Theor Appl Genet (1995) 91:522–528 

A. Gimelfarb · R. Lande

# Marker-assisted selection and marker-QTL associations in hybrid populations

Received: 19 December 1994 / Accepted: 5 January 1995

**Abstract** A detailed analysis is presented of the relationship between genetic markers and quantitative trait loci (QTLs) in the process of marker-assisted selection (MAS). We simulated MAS employing a multiple linear regression to chose from among all of the markers in the genome those to be utilized by selection and to estimate their associated effects on the trait. The simulations demonstrate that, even when such selection is quite effective, the markers utilized by selection are not necessarily the most tightly linked to the QTLs controling the trait. Moreover, the additive effects associated with the markers estimated by the regression may not accurately reflect the contributions to the trait by the most tightly linked QTLs.

**Key words** Marker-QTL association · Marker-assisted selection

## Introduction

For selection utilizing genetic markers to be effective in improving a quantitative trait, the following two processes are usually thought to be essential: detecting markers closely linked to QTLs (mapping QTLs) and determining the effects of these markers on the trait (the contributions to the trait by QTLs linked to the markers). A number of methods for mapping QTLs and estimating their effects have been suggested and investigated (Lander and Botstein 1988; Paterson et al. 1988, 1990; Knapp et al. 1990; Knott and Haley 1992; Haley and Knott 1992; Darvasi et al. 1993; Dudley 1993; Zeng 1993, 1994).

Lande and Thompson (1990) proposed a method of marker-assisted selection (MAS) in which genetic markers are not used to actually map QTLs, but rather a multiple linear regression of the phenotype on markers is employed to choose markers associated with QTLs and to estimate their effects on the character. The chosen markers are those that account for the major part of the regression variance, and the coefficient of a marker in the regression is regarded as its associated effect. In an infinitely large population, the markers chosen in this manner are expected to be the most closely linked to QTLs among all the markers in the genome; and their associated effects are expected to be estimates of the contributions to the character by QTLs. Lande and Thompson derived a selection index incorporating the individual's phenotype as well as the "molecular score", i.e., the sum of the associated effects of the chosen markers in the individual's genotype. Our previous simulations (Gimelfarb and Lande 1994a, b), as well as those of Zhang and Smith (1992, 1993), demonstrated that in populations obtained by a cross between inbred lines, such a selection index can be quite effective for a wide range of parameters, particularly when the population size is sufficiently large.

Results from previous simulations indicated that MAS in a finite population, even when it is quite efficient, does not necessarily utilize markers that are most closely linked to QTLs; and the effect of a marker obtained by multiple regression does not necessarily represent an estimate of the contribution to the character by a linked QTL. In many simulations, for example, the marker accounting for the largest proportion of the regression variance was not the nearest to a QTL; and the effects of markers included in the molecular score did not match even approximately the contributions to the character by QTLs. Also, increasing the number of markers per chromosome from 6 to 51, and, hence, allowing the regression to chose markers more tightly linked to QTLs, had relatively little effect on selection efficiency and could even produce a weaker response to selection.

In the present paper, we report the results of a more detailed investigation of the relationships between

Communicated by C. Smith

A. Gimelfarb · R. Lande (✉)
Department of Biology, University of Oregon Eugene, OR 97403-1210, USA

markers and QTLs, in populations of finite size, in the process of marker-assisted selection based on the method of Lande and Thompson (1990).

## Materials and methods

General aspects of simulations of marker-assisted selection are discussed in Gimelfarb and Lande (1994 a, b). The genetic map used in the majority of the simulations reported here had seven chromosomes of 100 centimorgans each. There were 15 QTLs contributing additively to a quantitative character. The distribution of the QTLs on the chromosomes can be seen in Figs. 1–4 where the position of a QTL is indicated by a diamond. Each QTL was assumed to have two alleles contributing either positively or negatively to the character. The absolute values of allelic contributions are shown in Figs. 1–4 under the QTL symbols. The total contribution by a chromosome (the sum of the allelic contributions by all QTLs on the chromosome) in the corresponding gametic phase is indicated in parentheses next to the chromosome number. All chromosomes have the same total contribution of 1.00 in the coupling gametic phase.

Besides QTLs, each chromosome also carried genetic markers. In the majority of simulations, the number of markers was the same on any chromosome, and they were equidistantly spaced along it. Simulations were also conducted with the number of markers on a chromosome equal to the number of QTLs with each marker occupying exactly the same position as a QTL. This case corresponds to a situation where a precise location of a QTL (but not its contribution to the character) is available prior to selection.

We also considered genetic maps with only one chromosome carrying QTLs on it (either one QTL or three). In some simulations, this was the only chromosome in the genotype, and it carried 11 genetic markers in addition to QTLs. In other simulations, the genotype consisted of the total of ten chromosomes with only one having a QTL on it and with the rest carrying genetic markers but no QTLs.

Recombination was modelled as described in our previous paper (Gimelfarb and Lande 1994a). No more than two crossovers were allowed for any pair of homologous chromosomes, but the positions of crossovers were random. Thus, there was interference with respect to the number of crossovers but not with respect to their relative position. The mapping function entailed by this recombination model was very close to Haldane's mapping function.

Selection was applied to an initial population representing the $F_2$ generation of a cross between two inbred lines. Consequently, the initial allelic frequencies were 0.5 for any QTL and any genetic marker. Two gametic phases of QTLs in the initial populations were considered: total coupling (the contributions to the character by adjacent QTLs on a chromosome are in the same direction) and total repulsion (the contributions by adjacent QTLs are in opposite directions). The phenotype, $Z$, of an individual was computed as

$$Z = g + e, \tag{1}$$

where $g$ is the individual's genotypic value, and $e$ is a random environmental contribution (normally distributed with zero mean and variance $v_e$). The genotypic value is the sum of the allelic contributions by all of the QTLs in the individual's genome. The environmental variance, $v_e$, was chosen so as to yield a given expected value of the heritability in a population with the allelic frequencies the same as in the initial population, but at linkage equilibrium. The heritability was $h^2 = 0.1$ in all simulations.

Selection was based on an index incorporating the additive effects of genetic markers as well as the phenotypic value of the trait. A fixed proportion (25%) of individuals with the highest value of the index were selected. Similar to our previous simulations with large population size (Gimelfarb and Lande 1994a, b), the relative weight of markers in the index was so high that selection was practically independent of phenotypes and based on markers only. The additive effects of markers were obtained by a two-stage linear regression of the phenotype on markers as described by Gimelfarb and Lande (1994a). In the first stage, a separate regression is fitted for each chromosome, and five markers with the highest contribution to the $R^2$ value of the regression are selected from a chromosome using stepwise multiple regression. In the second stage, a regression is fitted using all of the markers selected in the first stage from all chromosomes, and a fixed number of markers with the highest contribution to the $R^2$ value is chosen for inclusion in the selection index. The number of markers included in the index was limited to seven in the majority of simulations, although some simulations were conducted with 15 markers in the index. The regression coefficient of a marker included in the index was regarded as the additive effect associated with the marker.

Only one generation of selection is considered in this paper. A result reported for a given set of parameters represents an average over 100 replicated computer runs.

## Results and discussion

Table 1 shows responses in the first generation of MAS for a quantitative character controlled by 15 QTLs with the genetic map as in Figs. 1–4. The headings Coupling and Repulsion refer to total coupling and total repulsion initial gametic phases. Population size, N, refers to the number of individuals of each sex. Standard errors of the mean responses in Table 1 are in the range of 0.5% to 3% of the reported values.

Columns MQ15 and MQ7 in Table 1 represent simulations with the genetic map having each marker located at the same position as a QTL. This corresponds to the ultimate goal of QTL mapping utilizing genetic markers, namely to identify markers that are so tightly linked to QTLs that "selection will then be equivalent to selection on the QTLs themselves" (Zhang and Smith 1992). The difference between the columns MQ15 and MQ7 is the former case all 15 markers in the genome were used to compute the molecular score, whereas in the latter case the molecular score was computed based on the seven markers with the highest contribution to the $R^2$ value of the regression. It may seem obvious that responses in column MQ15 should represent the strongest possible response to MAS. It is important to keep in mind, however, that, even if all QTLs in a genome are

**Table 1** Response of the mean phenotype in the first generation of MAS (N is the number of individuals of each sex; MQ15 and MQ7 correspond to map with markers only at the exact position of QTLs; the number of markers included in the selection index is 7, except for Q15, in which case it is 15)

| N | MQ15 | MQ7 | Markers/chromosome | | | | |
|---|---|---|---|---|---|---|---|
| | | | 6 | 11 | 21 | 51 | 101 |
| Coupling | | | | | | | |
| 100 | 1.63 | 1.63 | 1.23 | 1.22 | 1.25 | 1.15 | 1.12 |
| 200 | 1.87 | 1.79 | 1.41 | 1.45 | 1.39 | 1.36 | 1.36 |
| 500 | 2.05 | 1.82 | 1.62 | 1.67 | 1.67 | 1.66 | 1.65 |
| 1000 | 2.12 | 1.98 | 1.74 | 1.74 | 1.82 | 1.80 | 1.80 |
| Repulsion | | | | | | | |
| 100 | 0.80 | 0.81 | 0.60 | 0.56 | 0.57 | 0.55 | 0.60 |
| 200 | 1.02 | 0.99 | 0.74 | 0.75 | 0.70 | 0.70 | 0.70 |
| 500 | 1.21 | 1.19 | 0.98 | 0.98 | 0.98 | 0.96 | 0.93 |
| 1000 | 1.30 | 1.27 | 1.10 | 1.11 | 1.16 | 1.13 | 1.13 |

mapped precisely by markers, selection is based not on individual genotype, but rather on the molecular score, i.e., the sum of effects of the markers estimated by the coefficients in the linear regression. The error in an estimate of a marker effect depends on the number of markers used to compute the molecular score: the more markers used the greater will be the error. Hence, it is not obvious that using fewer markers to compute the molecular score should yield a weaker response to MAS in a population of a given size, even if each marker is located at the same position as a QTL. Comparing columns MQ15 and MQ7, it can be seen that reducing the number of markers contributing to the molecular score from 15 to 7 generally produces a weaker response. However, the differences between the responses are small, particularly for the repulsion gametic phase, and there are no differences between the responses if the population size is 100 individuals of each sex.

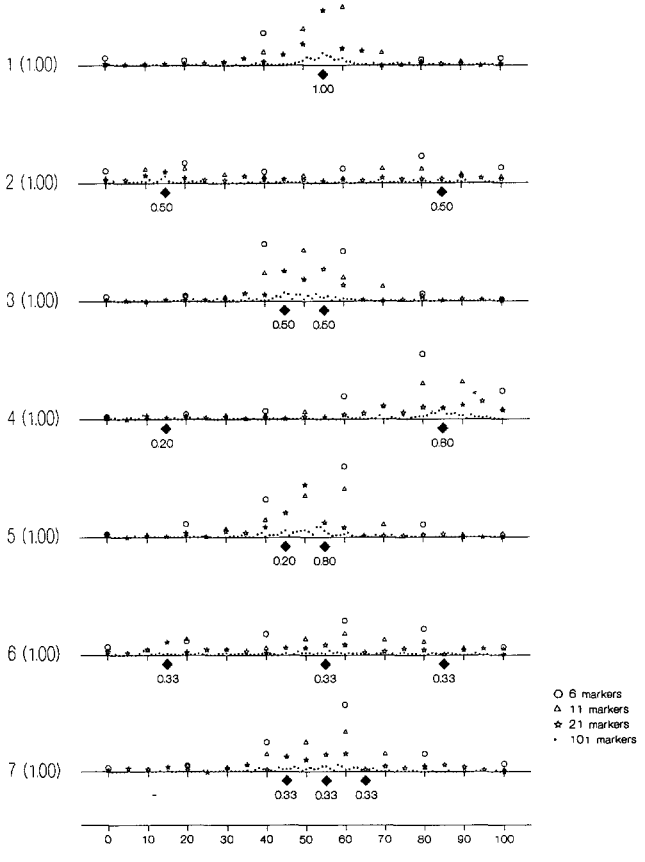
The last five columns in Table 1 represent a more realistic situation in which a precise map of QTLs is not available, and each chromosome carries the same number (6, 11, 21, 51 or 101) of equidistant genetic markers. A comparison between these columns reveals that increasing the number of markers on a chromosome does not necessarily lead to a stronger response to MAS based on the molecular score. There appears to be an optimum number of markers per chromosome producing the highest response in a population of a given size and initial gametic phase, but the differences between the columns are quite small and mostly not significant statistically.

The molecular score in the simulations represented by the last five columns in Table 1 was computed based on the seven markers with the highest contribution to the $R^2$ of the regression. In simulations (data not shown) either with only three or with 15 markers contributing to the molecular score, response to MAS were always weaker than the response in Table 1.

Figures 1 and 2 illustrate the probabilities that particular genetic markers are included, based on their contributions to the $R^2$ of the regression, among the seven markers contributing to the molecular score in a population of 500 individuals of each sex. Each marker is denoted by a symbol corresponding to a particular number of markers per chromosome. The coordinate on the horizontal axis of a marker indicates the position of the marker on the chromosome. The probability that a marker contributes to the molecular score is proportional to the distance from the symbol of the marker to the horizontal axis. The probability for a marker to be included in the molecular score is generally higher for markers located nearer to a QTL. However, this probability is not determined solely by the distance between the marker and the nearest QTL but is affected also by the total number of markers on the chromosome, as well as by the number of QTLs on the chromosome, by the effects of the QTLs on the character, by the distances between the QTLs, and by the gametic phase. Thus, the type of results illustrated in the figures should apply to outbreeding populations in general, but on a smaller scale of genetic distance over which there is significant linkage disequilibrium.

Figures 1 and 2 show clearly that the probability that a particular marker contributes to the molecular score declines rapidly if more markers are added on the chromosome. It is also of interest to know how adding markers on a chromosome affects the probability that a segment of the chromosome (rather than an individual marker) is chosen by the regression to contribute to the molecular score. Chromosomes in Figs. 3 and 4 are divided into segments of 10 cM each (except for the segments at both ends of a chromosome whose length is 5 cM). The scale at the bottom of the figures indicates positions of the mid-points of the segments. Each segment contains only one marker if the chromosome carries 11 markers, whereas there are ten markers in a segment (except for the two extreme segments) in the case of 101 markers per chromosome. The height of a bar in Figs. 3 and 4 is proportional to the probability that a chromosomal segment (i.e., for at least one marker from the segment) contributes to the molecular score, given that the seven markers with the highest contribu-

**Fig. 1** The probability that a marker on a chromosome in *coupling* phase is chosen to contribute to the molecular score (*diamonds* indicate QTLs with their allelic contributions shown below. The scale on the horizontal axes is in centimorgans. The distance between two adjacent horizontal axes corresponds to a probability of 0.8)



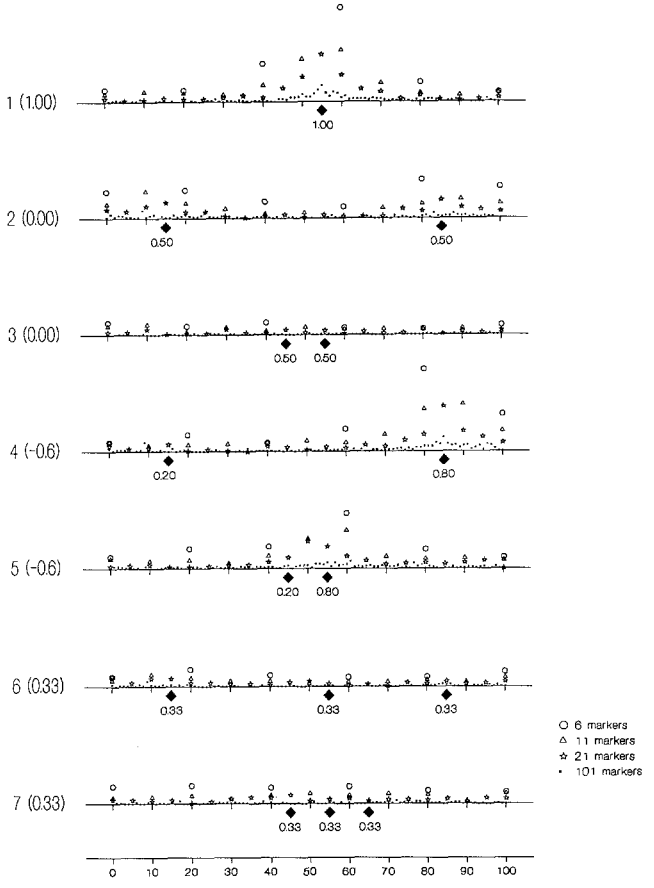



**Fig. 2** The probability that a marker on a chromosome in *repulsion* phase is chosen to contribute to the molecular score (symbols and scaling as in Fig. 1)

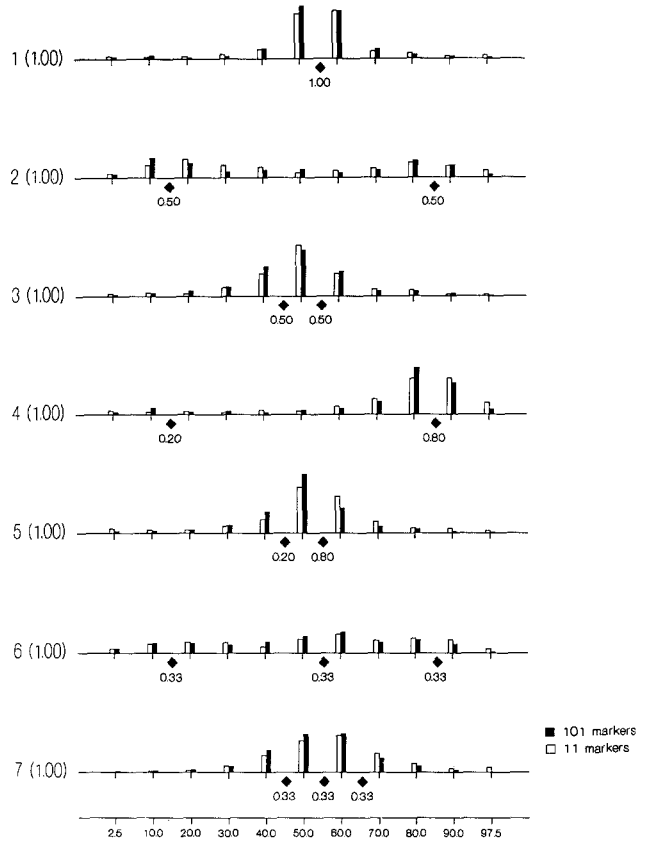



**Fig. 3** The probability that a segment of a chromosome in *coupling* phase is chosen to contribute to the molecular score (positions, in centimorgans, of the middle points of segments are shown on the lower scale. *Diamonds* indicate QTLs with their allelic contributions shown below. The distance between two adjacent horizontal axes corresponds to a probability of 0.8)

tion to the $R^2$ value of the regression are chosen. For a segment of 10 cM, this probability is not much affected by the number of markers in the segment.

An important conclusion from Figs. 1–4 is that, in a population of a given size, regression of the phenotype on genetic markers in the $F_2$ generation of a cross between two inbred lines is not capable of discriminating between different markers if the distance between them is less than a particular value (approximately 10 cM in a population of 500 individuals of each sex for the recombination model used in our simulations). Knott and Haley (1992) investigated maximum-likelihood methods for mapping of one and two QTLs on a chromosome with equidistant genetic markers. They found that there was very little difference in a population of 1000 individuals between estimates of parameters obtained with markers 20 cM apart and those obtained with markers 10 cM apart. In the work on mapping QTLs using likelihood methods reported by Darvasi et al. (1993), the power of a test for detecting a QTL was practically the same with markers 10 cM apart as with markers only 0.1 cM apart in populations of 500 and 1000 individuals.

The reason why a regression of the phenotype on genetic markers cannot discriminate between closely linked markers is that crossing two inbred lines creates linkage disequilibria not only between markers and QTLs but also between different markers. This causes colinearity in the regression, and increasing the number of markers per chromosome entails more colinearity. Consequently, even though adding more markers shortens the distance between a QTL and the nearest marker, this does not mean that the nearest marker will have the highest contribution to $R^2$, and, hence, will be chosen by the regression to contribute to the molecular score. Colinearity can be alleviated by allowing for a few generations of random mating prior to selection. However, random mating will also reduce the linkage disequilibria between the markers and QTLs that MAS utilizes. Simulations (data not shown) with 5, 10 and 20 generations of random mating before selection demonstrated that, regardless of the number of markers per chromosome, random mating prior to selection always reduces the response to MAS. A similar finding emerged in our previous simulations (Gimelfarb and Lande 1994a) as well as in those of Zhang and Smith (1992).

Table 2 shows the probabilities that a particular number of markers from a given chromosome are chosen by the regression among the seven markers

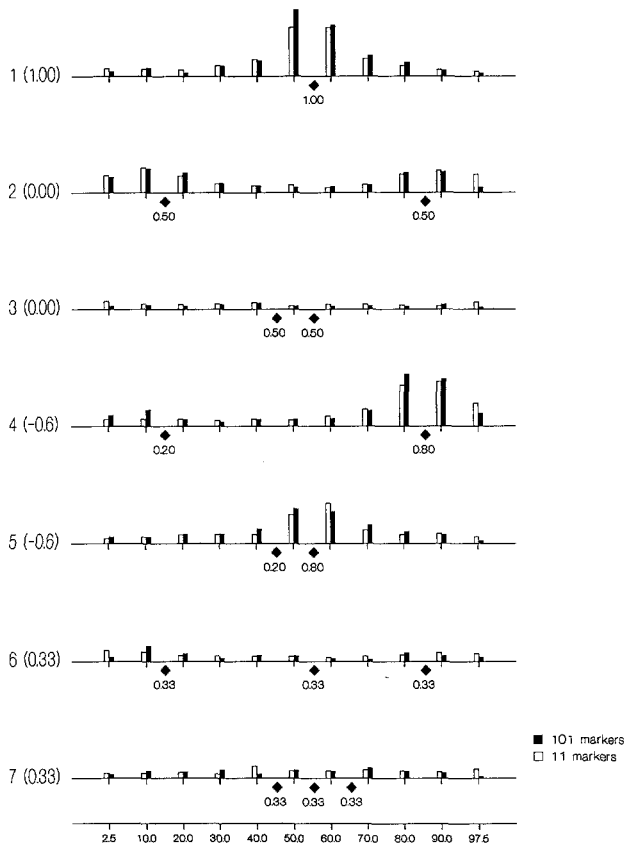



**Fig. 4** The probability that a segment of a chromosome in *repulsion* phase is chosen to contribute to the molecular score (symbols and scaling as in Fig. 3)

contributing to the molecular score. A comparison between the columns corresponding to 11 and 101 markers per chromosome reveals that the probability that a chromosome contributes markers to the molecular score is determined not so much by the number of markers on the chromosome as by the contribution to the phenotype by all QTLs on the chromosome. Indeed, any chromosome in coupling phase makes the same contribution to the phenotype (Fig. 1), and the probabilities in Table 2 are also roughly similar among chromosomes in coupling phase. On the other hand, chromosomes in repulsion phase contribute differentially to the phenotype (Fig. 2), and the probabilities in Table 2 differ substantially among such chromosomes. Moreover, the probability that a chromosome in repulsion phase contributes markers to the molecular score depends not only on the total contribution by the chromosome to the phenotype, but also on the relative position of the QTLs. For example, the contributions by chromosomes 1 and 2 in repulsion phase are, respectively, 1.00 and 0.00 (Fig. 2). Yet, the difference between the probabilities that the two chromosomes contribute a marker to the molecular score are not dramatically different (0.219 vs 0.153). On the other hand, the contribution by chromosome 3 in repulsion phase is zero, i.e., the same as by chromosome 2, yet the probability of contributing a marker to the molecular score is more than twice as high for chromosome 2 as for chromosome 3. This can be attributed to the fact that the distance between QTLs on chromosome 2 is greater than the distance between QTLs on chromosome 3 (Fig. 2).

The probabilities that two or more markers from a chromosome are included among the seven markers contributing to the molecular score are very low in coupling phase (Table 2). They are substantially higher in repulsion phase, although there does not seem to be a clear relationship between this probability and the number of QTLs on a chromosome.

Table 3 demonstrates some parameters of markers contributing to the molecular score. Columns denoted as *a1* show the average effect associated with a marker, if

**Table 2** The probability that a given number of markers from a chromosome is chosen to contribute to the molecular score

| Chrm. | 11 markers/chromosome | | | | 101 markers/chromosome | | | |
|---|---|---|---|---|---|---|---|---|
| | Markers in molecular score | | | | Markers in molecular score | | | |
| | ≥1 | 1 | 2 | ≥3 | ≥1 | 1 | 2 | ≥3 |
| | Coupling | | | | | | | |
| 1 | 0.147 | 0.140 | 0.006 | 0.001 | 0.152 | 0.139 | 0.013 | 0.000 |
| 2 | 0.128 | 0.121 | 0.007 | 0.000 | 0.123 | 0.114 | 0.009 | 0.000 |
| 3 | 0.148 | 0.143 | 0.005 | 0.000 | 0.154 | 0.146 | 0.008 | 0.000 |
| 4 | 0.143 | 0.139 | 0.004 | 0.000 | 0.139 | 0.130 | 0.008 | 0.001 |
| 5 | 0.152 | 0.145 | 0.006 | 0.000 | 0.152 | 0.143 | 0.009 | 0.000 |
| 6 | 0.132 | 0.127 | 0.005 | 0.000 | 0.126 | 0.113 | 0.012 | 0.001 |
| 7 | 0.150 | 0.146 | 0.003 | 0.001 | 0.154 | 0.144 | 0.009 | 0.001 |
| | Repulsion | | | | | | | |
| 1 | 0.219 | 0.160 | 0.050 | 0.009 | 0.245 | 0.158 | 0.065 | 0.022 |
| 2 | 0.153 | 0.078 | 0.064 | 0.011 | 0.144 | 0.062 | 0.065 | 0.017 |
| 3 | 0.067 | 0.050 | 0.014 | 0.003 | 0.048 | 0.028 | 0.014 | 0.006 |
| 4 | 0.196 | 0.132 | 0.057 | 0.007 | 0.216 | 0.124 | 0.065 | 0.023 |
| 5 | 0.172 | 0.131 | 0.035 | 0.006 | 0.175 | 0.104 | 0.051 | 0.020 |
| 6 | 0.091 | 0.075 | 0.014 | 0.002 | 0.080 | 0.055 | 0.018 | 0.007 |
| 7 | 0.102 | 0.086 | 0.014 | 0.002 | 0.092 | 0.065 | 0.021 | 0.006 |

**Table 3** Some parameters of markers contributing to the molecular score (*a1* effect of a single marker; *a2* summary effect of two markers; *d* distance between two markers; *cr* correlation between effects of two markers)

| Chrm | 11 markers/chromosome | | | | 101 markers/chromosome | | | |
|---|---|---|---|---|---|---|---|---|
| | *a1* | *a2* | *d* | *cr* | *a1* | *a2* | *d* | *cr* |
| | Coupling | | | | | | | |
| 1 | 1.05 | 1.17 | 40.0 | −0.98 | 1.10 | 0.86 | 19.2 | −0.98 |
| 2 | 0.91 | 1.52 | 63.8 | −0.89 | 0.98 | 1.21 | 32.8 | −0.99 |
| 3 | 1.04 | 0.87 | 39.1 | −0.88 | 1.09 | 0.88 | 14.5 | −0.99 |
| 4 | 0.99 | 1.30 | 56.0 | −0.92 | 1.04 | 1.02 | 33.0 | −0.97 |
| 5 | 1.04 | 0.83 | 44.3 | −0.79 | 1.11 | 0.99 | 25.7 | −0.91 |
| 6 | 0.90 | 1.27 | 52.7 | −0.92 | 0.99 | 0.94 | 15.8 | −0.96 |
| 7 | 1.02 | 1.20 | 32.5 | −0.94 | 1.10 | 0.85 | 15.4 | −0.96 |
| | Replusion | | | | | | | |
| 1 | 1.03 | 0.86 | 32.9 | −0.77 | 1.10 | 0.89 | 19.7 | −0.98 |
| 2 | 0.04 | 0.03 | 60.5 | −0.99 | 0.01 | 0.03 | 41.6 | −0.99 |
| 3 | 0.10 | 0.02 | 34.2 | −1.00 | −0.07 | 0.12 | 15.2 | −1.00 |
| 4 | −0.81 | −0.50 | 49.5 | −0.82 | −0.89 | −0.57 | 35.7 | −0.98 |
| 5 | −0.77 | −0.64 | 32.1 | −0.82 | −0.85 | −0.60 | 15.9 | −0.98 |
| 6 | 0.47 | 0.20 | 32.8 | −0.94 | 0.68 | 0.49 | 27.8 | −0.95 |
| 7 | 0.57 | 0.49 | 38.8 | −0.94 | 0.77 | 0.52 | 12.4 | −0.99 |

it is the only marker from the corresponding chromosome. These effects are quite similar to the total contributions by chromosomes to the phenotype (Figs. 1 and 2). The rest of the columns in the table show parameters of two markers contributing to the molecular score, if they come from the same chromosome. Parameter *a2* is the sum of the average effects associated with the markers, whereas parameters *d* and *cr* are, respectively, the average distance (in centimorgans) between the two markers and the average correlation between their associated effects. Similar to the effect associated with a single marker, the sum of the effects associated with two markers corresponds roughly to the total contribution by a chromosome to the phenotype. This does not imply, however, that the effect associated with each marker reflects the contribution by a QTL to the phenotype. Indeed, the correlation between the effects associated with two markers is close to −1 in almost all cases, meaning that the associated effects of the markers are usually large and of opposite sign, irrespective of the number of QTLs on the chromosome and their gametic phase. It should be noted that a similar pattern of the associated effects of markers occurred also in runs with maps having all markers located at exactly the same position as the QTL, and with the molecular score computed based on all of the markers in the genome (column MQ15 in Table 1). In one such run, for example, the associated effects were −3.15 and 2.14 for the markers located at the first and the second QTL, respectively, on chromosome 3. Thus, the associated effect of a marker included in the molecular score may not accurately estimate the contribution to the phenotype by a QTL.

So far we have considered genotypes with each chromosome carrying at least one QTL. This, however, may not always be the case, and there can be instances in which some chromosomes do not carry a QTL controlling a particular trait. To investigate the role of unlinked markers in marker-assisted selection we examined this case as well as the situation in which a chromosome carries QTLs but does not have any genetic markers. Table 4 presents responses to MAS in simulations with genotypes having one chromosome carrying QTLs (either one located at 50 cM, or three located at 15 cM, 50 cM and 85 cM). The contribution to the character was 1.0 by a unique QTL and 1/3 by each of the three QTLs. Headings of the columns in the table indicate the total number of chromosomes in the genotype (either 1 or 10) as well as the number of markers contributing to the molecular score (either 3 or 6). In simulations represented by columns 1 (3) and 1 (6), the chromosome with a QTL was the only chromosome in the genotype, and it carried 11 equidistant markers in addition to QTLs. In other simulations, the genotype consisted of ten chromosomes, only one of which had a QTL, whereas the rest carried 11 equidistant markers but no QTLs. In columns denoted 10 (3) and 10 (6), the chromosome with a QTL also carried 11 equidistant markers, whereas in columns 10 (3)* and 10 (6)*, the chromosome with a QTL did not have genetic makers on it. The length of

**Table 4** Response of the mean phenotype in the first generation of MAS by a character controlled by QTLs located on one chromosome (percentages refer to the proportion of the markers in the molecular score coming from chromosomes without QTLs)

| No. QTLs | Total number of chromosomes (markers in molecular score) | | | | | |
|---|---|---|---|---|---|---|
| | 1 (3) | 10 (3) | 10 (3)* | 1 (6) | 10 (6) | 10 (6)* |
| One QTL | 0.89 | 0.85 (58%) | 0.15 | 0.88 | 0.75 (72%) | 0.12 |
| Three QTLs, coupling | 0.69 | 0.67 (19%) | 0.26 | 0.69 | 0.65 (53%) | 0.22 |
| Three QTLs, repulsion | 0.26 | 0.18 (44%) | 0.02 | 0.26 | 0.17 (59%) | 0.02 |

* Markers are only on chromosomes without QTLs

any chromosome was 100 centimorgans. Percentages in the table refer to the proportion among the markers contribution to the molecular score of those coming from chromosomes that do not carry a QTL. This proportion is, obviously, 0% if there is only one chromosome in the genome, or 100% if genetic markers are located only on chromosomes without a QTL.

One of the consequences of increasing the number of markers in a genome, as discussed earlier, is more colinearity in the regression of the phenotype on markers due to linkage disequilibrium between markers on the same chromosome. Another consequence of increasing the number of markers is that in a population of finite size it will cause more associations between QTLs and markers that are not linked to any QTL at all. It seems obvious that such spurious associations between markers and QTLs represent "noise" and, hence, should reduce the response to MAS. Indeed, comparisons between columns 1 (3) and 10 (3), as well as between columns 1 (6) and 10 (6), in Table 4 show that adding markers on chromosomes not carrying a QTL generally decreases the reponse to MAS. However, the decrease in the response is not very large (indeed it is very small in the case of three QTLs in coupling phase). This is quite surprising, given that a substantial proportion of markers chosen to contribute to the molecular score in the case of ten chromosomes come from the chromosomes that do not carry a QTL. Zhivotovsky (1976) conducted computer simulations of selection based exclusively on genetic markers not linked to any QTL in a population representing a finite sample from a large population in linkage equilibrium. He found that such selection can produce a non-zero response. Columns10 (3)* and 10 (6)* in Table 4 also demonstrate that MAS can utilize the associations between genetic markers and QTLs that are not due to genetic linkage, but rather are caused by finite population size. Indeed, except for the case of three QTLs in repulsion phase, there is a non-zero response to selection based on markers that are located on chromosomes not carrrying a QTL. Unlike associations between QTLs and linked markers which can persist for many generations, associations between QTLs and unlinked markers decay within a few generations. This may be one of the reasons why the efficiency of MAS in generations following the first one is improved if the set of markers contributing to the molecular score, as well as their additive effects, are re-evaluated in each generation rather than being evaluated only in the first generation (Gimelfarb and Lande 1994a).

The main conclusion emerging from the present work is that, even though MAS based on the method of Lande and Thompson (1990) (employing a multiple linear regression of the phenotype on genetic markers) can be quite effective in finite populations, the markers chosen by the regression to be utilized by selection are not necessarily those closest to QTLs, and their associated effects may not correspond to the contributions to the trait by linked QTLs. Any marker contributing substantially to the $R^2$ value of the regression, even if it is not linked to a QTL, can be effectively utilized by selection. Increasing the number of markers per chromosome does not necessarily result in a stronger response to selection, since the regression does not discriminate among markers if the distance between them is less than a particular value (10 cM for an $F_2$ population of 500 individuals of each sex).

**Acknowledgements** This work was supported by U.S. Public Health Service grant GM27120.

## References

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943–951

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

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

Gimelfarb A, Lande R (1994a) Simulation of marker-assisted selection for non-additive traits. Genet Res 64:127–136

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Knapp SJ, Bridges WC, Birkes D (1990) Mapping quantitative trait loci using molecular-marker linkage maps. Theor Appl Genet 79:583–592

Knott SA, Haley CS (1992) Aspects of maximum-likelihood methods for the mapping of quantitative traits in line crosses. Genet Res 60:139–151

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

Lander ES, Botstein D (1988) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Paterson AH, Lander ES, Hewitt JD, Peterson SE, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using a selected overlapping recombinant chromosome in an interspecies cross of tomato. Genetics 124:735–724

Zeng Z-B (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zeng Z-B (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

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

Zhivotovsky LA (1976) Computer models of quantitative characters in genetics. V. The possibility of using genetic markers for early evaluation of animal productivity. Genetika 12:147–152 (in Russian with English summary)