreased in all schemes after the first selected generation. Only a few generations of selection were necessary to exhaust nearly all the disequilibrium; however, the rate of the reduction was different among the different selection schemes.

The results from Figure 1a show the effects of varying the parameter values of p and q: as the favorable marker and QTL allele frequencies increase, the disequilibrium decays faster. An approximate assessment of the rate of decays can be obtained by regressing $\log_e(D_t)$ on the generation number and expressing the rate as $1 - \exp(b)$, where b is the slope of the regression. If there is no selection, this rate of decays is equal to the recombination fraction r. For $p = q = 0.3, 0.5, 0.7$, the estimated rates of decays under MAS were 0.129, 0.185 and 0.224, respectively for a recombination fraction of 0.1. Thus with selection coefficients of $(s_1, s_2) = (0.3, 0.1)$ and high gene frequency, the decay in D was more than twice the recombination rate. Figure 1b shows that the effect of increasing allele frequency in Figure 1a is perhaps due more to the value of p, the frequency of the favorable marker allele, than q the favorable QTL allele.

The selection intensities at both the marker and QTL had a strong effect on decay of the disequilibrium as shown in Figure 1c. When $r = 0.1$, the observed decays were 0.186 and 0.371 for selection coefficients (0.3, 0.1) and (0.6, 0.2), respectively.

When the favorable marker and QTL alleles were recessive in fitness ($h_1 = h_2 = 1$), as found in Table 2, the decline in D was similar to the additive case initially but the decline became approximately linear (see Figure 1d) and more rapid in the later generations. The lowest rate of decline observed in the cases studied was for additive alleles. When $r = 0.5$, the measure of decay in disequilibrium was also 0.5, as expected whether or not there is selection.

Relative efficiency and accumulated genetic gain over multiple generations: Figure 2 presents theoretical and simulated efficiencies of MAS (*i.e.*, relative rate of gain of MAS over that for phenotypic selection) over 10 generations for the same schemes as those considered in Figure 1. It should be noted that these curves demonstrate not only changes in efficiency due to linkage disequilibrium, but also separate but related changes brought about through changes in q. As q increases to 0.5, genetic variance increases. However, when q increases beyond 0.5 the genetic variance decreases again. The response to both MAS and phenotypic selection

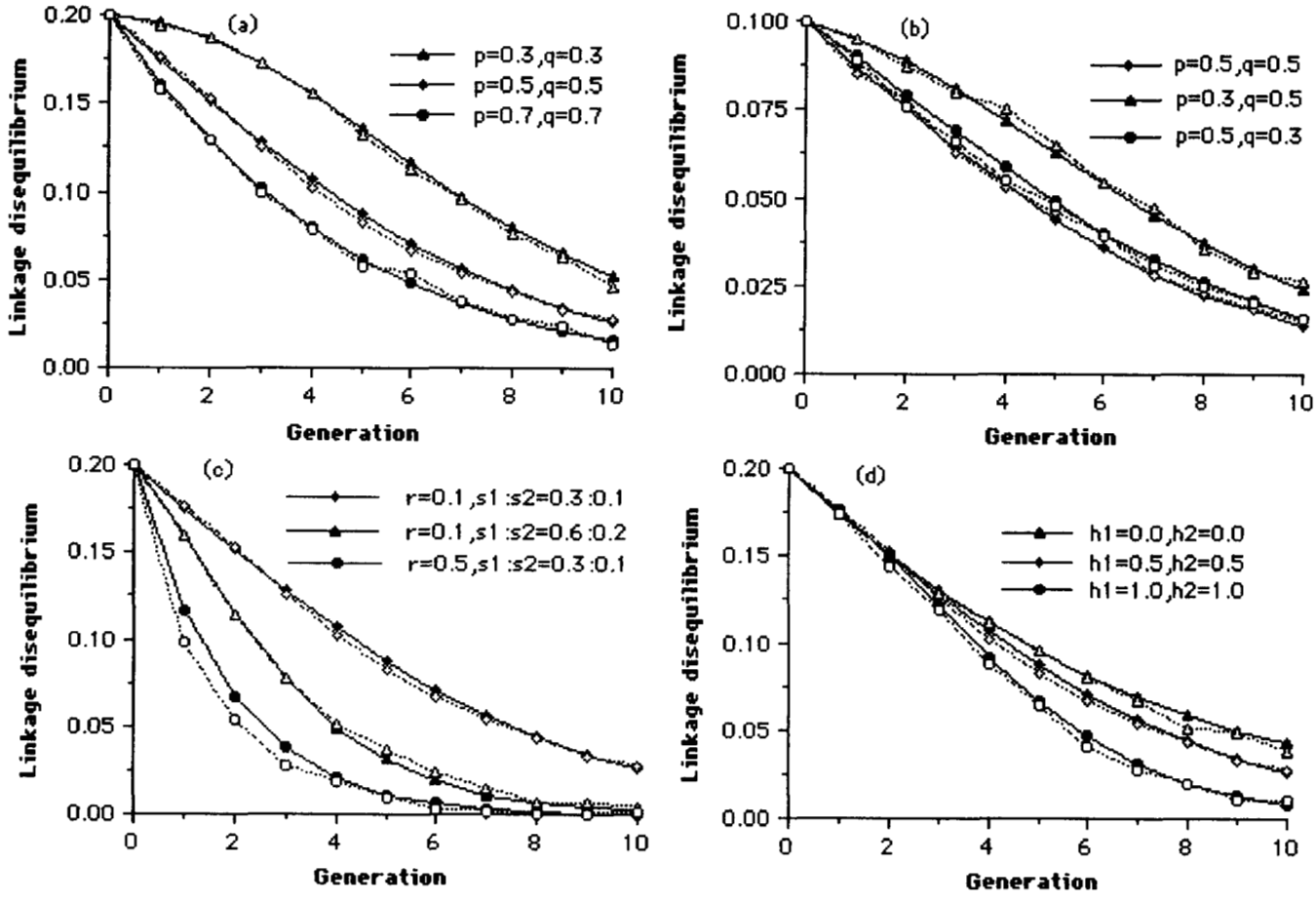

FIGURE 1.—Expected linkage disequilibrium under continuous marker-assisted selection, which were predicted from theoretical approximation (solid symbols and curves), and the corresponding simulated values (open symbols and dashed curves), which were averaged from 200 replicates, for different selection schemes, where N; π; r; p, q; h_1, h_2; s_1, s_2; and D are as follows: census population size of each sex; selection proportion; recombination frequency; gene frequencies at the marker and the QTL, respectively; selection coefficients at the marker and QTL, respectively; and the disequilibrium coefficient between the marker locus and QTL, respectively. In these schemes $N = 500$; $\pi = 10\%$; $r = 0.1$; $p = q = 0.5$; $h_1 = h_2 = 0.5$; $s_1{:}s_2 = 0.3{:}0.1$; and $D = 0.2$, unless otherwise indicated.

will reflect this pattern individually, but the differential response in the two schemes means that after the first generation the two selection procedures, when compared at a given generation, were operating under different conditions. This complicates interpretation of underlying factors affecting the dynamic change in the efficiency of MAS over several generations. To avoid this complexity and to explore contribution of remaining disequilibrium to continuous response of MAS, accumulated genetic gains after 10 generations of MAS in these schemes were predicted from the theoretical Equations 1 and were listed in Table 5 together with their corresponding simulated means.

It can be seen that theoretical analyses adequately predict the simulated dynamics of the efficiencies (Figure 2) and the simulated genetic gain accumulated over 10 generations of MAS (Table 5). It is clear from the figures that the efficiency of MAS declines more linearly than the linkage disequilibrium. The schemes in which the linkage disequilibrium is more persistent are not necessarily more efficient in the long term as defined here. This is due to the fact that the dynamics were driven by the two different selection procedures as explained above.

For $p = q$ (Figure 2a) MAS was more efficient and the accumulated gain was much larger when favorable allele was at lower frequency, in agreement with the persistence of the linkage disequilibrium. However, when the scheme with $p = 0.3$ and $q = 0.5$ is compared to that with $p = 0.5$ and $q = 0.3$, the differentiation in persistence of D (see Figure 2b) made no clear difference to the generation by generation efficiency in the long term. Both the schemes had the same initial value of ρ^2, but the accumulated responses over 10 generations were 0.26 and 0.21 for these two schemes, respectively, suggesting that the scheme with the more persistent disequilibrium ($p = 0.3$, $q = 0.5$) would achieve a greater response. Free recombination between the marker locus and the QTL ($r = 0.5$) destroyed the disequilibrium quickly and thus the response to continuous MAS. Moreover, stronger selection on both the marker locus and QTL ($s_1{:}s_2$ =

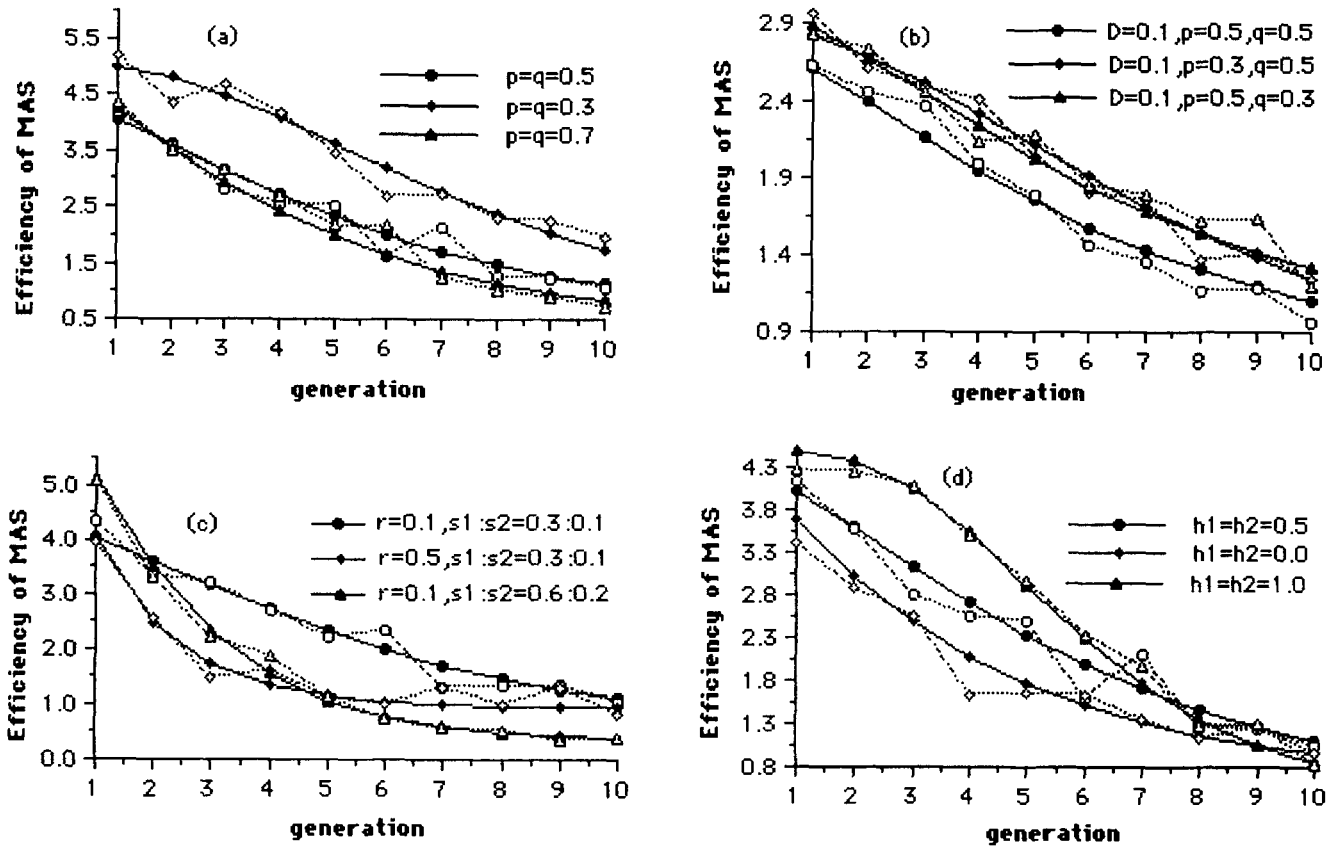

FIGURE 2.—Theoretical predictions of efficiency of continuous marker assisted selection (solid symbols and curves) and the corresponding simulated values (open symbols and dashed curves), which were averaged from 1000 replicates, for different selection schemes, where N; π; r; p, q; h_1, h_2; and s_1, s_2 are as follows: census population size of each sex; selection strength; recombination frequency; gene frequencies at the marker and the QTL, respectively; and selection coefficients at the marker and QTL, respectively. In these schemes $N = 500$; $\pi = 10\%$; $r = 0.1$; $p = q = 0.5$; $h_1 = h_2 = 0.5$; and $s_1:s_2 = 0.3:0.1$, unless otherwise indicated.

0.6:0.3) had also yielded a quick exhaustion in the disequilibrium and further using marker information in this scheme had lowered the response compared to phenotypic selection. However, a large genetic gain had accumulated during the consecutively strong MAS. Although decline in the disequilibrium was slowed down as degrees of dominance in the selection coefficients (h_1 and h_2) decreased (Figure 1d) the long-term efficiency and the accumulated genetic gain increased with increasing the dominance (Figure 2d). This reflects difference in the efficient use of the marker information among these schemes.

DISCUSSION

Many studies have been devoted to predicting genetic response of a quantitative trait to MAS (SMITH 1967; STAM 1986; LANDE and THOMPSON 1990; ZHANG and SMITH 1992). The predictions were usually made while assuming that both individual phenotypic value of the trait and selection index, which combines the phenotype of the trait and marker information, are normally distributed (for example, LANDE and THOMPSON 1990). The normality assumption requires that the quantitative genetic variation be under control of an infinite number of loci (BULMER 1980). The use of this assumption has led to avoidance of the sophistications involved in modeling the relationship between changes in allelic frequencies at both the markers and QTLs due to selection and the linkage disequilibria among the marker loci and QTLs, which are central driving factors in determining the amount of marker information usable for the MAS. In theory, it has been argued that only those QTLs with substantially detectable effects and closely associated with marker loci may contribute to genetic gains from MAS (LANDE and THOMPSON 1990). Moreover, recent data from plant and animal experiments indicated that a substantially large proportion of quantitative genetic variation could be explained by segregation at only a few marker loci and caused by a finite number of individual QTLs or DNA segments (PATERSON *et al.* 1988; ANDERSON *et al.* 1994; LAI *et al.* 1994). In light of these facts, the computer simulation study by GIMELFARB and LANDE (1994) investigated the robustness of theoretical prediction of genetic response to MAS proposed by LANDE and THOMPSON when a few QTLs were considered. Their simulations confirmed substantial extra genetic gain

TABLE 5

Theoretical predictions of accumulated genetic gains after 10 generations of marker-assisted selection and their corresponding simulation means for the 10 different schemes

Schemes	p	q	r	s_1	s_2	h_1	h_2	D	Δq_{10}	$\Delta \tilde{q}_{10}$
1	0.5	0.5	0.1	0.3	0.1	0.5	0.5	0.2	0.303	0.301
2	0.3	0.3	0.1	0.3	0.1	0.5	0.5	0.2	0.408	0.405
3	0.7	0.7	0.1	0.3	0.1	0.5	0.5	0.2	0.212	0.212
4	0.5	0.5	0.1	0.3	0.1	0.5	0.5	0.1	0.225	0.223
5	0.3	0.5	0.1	0.3	0.1	0.5	0.5	0.1	0.261	0.257
6	0.5	0.3	0.1	0.3	0.1	0.5	0.5	0.1	0.241	0.239
7	0.5	0.5	0.5	0.3	0.1	0.5	0.5	0.2	0.203	0.200
8	0.5	0.5	0.1	0.6	0.2	0.5	0.5	0.2	0.432	0.430
9	0.5	0.5	0.1	0.3	0.1	0.0	0.0	0.2	0.223	0.221
10	0.5	0.5	0.1	0.3	0.1	1.0	1.0	0.2	0.385	0.381

The standard errors of these simulated means for 500 replicates arranged from 0.003 to 0.005. p, q, r, s_1, s_2; h_1, h_2; and D are as follows: allelic frequencies at the marker and QTL, respectively; recombination fraction between the two loci; selection coefficients at the two loci; and the coefficient of linkage disequilibrium between the marker locus and QTL, respectively; Δq_{10}, accumulated genetic gains after 10 generations of marker-assisted selection; $\Delta \tilde{q}_{10}$, simulation means for genetic gains after 10 generations.

yielded using MAS over traditional phenotypic selection and revealed substantial bias in the theoretical prediction based on the infinite-loci model. This demonstrates that MAS theory as a whole is far from conclusive and that a better understanding of the genetic properties of MAS needs theoretical studies under a finite-loci assumption. The present study made a primary effort in this attempt and considered a simple case of the finite-loci MAS model that involved only a single marker locus and QTL. We hope that this will initiate development of more sophisticated models taking multiple markers and QTLs into account. However, analyzing these multiple-loci models will not be easy inasmuch as theoretical analysis of linkage disequilibria involving three diallelic marker loci has been proved to be very complicated (BROWN 1975).

The efficiency of MAS was compared to a traditional phenotypic selection scheme by deriving the change in frequency of the favorable allele at a QTL in the present study using a population genetics model. The model was extended to more than one generation of selection and its parameterization gives insight into the factors affecting the efficiency of MAS in both the short and long terms. The model presented here differs from others in several different aspects. The two-loci models of MAYNARD-SMITH and HAIGH (1974) and OHTA and KIMURA (1975) assumed one of the loci selectively neutral. The model of SMITH (1967) and STAM (1986) considered only one generation of selection using a single marker, which was also the QTL. EWENS (1979) studied a similar model, but did not derive Equations 1.1–1.4. Also he did not apply the model to the specific context of MAS. LANDE and THOMPSON (1990) constructed general statistical prediction of genetic response to a single generation of MAS but they made no effort to investigate more detailed genetic aspects of the selection scheme. The extension to multiple generations requires the modeling of the evolution of linkage disequilibrium attempted here.

The present study (i) has developed a two-loci population genetics model under which MAS based on indices of a marker locus and a QTL underlying a quantitative trait can be appropriately described; (ii) has shown that genetic response predicted from the two-loci model differs from that by using the formulas of LANDE and THOMPSON (1990); and (iii) has examined the factors that affects the dissipation of linkage disequilibrium during directional selection on the marker and QTL compared to the situation where there is no selection.

The system of Equations 1.1–1.4, derived from theoretically analyzing the genetic model of Table 1, gave accurate predictions, over one generation, of change in gene frequency at the QTL and at the marker, and the linkage disequilibrium (Figure 1). The selection coefficients required for applying Equations 1.1–1.4 were calculated using numerical methods as described in the study. The accuracy of the one generation predictions of the genetic parameters for describing the MAS model suggests that a recursive procedure, *i.e.*, estimation generation by generation based on these equations and methods, would also prove accurate for several generations.

When the general idea of calculating genetic progress in LANDE and THOMPSON (1990) was applied to the two-loci model presented here, the factors in determining the efficiency of MAS became to be easily elucidated. It is shown in the present study that in addition to the proportion of marker-associated additive genetic variance and heritability of the trait under selection, frequencies of the selectively favorable marker and QTL alleles play an important role in determining the efficiency in both the short and long terms. For a given proportion and heritability, our study showed that an

increase in the efficiency can be expected when both the frequencies p and q are low. This is not because the favorable QTL allele is more frequent within the chromosomes carrying the favorable marker allele when p and q are low. Indeed, a comparison between schemes 8 and 9 in Table 4 indicates that the frequency of the QTL allele among the chromosomes bearing the selectively favorable marker allele is greater for the scheme 9, in which a lower genetic progress was observed.

The discrepancies in prediction of the selection response between of the two-loci model described by Equations 1.1–1.4 and LANDE and THOMPSON's Equation 4 in the present notation may be explained by comparing the scheme 8 to the scheme 9. MAS in these two schemes has been sufficient to bring the favorable QTL allele close to fixation (the allele frequencies reached 0.917 and 0.993 in the two schemes, respectively; Table 4). This illustrates a nonlinearity between selection progress and selection progress that arose from the fixation of the QTL allele, thereby introducing an upper bound to selection progress. This nonlinearity can also be seen in the schemes 3 and 4 illustrated in Table 4. However, the approach of LANDE and THOMPSON (1990) was based on normal distribution theory of the trait phenotype under question and selection index, which is justified by the infinite-loci genetic control of the trait. Under the infinite genetic model, a linearity holds between selection intensity and selection progress as illustrated in the corresponding numerical results in Table 4 and as expected from the classic theory (FALCONER 1989). Therefore it is to be expected that the gain from MAS for a trait affected by a small number of QTLs with detectably large size of effects will not be accurately predicted by such a linear model. This study confirms this to be the case. It must be pointed out that the comparison between the present two-loci model and the infinite-loci model may be regarded unfair in the sense that the general approach of LANDE and THOMPSON (1990) is being applied beyond the circumstance for which it was developed. However, such a comparison at least indicates that further research efforts must be paid to derive more realistic predictions of genetic progress from MAS before the theory can be used.

Other attempts to solve the dynamics of linkage disequilibrium between two linked loci under directional selection in a finite population have problems to obtain a formal solution of a group of multidimensional diffusion equations (for example, in HILL and ROBERTSON 1966). OHTA and KIMURA (1969) suggested an approximate method using the Kolmogorov backward equation to overcome this difficulty but their approach is not appropriate when selection on the two loci is strong because the linkage disequilibrium is no longer a linear function of the variables of the model. The model developed in the present study was able to predict accurately the evolution of linkage disequilibrium between a marker locus and QTL under selection, recombination and random genetic drift over several generations. This provides a basis for predicting response for more than one generation of MAS. The numerical analyses showed that the dissipation of linkage disequilibrium depended on intensity of selection, recombination rate and allele frequencies at both the marker and QTL when the other parameters were fixed. The effect of selection intensity on the rate of the dissipation is found to be associated with frequencies of the selectively favored marker and QTL alleles. A better maintenance of the disequilibrium with low frequencies of the alleles has an advantage for a longer persistent response to MAS. A strong selection on both the marker and QTL is sufficient to bring the loci to be in linkage equilibrium very quickly.

The present study has been focused on modeling linkage disequilibrium between a polymorphic marker locus and a single QTL. This gives some insight about the potential and limitation of using marker information to improve selection response of a quantitative trait. To use of map information efficiently in MAS, it is necessary to extend the present model to multiple markers and QTLs. However, the question remains on how can the multilocus linkage disequilibria be modeled in MAS schemes and what is the most efficient strategy to relate the multidimensional linkage disequilibria among the marker loci and QTLs to genetic progress of MAS. We hope the present model and analysis will provide basis for a multilocus analysis and serve as an important building block for our understanding of the dynamic system under MAS.

We thank Professor W. G. HILL for his critical comments on an early version of the manuscript. Dr. Z-B. ZENG and two anonymous reviewers are acknowledged for their criticisms and comments, which have been very helpful in clarifying ambiguities and improving presentation in the article. Z.W.L. is indebted to Drs. M. J. KEARSEY, C. HALEY and S. BISHOP for their comments and encouragement. We are grateful to the Ministry of Agriculture, Fisheries and Food (U.K.) and the Natural Science Foundation of China for funding this study.

LITERATURE CITED

ANDERSSON, L., C. S. HALEY, H. ELLEGREN, S. A. KNOTT, M. JOHANSSON *et al.*, 1994 Genetic mapping of quantitative trait loci for growth and fatness in pigs. Science **263:** 1771–1774.

BROWN, A. H. D., 1975 Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8:** 184–201.

BULMER, M. G., 1980 *The Mathematical Theory of Quantitative Genetics.* Clarendon Press. Oxford.

DEKKERS, J. C. M., and M. R. DENTINE, 1991 Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor. Appl. Genet. **81:** 212–220.

DRAPER, N. R., and H. SMITH, 1966 *Applied Regression Analysis* John Wiley, New York.
EWENS, W. J., 1979 *Mathematical Population Genetics.* Springer-Verlag, Berlin.
FALCONER, D. S., 1989 *Introduction to Quantitative Genetics,* Ed. 3, Longman, New York.
GIMELFARB, A., and R. LANDE, 1994 Simulation of marker assisted selection in hybrid populations. Genet. Res. **63:** 39–47.
HAZEL, L. N., 1943 The genetic basis for constructing selection indices. Genetics **38:** 476–769.
HILL, W. G., 1969 On the theory of artificial selection in finite populations. Genet. Res. **13:** 143–163.
HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269–294.
HILL, W. G., and A. ROBERTSON, 1968 Linkage disequilibrium in finite population. Theor. Appl. Genet. **38:** 226–231.
KIMURA, M., and T. OHTA, 1971 *Theoretical Aspects of Population Genetics.* Princeton University Press, Princeton.
LAI, C., R. F. LYMAN, A. D. LONG, C. H. LANGLEY, T. F. MACKAY, 1994 Naturally occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila melanogaster.* Science **266:** 1697–1702.
LANDE, R., and R. THOMPSON, 1990 Efficiency of marker-assisted selection in improvement of quantitative traits. Genetics **124:** 743–756.
LUO, Z. W., and M. J. KEARSEY, 1989 Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. Heredity **63:** 401–408.
MAYNARD-SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.
NEIMANN-SORENSEN, A., and A. ROBERTSON, 1961 The association between blood groups and several production characteristics in three Danish cattle breeds. Acta Agric. Scand. **11:** 163–196.
OHTA, T., and M. KIMURA, 1969 Linkage disequilibrium due to random genetic drift. Genet. Res. **13:** 47–55.
OHTA, T., and M. KIMURA, 1975 The effect of selected linked locus on heterozygosity of neutral alleles (the hitch-hiking effect). Genet. Res. **25:** 313–326.
PATERSON, A. H., E. S. LANDER, J. D. HEWITT, S. PETERSON, S. E. LINCOLN *et al.,* 1988 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphism. Nature **335:** 721–726.
SALES, J., and W. G. HILL, 1967 Effects of sample errors on efficiency of selection indices. Anim. Prod. **23:** 1–14.
SMITH, C., 1967 Improvement of metric traits through specific genetic loci. Anim. Prod. **9:** 349–358.
STAM, P., 1986 The use of marker loci in selection for quantitative characters, pp. 170–182 in *Exploiting New Technologies in Livestock Improvement: Animal Breeding,* edited by C. SMITH, J. W. B. KING, and J. MCKAY. Oxford University Press, Oxford.
ZHANG, W., and C. SMITH, 1992 Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theor. Appl. Genet. **83:** 813–820.

Communicating Editor: Z-B. ZENG

APPENDIX

Following FALCONER (1989), the average additive genetic effects of the haplotype are given as follows:

Marker Allele	QTL Allele	Frequency	Average effect
M	A	pQ	α_A
	a	$p(1-Q)$	α_a
m	A	$(1-p)R$	α_A
	a	$(1-p)(1-R)$	α_a

By definition of the above average effects: $q\alpha_A + (1-q)\alpha_a = 0$ and $pQ + (1-p)R = q$, so

$$p[Q\alpha_A + (1-Q)\alpha_a] + (1-p) \times [R\alpha_A + (1-R)\alpha_a] = 0.$$

Therefore the average effects of the marker alleles are given by $[Q\alpha_A + (1-Q)\alpha_a]$ and $[R\alpha_A + (1-R)\alpha_a]$ for M and m, respectively. These average effects can be simplified into Dd/p and $-Dd/(1-p)$, respectively, by using the relationships $Q = q + D/p$ and $R = q - D/(1-p)$. Thus, the additive genetic variance at the marker locus is

$$\begin{aligned}\sigma_M^2 &= 2p[Q\alpha_A + (1-Q)\alpha_a]^2 + 2(1-p) \times [R\alpha_A + (1-R)\alpha_a]^2 \\ &= 2D^2d^2[p^{-1} + (1-p)^{-1}] \\ &= 2D^2[p(1-p)]^{-1}d^2 \\ &= D^2[p(1-p)q(1-q)]^{-1}\sigma_A^2, \end{aligned} \quad \text{(A1)}$$

where D is the coefficient of linkage disequilibrium between the marker and QTL. Thus the variance attributable to the marker is $\rho^2\sigma_A^2$, where ρ is the correlation between the marker locus and the QTL in the population since D is the covariance between the two loci, and $p(1-p)$ and $q(1-q)$ are the corresponding variances. The expression includes as a special case the value $(1-2r)^2$ given by DEKKERS and DENTINE (1991) for $p = q = 1/2$ for disequilibrium arising from hybridization since in that case $D = 1/4(1-2r)$.