Theor Appl Genet (2004) 109:370–376
DOI 10.1007/s00122-004-1636-6

ORIGINAL PAPER

Pengyuan Liu · Jun Zhu · Yan Lu

Marker-assisted selection in segregating generations of self-fertilizing crops

Received: 1 January 2004 / Accepted: 15 February 2004 / Published online: 7 April 2004

Abstract Computer simulations were used to study the efficiency of MAS for breeding self-fertilizing crops, based on a general model including additive, dominance and epistasis. It was shown that MAS not only gave larger genetic responses but also dramatically increased the frequencies of superior genotypes as compared with phenotypic selection. However, the advantages of MAS over phenotypic selection were considerably reduced when conducting selection in later generations. A modified method combining MAS in early generations with phenotypic selection in later generations was thus proposed from an efficiency standpoint. We also proposed a potential index to measure the probability of an individual showing superior genotypes under selfing. It was apparent that more superior genotypes could be derived from selection by using the potential index than by using other methods. The implications of these findings for plant breeding are discussed.

Introduction

Breeding strategies for self-fertilizing crops are often aimed at obtaining pure homozygous lines with superior genotypes. However, the likelihood of selecting superior genotypes is not high for low to moderately heritable traits (Johnson 1989). Plant breeders cope with this problem by producing and testing progeny from numerous crosses with low selection intensities, replicated testing, testing advanced generations, and recurrent selection (Hallauer and Miranda 1981). Marker-assisted selection (MAS) has emerged as an efficient strategy to increase accuracy of selection (Lande and Thompson 1990; Dudley 1993; Bouchez et al. 2002). MAS should be most effective in early generations of selection among progeny from crosses between inbred lines (Lande 1992; Zehr et al. 1992; Stromberg et al. 1994).

Recently, numerous papers have been published to evaluate MAS (e.g. Lande and Thompson 1990; Zhang and Smith 1992, 1993; Gimelfarb and Lande 1994; Whittaker et al. 1995; Hospital et al. 1997, 2000; Luo et al. 1997; Berloo and Stam 1998; Knapp 1998; Moreau et al. 1998; Ollivier 1998; Spelman and Bovenhuis 1998; Xie and Xu 1998; Charmet et al. 1999). However, all MAS efforts, except the two studies of Berloo and Stam (1998) and Charmet et al. (1999), have mainly focused on improving the population mean rather than on selecting homozygous superior genotypes. Even in the two exceptional cases (Berloo and Stam 1998; Charmet et al. 1999), the researchers investigated a specific strategy based on intercrosses of pairs of recombinant inbred lines for one or two generations only. Knapp (1998) presented an explicit expression of the probability of selecting one or more superior genotypes by MAS. This parameter is a function of heritability, the proportion of additive variance associated with markers, and the genotypic superiority. Knapp (1998) showed, in theory, that a breeder using phenotypic selection must test 1.0–16.7 times more progeny than a breeder using MAS to be assured of selecting one or more superior genotypes. Recently, Bouchez et al. (2002) reported marker-assisted introgression of favorable alleles at three QTLs for earliness and grain yield among maize elite lines. The study proved successful to manipulate genetic markers at the genotypic level to increase the efficiency of marker-assisted introgression.

In our recent study, we proposed a method for MAS based on QTLs with complicated epitasis (Liu et al. 2003). However, this study was also focused solely on population improvement, which is more relevant to open-pollinated outcrossing plants. The purpose of the present study was to investigate, by stochastic simulations, the effectiveness of MAS in comparison to the conventional selection method for breeding self-fertilizing crops. A secondary purpose was to examine the differential effi-

---

Communicated by H.C. Becker

P. Liu · J. Zhu (✉) · Y. Lu
Institute of Bioinformatics,
Zhejiang University,
Hangzhou, 310029 Zhejiang, China
e-mail: jzhu@zju.edu.cn
Tel.: +86-571-86971731
Fax: +86-571-86971498

ciency of MAS conducted in early segregating generations and advanced generations.

Materials and methods

Genetic model

Considering the case of a population derived from a cross between two homozygous lines, we evaluated the efficiency of MAS in improving the mean genotypic value of the breeding population and increasing the frequencies of superior homozygous genotypes in segregating generations of self-fertilizing crops. We adopted the model described by Liu et al. (2003), in which quantitative genetic variation was assumed to be under control of additive, dominant and epistatic effects. Suppose that there are n QTLs, and denote Q_i as the ith QTL. Each Q_i is bracketed by two flanking marker alleles M_{i-} and M_{i+}. When there are no *QE* interactions, the phenotypic value of individual k can be expressed as (for i, j=1, 2, ..., n),

$$\begin{aligned} y_k = & \mu + \sum_i \mathrm{a}_i x_{\mathrm{A}_{ik}} + \sum_i \mathrm{d}_i x_{\mathrm{D}_{ik}} + \sum_{i<} \sum_j \mathrm{aa}_{ij} x_{\mathrm{AA}_{ijk}} \\ & + \sum_{i<} \sum_j \mathrm{ad}_{ij} x_{\mathrm{AD}_{ijk}} + \sum_{i<} \sum_j \mathrm{da}_{ij} x_{\mathrm{DA}_{ijk}} \\ & + \sum_{i<} \sum_j \mathrm{dd}_{ij} x_{\mathrm{DD}_{ijk}} + \varepsilon_k \end{aligned} \quad (1)$$

where μ 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 coefficient $x_{\mathrm{A}_{ik}}$ and $x_{\mathrm{D}_{ik}}$ are 1 and 0 for genotype Q_iQ_i, $x_{\mathrm{A}_{ik}}$ and $x_{\mathrm{D}_{ik}}$ are −1 and 0 for genotype q_iq_i, $x_{\mathrm{A}_{ik}}$ and $x_{\mathrm{D}_{ik}}$ are 0 and 1 for genotype Q_iq_i; $x_{\mathrm{AA}_{ijk}} = x_{\mathrm{A}_{ik}} x_{\mathrm{A}_{jk}}$, $x_{\mathrm{AD}_{ijk}} = x_{\mathrm{A}_{ik}} x_{\mathrm{D}_{jk}}$, $x_{\mathrm{DA}_{ijk}} = x_{\mathrm{D}_{ik}} x_{\mathrm{A}_{jk}}$, $x_{\mathrm{DD}_{ijk}} = x_{\mathrm{D}_{ik}} x_{\mathrm{D}_{jk}}$; ϵ_k is the residual effect, $\varepsilon_k \sim (0,\ \sigma_\varepsilon^2)$.

In our previous study (Liu et al. 2003), we presented a Bayesian approach to estimate the probability of a particular QTL genotype conditioned on its marker type and phenotype. In the Bayesian framework, the probabilities of all possible QTL genotypes that each individual contains can be obtained given the observed data (i.e. marker type and phenotype) during selection procedures. These probabilities can then be used in calculating the following selection indices.

Potential index

In self-fertilizing crop breeding programs, those individuals that have the potential to propagate as many offspring containing superior genotypes as possible are retained for self-fertilization in the next generation. From this guideline, we propose a potential index to rank individuals as follows. The potential index can be viewed as the weighted sum of all possible superior homozygous genotypic values. Denote the probability of QTL genotype i as p_{ik} for individual k. Assuming that breeders intend to select the top t homozygous superior genotypes, we can write the potential index of individual k as (for i=1, ..., 3^n and j=1, ..., t)

$$I_k = \sum_i p_{ik} \sum_j f_{j(i)} g_j \quad (2)$$

where $f_{j(i)}$ is the frequency of the superior homozygous genotype j, which is produced from the self-fertilization of a specific individual containing QTL genotype i; g_j is the genotypic value of the superior homozygous genotype j.

In practice, however, we have no way of determining the gametal phase for heterozygotes. We thus proposed an algorithm to calculate multi-locus gamete frequencies for heterozygotes (shown in the Appendix). For individual k containing m heterozygous loci, Equation 2 is accordingly modified as

$$I_k = \sum_i p_{ik} \sum_l {}_if_l \sum_j f_{j(l)} g_j \quad (3)$$

where ${}_if_l$ is the frequency of the lth gametal phase of QTL genotype i, $\sum_l {}_if_l = 1$, l=1,...,2^{m-1}, $2 \le m \le n$; p_{ik} and $f_{j(l)}$ have the same meanings as in Equation 2. If the QTL genotypes are known, it is easy to calculate the potential index directly ($p_{ik} \equiv 1$, $i \equiv 1$).

Marker index

If the target gene is tightly linked to a marker and there is no crossover event in the bracket covering the target gene and marker, it is possible to infer the presence of a gene from the presence of a marker tightly linked to the gene. We can derive the marker index in terms of a potential index using Equation 3

$$M_k = \sum_l {}_if_l \sum_j f_{j(l)} g_j \quad (4)$$

Note that the marker index can also be viewed as the weighted sum of the superior homozygous genotypic values, but here the QTL genotype i of individual k is uniquely determined by its marker genotype.

Breeding value and genotypic value are also used as an index in selection. For instance, breeding value B_k of individual k can be calculated as (Liu et al. 2003)

$$B_k = \sum_{\mathbf{z}} p(\mathbf{z}|y_k, \mathbf{h}) \left(\sum_i \mathrm{a}_i x_{\mathrm{A}_{ik}} + \sum_{i<} \sum_j \mathrm{aa}_{ij} x_{\mathrm{AA}_{ijk}} \right) \quad (5)$$

where $p(\mathbf{z}|y_k, \mathbf{h})$ is the probability of QTL genotype $\mathbf{z}$ conditional on its marker type $\mathbf{h}$ and phenotype y_k.

Simulations

The genetic map of simulations were similar to those of Liu et al. (2003). However, this map had eight chromosomes with 12 diallelic QTLs and 24 digenic epistases assigned randomly on the map. This is because the fewer the number of QTLs, the quicker the alleles of the QTLs are fixed under self-fertilization. Additive and dominant effects at different loci, and epistatic interactions between loci were first generated by drawing a standard normal distribution. These genetic effects were then rescaled according to their relative contributions to the total genetic variation. For simplicity, the trait heritability h^2 was set as 0.4 and total phenotypic variance V_P was fixed at 1.0. The relative genetic contributions were set as V_A:V_D:V_AA:V_AD:V_DA:V_DD=4:2:2:1:1:2 where V_A, V_D, V_AA, V_AD, V_DA and V_DD are genetic variances due to additive, dominance, and epistatic interactions of additive×additive, additive×dominant, dominant×additive and dominant×dominant, respectively (Liu et al. 2003) and $V_\mathrm{A}+V_\mathrm{D}+V_\mathrm{AA}+V_\mathrm{AD}+V_\mathrm{DA}+V_\mathrm{DD}=h^2V_P$. The phenotypic value of individual k was calculated as

$$y_k = \mu + G_k + \varepsilon_k \quad (6)$$

where μ is the population mean; G_k is the genotypic value of individual k by summing all genetic effects within and between loci; ϵ_k is the residual effect and is obtained by generating a pseudo-random normal deviate with zero mean and known variance $(1-h^2)$ V_P. Detailed procedures for generating phenotypic data were given in our previous study (Liu et al. 2003).

Selection methods

Eight methods were compared using computer simulations. They were: (1) selection based on an individual's phenotypic value P; (2) selection based on the marker index M; (3) selection based on the potential index by using the true QTL genotype *I*; (4) selection based on the genotypic value of the true QTL genotype *G*; (5) selection based on the breeding value of the true QTL genotype *B*; (6) selection based on the estimated potential index $\hat{I}$; (7) selection based on the estimated genotypic value $\hat{G}$; and (8) selection based on the estimated breeding value $\hat{B}$.

Three breeding strategies were also examined: (1) early-generation selections in F_2, F_3 and F_4; (2) later-generation selections in F_5, F_6 and F_7; and (3) all-generation selections in generations in F_2 to F_7. An alternate method combining MAS on the estimated potential index $\hat{I}$ in early generations with phenotypic selection *P* in later generations ($\hat{I}$+*P*) was also simulated.

Outcomes of selection

Genetic response is usually used to measure the efficiency of selection. Cumulative genetic response achieved by selection in each generation is calculated as

$$\Delta G_{(t)} = \frac{\bar{G}_{(t)} - \bar{G}_{(0)}}{\sigma_{G_{(0)}}} \tag{7}$$

where $\bar{G}_{(t)}$ is the genetic mean of the population at generation *t* (t=0 representing the initial population F_2), and $\sigma_{G_{(0)}}$ is the genetic standard deviation in the F_2 (Liu et al. 2003).

We also complement another measurement, the frequency of the superior homozygous genotype, to evaluate the final outcome of selection. Suppose that a population is composed of *N* individuals, and there are N_j individuals containing homozygous genotype *j* within this population. The frequency of homozygous genotype *j* is thus

$$f'_j = \frac{N_j}{N} \tag{8}$$

We only calculated the frequencies of the superior homozygous genotypes in the final population, in which the selected individuals have been self-fertilized up to the 20th generation.

The main breeding objective is assumed to be to obtain the top 20 superior homozygous genotypes (i.e. t=20). The initial breeding population was composed of a total of 500 F_2 individuals derived from a cross between two homozygous lines. In each generation with selection, the top 20% of individuals were selected using a mass selection strategy and each selected individual was self-fertilized to produce five offspring. Note that the five offspring that were derived from the same parents could be all selected at the next generation and the size of the new breeding populations is also 500 individuals. In other generations without selection each individual was fertilized to produce one offspring. Selections were performed by the above methods for several generations and then repeatedly self-fertilized up to the 20th generation. Simulations were replicated 200 times for each case and the mean results of the 200 simulations were presented.

Results

Cumulative genetic responses in each generation are presented in Fig. 1 for the various methods of early, later, and all-generation selection. The cumulative frequencies of the top 20 superior genotypes in the final population is shown in Table 1. There was a general tendency; the cumulative responses to the various selection methods increased rapidly in selecting generations, but decreased thereafter until equilibrium was reached under inbreeding. Higher cumulative responses in the final population usually resulted from the higher frequencies of superior genotypes.

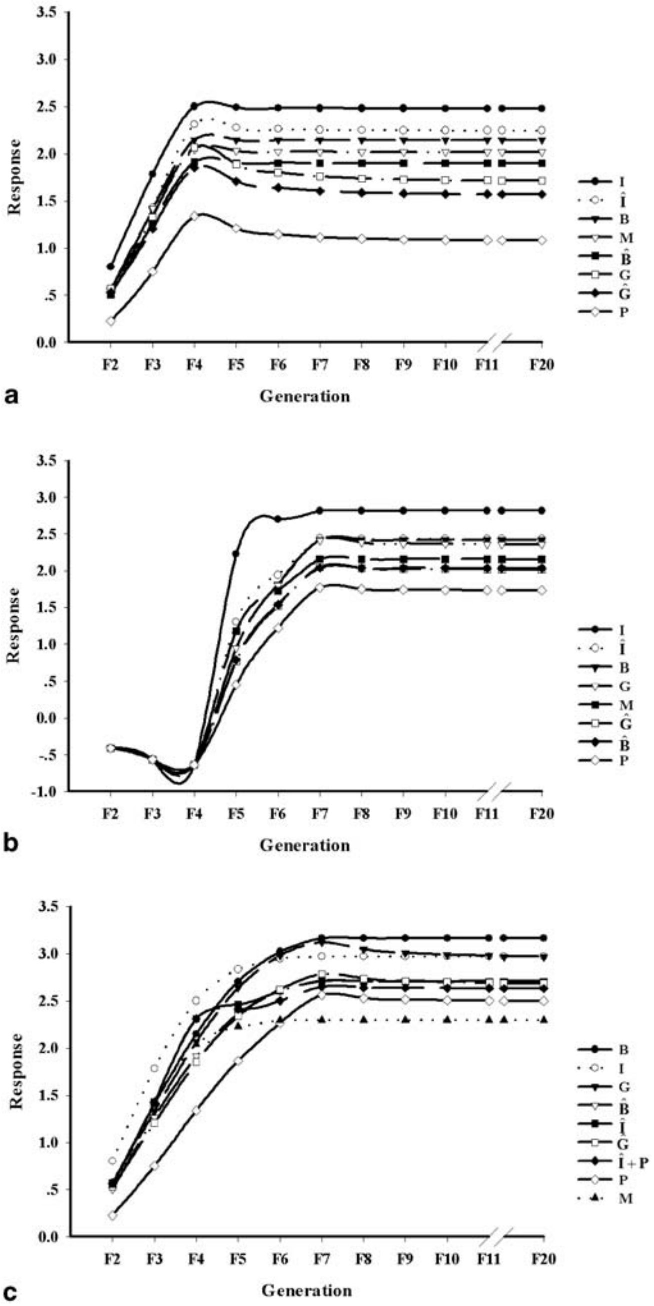

Fig. 1 Cumulative genetic response in each generation for various selection methods when selecting in early generations F_2 to F_4 (**a**), in later generations F_5 to F_7 (**b**), and in all generations F_2 to F_7 (**c**). *P* Phenotypic selection, *M* marker index selection, *I* potential index selection, *G* genotypic selection, *B* breeding value selection. $\hat{I}$, $\hat{G}$ and $\hat{B}$ indicate selection based on estimates of *I*, *G*, and *B*, respectively. $\hat{I}$+*P* indicates selection based on estimated potential index in early generation and phenotypic selection in later generations

Selection in early generations

Of all the selection methods, phenotypic selection yielded the smallest response and the lowest frequencies of superior genotypes. Selections using the known parameters without error (i.e. the selection methods *I*, *G* and *B*)

Table 1 Cumulative frequency of the top 20 superior homozygous genotypes in the final population for various selection methods. *P* Phenotypic selection, *M* marker index selection, *I* potential index selection, *G* genotypic selection, *B* breeding value selection. $\hat{I}$, $\hat{G}$ and$\hat{B}$ indicate selection based on estimates of *I*, *G*, and *B*, respectively

Selection generations	Selection methods							
	P	*M*	$\hat{I}$	$\hat{G}$	$\hat{B}$	*I*	*G*	*B*
F_2 to F_4	0.111	0.515	0.582	0.233	0.244	0.832	0.289	0.315
F_5 to F_7	0.196	0.554	0.629	0.335	0.307	0.990	0.511	0.511
F_2 to F_7	0.612	0.553	0.898	0.706	0.776	0.990	0.815	0.999

worked marginally better than those using the estimated parameters (i.e. the selection methods $\hat{I}$, $\hat{G}$ and $\hat{B}$), indicating that there is considerable room for improvement in the estimation of QTL parameters. Selections based on the QTL genotype tended to give higher selection efficiency than those based on the genotypic value. Selection on the marker index (i.e. the method *M*) also performed reasonably well in several early generations. It should be noted, however, that all target genes were tightly linked to a marker with a genetic distance less than 5.0 cM in the present study. These tight linkages are favorable for the method *M*.

In all the cases studied, selection based on the potential index using true QTL genotype (i.e. the method *I*) performed dramatically better than the other methods. Selection based on the estimated potential index (i.e. the method $\hat{I}$) ranked second among all the selection methods. The overall genetic response provided by this method was two times greater than that by phenotypic selection, while the cumulative frequency of the top 20 homozygous genotypes was more than five times higher than that of phenotypic selection in the final population. It is also worth noting that genetic responses to all the selection methods dropped slightly after the F_4 generation, mainly due to the disappearance of the remaining dominance and its epistatic interactions under selfing.

Selection in later generations

Selections in later generations provided larger genetic responses, especially in the first selected generation, and higher frequencies of superior homozygous genotypes in the final population than those in early-generation selections. However, the advantages of the method $\hat{I}$ over *P*, $\hat{G}$, and $\hat{B}$ were reduced. For instance, the ratio of the cumulative frequency of the method $\hat{I}$ to the cumulative frequency of phenotypic selection was halved in comparison with selection in early generations. These results were expected, since many alleles at QTLs were inclined toward homozygosity after selfing for several generations, and heritability of the improved trait was consequently increased. Another noticeable feature of Fig. 1b was that considerable inbreeding depression was observed to accompany self-fertilization in all cases. This result also provided indirect evidence suggesting the poor performances of such selection methods as *P*, *G* and $\hat{G}$ in early generations.

Selection in all generations

In comparison with the aforementioned two breeding strategies, there were considerable increases in the cumulative genetic responses, and in the cumulative frequency of the top 20 homozygous genotypes among all selection methods. This was due to the reason that selections in all generations were favorable for additional advances in the later three generations based on the past gains achieved in the early three generations.

Of all the selection methods, phenotypic selection received the maximum benefits from additional selections in later generations. The method *M* performed reasonably well in early generations, but poorly in later generations. It was expected that the linkage disequilibrium between the marker and the QTL would be gradually eroded by recombination and selection. It was a little surprising to find that selection on breeding value does so well relative to selection on genotype.

Efficiency of a modified method

Efficiency of a modified method combining MAS ($\hat{I}$) in early generations with phenotypic selection (*P*) in later generations was also presented in Fig. 1c. Results from selections in all generations have shown that phenotypic selection was reasonably good in the later generations, while $\hat{I}$ was quite good in the early generations. We thus proposed a modified method combining the method $\hat{I}$ in early generations with phenotypic selection (*P*) in later generations ($\hat{I}$+*P*), in which the cost of genotyping is hence reduced. Figure 1c shows that there was a slight decrease in the selection efficiency of the method $\hat{I}$+*P* as compared with the method $\hat{I}$ in all generations. The cumulative frequency of the top 20 superior homozygous genotypes in the final population for $\hat{I}$+*P* was 0.787, which was lower than that of $\hat{I}$ (0.898), but higher than those of $\hat{B}$ (0.706) and $\hat{G}$ (0.776) as shown in Table 1.

Discussion

Substantial genetic diversity due to recombination and segregation that can be explored by plant breeders occurs in early segregating generations. However, early generation selection has generally been found to be ineffective for conventional phenotypic selection (McGinnis and Shebeski 1968; Knott 1972; Depauw and Shebeski 1973; Hanson et al. 1979). This is in agreement with our results that phenotypic selection in early segregating generations leads to smaller genetic responses and lower frequencies of superior genotypes. In consideration of the limitations of early generation selection, plant breeders usually delay phenotypic selection until most alleles at QTLs tend to be fixed under self-fertilization. However, as shown in our results, even MAS conducted in early generations still obtains a larger genetic response and higher frequencies of superior genotypes than phenotypic selection in later generations. We also showed that, although MAS conducted in later generations works better than that in early generations, the advantages of MAS over phenotypic selection are reduced noticeably. Therefore, MAS is more suitable for selection in early segregating generations. Thus, a modified method with MAS in early generations combining phenotypic selection in later generations ($\hat{I}+P$) was recommended from an efficiency standpoint. This method does well relative to the method $\hat{I}$ in all-generation selections. As Valentine (1979) stated, if yield is to be maximized, no opportunity for selection in early generations should be lost. The $\hat{I}+P$ method could be an optimum breeding strategy for self-fertilizing crops, to maximize benefits from marker-based procedure.

It is apparent that more superior genotypes derive from selection by QTL genotype than that by phenotypic value and genotypic value. We proposed a potential index to measure the probability of an individual showing superior genotypes under a selfing system. This index depends on the number of the top homozygous genotypes that breeders intend to select and the genotypic values of these top genotypes. It appears that the smaller number of the top homozygous genotypes selected, the higher the frequencies of the top genotypes achieved. In our previous study (Liu et al. 2003), we also proposed a selection index termed as a “breeding value” for a random mating system. MAS using this breeding value usually performs better than other methods in population improvement (Liu et al. 2003). Interestingly, in the case of selection in all generations, this breeding value also does very well relative to the potential index under a selfing system. This is in part because the breeding value accounts for additive effects and additive × additive epistasis that can be gradually fixed under inbreeding (Liu et al. 2003).

For simplicity, we adopted a mass selection technique, and the size of selection population was fixed. In practice, for self-fertilizing crops the breeding methods commonly used are pedigree, bulk, backcross, and their modifications. In the case of the pedigree method, family data can be used to not only enhance the efficiency of phenotypic selection, but also to improve the precision of estimation of the QTL genotype. MAS in early generations allows breeders to discard a large number of lines with inferior genotypes in the inbreeding process and, at the same time, maintain a high probability of superior genotypes at homozygosity, as shown by our results. Conversely, plant breeders using phenotypic selection have to advance a considerable number of lines that are eventually discarded, to be assured of selecting one or more superior genotypes (Singh 1994). However, in the case of the bulk method, delaying selection in later generations is more favorable for phenotypic selection than for MAS. The superiority of MAS over phenotypic selection depends largely on the breeding methods used. The performance of these selection schemes requires further investigation. However, our previous conclusion that MAS is more cost-effective for selection in early segregating generations still holds.

In our previous study (Liu et al. 2003), we showed that errors in QTL detection may reduce genetic responses to MAS under population improvement. However, the MAS using inaccurate QTL estimates still gives better results than phenotypic selection in most cases. The inaccuracy of QTL estimation may affect MAS for breeding self-fertilizing crops in a way similar to that shown in our previous study (Liu et al. 2003). Therefore, verifications of the putative QTL and its magnitude of effect and accurate chromosome map location are also imperative to realize the potential of MAS for developing inbred lines. In addition, we assumed no *QE* interactions in the model in the present study. *QE* interaction is less favorable for phenotypic selection than MAS, especially in early segregating generations. We thus can expect that if there are *QE* interactions the superiority of MAS over phenotypic selection in segregating generations of self-fertilizing crops will increase.

Acknowledgements We greatly thank Dr. Xiang-yang Lou for his helpful discussions and comments regarding part of this paper, and David Bartsch for his careful reading of this manuscript. We also thank two anonymous reviewers for useful comments and suggestions on the earlier version of the manuscript. This research was supported in part by the National Natural Science Foundation of China.

Appendix

An algorithm to calculate multi-locus gamete frequencies for self-fertilization

Derivation of the frequency of each gametal type from a given parental genotype, when self-fertilizing, requires its gametal phase to be known. The frequency of each gametal type is easily calculated in the F_2. However, for subsequent generations, more work is required. An efficient algorithm to calculate multi-locus gamete frequencies is given as follows.

Suppose that there are n QTLs, and denote Q_i allele at the ith QTL. We then obtain $(n-1)$ possible heterozygotic types in total, each containing 2, 3, ..., n heterozygous loci

Table 2 The relative frequency of each gametal type in the *g*th generation for double heterozygotes. Note that $f_1(g)$ and $f_2(g)$ are the relative frequencies of the first (Q_1Q_2/q_1q_2) and second (Q_1q_2/q_1Q_2) types of informative gametes in the *g*th generation, respectively; r_{12} is the recombination fraction between Q_1 and Q_2

Freq.	Q_1Q_2/q_1q_2	Code	Freq.	Q_1q_2/q_1Q_2	Code
$f_1(1)$	$\frac{1}{2}(1-r_{12})^2$	0	$f_2(1)$	$\frac{1}{2}r_{12}^2$	1
$f_1(2)$	$\frac{1}{2}(1-r_{12})^2 f_1(1)+\frac{1}{2}r_{12}^2 f_2(1)$	(0),(1)	$f_2(2)$	$\frac{1}{2}r_{12}^2 f_1(1)+\frac{1}{2}(1-r_{12})^2 f_2(1)$	(1),(0)
⋮	⋮	⋮	⋮	⋮	⋮
$f_1(g)$	$\frac{1}{2}(1-r_{12})^2 f_1(g-1)+\frac{1}{2}r_{12}^2 f_2(g-1)$	(0),(1)	$f_2(g)$	$\frac{1}{2}r_{12}^2 f_1(g-1)+\frac{1}{2}(1-r_{12})^2 f_2(g-1)$	(1),(0)

respectively. Denote the *m*-locus heterozygote as H_m (m=2, ...,n, and each heterozygote H_m consists of 2^{m-1}) types of informative gametes. Note that the total number of gametes is 2^m but the number of informative gametes is only $2^m/2$, since one type of gamete is complementary to another. For simplicity, we present the exact expression of gamete frequencies for double heterozygotes. Assuming the original population to be entirely Q_1Q_2/q_1q_2, we can obtain the relative frequency of two types of informative gametes in the *g*th generation (Table 2). In summary, for *m*-locus heterozygotes, the relative frequency of each gametal type is analogically calculated by the following recurrence equation,

$$
{}_mf_u(g)=\frac{1}{2}\sum_v\prod_l\Big[(1-r_l)^{2C_{v\to u(l)}} + r_l^{2C_{v\to u(l)}} - 1\Big]{}_mf_v(g-1)
$$

where ${}_mf_u(g)$ is the frequency of the *u*th gametal type of *m*-locus heterozygote in the *g*th generation; $\mathrm{C}_{v\to u}$ is a transformation vector from C_v to C_u; C_u is a vector of crossover event(s) corresponding to the *u*th gametal type, as a result of C_v; the *u*th type of recombinant gamete is derived from its parental gamete type; $\mathrm{C}_u=(C_{u(1)}, \ldots C_{u(l)}, \ldots, C_{u(m)})$, $C_{u(l)}$ is 0 and 1 corresponding to the situations in that there is no crossover and a crossover occurred in a specific QTL bracket l respectively; r_l is the recombinant fraction of QTL bracket l; for m=2, ..., n, u=1, ..., 2^{m-1} and l=1, ..., m–1.

References

Berloo RV, Stam P (1998) Marker-assisted selection in autoamous RIL populations: a simulation study. Theor Appl Genet 96: 147–154

Bouchez A, Hospital F, Causse M, Gallais A, Charcosset A (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics 162:1945–1959

Charmet G, Robert N, Perretant MR, Gay G, Sourdille, Groos C, Bernard S, Bernard M (1999) Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. Theor Appl Genet 99:1143–1148

Depauw RM, Shebeski LH (1973) An evaluation of an early generation yield testing procedure in *Triticum aestivum* L. Can J Plant Sci 53:465–470

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33:660–668

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

Hallauer AR, Miranda F (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Hanson PR, Jenkins G, Westcott B (1979) Early generation selection in a cross of spring barley. Z Pflanzenzuecht 83:64–80

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

Hospital F, Goldringer I, Openshaw S (2000) Efficient marker-based recurrent selection for multiple quantitative trait loci. Genet Res 75:357–368

Johnson B (1989) The probability of selecting genetically superior S_2 lines from a maize population. Maydica 34:5–14

Knapp SJ (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci 38:1164–1174

Knott DR (1972) Effects of selection for F_2 plant yield on subsequent generations in wheat. Can J Plant Sci 52:721–726

Lande R (1992) Marker-assisted selection in relation to traditional methods of plant breeding. In: Beckman JS, Osborn TS (eds) Plant genomes: methods for genetic and physical mapping. Kluwer, Dordrecht, pp 209–305

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

Liu PY, Zhu J, Lou XY, Lu Y (2003) A method for marker-assisted selection based on QTLs with epistatic effects. Genetica 119:75–86

Luo ZW, Thompson R, Woolliams JA (1997) A population genetics model of marker-assisted selection. Genetics 146:1173–1183

McGinnis RC, Shebeski LH (1968) The reliability of single plant selection for yield in F_2. In: Proceedings of the 3rd international wheat genetics symposium, Canberra, 1997. Aust Acad Sci, Canberra pp 109–114

Moreau L, Charcrosset A, Hospital F, Gallais A (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

Singh SP (1994) Gamete selection for simultaneous improvement of multiple traits in common bean. Crop Sci 34:352–355

Spelman R, Bovenhuis H (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

Stromberg LD, Dudley JW, Rufener GK (1994) Comparing conventional early generation selection with molecular marker assisted selection in maize. Crop Sci 34:1221–1225

Valentine J (1979) The effect of competition and method of sowing on the efficiency of single plant selection for grain yield, yield components and other characters in spring barley. Z Pflanzenzuecht 83:193–204

Whittaker JC, Curnow RN, Haley CS, Thompson R (1995) Using marker maps in marker-assisted selection. Genet Res 66:255–265

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

Zehr BE, Dudley JW, Chojecki J (1992) Some practical considerations for using RFLP markers to aid in selection during inbreeding of maize. Theor Appl Genet 84:704–708

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

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