*Genetica* **119:** 75–86, 2003.



# A method for marker-assisted selection based on QTLs with epistatic effects

Pengyuan Liu, Jun Zhu*, Xiangyang Lou & Yan Lu
*Institute of Bioinformatics, Zhejiang University, Hangzhou 310029, China; *Author for correspondence: Department of Agronomy, Zhejiang University, Hangzhou, Zhejiang 310029, P. R. China (Phone: +86-571-86971731; Fax: +86-571-86971498; E-mail: jzhu@zju.edu.cn)*

Received 12 July 2002 Accepted 5 December 2002

*Key words:* breeding values, epistatic effects, marker-assisted selection, QTL

**Abstract**

A method for marker-assisted selection (MAS) based on quantitative trait loci (QTLs) with epistatic effects is proposed. The efficiency of such method is investigated by simulations under a wide range of situations. In the presence of epistasis, MAS generally yields longer persistence response than that based exclusively on additive or additive and dominance. Neglecting epistasis could result in considerable loss in response, and more pronounced at later generations. In addition to population size and trait heritability, genetic variance configurations play an important role in determining both the short- and long-term efficiencies of MAS. MAS using breeding values not only achieves higher response, but also tends to have smaller standard error than other methods in most cases. Errors in QTL detection cause distinct reductions in responses to MAS in most cases. It is thus concluded that verifications of putative QTL and its magnitude of effect and accurate map chromosome location are imperative to realize the potentials of MAS.

## Introduction

The advent of molecular marker technology in the 1980s has opened a new era for quantitative genetics studies and with an appealing prospect of its use in breeding program. By integrating marker information into artificial selection for polygenic traits, a system known as marker-assisted selection (MAS) has been evolved, which substantially increases the efficiency of selection, as shown by many researchers. The efficiency of such method has been investigated by either analytical approach (Lande & Thompson, 1990; Luo, Thompson & Woolliams, 1997; Knapp, 1998; Moreau et al., 1998; Ollivier, 1998; Xie & Xu, 1998) or computer simulation (Zhang & Smith, 1992, 1993; Gimelfarb & Lande, 1994; Ruane & Colleau, 1995; Whittaker et al., 1995; Hospital et al., 1997; Spelman & Bovenhuis, 1998). However, the practical utilization of information regarding epistasis in breeding is a very complicated issue that has not been adequately addressed.

The importance of epistasis has been strongly suggested from quantitative genetic studies (Mather & Jinks, 1982; Pooni, Coombs & Jinks, 1987). Results from recent QTL mapping studies have also provided strong evidence suggesting epistasis to be an important genetic basis of heterosis underlying complex quantitative traits such as plant height, grain yield and its components (e.g., Fu & Ritland, 1996; Li et al., 1997; Yu et al., 1997; Cao et al., 2001). Most studies of MAS in breeding are based on simple infinite-loci additive genetic models (e.g., Lande & Thompson, 1990) or two-locus genetic models (e.g., Luo, Thompson & Woolliams, 1997) under some restricted hypotheses. If there are additive × additive epistasis effects, additional genetic gain can be achieved in selection breeding. Nevertheless, the accurate quantification and characterization of MAS in the presence of epistasis remains largely unexplored.

The objective of the present research is (i) to propose a method for MAS based on QTLs with epistasis, applied to the improvement of a quantitative trait in

breeding, (ii) to assess the effectiveness of MAS compared to phenotypic selection, (iii) and to investigate the effects of inaccurate estimates of the location and effect of QTLs on the genetic response to MAS.

## Methods

For simplicity, we employed a cross between two inbred lines, each assumed to be homozygous for different alleles at all loci, to initiate a selection. The observable data in a typical MAS breeding program are phenotypic value, marker genotype of each individual, and genetic architecture of QTLs involved in the improved trait. We first focused on the efficiency of QTL selection. Namely, we investigated the efficiency of MAS when QTLs have been known without error. Then, we evaluated the impacts of QTL detection, in which the location and effect of QTLs were estimated inaccurately, on the genetic response to MAS.

### *Genetic model*

Suppose that there are $n$ QTLs, and denote $Q_i$ as the $i$th QTL. Each $Q_i$ is bracketed by two flanking marker alleles $M_{i-}$ and $M_{i+}$. According to the definitions of genetic effects (additive, dominance and digenic epistasis of additive × additive, additive × dominant, dominant × additive, and dominant × dominant) given by Mather and Jinks (1982), under the assumption of no genotype × environment interaction, phenotypic value of individual $k$ can be expressed as

$$\begin{aligned} y_k &= \mu + \sum_i a_i x_{A_{ik}} + \sum_i d_i x_{D_{ik}} + \\ &+ \sum_{i<} \sum_j aa_{ij} x_{AA_{ijk}} + \\ &+ \sum_{i<} \sum_j ad_{ij} x_{AD_{ijk}} + \\ &+ \sum_{i<} \sum_j da_{ij} x_{DA_{ijk}} + \\ &+ \sum_{i<} \sum_j dd_{ij} x_{DD_{ijk}} + \varepsilon_k \end{aligned} \tag{1}$$

where $\mu$ is the population mean; $a_i$ and $d_i$ are the additive and dominant effects of $Q_i$, respectively; $aa_{ij}$, $ad_{ij}$, $da_{ij}$ and $dd_{ij}$ are the epistatic effects of additive × additive, additive × dominant, dominant × additive and dominant × dominant between $Q_i$ and $Q_j$, respectively; the coefficients $x_{A_{ik}}$ and $x_{D_{ik}}$ are 1 and 0 for genotype $Q_iQ_i$, $x_{A_{ik}}$ and $x_{D_{ik}}$ are $-1$ and 0 for genotype $q_iq_i$, $x_{A_{ik}}$ and $x_{D_{ik}}$ are 0 and 1 for genotype $Q_iq_i$; $x_{AA_{ijk}} = x_{A_{ik}}x_{A_{jk}}$, $x_{AD_{ijk}} = x_{A_{ik}}x_{D_{jk}}$, $x_{DA_{ijk}} = x_{D_{ik}}x_{A_{jk}}$, $x_{DD_{ijk}} = x_{D_{ik}}x_{D_{jk}}$; $\varepsilon_k$ is the residual effect, $\varepsilon_k \sim N(0, \sigma_\varepsilon^2)$; for $i, j = 1, 2, \ldots, n$.

### *Estimation of breeding value*

If we know QTL effects and genotype of an individual, we can write breeding value $B$ of that individual $k$ as $\sum_i a_i x_{A_{ik}} + \sum_{i<} \sum_j aa_{ij} x_{AA_{ijk}}$. In an actual situation of breeding practice, the true QTL genotype of an individual is not available, only its trait phenotype and marker genotype are observable. However, we can calculate the probability of a particular QTL genotype conditioned on its trait phenotype and marker genotype, and hence estimation of breeding value of an individual is the weighted sum of breeding value of all possible genotypes,

$$\begin{aligned} \hat{B}_k &= E(B_k | y_k, \mathbf{h}) \\ &= \sum_{\mathbf{z}} p(\mathbf{z}|y_k, \mathbf{h}) \Bigg( \sum_i a_i x_{A_{ik}} + \\ &\quad + \sum_{i<} \sum_j aa_{ij} x_{AA_{ijk}} \Bigg) \end{aligned} \tag{2}$$

where $\mathbf{z}$ is a vector of QTL genotype, $\mathbf{z} = (z_1, z_2, \ldots, z_n)$, $z_i$ is the genotype of $Q_i$; $y_k$ is the phenotypic value; $\mathbf{h}$ is a vector of marker genotype, $\mathbf{h} = (h_1, h_2, \ldots, h_n)$, $h_i$ is the genotype of two flanking markers bracketing $Q_i$; $p(\mathbf{z}|y_k, \mathbf{h})$ is the probability of getting a particular $\mathbf{z}$ given the trait phenotype and marker genotype. In order to use the above formulae, $p(\mathbf{z}|y_k, \mathbf{h})$ needs to be known,

$$p(\mathbf{z}|y_k, \mathbf{h}) = \frac{f(y_k|\mathbf{z})\, p(\mathbf{z}|\mathbf{h})}{\sum_{\mathbf{z}} f(y_k|\mathbf{z})\, p(\mathbf{z}|\mathbf{h})} \tag{3}$$

where $f(y_k|\mathbf{z})$ is the probability density function of phenotype conditioned on $\mathbf{z}$, and has normal distribution with known variance,

$$\begin{aligned} y_k|\mathbf{z} \sim N \Bigg( & \mu + \sum_i a_i x_{A_{ik}} + \sum_i d_i x_{D_{ik}} + \\ & + \sum_{i<} \sum_j aa_{ij} x_{AA_{ijk}} + \sum_{i<} \sum_j ad_{ij} x_{AD_{ijk}} + \\ & + \sum_{i<} \sum_j da_{ij} d_{DA_{ijk}} + \\ & + \sum_{i<} \sum_j dd_{ij} x_{DD_{ijk}}, \sigma_\varepsilon^2 \Bigg) \end{aligned} \tag{4}$$

where $p(\mathbf{z}|\mathbf{h})$ is the probability of getting a particular $\mathbf{z}$ given the marker genotype $\mathbf{h}$. We assume

independence of recombination events, that is, Haldane's (1919) mapping function,

$$p(\mathbf{z}|\mathbf{h}) = \prod_{i=1}^{n} p(z_i|h_i) \tag{5}$$

Given the location of QTL relative to the marker bracket $p(z_i|h_i)$ is easily calculated in the $F_2$, but difficult in subsequent generations, since selection and recombination will change $p(z_i|h_i)$ in a very complicated mode. We adopted a method proposed by Whittaker et al. (1995) with some modifications to estimate $p(z_i|h_i)$ from marker frequencies in subsequent generations (see Appendix A).

Similarly, additive effects $\hat{A}$ of individual $k$ can be calculated as $\sum_z p(\mathbf{z}|y_k, \mathbf{h}) \sum_i a_i x_{A_{ik}}$. Other genetic effects, such as dominance and epistasis, and genotypic value are also estimated in the same way.

## Simulation

### *Genetic map*

The genetic map of simulation had five chromosomes, each 100 cM long, on which molecular markers were evenly spread with 10 cM, and with one marker at each chromosome end. A total of eight putative diallelic QTLs were randomly assigned among chromosomes (Figure 1). Then, a total of 16 digenic epistases were produced randomly from combinations of any two of these QTLs. The effects of QTLs were drawn from a normal distribution with zero mean and different known variance ($V_A$, $V_D$, $V_{AA}$, $V_{AD}$,$V_{DA}$ and $V_{DD}$), respectively. The recombination fraction was derived from the map distance ($d$) as $r = 0.5(1 - e^{-2d})$ (Haldane, 1919).

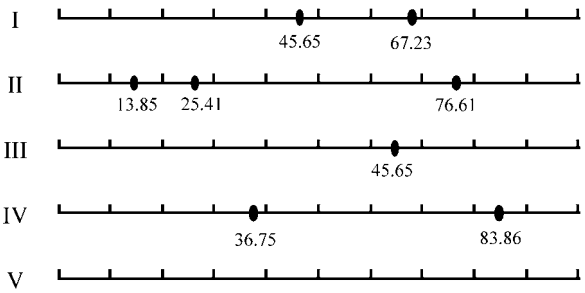


*Figure 1.* Genetic map with eight digenic QTLs randomly assigned among five chromosomes and 11 evenly spaced genetic markers per chromosome. The ellipses represent QTLs and numbers under QTLs indicate their distances on chromosomes.

### *Generating phenotypic data*

For generating each data set based on the above genomic information, the expected genetic variance $V_G$ in initial population ($F_2$) was calculated first:

$$\begin{aligned}
V_G = {} & \frac{1}{2}\sum_i\sum_j \rho_{ij} a_i a_j + \frac{1}{4}\sum_i\sum_j \rho_{ij}^2 d_i d_j + \\
& + \sum_{i<}\sum_j\sum_{i'<}\sum_{j'} \omega_{ij\cdot i'j'} aa_{ij} aa_{i'j'} + \\
& + \sum_{i<}\sum_j\sum_{i'<}\sum_{j'} \varpi_{ij\cdot i'j'} dd_{ij} dd_{i'j'} + \\
& + \sum_{i<}\sum_j\sum_{i'<}\sum_{j'} \varphi_{ij\cdot i'j'} ad_{ij} ad_{i'j'} + \\
& + \sum_{i<}\sum_j\sum_{i'<}\sum_{j'} \psi_{ij\cdot i'j'} da_{ij} da_{i'j'} + \\
& + \sum_{i<}\sum_j\sum_{i'<}\sum_{j'} (\xi_{ij\cdot i'j'} aa_{ij} dd_{i'j'} + \\
& + \zeta_{ij\cdot i'j'} dd_{ij} aa_{i'j'}) + \\
& + \sum_{i<}\sum_j\sum_{i'<}\sum_{j'} (\tau_{ij\cdot i'j'} ad_{ij} da_{i'j'} + \\
& + \zeta_{ij\cdot i'j'} da_{ij} ad_{i'j'}) + \\
& + \sum_{k}\sum_{k'<}\sum_{l'} \theta_{k\cdot k'l'} a_k ad_{k'l'} + \\
& + \sum_{l}\sum_{k'<}\sum_{l'} \vartheta_{l\cdot k'l'} a_l ad_{k'l'} + \\
& + \sum_{k}\sum_{k'<}\sum_{l'} \lambda_{k\cdot k'l'} d_k aa_{k'l'} + \\
& + \sum_{k}\sum_{k'<}\sum_{l'} \gamma_{k\cdot k'l'} d_k dd_{k'l'}
\end{aligned} \tag{6}$$

where $i, j, i', j', k, l, k', l' = 1, \ldots, n$, $\rho_{ij} = 1 - 2r_{ij}$, $\delta_{ij} = r_{ij}^2 + (1 - r_{ij})^2$, $\eta_{ij} = r_{ij}(1 - r_{ij})$,

$$\omega_{ij\cdot i'j'} = \frac{1}{4}(2\rho_{ab}\delta_{bc}\rho_{cd} - \rho_{ij}\rho_{i'j'}),$$

$$\varpi_{ij\cdot i'j'} = \frac{1}{4}(2\delta_{ab}\delta_{bc}\delta_{cd} - \delta_{ij}\delta_{i'j'}),$$

$$\varphi_{ij\cdot i'j'} = \begin{cases} \rho_{ab}\eta_{bc}\delta_{cd} & i \le i' < j \text{ or } i' \le i < j' \\ 0 & \text{otherwise} \end{cases},$$

$$\psi_{ij\cdot i'j'} = \begin{cases} \delta_{ab}\eta_{bc}\rho_{cd} & i \le i' < j \text{ or } i' \le i < j' \\ 0 & \text{otherwise} \end{cases},$$

$$\xi_{ij\cdot i'j'} = \begin{cases} \rho_{ab}\eta_{bc}\delta_{cd} - \frac{1}{4}\rho_{ij}\delta_{i'j'} & j \le i' \\ \delta_{ab}\eta_{bc}\rho_{cd} - \frac{1}{4}\rho_{ij}\delta_{i'j'} & j' \le i \\ -\frac{1}{4}\rho_{ij}\delta_{i'j'} & i < i' < j' < j \\ 2\eta_{ab}\rho_{bc}\eta_{cd} - \frac{1}{4}\rho_{ij}\delta_{i'j'} & i' < i < j < j' \\ -\frac{1}{4}\rho_{ij}\delta_{i'j'} & \text{otherwise} \end{cases},$$

$$\zeta_{ij\cdot i'j'} = \begin{cases} \delta_{ab}\eta_{bc}\rho_{cd} - \frac{1}{4}\delta_{ij}\rho_{i'j'} & j \le i' \\ \rho_{ab}\eta_{bc}\delta_{cd} - \frac{1}{4}\delta_{ij}\rho_{i'j'} & j' \le i \\ 2\eta_{ab}\rho_{bc}\eta_{cd} - \frac{1}{4}\delta_{ij}\rho_{i'j'} & i < i' < j' < j \\ -\frac{1}{4}\delta_{ij}\rho_{i'j'} & i' < i < j < j' \\ -\frac{1}{4}\delta_{ij}\rho_{i'j'} & \text{otherwise} \end{cases},$$

$$\tau_{ij\cdot i'j'} = \begin{cases} 2\eta_{ab}\rho_{bc}\eta_{cd} & i' < i \text{ and } j' < j \\ 0 & \text{otherwise} \end{cases},$$

$$\zeta_{ij\cdot i'j'} = \begin{cases} 2\eta_{ab}\rho_{bc}\eta_{cd} & i < i' \text{ and } j < j' \\ 0 & \text{otherwise} \end{cases},$$

$$\theta_{k\cdot k'l'} = \begin{cases} 2\rho_{AB}\eta_{BC} & k \le l' \\ 0 & \text{otherwise} \end{cases},$$

$$\vartheta_{l\cdot k'l'} = \begin{cases} 2\eta_{AB}\rho_{BC} & l \ge k' \\ 0 & \text{otherwise} \end{cases},$$

$$\lambda_{k\cdot k'l'} = \begin{cases} 2\eta_{AB}\rho_{BC} - \frac{1}{2}\rho_{k'l'} & k \le k' \\ 2\rho_{AB}\eta_{BC} - \frac{1}{2}\rho_{k'l'} & k \ge l' \\ -\frac{1}{2}\rho_{k'l'} & \text{otherwise} \end{cases},$$

$$\gamma_{k\cdot k'l'} = \delta_{AB}\delta_{BC} - \frac{1}{2}\delta_{k'l'},$$

[$a \ldots b \ldots c \ldots d$ are ordered as $a \ldots b \ldots c \ldots d$ on the genome; $i, j, i', j' \subset (a, b, c, d)$. Similarly, $A \ldots B \ldots C$ are ordered as $A \ldots B \ldots C$ on the genome; $k$, $k'$, $l'$ or $l$, $k'$, $l'' \subset (A, B, C)$.] under the Haldane's mapping function (Haldane, 1919). $r_{ij}$ is the recombination fraction between $Q_i$ and $Q_j$.

Additive and dominant effects at different loci and epistatic interactions between loci were first generated by drawing a standard normal distribution. Formula (6) was used to calculate total genetic variance ($V'_{\mathrm{G}}$) and variance components ($V_C$) based on these genetic effects and the above genetic map (Figure 1). To make sure different genetic variance components equal a given ratio $w_C(\sum_C w_C = 1)$, these effects were re-scaled by setting $k' = \sqrt{V'_{\mathrm{G}} w_C / V_C}$. Note that in this case, covariances between each term of genetic effect are usually over one order of magnitude less than those of variances in most cases and hence, were neglected. We chose $V_{\mathrm{G}}$ to yield the desired heritability, given that the phenotypic variance was fixed at 1 in all simulations. The genetic variance $V_{\mathrm{G}}$ was, therefore, calculated as $H^2 V_P$. $H^2$ is the heritability in the broad sense and defined as $V_{\mathrm{G}}/V_P$ while $h^2$ is heritability in the narrow sense and defined as $(V_A + V_{AA})/V_P$ (for details, see Zhu, 1997). Finally, formula (6) was used to compute again the total genetic variance ($V''_{\mathrm{G}}$) based on the above rescaled genetic effects. These genetic effects could then be readjusted again by setting $k'' = \sqrt{V_{\mathrm{G}}/V''_{\mathrm{G}}}$ and kept constant throughout selection. The genotypic value $G_k$ of individual $k$ was obtained by summing all genetic effects within and between loci (Mather & Jinks, 1982). The phenotypic value of individual $k$ was calculated as follows,

$$y_k = \mu + G_k + \varepsilon_k \tag{7}$$

where $\mu$ is the population mean; $\varepsilon_k$ was obtained by generating a pseudo-random normal deviate with zero mean and known variance $(1 - H^2)V_P$.

*Effects of inaccurate parameter estimates on the genetic response to MAS*

The parameters used to characterize QTL are the effects of QTL such as additive, dominance, epistasis, and QTL location relative to marker bracket. Simulations were conducted to study the genetic consequence of overestimation and underestimation of QTL effects. Overestimation of QTL effects accounted for phenotypic variation of 10, 20 and 30%, respectively, and underestimation for 10, 20 and 30%, respectively. Simulations were also undertaken to study the genetic consequence of the putative QTL location departing from the true QTL location with the distance of 1.0, 5.0 and 10.0 cM, respectively. Moreover, one to three nonexistent QTLs and two to six 'ghost' epistases accounting for 10–30% of phenotypic variation were added to the genome randomly to study the effect of false positive QTL on efficiency of MAS.

*Selection*

In each generation of sample size $N$, the top 30% of individuals were selected (i.e., composing of a parental pool) and then mated at random to produce $N$ offspring. Selections were performed on phenotypic value ($P$), breeding value ($\hat{B}$), and total genotypic value ($\hat{G}$), respectively. Selections were also performed on additive effects ($\hat{A}$), additive and dominant effects ($\hat{A} + \hat{D}$) to study the genetic consequence of MAS when ignoring epistatic effects. Calculation of the four selection indices ($\hat{B}$, $\hat{G}$, $\hat{A}$ and $\hat{A} + \hat{D}$) was described in Methods section, using QTL map information obtained in previous experiments. The corresponding cumulative genetic responses were calculated as

$$\Delta G_{(t)} = \frac{\overline{G}_{(t)} - \overline{G}_{(0)}}{\sigma_{G_{(0)}}} \tag{8}$$

*Table 1.* Cumulative responses to selection based on phenotypic value, additive effect, additive and dominant effect, genotypic value, and breeding value with $H^2 = 0.4$ and $h^2 = 0.2$ for different population sizes

| Generation | $N = 100$ | | | | | $N = 500$ | | | | | $N = 1000$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $P$ | $\hat{A}$ | $\hat{A}+\hat{D}$ | $\hat{G}$ | $\hat{B}$ | $P$ | $\hat{A}$ | $\hat{A}+\hat{D}$ | $\hat{G}$ | $\hat{B}$ | $P$ | $\hat{A}$ | $\hat{A}+\hat{D}$ | $\hat{G}$ | $\hat{B}$ |
| 1 | 0.38 | 0.67 | 0.56 | 0.62 | 0.69 | 0.38 | 0.68 | 0.58 | 0.62 | 0.69 | 0.38 | 0.68 | 0.58 | 0.62 | 0.69 |
| 2 | 0.71 | 1.14 | 1.03 | 1.15 | 1.24 | 0.73 | 1.16 | 1.04 | 1.14 | 1.25 | 0.73 | 1.18 | 1.04 | 1.14 | 1.25 |
| 3 | 1.02 | 1.48 | 1.37 | 1.57 | 1.72 | 1.05 | 1.50 | 1.40 | 1.58 | 1.76 | 1.05 | 1.52 | 1.41 | 1.58 | 1.78 |
| 4 | 1.31 | 1.70 | 1.65 | 1.88 | 1.93 | 1.33 | 1.73 | 1.70 | 1.91 | 1.95 | 1.33 | 1.75 | 1.71 | 1.92 | 1.96 |
| 5 | 1.53 | 1.88 | 1.78 | 2.04 | 2.04 | 1.57 | 1.91 | 1.86 | 2.08 | 2.04 | 1.59 | 1.93 | 1.87 | 2.09 | 2.04 |
| 6 | 1.73 | 1.96 | 1.87 | 2.15 | 2.16 | 1.77 | 1.98 | 1.93 | 2.18 | 2.16 | 1.78 | 1.99 | 1.95 | 2.20 | 2.17 |
| 7 | 1.88 | 2.00 | 1.93 | 2.20 | 2.26 | 1.92 | 2.01 | 1.99 | 2.24 | 2.31 | 1.94 | 2.02 | 2.00 | 2.27 | 2.34 |
| 8 | 1.97 | 2.03 | 1.98 | 2.24 | 2.33 | 2.04 | 2.03 | 2.02 | 2.28 | 2.40 | 2.05 | 2.03 | 2.04 | 2.30 | 2.42 |
| 9 | 2.04 | 2.03 | 2.00 | 2.26 | 2.37 | 2.12 | 2.04 | 2.05 | 2.30 | 2.44 | 2.13 | 2.04 | 2.07 | 2.32 | 2.46 |
| 10 | 2.10 | 2.04 | 2.03 | 2.28 | 2.40 | 2.17 | 2.05 | 2.06 | 2.32 | 2.46 | 2.19 | 2.05 | 2.08 | 2.33 | 2.47 |
| ⋮ | | | ⋮ | | | | | | ⋮ | | | | | | ⋮ |
| 20 | 2.30 | 2.06 | 2.07 | 2.35 | 2.47 | 2.35 | 2.09 | 2.09 | 2.36 | 2.51 | 2.36 | 2.10 | 2.09 | 2.37 | 2.52 |

Note: Results from generations 11 to 19 were not presented in order to save space.
$P$, phenotypic value; $\hat{A}$, additive effect; $\hat{A}+\hat{D}$, additive and dominant effect; $\hat{G}$, additive and dominant and epistatic effect, that is, genotypic value; $\hat{B}$, breeding value; $N$, population size.
$V_A{:}V_D{:}V_{AA}{:}V_{AD}{:}V_{DA}{:}V_{DD} = 4{:}2{:}2{:}1{:}1{:}2$.

where $\overline{G}_{(t)}$ is the genetic mean of the population at generation $t$ ($t = 0$, standing for the initial population $F_2$), and $\sigma_{G_{(0)}}$ is the genetic standard deviation in the $F_2$.

Simulations were run under two levels of heritability (0.4 and 0.8), and three population sizes (100, 500, 1000). MAS was undertaken for 20 generations in total. Simulations were replicated 200 times for each case and the mean results of 200 simulations were presented.

## Results

### *Genetic response to MAS with epistasis*

#### *Genetic response*

Results show, as expected, that the maximum ratio between response to MAS and phenotypic selection records at early generations and declines rapidly thereafter; the greater response is achieved with larger population size and higher heritability (Tables 1 and 2). However, neglecting epistatic effects underlying selection results in considerable loss in genetic response to MAS, especially at later generations. For example, with heritability in the broad sense 0.4 and population size 500, the responses to MAS based only on additive effects and additive and dominant effects are decreased after 20 generations by up to 17%, as compared with MAS on breeding values. When including epistasis, selection also yields longer persistence response than that based exclusively on additive or additive and dominance. It is expected that epistatic genetic variation is not susceptible to be exhausted during artificial selection and thus considerable variation remains after so many generations of selection, which will, no doubt, help to increase long-term selection efficiency.

#### *Breeding value as a measure for selection*

Note that the breeding value in the context is referred to the 'additive genotype', and variation in the breeding values ascribed to the 'additive effects' of genes and their additive × additive interactions. It is shown in Tables 1 and 2 that breeding value is a more appropriate measure used in selection than genotypic value and other selection indices in most cases. Selection based on breeding value usually produces marginally extra genetic gains in comparison to that on genotypic value, and are more pronounced at early generations. Also, selection based on breeding value tends to have smaller standard error than other methods in most cases (results not shown). Because of this, the remaining results in the paper are reported only for MAS based on breeding value.

*Table 2.* Cumulative responses to selection based on phenotypic value, additive effect, additive and dominant effect, genotypic value, and breeding value with $H^2 = 0.8$ and $h^2 = 0.4$ for different population sizes

| Generation | $N = 100$ | | | | | $N = 500$ | | | | | $N = 1000$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $P$ | $\hat{A}$ | $\hat{A}+\hat{D}$ | $\hat{G}$ | $\hat{B}$ | $P$ | $\hat{A}$ | $\hat{A}+\hat{D}$ | $\hat{G}$ | $\hat{B}$ | $P$ | $\hat{A}$ | $\hat{A}+\hat{D}$ | $\hat{G}$ | $\hat{B}$ |
| 1 | 0.56 | 0.68 | 0.57 | 0.64 | 0.70 | 0.56 | 0.69 | 0.59 | 0.63 | 0.70 | 0.56 | 0.69 | 0.59 | 0.63 | 0.71 |
| 2 | 1.06 | 1.16 | 1.06 | 1.18 | 1.28 | 1.06 | 1.20 | 1.07 | 1.16 | 1.28 | 1.06 | 1.21 | 1.07 | 1.17 | 1.28 |
| 3 | 1.47 | 1.54 | 1.44 | 1.62 | 1.82 | 1.48 | 1.57 | 1.46 | 1.63 | 1.85 | 1.48 | 1.57 | 1.47 | 1.64 | 1.86 |
| 4 | 1.78 | 1.80 | 1.76 | 1.93 | 2.04 | 1.81 | 1.83 | 1.78 | 1.96 | 2.03 | 1.80 | 1.83 | 1.79 | 1.97 | 2.04 |
| 5 | 1.99 | 1.96 | 1.93 | 2.10 | 2.13 | 2.02 | 1.99 | 1.96 | 2.13 | 2.11 | 2.02 | 1.99 | 1.98 | 2.13 | 2.12 |
| 6 | 2.12 | 2.02 | 2.01 | 2.21 | 2.24 | 2.15 | 2.02 | 2.04 | 2.23 | 2.24 | 2.15 | 2.03 | 2.05 | 2.23 | 2.26 |
| 7 | 2.19 | 2.03 | 2.05 | 2.26 | 2.35 | 2.23 | 2.04 | 2.07 | 2.28 | 2.40 | 2.23 | 2.04 | 2.08 | 2.29 | 2.43 |
| 8 | 2.23 | 2.04 | 2.07 | 2.30 | 2.41 | 2.28 | 2.05 | 2.09 | 2.32 | 2.46 | 2.28 | 2.05 | 2.10 | 2.33 | 2.48 |
| 9 | 2.27 | 2.05 | 2.08 | 2.32 | 2.45 | 2.31 | 2.05 | 2.10 | 2.33 | 2.48 | 2.32 | 2.06 | 2.11 | 2.35 | 2.49 |
| 10 | 2.30 | 2.04 | 2.09 | 2.33 | 2.46 | 2.33 | 2.06 | 2.10 | 2.35 | 2.50 | 2.34 | 2.07 | 2.11 | 2.36 | 2.50 |
| ⋮ | | | ⋮ | | | | | | ⋮ | | | | | | ⋮ |
| 20 | 2.37 | 2.05 | 2.11 | 2.37 | 2.49 | 2.37 | 2.09 | 2.11 | 2.36 | 2.52 | 2.37 | 2.11 | 2.11 | 2.37 | 2.53 |

*Genetic variance configurations*

The effects of various genetic variance configurations on the short- and long-term selection responses are investigated (Figure 2). It is observed that MAS yields a higher selection response compared with phenotypic selection at least at early generations in all genetic variance configurations studied. However, the relative efficiency between MAS and phenotypic selection and aggregated gain obtained after 20 generations are quite different among different genetic variance configurations even when the total genetic variance is fixed. This is also observed in phenotypic selection.

Higher response to MAS is achieved when additive and additive × additive variances predominate in the total genetic variance. However, MAS gives lower response and becomes less effective at later generations when dominance and its epistatic interactions account for major part of the total genetic variation. It is apparent that overall genetic response varies proportionately to the amount of additive and additive × additive genetic variations existing in breeding population. On the other hand, the response to MAS is more sustained and effective when epistatic variance constitutes the major part of the total genetic variance. For instance, at generation six, MAS could have reached a genotypic value not achieved by phenotypic selection for 40 or more generations (result not shown). This may be the reason why the surprisingly high efficiency of MAS still holds in the long term. For traits controlled largely by non-allelic gene interactions, MAS not only substantially increases the efficiency of selection but, most importantly, realizes the major part of cumulative response in short term. MAS is, however, less persistent but more effective at early generations when epistatic effect accounts for minor genetic variation involved in the selected trait. In this situation, epistasis is negligible and selection solely based on additive or additive and dominance model consequently produces similar results as have been shown by many researchers (e.g., Zhang & Smith, 1992, 1993).

*Effects of inaccurate estimates on the genetic response to MAS*

In practice, the overall efficiency of MAS should be a combination of the efficiency of QTL detection and marker-based selection. Different scenarios to study the sensitivity of genetic response to errors in QTL mapping are investigated (Figure 3–5).

*QTL position*

The larger the distance deviated from the true QTL position, the lower the efficiency of MAS relative to phenotypic selection. A QTL-marker bracket was used to infer the transmission of parental QTL allele during selection in our research. As a result of this, when the number of putative QTLs, which were mapped into a neighboring marker bracket rather than their real marker bracket, increases, the efficiency of MAS

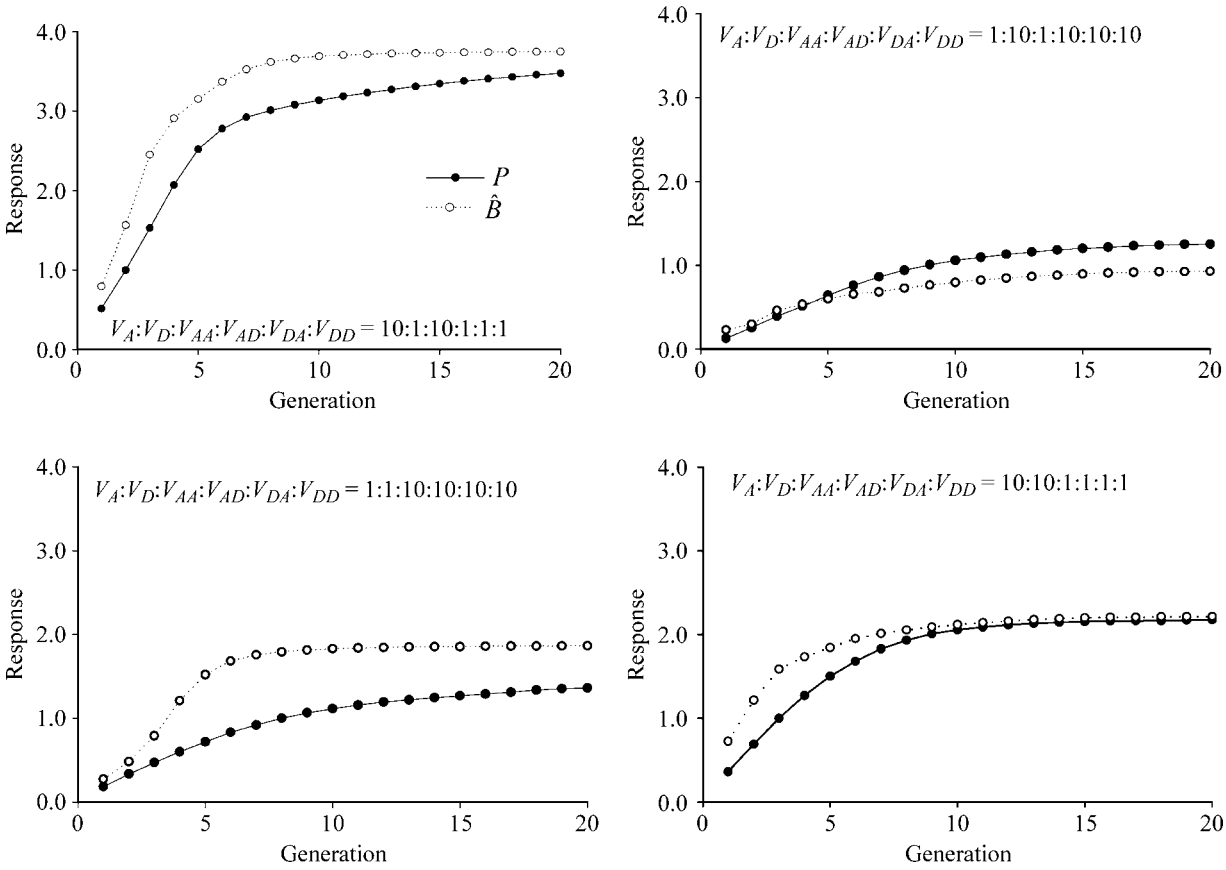


*Figure 2.* Cumulative responses to phenotypic and marker-assisted selection under various genetic variance configurations. Different variance configurations of each class are given in plot. *P*, phenotypic selection; $\hat{B}$, marker-assisted selection based on breeding value. $N = 500$, $H^2 = 0.4$.

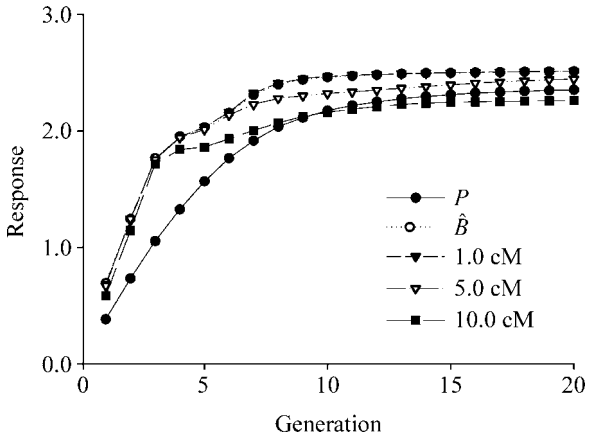


*Figure 3.* Effect of incorrect position of QTLs on cumulative genetic response. The distance deviating from the true QTL position of each class is give in plot. $N = 500$, $H^2 = 0.4$, $h^2 = 0.2$, $V_A$:$V_D$:$V_{AA}$:$V_{AD}$:$V_{DA}$:$V_{DD}$ = 4:2:2:1:1:2.

drops rapidly. However, the difference in genetic gain with MAS is negligible when the postulated position of QTL is less than 5.0 cM away from the true QTL position. In such a situation, most postulated QTLs still reside within their real marker bracket. This suggests that the inaccuracies in estimating QTL position within its real marker bracket have little impact on genetic response to MAS.

*QTL effect*

Lower genetic response to MAS is observed when inflating or deflating QTL effects. A reduction in the aggregated genetic gains in MAS is a function of the proportion of biased estimated variations. The impact on the genetic response to underestimation of QTL effect was less substantial than for overestimation. However, even when the larger overestimated or un-

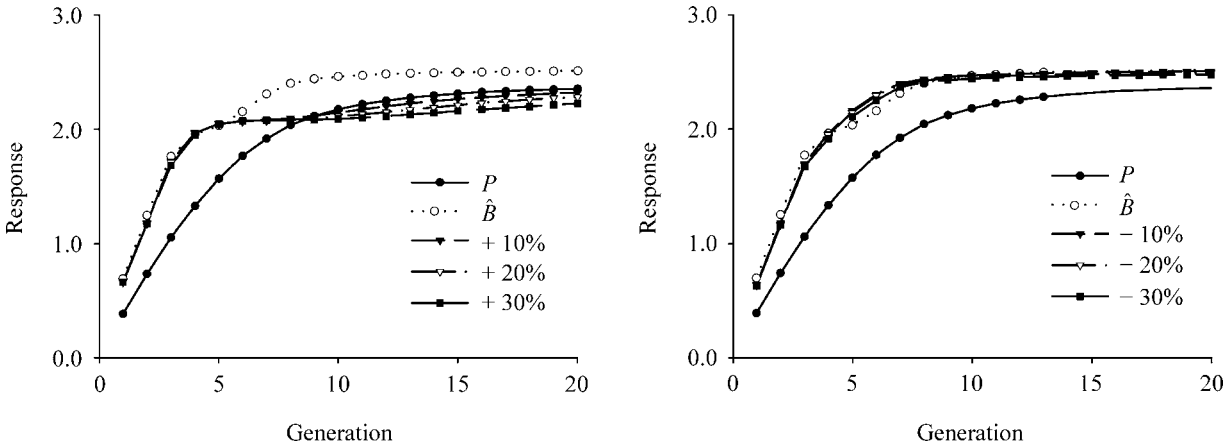


*Figure 4.* Effect of incorrect effect of QTLs on cumulative genetic response. The percentage of phenotypic variation accounted biased by these QTLs of each case is given in plot, +, overestimation of QTLs effect; −, underestimation of QTLs effect. $N = 500$, $H^2 = 0.4$, $h^2 = 0.2$, $V_A{:}V_D{:}V_{AA}{:}V_{AD}{:}V_{DA}{:}V_{DD} = 4{:}2{:}2{:}1{:}1{:}2$

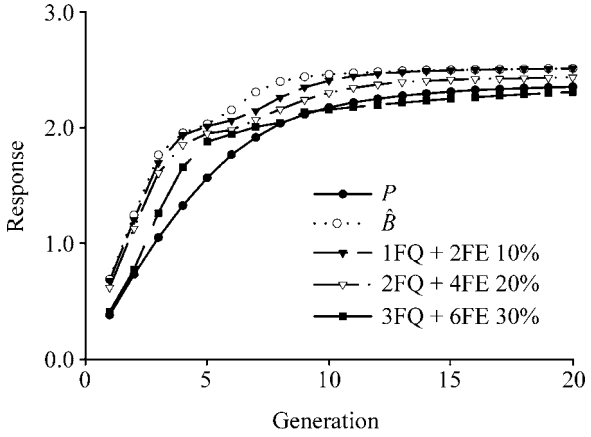


*Figure 5.* Effect of false positive QTLs on cumulative genetic response. The number of false positive QTLs and epistases, and the percentage of phenotypic variation accounted by these QTLs of each case are given in plot. FQ, false positive QTL; FE, false positive epistasis. $N = 500$, $H^2 = 0.4$, $h^2 = 0.2$, $V_A{:}V_D{:}V_{AA}{:}V_{AD}{:}V_{DA}{:}V_{DD} = 4{:}2{:}2{:}1{:}1{:}2$.

derestimated variation explained by those QTLs were involved in selection, MAS still gives a higher selection response as compared with phenotypic selection especially in the short term. It seems that selection is reasonably robust to errors in estimating QTL effects in our research.

*False positive QTL*

Genetic response to MAS declines dramatically with an increase in the number of 'ghost' QTLs and the magnitude of variations explained by those QTLs. This differs from the results of Berloo and Stam (1998) in which introduction of false positive QTLs did not significantly affect the MAS selection results. It should be noted however that Berloo and Stam (1998) focused on selecting superior genotypes only for one generation rather than improving population mean for several consecutive generations. Therefore, for a given breeding scheme, breeders should be cautious to include those QTLs with lower probability threshold into the selection index.

## Discussion

Recent studies of MAS have tended to focus on the use of multiple regression and mixed model approaches to investigate the efficiency of MAS (e.g., animal model for BLUP, Fernando & Grossman, 1989). These methods appear to have failed to fully utilize QTL map information that can be used to improve the performance of MAS. Whittaker et al. (1995) proposed to take into account the position of markers when estimating breeding value. However, these authors showed that this only slightly improves the efficiency of MAS. Extending preliminary analysis of Whittaker et al. (1995), we proposed a method for MAS based on QTLs with epistasis. This allows us to gain more insight into the characterization and potential of MAS for the improvement of a quantitative trait. However, the perspective here is quite different in that we evaluated the efficiency of MAS assuming that we had obtained knowledge about QTLs affecting the trait studied, then investigated the effects of inaccurate estimates of QTL detection on genetic response to MAS. It should be also noted that we concentrated more on identifying QTL genotype rather than estimating

QTL allelic effects associated with each marker allele. To estimate $p(z_i|h_i)$ over successive generations, we also employed an approximate method provided by Whittaker et al. (1995) but with some modifications: (i) estimating the frequencies of marker haplotype by maximum likelihood approach, and (ii) deriving the frequencies of QTL allele from the parental pool directly (see Appendix). Our method usually worked better in most cases (results not shown) and is easily extended to track changes in QTL alleles in pedigree breeding for self-fertilizing crops.

Complex quantitative traits consist of genetic effects more than simple additive and dominant effects. The most striking feature of our research is the use of a more sophisticated biological model by including epistatic effects, which gives a greater ability to optimize long-term selection response and provides a better understanding of genetic base underlying improvement of quantitative trait. We demonstrated that, if there are epistatic effects, MAS generally bears longer persistence response than those reported in previous studies (e.g., Zhang & Smith, 1992, 1993; Gimelfarb & Lande, 1994), in which authors used an additive genetic model. Ignoring epistatic effects, however, leads to considerable loss in genetic gain provided by MAS. It is indicated that epistasis is an important component of genetic basis on which to base selection merits serious consideration in breeding. On the other hand, it was shown in the present study that in addition to the population size and heritability of the trait under selection, genetic variance configurations play an important role in determining the efficiency in both short and long terms. Firstly, cumulative genetic response to selection varies proportionately to the amount of additive and additive $\times$ additive genetic variations involved in the selected trait. Secondly, the larger epistatic genetic variation under selection, the higher efficiency of MAS retains in the long term. This implies that different types of genetic effects contributing to expression of the selected trait produce notable effects not only on efficiency of MAS and thus on the gain per unit cost, but also on the selection limit reached.

In addition, we proposed breeding value as a measure for identifying desired individuals on which MAS bases. Here, we extended the original connotation of breeding value by including epistasis of additive $\times$ additive (Falconer, 1981). In the results, selection on breeding value not only obtained higher genetic response, but also tended to have smaller standard error than other methods in most cases. To our knowledge, additive and additive $\times$ additive epistasis can be gradually fixed under directional selection. It was observed that the favorable alleles at QTLs tend to be faster toward fixation under MAS than under phenotypic selection, in which quantitative genetic variation was assumed to be under control for simple additive effects (Zhang & Smith, 1992; Hospital et al., 1997). We might envisage then that selection using breeding values, which usually works better than other methods, might also fix favorable binate alleles among QTLs more frequently. However, the dynamics of fixation of binate QTLs was not investigated in the present paper and requires further investigation.

In our study, we formally treated QTL detection and QTL selection as each independent phase. In reality, the overall efficiency of MAS will be a combination of the efficiency of QTL detection and marker-based selection, and hence probably below the sole efficiency of MAS investigated in the present paper. However, the responses achieved are reasonable in consideration of complex epistasis involved in selection. We pointed out that errors in QTL detection cause a distinct reduction in response to MAS in most cases, except when the location of QTL is estimated inaccurately but still resides within its real marker bracket. It is thus concluded that verification of putative QTL, which mainly includes two aspects: (i) magnitude of QTL effect, (ii) and accurate chromosome map location, is imperative to realize the potentials of MAS. Beavis (1994) showed that the effects associated with QTLs are more overestimated when the population size is small, and that this bias cannot be neglected even for a population of 500 individuals. A related problem was also investigated by Wang (2000), who concluded that it is common to exaggerate the effects of those QTLs detected in a single experiment in practice. As such, it is necessary to shrink the estimate for optimizing the long-term gains. On the other hand, false positive QTLs have a great influence on genetic response to MAS. Higher probability thresholds for declaring a QTL effect significantly reduce the chances of spurious QTLs being reported, but also reduce chances of detecting QTLs with small effects (Wang, 2000). It should be pointed out that QTLs with large effects are easily manipulated by traditional breeding practices, and may already be at high frequencies or fixed in many breeding lines and populations (Zhang & Smith, 1992). Therefore, MAS may be more productive if one places greater emphasis on those QTLs with small effects rather than a few QTLs with large effects. It suggests that new breeding strategies based on QTL evaluation with a large

population will be essential to realize the potentials of MAS.

Our approach aims at utilizing QTL map information obtained at previous generations. In this method, only the markers bracketing QTL, which account for phenotypic variation of the trait studied, are chosen with phenotypic evaluation and genotyped when trait is submitted to selection. The cost of genotyping is hence reduced. A prerequisite is that genetic architectures of QTLs affecting the improved trait are available (i.e., breeding population should have been screened for QTLs and markers) and hence QTLs location and effects are estimated. However, the problem related to mapping QTL with epistatic effects is a complicated issue and will be found elsewhere (Zhu & Weir 1998; Wang et al., 1999; Gao, 2001).

To maximize the benefits from marker-based procedures, a proper strategy in the implementation of MAS based on our conclusions, also mentioned by Hospital et al. (1997), should receive due consideration. It consists of two phases: (i) large-scale screening of the breeding population to determine the precise location of individual QTL and to quantify gene action and interaction among genes. In this phase, other results from such fine mapping and comparative mapping should be integrated; and (ii) performing selection to increase the frequencies of favorable alleles at the QTLs with as small a population as possible ($\sim$100).

**Acknowledgements**

We greatly thank J.C. Whittaker for his helpful discussions and comments of part work of this paper. This research was supported in part by the National Natural Science Foundation of China.

**Appendix**

Whittaker et al. (1995) described an approximation for estimating $p(z_i|h_i)$ in subsequent generations by reference to marker genotypes. The probability of the $i$th QTL allele ($Q_i$) given the marker haplotype $M_{i_l}M_{i_r}$ is

$$p_t(Q_i|M_{i_l}M_{i_r}) \approx \frac{1}{p_t(M_{i_l}M_{i_r})} \times \times \left( p_t(M_{i_l})p_t(Q_i)p_t(M_{i_r}) + p_t(M_{i_l})L_t(i) + + L_t(i)p_t(M_{i_r}) + \frac{L_t(i)^2}{p_t(Q_i)} \right) \quad (A.1)$$

where the subscript $t$ means that these are the probabilities in the $t$th generation; $p_t(M_{i_l})$ and $p_t(M_{i_r})$ are the frequencies of marker alleles $M_{i_l}$ and $M_{i_r}$ (known), respectively; $p_t(Q_i)$ is the frequency of QTL allele $Q_i$; $\hat{L}_t(i)$ is the linkage disequilibrium between the $i$th QTL locus and either the left flanking marker $M_{i_l}$ or the right flanking marker $M_{i_r}$,

$$\hat{L}_t(i) = \sqrt{p_t(Q_i)p_t(q_i)\hat{D}_t(i)} \quad (A.2)$$

where $\hat{D}_t(i)$ is the linkage disequilibrium between the flanking markers $M_{i_l}$ and $M_{i_r}$,

$$\hat{D}_t(i) = p_t(M_{i_l}M_{i_r})p(m_{i_l}m_{i_r}) - - p_t(M_{i_l}m_{i_r})p(M_{i_l}m_{i_r}) \quad (A.3)$$

where $p_t(M_{i_l}M_{i_r})$, $p(m_{i_l}m_{i_r})$, $p_t(M_{i_l}m_{i_r})$ and $p(m_{i_l}M_{i_r})$ are the frequencies of marker haplotypes $M_{i_l}M_{i_r}$, $m_{i_l}m_{i_r}$, $M_{i_l}m_{i_r}$ and $m_{i_l}M_{i_r}$, respectively (for details, see Whittaker et al., 1995).

To use the above expressions we need to know $p_t(Q_i)$ and $p_t(M_{i_l}M_{i_r})$. We can estimate $p_t(Q_i)$ from the parental pool in previous generation $t-1$. Suppose there are $N$ individuals composing of the parental pool and $n$ QTLs affecting the selected trait. Each individual from the parental pool has possible $3^n$ genotypes and the probability of each genotype can be calculated from formula (3) in text. Denoting $p_t(Q_i)_k$ the probability that parent $k(k = 1, \ldots, N)$ possesses $Q_i$ in generation $t$, we then obtain

$$p_t(Q_i)_k = \sum_{j=1}^{3^n} p_{t-1}(\mathbf{z}_j|y_k, \mathbf{h}_k)\vartheta \quad (A.4)$$

where $y_k$ and $\mathbf{h}_k$ are the phenotypic value and marker genotype of parent $k$ in generation $t-1$, respectively; $\mathbf{z}_j$ are the $j$th possible QTL genotypes, $\mathbf{z}_j = (z_1, z_2, \ldots, z_n)_j$, and their meanings are defined the same as Methods section in text; the coefficient $\vartheta$ is 1, 0.5 and 0 for genotype $Q_iQ_i$, $Q_iq_i$ and $q_iq_i$, respectively. Thus,

$$p_t(Q_i) = 1/N \sum_k p_t(Q_i)_k \quad (A.5)$$

Since the individuals from the parental pool are employed to mating at random to produce the offspring population, genotypic frequencies are the products of gametic frequencies and the maximum likelihood approach may be employed (Weir, 1996). The log-likelihood for multinomial marker data

can be expressed as

$$\begin{aligned}
&\ln L(p(M_{i_l}),\, p(M_{i_r}),\, p(M_{i_l}M_{i_r})) = \\
&\quad \text{Constant} + (2n_{M_{i_l}M_{i_l}M_{i_r}M_{i_r}} + n_{M_{i_l}M_{i_l}M_{i_r}m_{i_r}} \\
&\quad + n_{M_{i_l}m_{i_l}M_{i_r}M_{i_r}})\ln(p(M_{i_l}M_{i_r})) + \\
&\quad + (2n_{M_{i_l}M_{i_l}m_{i_r}m_{i_r}} + n_{M_{i_l}M_{i_l}M_{i_r}m_{i_r}} + \\
&\quad + n_{M_{i_l}m_{i_l}m_{i_r}m_{i_r}})\ln(p(M_{i_l}) - p(M_{i_l}M_{i_r})) + \\
&\quad + (2n_{m_{i_l}m_{i_l}M_{i_r}M_{i_r}} + n_{M_{i_l}m_{i_l}M_{i_r}M_{i_r}} + \\
&\quad + n_{m_{i_l}m_{i_l}M_{i_r}m_{i_r}})\ln(p(M_{i_r}) - p(M_{i_l}M_{i_r})) + \\
&\quad + (2n_{m_{i_l}m_{i_l}m_{i_r}m_{i_r}} + n_{M_{i_l}m_{i_l}m_{i_r}m_{i_r}} + \\
&\quad + n_{m_{i_l}m_{i_l}M_{i_r}m_{i_r}})\ln(1 - p(M_{i_l}) - p(M_{i_r}) + \\
&\quad + p(M_{i_l}M_{i_r})) + n_{M_{i_l}m_{i_l}M_{i_r}m_{i_r}} \\
&\quad \ln[p(M_{i_l}M_{i_r})(1 - p(M_{i_l}) - p(M_{i_r}) + \\
&\quad + p(M_{i_l}M_{i_r})) + (p(M_{i_l}) - \\
&\quad - p(M_{i_l}M_{i_r}))(p(M_{i_r}) - p(M_{i_l}M_{i_r}))] \qquad \text{(A.6)}
\end{aligned}$$

where $p_t(M_{i_r})$ and $p_t(M_{i_l}M_{i_r})$ have the same meanings as in the formulae (A.1) and (A.3), respectively; $n\ldots$ are the corresponding count of the marker genotype $M_{i_l}M_{i_l}M_{i_r}M_{i_r}$ and so on, respectively. Numerical methods are used to solve the above cubic equation, each of the three roots checked to make sure that all gametic frequencies are between zero and appropriate allele frequencies, and then the three likelihoods are computed. The valid solution that maximizes the likelihood is the required maximum likelihood estimator for $p(M_{i_l}M_{i_r})$.

As $p_1(Q_i)$ is known in $F_2$, this value and marker frequencies such as $p_1(M_{i_l}M_{i_r})$ can be used to calculate $p_1(z_i|h_i)$ with formula (A.1), in turn the $p_1(\mathbf{z}|y_k, \mathbf{h})$ for each individual from the parental pool used to calculate $p_2(Q_i)$ with formula (A.4). Continuing the process allows us to calculate the conditional probabilities of QTL genotypes on marker genotypes in advanced generations.

## References


Beavis, W.D., 1994. The power and deceit of QTL mapping experiments: lessons from comparative QTL studies. Proc. Annu. Corn Sorghum Res. Ind. Conf. 49: 250–265.

Berloo, R.V. & P. Stam, 1998. Marker-assisted selection in autoamous RIL populations: a simulation study. Theor. Appl. Genet. 96: 147-154.

Cao, G.Q., J. Zhu, C.X. He, Y.M. Gao & J.Q. Yan, 2001. Impacts of epistasis and QTL × environment interaction for developmental behavior of plant height in rice (*Oryza sativa* L.). Theor. Appl. Genet. 103: 153–160.

Falconer, D.S., 1981. Introduction to Quantitative Genetics. Longman Press, New York, 2nd edn.

Fernando, R.L. & M. Grossman, 1989. Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21: 467–477.

Fu, Y.B. & K. Ritland, 1996. Marker-based inferences about epistasis for genes influencing inbreeding depression. Genetics 144: 339–348.

Gao, Y.M., 2001. Studies on approaches for mapping QTLs with complicated epistasis and their interactions with environment and methods for heterosis prediction. PhD Dissertation, Zhejiang University, Hangzhou, China.

Gimelfarb, A. & R. Lande, 1994. Simulation of marker-assisted selection in hybrid populations. Genet. Res. 63: 39–47.

Haldane, J.B.S., 1919. The combination of linkage values and the calculation of distance between the loci of linked factors. J. Genet. 8: 299–309.

Hospital, F., L. Moreau, F. Lacoudre, A. Charcosset & A. Gallais, 1997. More on the efficiency of marker-assisted selection. Theor. Appl. Genet. 95: 1181–1189.

Knapp, S.J., 1998. Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci. 38, 1164–1174.

Lande, R. & R. Thompson, 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124: 743–756.

Li, Z.K., S.R.M. Pinson, W.D. Park, A.H. Paterson & J.W. Stansel, 1997. Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145: 453–465.

Luo, Z.W., R. Thompson & J.A. Woolliams, 1997. A population genetics model of marker-assisted selection. Genetics 146: 1173–1183.

Mather, K. & J.L. Jinks, 1982. Biometrical Genetics. Chapman and Hall, London, 3rd edn.

Moreau, L., A. Charcrosset, F. Hospital & A. Gallais, 1998. Marker-assisted selection efficiency in populations of finite size. Genetics 148: 1353–1365.

Ollivier, L., 1998. The accuracy of marker-assisted selection for quantitative traits within populations in linkage equilibrium. Genetics 148: 1367–1372.

Pooni, H.S., D.J. Coombs & P.S. Jinks, 1987. Detection of epistasis and linkage of interacting genes in the presence of reciprocal differences. Heredity 58: 257–266.

Ruane, J. & J.J. Colleau, 1995. Marker-assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res. 66: 71–83.

Spelman, R. & H. Bovenhuis, 1998. Genetic response from marker-assisted selection in an outbred population for differing marker bracket sizes and with two identified quantitative trait loci. Genetics 148: 1389–1396.

Wang, D.L., J. Zhu, Z.K. Li & A.H. Paterson, 1999. Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor. Appl. Genet. 99: 1255–1264.

Wang, S.C., 2000. Simulation study on the methods for mapping quantitative trait loci in inbred line cross. PhD Dissertation, Zhejiang University, Hangzhou, China.

Weir, B.S., 1996. Genetic Data Analysis. Sinauer Associates Inc., USA, 2nd edn.

Whittaker, J.C., R.N. Curnow, C.S. Haley & R. Thompson, 1995. Using marker-maps in marker-assisted selection. Genet. Res. 66: 255–265.

Xie, C. & S. Xu, 1998. Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. Heredity 80: 489–498.

Yu, S.B., J.X. Li, C.G. Xu, Y.F. Tan, Y.J. Gao, X.H. Li, Q. Zhang & M.A. Saghaimaroof, 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. USA 94: 9226–9231.

Zhang, W. & C. Smith, 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theor. Appl. Genet. 83: 813–820.

Zhang, W. & C. Smith, 1993. Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. Theor. Appl. Genet. 86: 492–496.

Zhu, J., 1997. Analysis Methods for Genetic Models. Agricultural Publishing House of China, Beijing.

Zhu, J. & B.S. Weir, 1998. Mixed-model approaches for genetic analysis of quantitative traits, pp. 321–330 in Advanced Topics in Biomathematics, edited by L.S. Chen, S.G. Ruan & J. Zhu. World Scientific Publishing Co., Singapore.