© Springer-Verlag 1999

R. Bernardo

Two-trait selection response with marker-based assortative mating

Abstract Marker-based assortative mating (MAM) – the mating of individuals that have similar genotypes at random marker loci – can increase selection response for a single trait by 3–8% over random mating (RM). Genetic gain is usually desired for multiple traits rather than for a single trait. My objectives in this study were to (1) compare MAM, phenotypic assortative mating (PAM), and RM of selected individuals for improving two traits and (2) determine when MAM will be most useful for improving two traits. I simulated 20 generations of selecting 32 out of 200 individuals in an F_2 population. The individuals were selected based on an index (SI) of two traits and were intermated by MAM, PAM, or RM. I studied eight genetic models that differed in three contrasts: (1) weight, number of quantitative trait loci (QTL), and heritability (h^2) for each trait; (2) linkage of QTL for each trait; and (3) trait means of the inbred parents of the F_2. For SI and the two component traits, MAM increased short-term selection response by 5–8% in six out of the eight genetic models. The MAM procedure was least effective in two genetic models, wherein the QTL for one trait were unlinked to the QTL for the other trait and the parents of the F_2 had divergent means for each trait. The loss of QTL heterozygosity was much greater with MAM than with PAM or RM. Consequently, the advantage of MAM over RM dissipated after 5–7 generations. Differences were small between selection responses with PAM and RM. The MAM procedure can enhance short-term selection response for two traits when selection is not stringent, h^2 is low, and the means of the parents of the F_2 are equal for each trait.

Key words Assortative mating · Index selection · Molecular markers · Quantitative trait loci

Communicated by P. M. A. Tigerstedt

R. Bernardo
Department of Agronomy, Purdue University, 1150 Lilly Hall of Life Sciences, West Lafayette, IN 47907-1150, USA
Fax: +1 765 496 2926
E-mail: bernardo@purdue.edu

Introduction

Genetic gain is achieved by selecting superior individuals in a population and intermating them to form the next generation. Random mating (RM) is the most common method of intermating selected individuals in plant breeding programs. Fisher (1918) found that phenotypic assortative mating (PAM) – the mating of individuals similar in phenotype – increases the additive variance (V_A) in unselected populations. This increase in V_A is due to a positive correlation between additive genetic values of mated individuals (r_M). Selection response could consequently increase due to larger V_A with PAM. McBride and Robertson (1963) found that PAM increased selection response for abdominal bristle number in *Drosophila melanogaster* by about 5% in most instances. Baker (1973) found that truncation selection reduces the usefulness of PAM. From theory he concluded that PAM will increase selection response by a maximum of 10% when heritability (h^2) is high and selection pressure is low, and by 4–5% under the conditions met in most breeding programs.

Molecular markers enable another form of assortative mating: marker-based assortative mating (MAM), the mating of individuals that have similar genotypes at random marker loci. Selection in plants is often done in populations generated from a cross between two inbreds. Assume there are many markers with unknown linkage to quantitative trait loci (QTL) and that a trait of interest is controlled by many QTL. The MAM procedure may cause a positive r_M due to chance linkage disequilibrium between markers and

QTL in biparental crosses. I found in a simulation study (Bernardo 1998) that MAM increased short-term selection response by 3–8% when h^2 was low (0.20) and when selection was not stringent (i.e., 32 individuals selected out of 200). In contrast, PAM increased selection response by a maximum of 4% when $h^2 = 0.80$.

Genetic gain is usually desired for multiple traits rather than for a single trait. The usefulness of MAM and PAM for increasing selection response has not been investigated for more than one trait. My objectives in this study were to (1) compare MAM, PAM, and RM of selected individuals for the simultaneous improvement of two traits in an F_2 population, and (2) determine the conditions in which MAM will be most useful for enhancing selection response for two traits.

Methods

Genetic models

I wrote a Fortran program to simulate two-trait selection followed by MAM, PAM, and RM. The initial population (Generation 0) comprised 200 individuals in an F_2 derived from crossing two inbreds. The selection index (SI) was a base index in which the economic weights for Trait 1 (w_1) and Trait 2 (w_2) were directly used as the trait weights in SI (Williams 1962). The numbers of QTL were q_1 for Trait 1 and q_2 for Trait 2. I studied eight genetic models that represented the factorial combinations of three factors, each with two levels. The three factors were the (1) weight, number of QTL, and h^2 for each trait; (2) linkage of QTL controlling each trait; and (3) means of the inbred parents (μ_{P1} for Parent 1 and μ_{P2} for Parent 2) of the F_2 (Table 1). In Models A1–A4, Traits 1 and 2 had equal weight in SI ($w_1 = w_2 = 1$), equal number of QTL ($q_1 = q_2 = 50$), and equal h^2 (0.20). In Models B1–B4, Trait 1 was more important ($w_1 = 1$ and $w_2 = 0.25$), was controlled by more QTL ($q_1 = 80$ and $q_2 = 20$), and had lower h^2 (0.20 for Trait 1 and 0.80 for Trait 2) than Trait 2. Models B1–B4 typified selection for yield (Trait 1) and moisture (Trait 2) in grain crops. The QTL for Trait 1 were either linked (Models A1, A2, B1, and B2) or unlinked (Models A3, A4, B3, and B4) to the QTL for Trait 2. For each trait,

the means of the two parental inbreds were either nearly equal (Models A1, A3, B1, and B3) or divergent (Models A2, A4, B2, and B4). With divergent parental means, $\mu_{P1} > \mu_{P2}$ for Trait 1 and $\mu_{P1} < \mu_{P2}$ for Trait 2.

An individual had ten pairs of chromosomes, each 100 cM long. All markers and QTL were segregating in the F_2. A total of 100 codominant marker loci were randomly distributed across the ten pairs of chromosomes. When QTL for Trait 1 were linked to QTL for Trait 2, all 100 QTL were randomly distributed throughout the genome. In Models A3 and A4, the 50 QTL for Trait 1 were randomly distributed on chromosomes 1–5, whereas the 50 QTL for Trait 2 were randomly distributed on chromosomes 6–10. In Models B3 and B4, the 80 QTL for Trait 1 were randomly distributed on chromosomes 1–8, whereas the 20 QTL for Trait 2 were distributed on chromosomes 9 and 10. When the inbred parents had nearly equal means for each trait, the first parent of the biparental cross was homozygous for the favorable allele at the odd-numbered QTL for Traits 1 and 2. The second parent was homozygous for the favorable allele at the even-numbered QTL for Traits 1 and 2. But when parental means were divergent, the first parent was homozygous for the favorable allele at all q_1 QTL for Trait 1, whereas the second parent was homozygous for the favorable allele at all q_2 QTL for Trait 2.

Genetic and phenotypic values of individuals were simulated for Trait 1 and Trait 2. The marker loci per se had no effects on the expression of the trait, whereas the effects of the QTL were exponential. For each trait, the homozygote for the less favorable allele, the kth QTL, had a value of zero, whereas the homozygote for the favorable allele had a value of 0.98^k. Dominance and epistasis were absent, and the heterozygote at the kth QTL had a value of $1/2(0.98^k)$. For the ith ($= 1, 2$) trait, the genetic value of each individual for each trait was equal to the sum of the genotypic effects across the q_i QTL. Additive variance assuming linkage equilibrium was $V_A = 1/2 \sum_{k=1}^{qi} [(1/2)(0.98^k)]^2$ in Generation 0. The phenotypic value of an individual for each trait was equal to its genetic value plus a random nongenetic effect. Nongenetic effects were normally and independently distributed with a mean of zero and variance of V_E. The value of V_E was calculated based on h^2 of each trait in Generation 0. The value of V_E remained constant across generations of selection.

Selection response

The phenotypic value of an individual for SI was equal to w_1 (phenotypic value for Trait 1) + w_2 (phenotypic value for Trait 2).

Table 1 Features of eight genetic models for MAM, PAM, and RM, and the average (from 1000 repeats) genetic correlation between Trait 1 and Trait 2 (r_{12})

Model	Trait weight; number of QTL; heritability		Linkage between QTL for each trait			Mean of Parent 1 (μ_{P1}) and Parent 2 (μ_{P2})		r_{12}
	Trait 1	Trait 2	QTL for Trait 1 and for Trait 2	Chromosomal locations of QTL		Trait 1	Trait 2	
				Trait 1	Trait 2			
A1	1; 50; 0.20	1; 50; 0.20	Linked	1–10	1–10	$\mu_{P1} \cong \mu_{P2}$	$\mu_{P1} \cong \mu_{P2}$	0.00 (-0.60, 0.53)[a]
A2	1; 50; 0.20	1; 50; 0.20	Linked	1–10	1–10	$\mu_{P1} > \mu_{P2}$	$\mu_{P1} < \mu_{P2}$	-0.73 (-0.88, -0.41)
A3	1; 50; 0.20	1; 50; 0.20	Unlinked	1–5	6–10	$\mu_{P1} \cong \mu_{P2}$	$\mu_{P1} \cong \mu_{P2}$	0.01 (-0.21, 0.25)
A4	1; 50; 0.20	1; 50; 0.20	Unlinked	1–5	6–10	$\mu_{P1} > \mu_{P2}$	$\mu_{P1} < \mu_{P2}$	-0.00 (-0.23, 0.53)
B1	1; 80; 0.20	0.25; 20; 0.80	Linked	1–10	1–10	$\mu_{P1} \cong \mu_{P2}$	$\mu_{P1} \cong \mu_{P2}$	-0.00 (-0.57, 0.61)
B2	1; 80; 0.20	0.25; 20; 0.80	Linked	1–10	1–10	$\mu_{P1} > \mu_{P2}$	$\mu_{P1} < \mu_{P2}$	-0.66 (-0.86, -0.30)
B3	1; 80; 0.20	0.25; 20; 0.80	Unlinked	1–8	9–10	$\mu_{P1} \cong \mu_{P2}$	$\mu_{P1} \cong \mu_{P2}$	0.00 (-0.25, 0.20)
B4	1; 80; 0.20	0.25; 29; 0.80;	Unlinked	1–8	9–10	$\mu_{P1} > \mu_{P2}$	$\mu_{P1} < \mu_{P2}$	-0.00 (-0.25, 0.22)

[a] Lowest and highest values of r_{12} among 1000 repeats are in parentheses

Likewise, the genetic value for SI was equal to w_1 (genetic value for Trait 1) + w_2 (genetic value for Trait 2). The 200 F_2 individuals in Generation 0 were ranked according to their phenotypic values for SI. The top $n = 32$ individuals were selected to form the next generation. Two hundred random crosses were made among the 32 selected individuals, and each cross contributed 1 individual to the next generation.

With PAM, crosses were made between the individuals with the highest and second highest SI, between the individuals with the third and fourth highest SI, and so on. With MAM, Rogers distance (Rogers 1972) was calculated across the 100 marker loci as $\sum_{k=1}^{100}[\sum_{l=1}^{2}(p_{ikl} - p_{jkl})^2]/200$, where: p_{ikl} = frequency of the lth allele at the kth marker locus in individual i; and p_{jkl} = frequency of the lth allele at the kth marker locus in individual j. Crosses were made between the pair of individuals with the smallest Rogers distance, between the pair of individuals with the next smallest Rogers distance, and so on. With both PAM and MAM, the number of individuals contributed by each of the $n/2 = 16$ crosses to the next generation was kept as uniform as possible.

I simulated recombination between loci (markers or QTL) during meiosis. Recombination frequencies were calculated from map distances with the Haldane mapping function (Haldane 1919). Interference was absent. Individuals in the next generation were formed by the pairing of homologous chromosomes, each homologue being randomly selected from each parent. Genetic values of the individuals in the resulting generation were equal to the sum of genotypic effects across all QTL for each trait, and phenotypic values were equal to the genetic effect plus a random nongenetic effect.

Individuals were selected and intermated by MAM, PAM, or RM for 20 generations for each genetic model. The selection experiment was repeated 1000 times. Generation 0 had a different random arrangement of marker loci and QTL in the genome in each of the 1000 repeats. Cumulative selection response in each generation was averaged across the 1000 repeats. Average cumulative selection responses with MAM, PAM, and RM were compared by paired t-tests at the $P = 0.05$ significance level. For each method of intermating and for each trait, the weighted (i.e., by the magnitude of QTL effects) loss in heterozygosity in generation t was calculated as $F = 1 - [\sum_{k=1}^{100}(0.98^k H_{kt})/(\sum_{k=1}^{100} 0.98^k)]/H_0$, where: H_{kt} = frequency of the heterozygote at the kth QTL in generation t; and $H_0 = 0.5$ = expected frequency of the heterozygote at any QTL in the initial generation. The intraclass correlation for additive genetic values of mated individuals (r_M) was calculated (Steel et al. 1997; p 299) for SI, Trait 1, and Trait 2. For each genetic model, the correlation between additive genetic values for Trait 1 and Trait 2 (r_{12}) was calculated in Generation 0 and averaged across the 1000 repeats.

I expressed the efficiency of assortative mating relative to RM as the average cumulative selection response with PAM or MAM, divided by the average cumulative selection response with RM. Efficiencies were calculated for SI, Trait 1, and Trait 2.

Results and discussion

Models A1–A4: two traits with equal weight, equal number of QTL, and equal h^2

Changes in allele frequency due to selection in Generation 0 are not affected by the subsequent method of intermating. Thus, the mean in Generation 1 with MAM, PAM, or RM is not expected to vary. The assortative mating of individuals in Generation 0 is expected to increase V_A in Generation 1. Consequently, any increase in selection response due to MAM or PAM would first be evident in Generation 2 (Bernardo 1998). In Generation 2, the relative efficiency of MAM

for improving SI was 1.05 ($P < 0.05$) in Model A1, 1.05 ($P < 0.05$) in Model A2, 1.06 ($P < 0.05$) in Model A3, and 1.03 ($P < 0.05$) in Model A4 (Fig. 1). The MAM procedure was most useful in Model A3, wherein the highest relative efficiency for SI, Trait 1, and Trait 2 was 1.08 ($P < 0.05$, Generation 3). In contrast, MAM was least effective in Model A4, wherein the highest relative efficiency for SI was only 1.03 ($P < 0.05$, Generation 3).

The MAM procedure enhanced selection response only in the first 5–7 generations (Fig. 1). As with selection for a single trait (Bernardo 1998), the advantage of MAM over RM dissipated in the long term. By Generation 20, the relative efficiency of MAM for improving SI decreased to less than 0.85 in Models A1–A3 and 0.95 in Model A4. The enhanced short-term selection response was attributed to a positive r_M induced by MAM. For example, in Model A3, r_M for Trait 1 and Trait 2 increased from about 0.23 in Generation 0 to nearly 0.60 by Generation 10 (Fig. 2). But such increases in r_M were counteracted by a rapid loss of QTL heterozygosity (F). The higher values of F with MAM than with RM (Fig. 2) caused the smaller long-term selection response with MAM.

The average genetic correlation between Trait 1 and Trait 2 was -0.73 in Model A2 and nearly zero in Models A1, A3, and A4 (Table 1). Thus, the negative correlation between traits had little effect on the usefulness of MAM when the two traits had equal weight, equal number of QTL, and equal h^2.

The PAM procedure led to relative efficiencies of only 1.00 to 1.02 for SI, Trait 1, and Trait 2 in Models A1–A3 (Fig. 1). The highest relative efficiency of PAM was 1.04 ($P < 0.05$), observed for SI, Trait 1, and Trait 2 in Generation 3, Model A4. The PAM procedure failed to induce high r_M values and, consequently, was less effective than MAM for enhancing selection response for two traits. In Model A3, for example, r_M with PAM was close to zero across generations (Fig. 2). The small differences in both r_M and F between PAM and RM (Fig. 2) explain the lack of any substantial advantage of PAM for improving more than one trait. The results for MAM and PAM suggest that, as with single-trait selection (Bernardo 1998), selected individuals should be intermated by RM if the objective is long-term improvement of more than one trait.

Models B1–B4: two traits with unequal weight, unequal number of QTL, and unequal h^2

The MAM procedure was most useful when the parental means were nearly equal for each trait (Models B1 and B3; Fig. 3). In Generation 2, the relative efficiency of MAM for improving SI was 1.06 ($P < 0.05$) in Model B1 and 1.05 ($P < 0.05$) in Model B3. The highest relative efficiencies of MAM for improving SI were 1.07 ($P < 0.05$) in Model B1 and 1.08 ($P < 0.05$) in
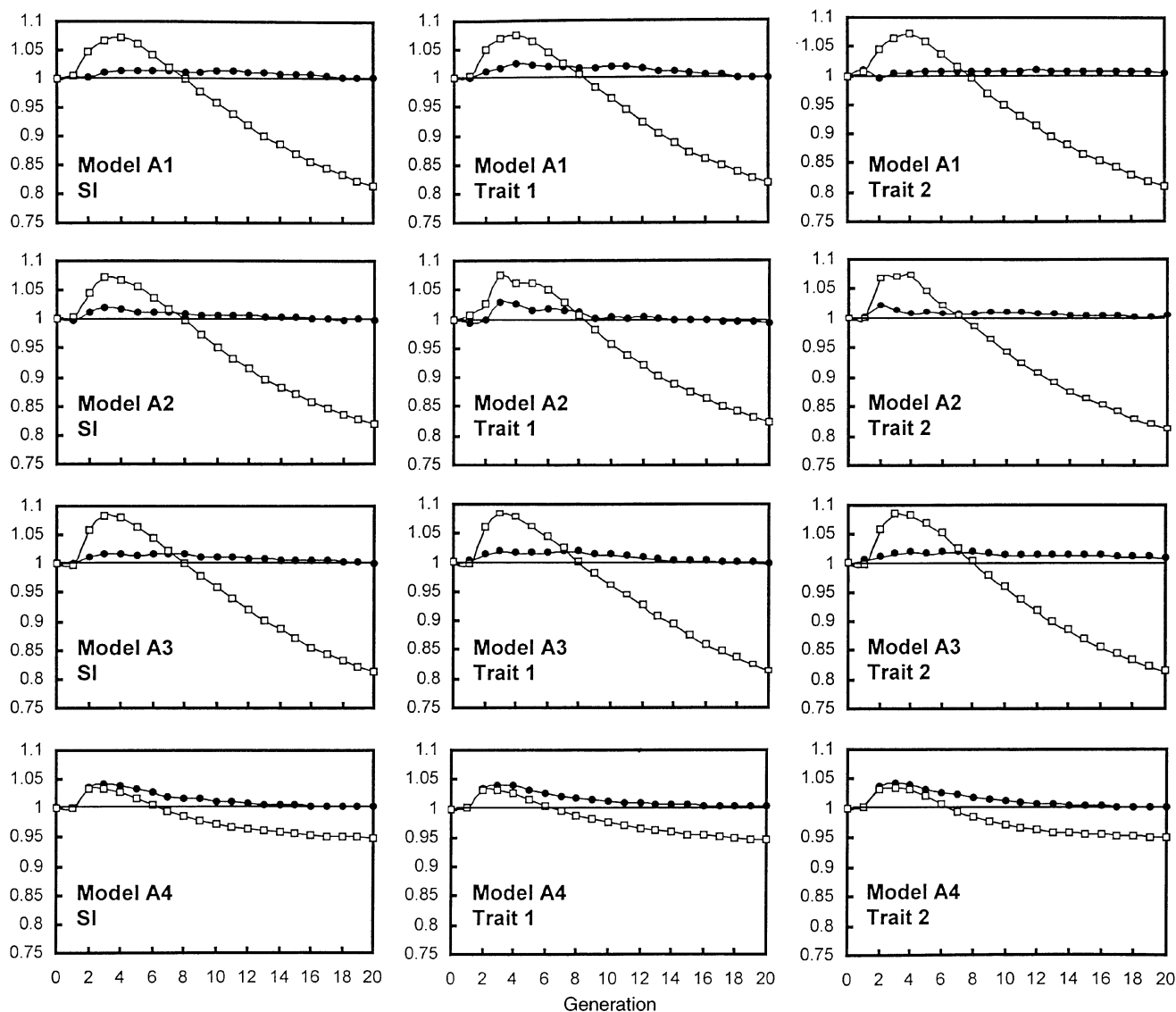
Fig. 1 Efficiency of MAM (*unshaded squares*) and PAM (*shaded circles*) relative to RM in Models A1–A4. Relative efficiencies are for the selection index (SI), Trait 1, and Trait 2

Model B3 (both in Generation 3). The highest relative efficiency for improving any trait was 1.10 ($P < 0.05$) for Trait 2, Generation 3, Model B3. Selection responses for both Trait 1 and Trait 2 were positive (i.e., favorable) in Models B1, B3, and B4. Unlike in Model A2, the negative genetic correlation between traits of − 0.66 in Model B2 affected the direction of selection response. In Generations 2–4, changes in both SI and Trait 1 with Model B2 were positive, and the relative efficiencies of MAM were 1.04 to 1.05. But for Trait 2, which was given less importance than Trait 1, changes in the mean were negative not only for MAM but for PAM and RM as well (Fig. 3). Hence, the unfavorable response for Trait 2 in Model B2 was a consequence of

the model and was not due to the method of intermating. In Model B4, the relative efficiency of MAM in Generations 2–4 was greater for Trait 2 (1.07–1.08) than for SI or Trait 1 (1.04–1.05).

For SI, the PAM procedure led to maximum relative efficiencies of only 1.02 in Models B1–B3 and 1.04 in Model B4 (Fig. 3). These results were consistent with those in Models A1–A4 and indicated that, for the genetic models studied, MAM was superior to PAM in enhancing simultaneous selection response for two traits. Enhanced response with MAM was achieved in the first few generations and, as with Models A1–A4, such advantage dissipated by Generations 7 and 8.

Usefulness of MAM for improving two traits

The MAM procedure can enhance short-term selection response for two traits by 5–8% over RM when three

Fig. 2 Correlation between additive genetic values of mated individuals (r_M) and loss of QTL heterozygosity (F) with MAM (*unshaded squares*), PAM (*shaded circles*), and RM (*shaded triangles*). Values of r_M and F are for Trait 1 and Trait 2 in Model A3. The values of r_M and F in the other genetic models were, in general, similar to those in Model A3
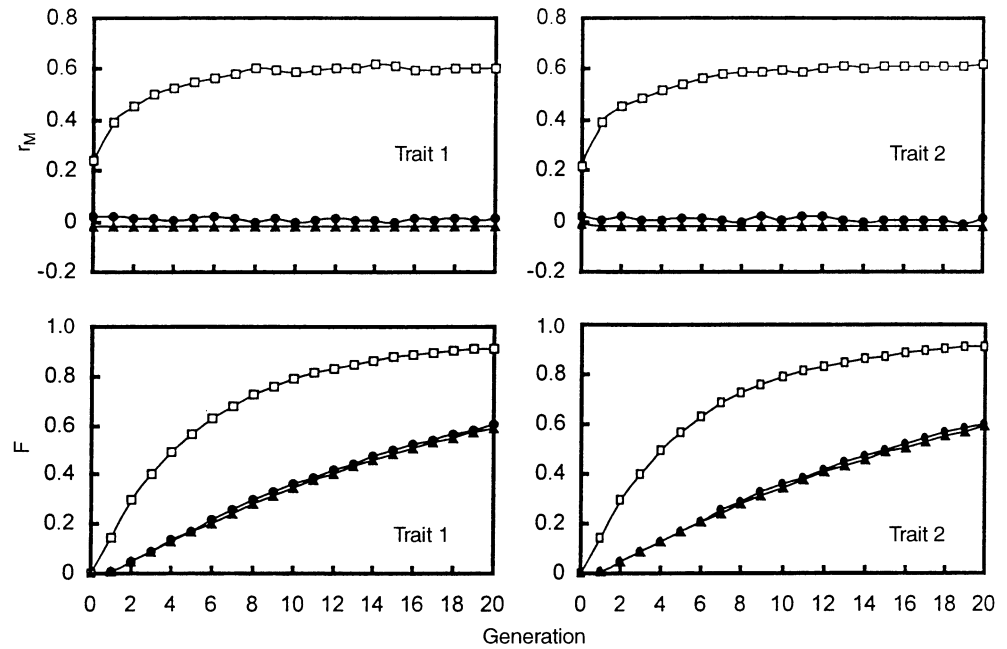
conditions are met: (1) selection is not stringent; (2) h^2 is low; and (3) the means of the parents of the F_2 are nearly equal for each trait. The first two conditions are extensions of the results for MAM with a single trait (Bernardo 1998).

To verify that MAM is most effective when n is large, I investigated MAM for improving two traits when the top $n = 8$ or 16 individuals were selected out of 200. Compared with the relative efficiency of MAM for improving SI when $n = 32$, the gain in relative efficiency in Generations 1–5 was about 5% less when $n = 8$, and about 3% less when $n = 16$ (results not shown).

Given a fixed size of the total population, a higher value of n implies lower selection pressure. Multiple-trait selection typically lowers the selection pressure for each trait. I therefore speculated that MAM would be more useful when breeding for multiple traits rather than for a single trait. Relative efficiencies of MAM were 3–8% for single-trait selection (Bernardo 1998) and 5–8% in this study. The lack of an increase in the effectiveness of MAM for two traits may have been due to a reduction in linkage disequilibrium between the markers and QTL. Because the tails of an F_2 population contribute the most towards linkage disequilibrium between markers and QTL, low selection pressures due to non-truncation selection for each trait may not necessarily increase r_M. Perhaps an alternative is to increase n, while maintaining a fixed proportion of individuals selected, by increasing the size of the total population. However, growing and selecting larger numbers of progenies may not be feasible due to resource limitations in breeding programs.

The PAM procedure is most useful for enhancing short-term selection response when h^2 is high (Baker 1973). The ineffectiveness of PAM in this study could have been due to (1) the low h^2 for Trait 1 and Trait 2 in Models A1–A4, or (2) the low h^2 for Trait 1, which was given more importance than Trait 2, in Models B1–B4. I investigated an additional genetic model similar to A3, except that h^2 was 0.80 instead of 0.20 for both Trait 1 and Trait 2. With $h^2 = 0.80$ for each trait, the maximum relative efficiency for SI was 1.04 in Generation 4 with PAM, and 1.01 in Generation 2 with MAM (results not shown). These results confirm that MAM is most effective for enhancing two-trait selection response when h^2 is low. They also suggest that, under the most favorable conditions for each form of assortative mating (i.e., high h^2 for PAM and low h^2 for MAM), PAM is less effective than MAM for enhancing selection response for two traits.

The relative efficiency of MAM was higher when parental means for each trait were nearly equal (Models A1, A3, B1, and B3) rather than divergent (Models A2, A4, B2, and B4). Thus, the level of divergence of parental means primarily determines the effectiveness of MAM for improving two traits simultaneously. Parental means will be divergent when two parents are chosen to correct each other's deficiencies, e.g., a high-yielding, low-protein inbred crossed with a low-yielding, high-protein inbred. But parental means are often equal when breeding with elite germplasm. Bailey and Comstock (1976) found that the probability of recovering an inbred superior to either parent is higher when parental means for a trait are equal rather than unequal. Linkage between QTL for Trait 1 and QTL for Trait 2 may allow MAM to retain its effectiveness even with divergent parental means (i.e., Model A2). But given that breeders have more
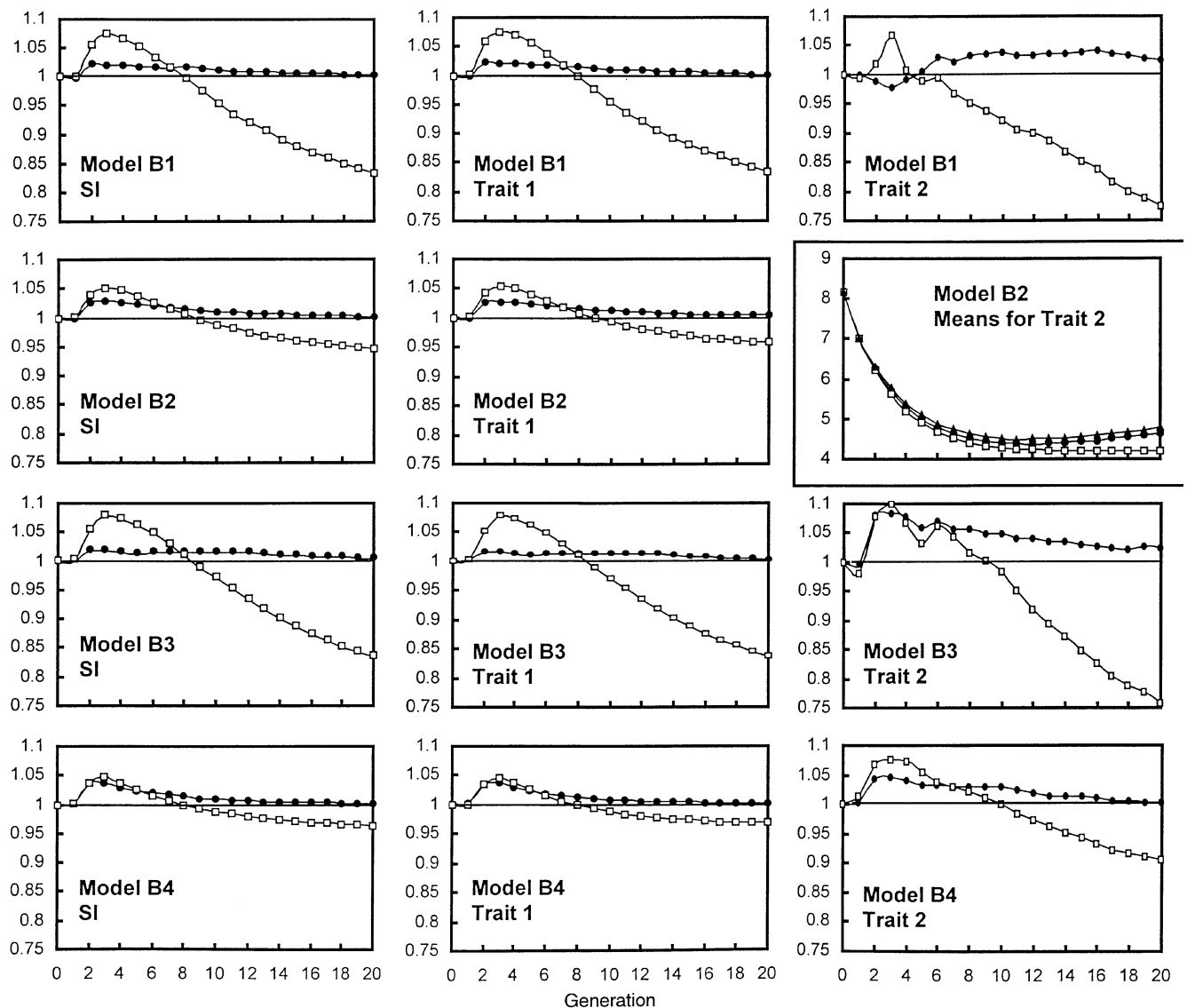
Fig. 3 Efficiency of MAM (*unshaded squares*) and PAM (*shaded circles*) relative to RM in Models B1–B4. Relative efficiencies are for the selection index (SI), Trait 1, and Trait 2. Due to response in the unfavorable (i.e., negative) direction, actual means instead of efficiencies are given for Trait 2 in Model B2

control over parental means than over the linkage of QTL for each trait, the MAM procedure should be used mainly when parental means are equal for each trait.

References

Bailey TB Jr, Comstock RE (1976) Linkage and the synthesis of better genotypes in self-fertilizing species. Crop Sci 16:363–370

Baker RJ (1973) Assortative mating and artificial selection. Heredity 31:231–238

Bernardo R (1999) Selection response with marker-based assortative mating. Crop Sci 39:69–73

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinburgh 52:399–433

Haldane JBS (1919) The combination of linkage values and the calculation of distance between loci of linked factors. J Genet 8:299–309

McBride G, Robertson A (1963) Selection using assortative mating in *Drosophila melanogaster*. Genet Res 4:356–369

Rogers JS (1972) Measures of genetic similarity and genetic distance. Studies in genetics VII. University of Texas Publ 7213:145–153

Steel RGD, Torrie JH, Dickey DA (1997) Principles and procedures of statistics, 3rd edn. McGraw-Hill, New York

Williams JS (1962) The evaluation of a selection index. Biometrics 18:375–393