Simulation of marker-assisted recurrent selection in autogamous species

Eduardo Bearzoti[*1] and Roland Vencovsky[2]
[1]Departamento de Ciências Exatas, Universidade Federal de Lavras, PO Box 37, CEP 37200-000, Lavras, MG, Brazil; [2]Departamento de Genética, ESALQ/USP, PO Box 83, CEP 13400-970, Piracicaba, SP, Brazil. (* Corresponding Author. E-mail: bearzoti@ufla.br)

ABSTRACT

This work used computer simulation to evaluate marker-assisted selection (MAS) in recurrent selection of autogamous species. Base populations were simulated by crossing each of five inbred lines with two others. Within each cross, 400 S_0 plants were genotyped in relation to 130 marker loci spread over a fictitious genome with 50 quantitative trait loci (QTLs) randomly located. Progenies derived from such plants were taken as selection units. MAS was applied using the selection index of Lande and Thompson (1990). For most situations considered, MAS was not efficient after the first selection cycle. When considering a single cycle of selection, MAS was efficient, but strongly affected by population size and heritability. Selection of a single progeny per cross did increase linkage disequilibrium, but caused very high levels of fixation rates, which were not reduced neither with the use of 200 marker loci, nor with 20 inbred lines to form base populations.

KEY WORDS: Marker-assisted selection, recurrent selection, molecular markers, QTL.

INTRODUCTION

Recurrent selection was proposed in autogamous species early in the 1960´s (Khadr and Frey, 1965; Matzinger and Wernsman, 1968), and its use in many crops has increased ever since (Miller and Fehr, 1979; Payne et al., 1986; McFerson and Frey, 1991; Beaver and Kelly, 1994). Although its effectiveness has been widely recognized, the technique is difficult to implement, mainly because recombination generally needs to be performed manually. Therefore, methodologies to enhance its efficiency are always sought, like male sterility (Werner and Wilcox, 1990) and the use of hill plots (Pomeranke and Stuthman, 1992). Marker-assisted selection (MAS) can be another such methodology. Use of molecular information to help identify superior genotypes (Lande and Thompson, 1990) can improve gain from selection. In autogamous species, the degree of polymorphism with molecular markers has sometimes been considered low (e.g. Apuya et al., 1988), but the development of highly efficient PCR-based techniques, such as AFLP and VNTR (Kochert, 1994), has broadened the usefulness of molecular markers in such species. Regarding RFLP and RAPD markers, molecular maps are already available for a variety of crops; a detailed list was presented by Phillips and Vasil (1994).

Lande and Thompson (1990) proposed the use of MAS for quantitative traits by means of a selection index which combines phenotypic and molecular information. The authors showed that the expected gain from selection under such a technique is at least equal to or higher than purely phenotypic selection. This efficiency increases as heritability (h^2) associated with the trait decreases and/or the proportion of additive genetic variance explained by molecular markers (p) increases.

For MAS to be successful, some linkage disequilibrium is necessary in the population being bred in order to increase the associations between marker loci and quantitative trait loci (QTLs). In autogamous species breeding, some level of linkage disequilibrium will be restored during each selection cycle, when selection units with some degree of inbreeding are used. Consequently, MAS is potentially superior in autogamous species. In outcrossing species, linkage disequilibrium is progressively dissipated in the recombination process across selection cycles. On the other hand, if highly inbred lines were used as selection units, the corresponding heritability could increase so much that MAS efficiency might be severely lowered (Lande, 1992). According to Lande and Thompson (1990), another advantage of MAS in selfing crops is the relatively lower amount of molecular markers needed,

due to their reproduction system.

Despite the theoretical efficiency of MAS, it is always desirable to evaluate its performance under different conditions relative to many factors, such as breeding scheme, number of polymorphic marker loci, trait heritability, sample size, and selection intensity among others. Computer simulation has been proposed as an efficient tool to evaluate selection strategies since the 1950s (Fraser, 1957), because it mimics real processes that would take many years to accomplish. Actually, it has already been widely used to evaluate marker-assisted recurrent selection (Zhang and Smith, 1992; Edwards and Page, 1994; Gimelfarb and Lande, 1994a; 1994b). Such studies have, in general, verified expressive efficiencies of MAS, particularly if it combines phenotypic and molecular information. However, it must be noted that all these reports considered a base population formed by crossing only two inbred lines, and so gametic phases between molecular markers and QTLs remained essentially the same across selection cycles. In recurrent selection of autogamous species, it is common to use more then two inbred lines to form the base population (Miller and Fehr, 1979; Payne et al., 1986; McFerson and Frey, 1991; Beaver and Kelly, 1994), and thereby posing a crucial problem for QTL detection, since a "mixture" of gametic phases between QTLs and molecular loci is almost sure to take place. Considering these points, the objective of this study was to evaluate marker-assisted recurrent selection in autogamous species through computer simulation, assuming genetically broad base populations, under different genetic and non-genetic parameter conditions.

MATERIAL AND METHODS

A fictitious diploid species, with complete selfing, was simulated considering a genome with 10 chromosome pairs of 1 Morgan each. To better simulate an actual crossing-over process, the mapping function proposed by Owen (1949) was considered, which is a chi-square density function with 4 degrees of freedom, divided by 4. This model takes interference (equal to 0.5) into account (among crossing points and also of crossing points with the centromere). Centromeres were randomly placed on the middle third of the chromosomes, since this seems to be the case for most cultivated autogamous species (Stebbins, 1971). In this genome, 50 fictitious QTLs were randomly distributed. Genetic effects (that is, half the difference between homozygotes) were generated using an exponential density function, with its parameter equal to 1. Only additive genetic action was considered. In this same genome, 13 marker loci were systematically distributed in each chromosome, totaling 130 loci in most situations.

The recurrent selection scheme was as follows: five inbred lines were simulated and crossed individually with two others (see Figure 1) in a circulant diallel fashion, thus forming a base-population composed of five crosses (subpopulations). Inbred lines had 20% of the loci with favorable alleles, and crosses were such that genetic dissimilarity among each pair of lines was held constant. This aimed at mimicking real situations, in the sense that lines of different genetic backgrounds are generally used. Within each cross, 400 S_0 plants were genotyped for the 130 marker loci. Progenies derived by selfing such plants were replicated twice and evaluated with respect to the simulated trait. Heritability of reference (h^2) was considered at S_0 plant level. For a given value of h^2, environmental effects were simulated, splitting environmental variance into, among and within-plot components. The relative magnitudes of such components were assigned considering an experimental area of intermediate heterogeneity, taking the coefficient of Smith (1938) of soil heterogeneity equal to 0.5.

Selection was carried out on an intra-subpopulational basis, and S_1 progenies were used as selection units (early selection), since this is a common practice in recurrent selection of autogamous species (Brim and Burton, 1979; Kenworthy and Brim, 1979; Miller and Fehr, 1979; Prohasca and Fehr, 1981; Busch and Kofoid, 1982; McFerson and Frey, 1991). Small experimental units were simulated (10 plants), since seed availability is generally low in early selection. Descendants from selected progenies formed 5 new subpopulations in the next selective cycle, by means of a systematic recombination process, to minimize genetic drift, as illustrated in Figure 1. For each situation, 6 cycles were simulated with 10 replications.

In a given subpopulation, MAS was always accomplished by means of the selection index of Lande and Thompson (1990):

$$I_i = b_Z Z_i + b_M M_i$$

where I_i is the index value of the ith progeny; Z_i is its average phenotypic value and M_i, called 'molecular net score' by Lande and Thompson (1990), is the sum

of the additive effects associated with alleles at marker loci, accounted for in S_0 plants which originated the S_1 progenies. The molecular score estimates were the predicted progeny values, considering the multiple regression of Z_i on the number of one of the alleles of each marker locus. The coefficients b_Z and b_M are relative weights given to Z_i and M_i, respectively; b_Z was taken as 1 and b_M as:

$$b_M = \frac{\frac{1}{H^2} - 1}{1 - p}$$

as suggested by Gimelfarb and Lande (1994a). In the above expression, H^2 is the heritability at the level of progeny means, and p is the proportion of additive genetic variance explained by molecular markers. A backward elimination procedure (Draper and Smith, 1981) was used to construct the multiple regression model of Z_i on molecular markers. Estimation of p was carried out as presented by Bearzoti and Vencovsky (1998).

MAS efficiency was measured by comparing the average progress attained to that found by selection based solely on phenotypic progeny means, considering the ratio of the observed gains as well as the significance (using Student's test) of the differences between them. The standard error of the differences were calculated using observed variances among the 10 replications of a given simulation. Linkage disequilibrium in a given cycle was quantified by the average of the absolute values of the disequilibrium between all pairs of polymorphic marker loci, 16.67 cM apart from one another, in S_0 generation. This average, in cycle 0, was actually the value observed in F_2 generations, and so it was taken as a reference value.

From this basic scheme, variants were considered, concerning heritability at S_0 plant level (0.1, 0.05 and 0.025); subpopulation size (400, 300 and 160 progenies); number of marker loci (130 and 200);

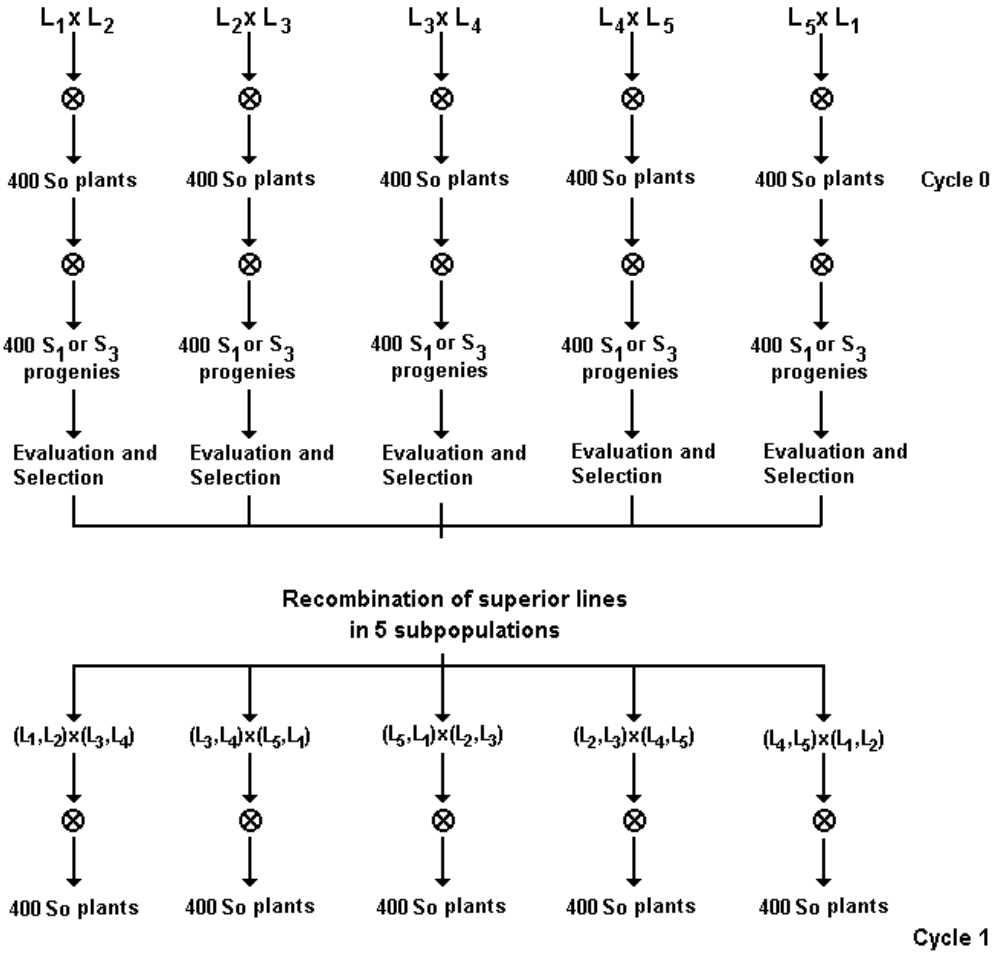

Figure 1. Recurrent selection scheme used in simulations, considering 5 inbred lines L_1, L_2, L_3, L_4 and L_5 and 5 crosses (subpopulations), and selective cycle 1. Later cycles were similarly structured (⊗: self-fertilization).

Smith's coefficient of soil heterogeneity (0.5 and 0.2); and number of inbred lines to initiate the recurrent selection program (5 and 20).

RESULTS AND DISCUSSION

Table 1 shows results of 6 simulated MAS selection cycles and the MAS efficiency over phenotypic selection for heritability at S_0 plant level (h^2) equal to 0.05, based on 400 S_1 progenies as selection units and 5% selected progenies per cycle. In cycle 0, heritability at the level of progeny means (H^2) averaged 0.231. This value was higher than h^2, as expected, due to the use of replications. Proportion p of genetic variance explained by molecular markers was considerably high in all selective cycles (Table 1). However, significant MAS efficiency was observed only in cycle 1. In subsequent cycles, it was close to 1. Linkage disequilibrium values are also presented in Table 1, and explain most of this lower efficiency. The population in cycle 0 consisted of inbred lines descendant from 5 crosses; therefore, these were actually 5 F_2 generations. The average linkage disequilibrium value observed in such generations (0.160) was taken as a reference (100%). Amongst the initial 5 crosses, different gametic phases between QTLs and marker loci must have occurred. In cycles 1 to 5, the recombination of selected progenies from different crosses might have caused mixtures of different gametic phases (repulsion and coupling), and this was probably the reason of the quite lower values of linkage disequilibrium in later cycles (close to 10%). It is remarkable that expected MAS gains, considering the values of H^2 and p in Table 1, are high in all cycles (values not shown). This expected efficiency would only be achieved if such values of H^2 and p were true; since the components of variance used in the selection index of Lande and Thompson (1990) are estimated, gain reductions are expected to occur.

The number of QTLs with unfavorable fixed alleles were equal to 0.60 in cycle 0 and 0.20 in cycle 1 (Table 1) because the recombination (Figure 1) and simulation were planned so that all QTLs could only show simultaneous segregation in cycle 2 or later. This was also valid for marker loci (Table 1). In later cycles, allele fixation rates were very low, and they must not have been an important factor for decreased selection gain. Frequencies of QTLs with favorable alleles fixed were also very low.

Tables 2 and 3 show the results of simulated MAS in 6 cycles of selection, considering the same conditions as Table 1, except that heritability h^2 at the S_0 plants level was equal to 0.025 and 0.1, respectively. It can be seen that values of H^2 increased accordingly to those of h^2. Again, gains from selection with MAS were not significantly different from those of phenotypic selection, in cycles 2 to 6. In cycle 1, however, MAS was efficient, with efficiencies being

Table 1. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.05. Populations consisted of 5 crosses (subpopulations); 400 S_1 progenies were evaluated in each subpopulation with 5% selection per cycle[1/].

Cycle	H^2	p	Ef	d	UF	FF	MF
0	0.231	0.881	—	0.160 (100%)	0.600	0.000	0.600
1	0.382	0.732	1.491[2/]	0.024 (15%)	0.200	0.000	0.200
2	0.435	0.745	1.074	0.017 (11%)	0.003	0.000	0.001
3	0.442	0.691	0.987	0.016 (10%)	0.001	0.000	0.013
4	0.386	0.750	0.970	0.016 (10%)	0.006	0.000	0.036
5	0.367	0.759	0.947	0.016 (10%)	0.001	0.003	0.074
6	0.297	0.794	0.935	0.017 (11%)	0.018	0.012	0.112

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed; [2/] significantly different from 1, at the 5% level.

equal to 1.735 and 1.343, for h^2 of 0.025 and 0.1, respectively. This strongly confirms that MAS may only be useful under low heritability, as theoretically expected (Lande, 1992). This trend was also true for identifying superior recombinant inbred lines for crossing, in the simulation study of Van Berloo and Stam (1998). The behavior of linkage disequilibrium and fixation rates across cycles were quite similar to those in Table 1.

Results in Tables 4 and 5 refer to the same conditions as those in Table 2; however, 300 and 160 progenies were used per cross in all cycles, respectively, instead of 400. Similar results regarding MAS efficiency can be seen in cycles 2 to 6. In cycle 1, gains from selection, though significantly higher than those of phenotypic selection, were progressively lower, as the number of evaluated progenies decreased (comparing Tables 2, 4 and 5). Similar results were obtained by Gimelfarb and Lande (1994a; 1994b), who considered population size the most important factor in MAS efficiency. The reason for this must be that it strongly affects the precision of estimating the coefficients in the index of Lande and Thompson (1990), since they are functions of variance components. Therefore, it can be said that, even in a single selection cycle, the use of MAS requires additional costs, not only for molecular marking, but also because of the large sample of individuals or progenies needed for evaluation. Indeed, required population sizes can be larger, perhaps, than those considered reasonable in conventional breeding of autogamous species. Fouilloux and Bannerot (1988), for example, showed that if one uses completely inbred lines as selection units in recurrent or non-recurrent selection for quantitative traits, sample sizes as large as those used in this study seem unjustified in most typical situations.

In cycles 2 to 6 the levels of linkage disequilibrium decreased (Tables 4 and 5), though to a lesser extent than those in Table 2. As population size decreased, fixation rates in QTLs and marker loci were higher than those in Table 2, wich may have contributed to the lowering of MAS efficiency.

One way to minimize the mixture of gametic phases between QTLs and marker loci in later cycles is the selection of a single progeny per subpopulation, since each cross of the following cycle would more closely resemble the base-population. This situation was tested considering h^2 equal to 0.025. Results are presented in Table 6. In cycle 1, the ratio between gains from selection with and without molecular information was equal to 1.304, which was lower than that observed in Table 2. The reason for this is not immediately clear, since reduction in number of progenies to sample a given population tail should lower gain from selection in both MAS and phenotypic selection. However, as the number of progenies progressively decreased, it was common to observe ties between their molecular scores (data not shown), which precluded identification of superior genotypes based on molecular information. Gimelfarb and Lande (1994 a), using a population of 1,000 individuals and 6 markers, also observed efficiency reductions when the selection proportion was reduced from 25% to 10%.

Table 2. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.025. Populations consisted of 5 crosses (subpopulations); 400 S_1 progenies were evaluated in each subpopulation with 5% selection per cycle[1/].

Cycle	H^2	p	Ef	d	UF	FF	MF
0	0.124	0.826	—	0.160 (100%)	0.600	0.000	0.600
1	0.219	0.815	1.735[2/]	0.024 (15%)	0.200	0.000	0.201
2	0.292	0.847	1.020	0.017 (11%)	0.002	0.000	0.010
3	0.274	0.880	1.016	0.017 (11%)	0.002	0.000	0.011
4	0.246	0.863	0.943	0.017 (11%)	0.003	0.000	0.028
5	0.247	0.876	0.897	0.017 (11%)	0.006	0.000	0.053
6	0.216	0.857	0.935	0.017 (11%)	0.012	0.002	0.083

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed;
[2/] significantly different from 1, at the 1% level.

Table 3. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.1. Populations consisted of 5 crosses (subpopulations); 400 S_1 progenies were evaluated in each subpopulation with 5% selection per cycle[1/].

Cycle	H^2	p	Ef	d	UF	FF	MF
0	0.387	0.862	—	0.160 (100%)	0.600	0.000	0.600
1	0.572	0.638	1.343[2/]	0.024 (15%)	0.200	0.000	0.200
2	0.600	0.647	0.998	0.017 (11%)	0.002	0.000	0.010
3	0.585	0.624	1.017	0.016 (10%)	0.001	0.000	0.014
4	0.517	0.649	0.962	0.016 (10%)	0.006	0.000	0.042
5	0.485	0.665	0.942	0.017 (11%)	0.010	0.007	0.080
6	0.403	0.683	0.975	0.017 (11%)	0.013	0.037	0.111

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed;
[2/] significantly different from 1, at the 1% level.

In Table 6, it can also be seen that use of a single selected progeny per subpopulation increased linkage disequilibrium (compared with Table 2), reaching values higher than 40% of the reference value. Nevertheless, this was not sufficient to make MAS efficient in cycles other than cycle 1. Actually, reduction in the proportion of selection led to higher fixation rates in QTLs and marker loci (Table 6). This may have precluded maintenance of further gains, by fixing unfavorable alleles at QTLs, and reduced the degree of polymorphism in marker loci, making it difficult to detect QTLs. Accordingly, proportions of genetic variance explained by molecular markers in Table 6 were lower than those in Table 2.

Two possible ways to minimize the impact of fixation when using single selected progenies per subpopulation are the use of a more saturated marker map and more lines to compose base-population (increase in effective number). A map of 200 marker loci systematically spaced and a base-population of 20 inbred lines were tested. Results are shown in Tables 7 and 8, respectively. In the latter situation, Smith's coefficient of soil heterogeneity was taken as 0.2. This reduced

Table 4. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.025. Populations consisted of 5 crosses (subpopulations); 300 S_1 progenies were evaluated in each subpopulation with 5% selection per cycle[1/].

Cycle	H^2	p	Ef	d	UF	FF	MF
0	0.167	0.672	—	0.160 (100%)	0.600	0.000	0.600
1	0.240	0.860	1.656[2/]	0.026 (16%)	0.200	0.000	0.200
2	0.265	0.963	1.032	0.019 (12%)	0.003	0.000	0.014
3	0.274	0.923	0.984	0.019 (12%)	0.002	0.000	0.020
4	0.259	0.901	0.953	0.019 (12%)	0.008	0.000	0.046
5	0.248	0.942	0.923	0.020 (13%)	0.014	0.000	0.075
6	0.230	0.907	0.982	0.021 (13%)	0.025	0.005	0.111

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed; [2/] significantly different from 1, at the 1% level.

the magnitude of environmental variance among plots, in an attempt to decrease H^2, and so establish a more favorable condition for MAS to be efficient. The use of 200 marker loci has indeed improved the proportion of genetic variance explained by molecular markers in all cycles, increasing MAS efficiency in cycle 1 from 1.304 to 1.425 (comparing Tables 6 and 7). However, in subsequent cycles, gain attained with MAS was not significantly different from that using phenotypic selection. Fixation rates were similar to those in Table 6 for QTLs and marker loci, but the increase in the absolute number of polymorphic markers was not sufficient to maintain MAS gains superior to those of phenotypic selection. Results from using 20 inbred lines and taking Smith's coefficient equal to 0.2 are presented in Table 8. The lower Smith's coefficient value reduced H^2 in all cycles, which improved MAS efficiency in cycle 1 (comparing Tables 6 and 8). The use of 20 inbred lines to form the base-population actually contributed toward reducing fixation in marker loci in late cycles, but it was still considerably high (above 50% in marker loci, from cycle 2 to 6). This may have been the main factor that made MAS non-efficient in selective cycles other than cycle 1.

Table 5. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.025. Populations consisted of 5 crosses (subpopulations); 160 S_1 progenies were evaluated in each subpopulation with 5% selection per cycle[1/].

Cycle	H^2	p	Ef	d	UF	FF	MF
0	0.155	0.715	—	0.160 (100%)	0.600	0.000	0.600
1	0.252	0.957	1.307[2/]	0.028 (18%)	0.201	0.000	0.203
2	0.242	0.966	1.009	0.025 (16%)	0.027	0.000	0.061
3	0.260	0.938	1.020	0.026 (16%)	0.033	0.000	0.089
4	0.279	0.977	1.118	0.027 (17%)	0.056	0.001	0.147
5	0.215	0.920	0.929	0.028 (18%)	0.069	0.004	0.182
6	0.190	0.947	0.933	0.030 (19%)	0.079	0.024	0.233

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed; [2/] significantly different from 1, at the 5% level.

Table 6. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.025. Populations consisted of 5 crosses (subpopulations); 400 S_1 progenies were evaluated in each subpopulation with 0.25% selection per cycle[1/].

Cycle	H^2	p	Ef	d	UF	FF	MF
0	0.125	0.727	—	0.160 (100%)	0.600	0.000	0.600
1	0.195	0.798	1.304[2/]	0.056 (35%)	0.340	0.000	0.424
2	0.207	0.697	1.121	0.072 (45%)	0.354	0.038	0.522
3	0.178	0.722	0.945	0.070 (44%)	0.357	0.075	0.570
4	0.141	0.629	0.0958	0.067 (42%)	0.359	0.143	0.628
5	0.114	0.532	0.835	0.060 (38%)	0.359	0.198	0.675
6	0.094	0.397	0.996	0.055 (34%)	0.369	0.262	0.715

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed; [2/] significantly different from 1, at the 5% level.

Table 7. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.025. Populations consisted of 5 crosses (subpopulations); 400 S_1 progenies were evaluated in each subpopulation with 0.25% selection per cycle, and 200 marker loci were used[1/].

Cycle	H^2	p	Ef	D	UF	FF	MF
0	0.122	0.793	—	0.161 (100%)	0.600	0.000	0.600
1	0.200	0.918	1.425[2/]	0.058 (36%)	0.354	0.000	0.419
2	0.220	0.848	1.122	0.068 (42%)	0.350	0.040	0.519
3	0.189	0.892	1.016	0.069 (43%)	0.369	0.087	0.586
4	0.150	0.653	1.083	0.059 (37%)	0.374	0.160	0.654
5	0.119	0.664	1.019	0.054 (34%)	0.374	0.210	0.698
6	0.094	0.556	0.883	0.046 (29%)	0.368	0.264	0.732

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed; MF and frequency of marker loci with one of the allelles fixed;
[2/] significantly different from 1, at the 1% level.

Table 8. Estimates of parameters observed in 10 replications of simulated MAS, considering h^2 equal to 0.025. Populations consisted of 20 crosses (subpopulations); 400 S_1 progenies were evaluated in each subpopulation with 0.25% selection per cycle[1/].

Cycle	H^2	p	Ef	D	UF	FF	MF
0	0.079	0.821	—	0.160 (100%)	0.600	0.000	0.600
1	0.109	0.806	1.681[2/]	0.058 (36%)	0.378	0.003	0.461
2	0.100	0.764	1.039	0.075 (47%)	0.435	0.030	0.526
3	0.109	0.726	1.006	0.074 (46%)	0.413	0.050	0.551
4	0.108	0.722	1.182	0.074 (46%)	0.365	0.062	0.551
5	0.100	0.739	0.760	0.070 (43%)	0.351	0.090	0.566
6	0.086	0.767	0.914	0.068 (43%)	0.353	0.121	0.583

[1/] H^2: heritability at progeny mean level; p: proportion of genetic variance explained by molecular markers; Ef: efficiency of MAS in relation to phenotypic selection; d: linkage disequilibrium; UF: frequency of QTL with unfavorable allelle fixed; FF: frequency of QTL with favorable allelle fixed and MF: frequency of marker loci with one of the allelles fixed;
[2/] significantly different from 1, at the 1% level.

The results in this study partially disagree with those of previous works, which attested to the high efficiency of MAS (e.g. Edwards and Page, 1994; Gimelfarb and Lande, 1994a; 1994b). The main reason for the disparity was that herein more than two inbred lines have been considered to form base populations, as usual in recurrent selection of autogamous species. In such situations, it was difficult to avoid the occurrence of mixtures of gametic phases among QTLs and marker loci, and consequently of linkage disequilibrium reduction, except when the population was formed by crosses of only 2 progenies selected in the previous cycle. However, in such cases the necessity of large population sizes in MAS caused very strong selection intensities with consequent high levels of fixation in QTLs and marker loci. The impact of high fixation rates was not diminished even by using a more saturated map or more inbred lines to form base-population, in an attempt to increase effective

size. These results clearly showed that finding situations for MAS to be feasible in recurrent selection of autogamous species is a difficult task.

ACKNOWLEDGEMENTS

This work was supported by CNPq.

RESUMO

Simulação de seleção recorrente assistida por marcadores em espécies autógamas

Este estudo utilizou simulação em computador para avaliar a seleção assistida por marcadores (SAM) na seleção recorrente em espécies autógamas. Populações base foram simuladas considerando o cruzamento de cada uma de cinco linhagens endogâmicas com duas outras. Em cada cruzamento, simulou-se a genotipagem de 400 plantas S_0 em relação a 130 locos marcadores distribuídos em um genoma fictício contendo 50 locos controladores de uma característica quantitativa (QTLs) aleatoriamente dispostos. Progênies derivadas de tais plantas foram tomadas como unidades de seleção. A SAM era implementada pelo índice de seleção de Lande e Thompson (1990). Em geral, a SAM não foi eficiente após o primeiro ciclo de seleção. Neste, contudo, a SAM foi eficiente, mas marcadamente afetada pelo tamanho da população e pela herdabilidade. A seleção de uma única progênie por cruzamento possibilitou o aumento do desequilíbrio de ligação, mas causou taxas de fixação gênica muito elevadas, não diminuídas nem pelo uso de 200 locos marcadores, nem pelo de 20 linhagens endogâmicas.

REFERENCES

Apuya, N.; Frazier, B.L.; Keim, P.; Roth, E.J. and Lark, K.G. 1988. Restriction length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merr. Theoretical and Applied Genetics. 75:889-901.

Bearzoti, E. and Vencovsky, R. 1998. Estimation of the proportion of genetic variance explained by molecular markers. Genetics and Molecular Biology. 21:557-566.

Beaver, J.S. and Kelly, J.D. 1994. Comparison of selection methods for dry bean populations derived from crosses between gene pools. Crop Science. 34:34-37.

Brim, C.A. and Burton, J.W. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Science. 19:494-498.

Busch, R.H. and Kofoid, K. 1982. Recurrent selection for kernel weight in spring wheat. Crop Science. 22:568-572.

Draper, N.R. and Smith, H. 1981. Applied regression analysis. John Wiley, New York.

Edwards, M.D. and Page, N.J. 1994. Evaluation of marker-assisted selection through computer simulation. Theoretical and Applied Genetics. 88:376-382.

Fouilloux, G. and Bannerot, H. 1988. Selection methods in the common bean (*Phaseolus vulgaris*). p. 503-542. In: Gepts, P. (Ed.). Genetic Resources of *Phaseolus* Beans Kluwer. Academic Publishers, Dordrecht.

Fraser, A.S. 1957. Simulation of genetic systems by automatic digital computers. I. Introduction. Australian Journal of Biological Science. 10:484-491.

Gimelfarb, A. and Lande, R. 1994a. Simulation of marker assisted selection in hybrid populations. Genetical Research. 63:39-47.

Gimelfarb, A. and Lande, R. 1994b. Simulation of marker assisted selection for non-additive traits. Genetical Research. 64:127-136.

Kenworthy, W.J. and Brim, C.A. 1979. Recurrent selection in soybeans. I. Seed yield. Crop Science. 19:315-318.

Khadr, F.H. and Frey, K.J. 1965. Effectiveness of recurrent selection in oat breeding (*Avena sativa* L.). Crop Science. 5:349-354.

Kochert, G. 1994. RFLP technology. DNA-based markers in plants. p.8-38. In: Phillips, R.L. and Vasil, I.K. (Eds.). Kluwer Academic Publishers, Dordrecht.

Lande, R. 1992. Marker-assisted selection in relation to traditional methods of plant breeding. p.437-451. In: Stalker, H.T. and Murphy, J.P. (Eds.). Plant Breeding in the 1990's. CAB International, Wallingford.

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

Matzinger, D.F. and Wernsman, E.A. 1968. Four cycles of mass selection in a synthetic variety of an autogamous species *Nicotiana tabacum* L. Crop Science. 8:239-243.

McFerson, J.K. and Frey, K.J. 1991. Recurrent selection for protein yield of oat. Crop Science. 31:1-8.

Miller, J.E. and Fehr, W.R. 1979. Direct and indirect selection for protein in soybeans. Crop Science. 19:101-106.

Owen, A.R.G. 1949. The theory of genetical recombination. I. Long-chromosome arms. Proceedings of the Royal Statistical Society Series B. 136:67-94.

Payne, T.S.; Stuthman, D.D.; McGraw, R.L. and Bregitzer, P.P. 1986. Physiological changes associated with three cycles of recurrent selection for grain yield improvement in oats. Crop Science. 26:734-736.

Phillips, R.L. and Vasil, I.K. 1994. DNA-based markers in plants. Kluwer Academic Publishers, Dordrecht.

Pomeranke, G.J. and Stuthman, D.D. 1992. Recurrent selection for increased grain yield in oat. Crop Science. 32:1184-1187.

Prohasca, K.R. and Fehr, W.R. 1981. Recurrent selection for resistance to iron deficiency chlorosis in soybeans. Crop Science. 21:524-526.

Smith, H.F. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. Journal of Agricultural Science. 28:1-23.

Stebbins, G.L. 1971. Chromosomal evolution in higher plants: a series of student texts in contemporary biology. Addison-Wesley, Reading.

Van Berloo, R. and Stam, P. 1998. Marker-assisted selection in autogamous RIL populations: a simulation study. Theoretical and and Applied Genetics. 96:147-154.

Werner, B.K. and Wilcox, J.R. 1990. Recurrent selection for yield in *Glycine max* using genetic male-sterility. Euphytica. 50:19-26.

Zhang, W. and Smith, C. 1992. Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theoretical and Applied Genetics. 83:813-820.

Received: May 17, 2001;

Accepted: December 17, 2001.