Theor Appl Genet (1992) 83:813–820

Computer simulation of marker-assisted selection utilizing linkage disequilibrium

W. Zhang and C. Smith *

Center for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Received June 5, 1991; Accepted June 21, 1991
Communicated by K. Sittmann

Summary. The value of marker-assisted selection (MAS) using linkage disequilibrium between genetic markers and quantitative trait loci (QTL) was examined. To simulate the disequilibrium, four base populations were created, F_2, F_5, F_{10} and F_{20}, by random mating from a cross between two inbred lines. Selections were on breeding values estimated from: (1) marker QTL (MQTL) associations (MAS); (2) conventional best linear unbiased prediction (BLUP) methods; and (3) a combination of 1 and 2 (COMB). Alternative cases were studied by varying the parameters (heritability, initial linkage disequilibrium, and distribution of QTL effects). A genome with 100 QTL and 100 markers randomly (but equally) spread over 20 chromosomes, each 100 centiMorgans (cM) in length, was generated. Linkage disequilibrium (over 30 replicates) of QTLs with their nearest marker averaged 0.153, 0.104, 0.068, and 0.047 for the four base populations, and fell to 0.035, 0.025, 0.021, and 0.018, respectively, after ten generations of MAS selection (heritability 0.25). The initial linkage disequilibrium had the greatest effect on the genetic gain by MAS with the responses for the base populations $F_2 > F_5 > F_{10} > F_{20}$. Genetic gains by conventional BLUP selection were usually greater than by MAS. However, MAS contributed to the combined selection (COMB) to give appreciably higher genetic responses. Hybridization of selected lines after several generations of selection contributed little to generating further linkage disequilibrium. Detection of markers closer to the QTL will increase the linkage disequilibrium available for selection. Eventually with very close linkage each QTL allele can be uniquely identified in selection, and selection will then be equivalent to selection on the QTLs themselves.

Key words: Marker-assisted selection – Linkage disequilibrium – BLUP

Introduction

Much of the animal breeding work on marker-assisted selection (MAS) has been concerned with the difficulties in selection within families in outbreeding populations at linkage equilibrium. This is because the linkage phase and the zygosity state for each individual needs to be assessed and monitored in transmission to progeny (Beckmann and Soller 1983; Soller and Beckmann 1983; Smith and Simpson 1986; Kashi et al. 1990; Dekkers and Dentine 1991). Another approach, using an animal model for BLUP (Fernando and Grossman 1989; Fernando 1990; Goddard 1992), estimates an effect associated with each marker allele in each individual, and is computationally demanding.

An alternative approach, proposed by Lande and Thompson (1990), is based on linkage disequilibrium, and allows selection across the whole population. With the possibility of a very large number of DNA markers, linkage disequilibrium between QTLs and their closely linked markers can be detected in population studies, even in outbreeding populations, and can be used in selection. The objective of this research was to study, by stochastic computer simulation, the effectiveness of such selection, in comparison to conventional selection methods.

Methods

The method was to generate base populations with known genetic markers in linkage disequilibrium with quantitative trait

* To whom correspondence should be addressed

loci (QTL) with known effects. From the base populations, estimates of marker QTL associations (MQTL effects) due to linkage disequilibrium of the markers and the QTLs were derived by BLUP. These regression estimates were used to calculate estimated breeding values (EBV) for all individuals and the EBVs were then used as the basis for selection. The responses to this marker-assisted selection (MAS) (without information on phenotype) were compared with responses to conventional selection on BLUP using an animal model on phenotypes and phenotypes of relatives, and with the response to selection on the two sources of information (MAS and BLUP) combined (COMB).

A genome of 20 Morgans was simulated, made up of 20 chromosomes, each 1 Morgan in length. A total of 100 additive QTLs was simulated, accounting for all the genetic variance (V_A) associated with the trait being considered. The QTL loci were diallelic and were allocated at random 5 per chromosome and at random along each chromosome. An example is given for one chromosome in Table 1. The initial gene frequencies (p) in the base population averaged 0.5. The gene effects (a_i) were normally distributed with a mean of 0 and a variance such that

$$V_A = h^2\sigma^2 = \sum_{i=1}^{100} 2p(1-p)\,a_i^2 = \sum 0.5\,a_i^2 \quad \text{at} \quad p=0.5,$$

where h^2 is the heritability and σ^2 is the phenotypic variance for the trait. Examples of the distribution of QTL effects are given in Fig. 1. The genome (QTLs and markers) simulated was the same for all replicates (30) at a given heritability.

The genotypic value (and breeding value) for the m^{th} individual was derived as

$$G_m = \sum_{i=1}^{100} (a_{ij} + a_{ik}),$$

where a_{ij} and a_{ik} are the effects of the two alleles j and k at the i^{th} locus. The phenotypic value is

$$Y_m = G_m + E_m,$$

where E_m is an environmental contribution, $N(0,(1-h^2)\sigma^2)$. For 100 QTLs, with normally distributed effects, the effects of many of the QTLs will be quite small and there will be a smaller number of QTLs with larger effects. It is sensible, if possible, to concentrate selection on the QTLs with the larger effects.

In addition, 100 markers were simulated. Five were assigned at random per chromosome. They also were diallelic with an average initial gene frequency of 0.5.

Four levels of initial linkage disequilibrium between the markers and the QTLs were simulated as follows. A population at complete disequilibrium was generated as the cross between two inbred lines. The QTL effects in the inbred lines were assigned at random with both positive and negative effects in each line. If one line was homozygous (AA) for the positive allele, the other line was homozygous (aa) for the negative allele, and vice versa. Similarly for the markers, if one line was MM, the other was mm. All F_1 individuals then have the same genotype. The F_1 population was allowed to mate at random for several (2–20) generations with recombination, which gradually breaks down the linkage disequilibrium over generations and allows gene frequencies to drift from 0.5. The average linkage disequilibrium between QTLs and their closest marker, measured as $D = ru - st$ [where r and u are the frequencies of the original gametes (AM and am) and s and t are the frequencies of the recombinant gametes (Falconer 1989)], was recorded at each generation. The levels of linkage disequilibrium for starting a MQTL selection simulation from generations F_2, F_5, F_{10} and F_{20} were chosen for analysis. The map positions of the marker and the QTL are known. The recombination rate was derived from the map distance (d) as $r = 0.5(1-e^{-2d})$ (Haldane 1919).

Evaluating MQTL effects

The associations between the markers and the QTL, called MQTL associations, were estimated in base populations (F_2, F_5, F_{10} and F_{20}) of 500 males and 500 females. The MQTL associations were estimated by BLUP by regression of individual phenotype on the number (0, 1, or 2) of a marker allele at a given locus, for all marker loci simultaneously.

The BLUP procedure was a version of the model given by Goddard (1992). The mixed model equations are:

$$\begin{bmatrix} \hat{\mu} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\varepsilon \end{bmatrix}^{-} \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{Z}'\mathbf{Y} \end{bmatrix}$$

where $\mathbf{X}$ is the design matrix ($1{,}000 \times 1$) for fixed effects [the mean (μ) in this case]; $\mathbf{Z}$ is the design matrix ($1{,}000 \times 100$) for marker genotypic scores; $\mathbf{A}^{-1}$ is the inverse of the variance-covariance matrix (100×100) among markers; and $\varepsilon = \sigma_e^2/\sigma_\beta^2$. $\mathbf{A}$ has the following structure (Goddard 1992):

$$\begin{bmatrix} 2/3 & 1/6 & & & & 0 \\ 1/6 & 2/3 & 1/6 & & & \\ & \cdot & \cdot & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & 1/6 & 2/3 & 1/6 \\ 0 & & & & 1/6 & 2/3 \end{bmatrix}$$

The estimated breeding values (EBVs) for MAS are then

$$\text{EBV} = \mathbf{Z}\hat{\beta}.$$

An example of the MQTL effects ($\hat{\beta}$) for one chromosome in one replicate is given in Table 1.

Many of the MQTL have very small estimated effects and tend to add noise to the system. Based on the correlation between true and estimated breeding values for differing numbers of the top MQTLs, the top 20 MQTL estimates were selected and the others omitted. The BLUP analysis was then re-run on an independent set of data derived stochastically from the same population to produce a set of unbiased MQTL effects, as suggested by Lande and Thompson (1990). These were used to

Table 1. An example of an original chromosome simulated, allocating markers (M) and QTLs (Q) at random. The QTL effects are drawn from a normal distribution of gene effects (both positive and negative) contributing to the total additive genetic variance. The marker (MQTL) effects estimated in a F_2 base population are also given

True QTL gene effects (SD units)	Locus	Map position (Morgans)	Estimated MQTL effect (SD units)
	M	0.08	0.00
	M	0.19	0.04
0.15	Q	0.22	
	M	0.23	0.10
−0.05	Q	0.25	
0.02	Q	0.39	
	M	0.49	0.04
0.02	Q	0.73	
−0.06	Q	0.95	
	M	0.99	−0.06

estimate breeding values for MAS. The same MQTL estimates were used throughout the selection work and were not re-estimated over the generations.

Selection on phenotype and combined selection

For comparison with MAS selection, selection was also carried out using conventional BLUP selection on phenotypes without marker information. Selection was on estimated breeding value (EBV) derived by animal model BLUP (Quaas and Pollak 1980; Kennedy and Sorensen 1988) using the phenotypes of individuals and their relatives.

For the combined selection (COMB) combining phenotype and MQTL effects, a modification of the procedure proposed by Lande and Thompson (1990) was used. The combined index $\mathbf{I}=\mathbf{H}\hat{\boldsymbol{b}}$ in which $\mathbf{H}$ has two column vectors. One is the breeding values estimated from phenotypes by a BLUP animal model (EBV_p). The other is the EBVs estimated from MAS (EBV_m). $\hat{\boldsymbol{b}}$ is a column vector of coefficients with two elements and is obtained by:

$$\hat{\boldsymbol{b}}=\mathbf{P}^{-1}\mathbf{G}$$

with

$$\mathbf{P}=\begin{bmatrix}\sigma^2_{EBV_p} & \mathrm{Cov}(EBV_p, EBV_m)\\ \mathrm{Cov}(EBV_p, EBV_m) & \sigma^2_{EBV_m}\end{bmatrix}$$

$$\mathbf{G}=\begin{bmatrix}\sigma^2_{EBV_p}\\ \sigma^2_{EBV_m}\end{bmatrix}$$

where $\sigma^2_{EBV_p}$ and $\sigma^2_{EBV_m}$ are the variances of EBV from the BLUP animal model for phenotype and from the MQTL effects in MAS, respectively. $\mathrm{Cov}(EBV_p, EBV_m)$ is their covariance.

Selection

Selection was based on the EBV derived from the three methods (MAS, BLUP, and COMB), selecting 30 males and 30 females for breeding the next generation. Mating was at random, with equal family size, producing 500 male and 500 female offspring. Crossovers occurred at random according to the recombination frequency. The individual chromosomes were followed so the genotypes of marker and QTL loci could be studied and the gene frequency and other parameters could be monitored. Selection on EBVs was continued for ten generations and the genetic responses were evaluated.

Parameters for the standard case studied are listed in Table 2. Alternative parameter sets and models are also given. Three levels of heritability for the trait selected were used, namely, 0.1, 0.25, and 0.5. It was assumed that selection was possible in both sexes at reproductive age.

In addition to a normal distribution of QTL gene effects, a Gamma distribution was also studied, namely:

$$f(x)=x^{C-1}\,e^{-Dx}\,D^C/\Gamma_C\,,$$

with $C=1/9$ and D chosen to satisfy V_A. Examples of a Normal- and of a Gamma-simulated distribution of gene effects are given in Fig. 1.

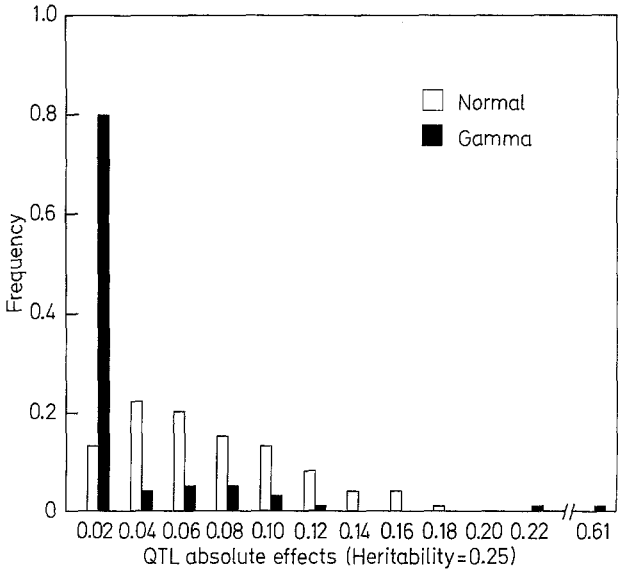

Fig. 1. Histograms of the number of QTL loci with different (absolute) size of effects, simulated for a Normal and a Gamma distribution, for a heritability of 0.25

Table 2. Parameter symbols and values used in the simulation. Alternative values for other cases studied are given in brackets

Parameter	Symbol	Standard Value	Alternative Values
Additive genetic variation	V_A	0.25	(0.1, 0.5)
Environmental variance	ε^2	0.75	(0.9, 0.5)
Phenotypic variance	σ^2	1.0	
Genome length (20 chromosomes, each 100 cM)		2000 cM	
Number of quantitative trait loci (QTL)		100	
Distribution of QTL effects (five QTLs per chromosome)		Normal	(Gamma)
Approximate initial frequency of QTL alleles (diallelic)		0.5	
Number of marker loci on the genome (five per chromosome)		100	
Approximate initial frequency of marker locus alleles (diallelic)		0.5	
Selection method		MAS	(BLUP, COMB)[a]

[a] MAS: marker-assisted selection; BLUP: best linear unbiased prediction on phenotypes; COMB: combines MAS and BLUP

Results

The individual QTLs contribute differentially to the total genetic variance (V_A). Thus, it will be practical (to reduce the marker testing effort) to identify and concentrate MAS selection on the MQTLs with the largest effects and to omit the others which add (in estimation) more noise than information to the system. The top 20 QTL (out of 100 simulated) accounted for about 67% of V_A (normal distribution of QTL effects and a heritability of 0.25), while the top 40 QTLs accounted for about 87%. Table 3 shows the correlation between the true breeding value

Table 3. Correlation (r) between the true (A) and estimated breeding values ($\hat{A}$) for a Normal distribution of QTL gene effects. The base populations (F_2, F_5, F_{10}, and F_{20}) correspond to 2, 5, 10, and 20 generations of random mating following the crossing of two inbred lines

Top QTL loci (%)	r	Top MQTL loci (%)	Base population			
			F_2 r	F_5 r	F_{10} r	F_{20} r
1	0.28	1	0.22	0.19	0.12	0.10
5	0.56	5	0.46	0.39	0.21	0.19
10	0.68	10	0.56	0.48	0.26	0.23
20	0.82	20	0.64	0.55	0.35	0.27
40	0.93	40	0.76	0.68	0.40	0.32
60	0.98	60	0.78	0.69	0.42	0.33
80	1.00	80	0.81	0.69	0.44	0.34
100	1.00	100	0.79	0.71	0.43	0.34

$V_A = 0.25$

Table 4. Average linkage disequilibrium (30 replicates) of QTL and their nearest marker at the start and after ten generations of selection on marker-assisted selection (MAS) and combined index (COMB) for a Normal distribution of QTL gene effects in four base populations (F_2, F_5, F_{10}, F_{20})

Population	Before selection	MAS After selection	COMB After selection
$h^2 = 0.1$			
F_2	0.153	0.036	0.046
F_5	0.104	0.034	0.034
F_{10}	0.068	0.022	0.028
F_{20}	0.047	0.017	0.026
$h^2 = 0.25$			
F_2	0.153	0.035	0.040
F_5	0.104	0.025	0.034
F_{10}	0.068	0.021	0.024
F_{20}	0.047	0.018	0.022
$h^2 = 0.50$			
F_2	0.153	0.028	0.031
F_5	0.104	0.024	0.030
F_{10}	0.068	0.020	0.025
F_{20}	0.047	0.016	0.021

(A) and the EBV ($\hat{A}$) estimated from the QTLs themselves, and for the MQTL associations for different percentages of the top effects and for four base populations. The correlations of A and $\hat{A}$ are initially high and gradually fall as linkage disequilibrium (Table 4) decreases. To make the selection on MQTL effects tractable, in view of marker testing and recording costs, the top 20 MQTLs were used in selection.

With a high level of disequilibrium in the F_2 base population (Table 4), good rates of genetic response were obtained with marker-assisted selection (Table 5), al-

Table 5. Average genetic mean (in genetic standard deviation units) and standard deviation from 30 replicates for selection by marker-assisted selection (MAS), best linear unbiased prediction of breeding value (BLUP), and combination (COMB) of the two methods; for different heritability (h^2), with a Normal distribution of QTL effects, base population with two generations of random mating (F_2)

	Generation	MAS		BLUP		COMB	
		Mean	SD	Mean	SD	Mean	SD
$h^2 = 0.1$	1	0.00	±0.05	0.00	±0.06	0.00	±0.05
	2	0.93	±0.10	1.04	±0.24	1.17	±0.17
	3	1.45	±0.10	1.62	±0.25	1.93	±0.17
	4	1.79	±0.11	2.12	±0.25	2.42	±0.17
	5	2.08	±0.11	2.27	±0.26	2.87	±0.18
	6	2.29	±0.13	2.40	±0.26	3.13	±0.18
	7	2.44	±0.14	2.76	±0.27	3.41	±0.18
	8	2.53	±0.16	3.01	±0.27	3.64	±0.18
	9	2.58	±0.16	3.24	±0.28	3.80	±0.17
	10	2.61	±0.16	3.59	±0.27	3.95	±0.16
$h^2 = 0.25$	1	0.00	±0.08	0.00	±0.10	0.00	±0.08
	2	1.03	±0.14	1.10	±0.30	1.42	±0.21
	3	1.62	±0.17	1.90	±0.31	2.18	±0.23
	4	2.08	±0.22	2.41	±0.32	2.98	±0.25
	5	2.38	±0.24	2.81	±0.31	3.60	±0.29
	6	2.58	±0.27	3.11	±0.32	3.90	±0.26
	7	2.75	±0.27	3.42	±0.34	4.10	±0.24
	8	2.88	±0.28	3.69	±0.34	4.23	±0.24
	9	2.96	±0.32	3.95	±0.36	4.34	±0.25
	10	3.03	±0.32	4.11	±0.34	4.42	±0.26
$h^2 = 0.50$	1	0.00	±0.12	0.00	±0.17	0.00	±0.12
	2	1.24	±0.22	1.30	±0.41	1.50	±0.25
	3	1.93	±0.25	2.13	±0.40	2.41	±0.29
	4	2.44	±0.26	2.70	±0.42	3.27	±0.26
	5	2.84	±0.32	3.12	±0.43	3.85	±0.26
	6	3.21	±0.39	3.55	±0.42	4.30	±0.32
	7	3.49	±0.44	3.80	±0.43	4.44	±0.32
	8	3.63	±0.45	4.21	±0.43	4.80	±0.33
	9	3.74	±0.45	4.52	±0.44	4.98	±0.32
	10	3.86	±0.41	4.81	±0.42	5.09	±0.29

though the response tended to decrease over the generations as the linkage disequilibrium was exploited, and as recombination reduced the value of the MQTL association information. In this simulation, selection on phenotypes by BLUP always gave higher responses than selection using markers (MAS), and the responses continued longer. As expected, with a fixed number of loci, compared with an infinite number, the genetic variation decreases steadily as the gene frequencies moved away from 0.5, and some loci were fixed. Combined selection, combining MAS and BLUP in an index, always gave a further increased response. The differences between the selection methods were well estimated with 30 replicates and the averages showed a consistent pattern. The response with MAS declined as the initial disequilibrium (Table 4) fell in the base population (Tables 5, 6, 7,

Table 6. Average genetic mean (in genetic standard deviation units) and standard deviation from 30 replicates for selection by marker-assisted selection (MAS), and combination (COMB) of MAS and BLUP selection; for different heritability (h^2), with a Normal distribution of QTL effects, base population with five generations of random mating (F_5)

	Generation	MAS	COMB
		Mean SD	Mean SD
$h^2=0.1$	1	0.00±0.06	0.00±0.05
	2	0.64±0.09	1.09±0.13
	3	0.87±0.11	1.81±0.15
	4	0.98±0.15	2.40±0.15
	5	1.11±0.16	2.79±0.17
	6	1.26±0.17	3.06±0.18
	7	1.36±0.19	3.31±0.20
	8	1.42±0.20	3.54±0.22
	9	1.48±0.21	3.70±0.22
	10	1.51±0.22	3.85±0.23
$h^2=0.25$	1	0.00±0.09	0.00±0.09
	2	0.88±0.18	1.21±0.19
	3	1.17±0.22	2.07±0.23
	4	1.43±0.25	2.83±0.26
	5	1.66±0.26	3.25±0.29
	6	1.82±0.29	3.52±0.28
	7	1.88±0.33	3.73±0.28
	8	1.93±0.36	3.93±0.29
	9	1.99±0.38	4.11±0.30
	10	2.03±0.40	4.27±0.29
$h^2=0.50$	1	0.00±0.12	0.00±0.13
	2	1.11±0.23	1.28±0.20
	3	1.55±0.27	2.20±0.22
	4	1.89±0.29	3.01±0.28
	5	2.18±0.37	3.54±0.29
	6	2.39±0.39	3.92±0.31
	7	2.53±0.46	4.24±0.34
	8	2.66±0.45	4.50±0.34
	9	2.74±0.47	4.74±0.34
	10	2.80±0.49	4.96±0.32

Table 7. Average genetic mean (in genetic standard deviation units) and standard deviation from 30 replicates for selection by marker-assisted selection (MAS), and combination (COMB) of MAS and BLUP selection; for different heritability (h^2), with a Normal distribution of QTL effects, base population with ten generations of random mating (F_{10})

	Generation	MAS	COMB
		Mean SD	Mean SD
$h^2=0.1$	1	0.00±0.06	0.00±0.06
	2	0.35±0.13	1.04±0.15
	3	0.45±0.15	1.64±0.17
	4	0.60±0.18	2.14±0.17
	5	0.73±0.18	2.49±0.18
	6	0.83±0.18	2.79±0.18
	7	0.89±0.19	3.04±0.19
	8	0.94±0.21	3.27±0.19
	9	1.00±0.20	3.49±0.20
	10	1.03±0.20	3.68±0.20
$h^2=0.25$	1	0.00±0.09	0.00±0.10
	2	0.56±0.18	1.20±0.23
	3	0.83±0.22	2.00±0.24
	4	0.98±0.23	2.58±0.25
	5	1.11±0.26	3.01±0.25
	6	1.20±0.26	3.37±0.27
	7	1.34±0.31	3.66±0.25
	8	1.43±0.31	3.86±0.25
	9	1.48±0.35	4.04±0.25
	10	1.52±0.39	4.21±0.27
$h^2=0.50$	1	0.00±0.13	0.00±0.13
	2	0.70±0.27	1.27±0.20
	3	0.98±0.33	2.24±0.22
	4	1.18±0.41	2.83±0.28
	5	1.37±0.46	3.29±0.29
	6	1.55±0.49	3.67±0.31
	7	1.64±0.52	4.00±0.34
	8	1.72±0.53	4.29±0.34
	9	1.79±0.59	4.56±0.34
	10	1.85±0.60	4.82±0.32

and 8). The linkage disequilibrium fell faster with selection than with random mating (Table 4), showing that the disequilibrium was being used by the selection. The response to combined selection also fell with the initial disequilibrium but still exceeded BLUP selection on its own (Tables 5, 6, 7, and 8). The gene frequencies of the 20 QTLs with the largest effects increased with MAS, most with high initial linkage disequilibrium and with high heritability (Table 9), possibly because the top 20 MQTLs are then more likely to be more closely associated with the top 20 QTLs. The gene frequencies of the top 20 QTLs also increased with BLUP and with combined selection, showing how the genetic variance with a fixed number of loci was steadily being used up.

The simulations were repeated for a Gamma distribution of gene effects with a few large effects and many small effects. The top 10% of QTLs contributed over 96 percent of the additive genetic variation and the correlations between true and estimated breeding value from the top QTL and MQTL effects were higher than for the normal distribution of effects (Table 10). Faster early responses were obtained with MAS than for the normal distribution, but the responses plateaued earlier and at lower levels (Table 11). The long-term (10 generations) responses with BLUP and combined selection, and the linkage disequilibrium (Tables 12 and 13), were also less than for the normal distribution of gene effects, as is expected with fewer and larger QTL effects.

Discussion

As shown theoretically by Lande and Thompson (1990), combined selection on markers and on phenotype can

Table 8. Average genetic mean (in genetic standard deviation units) and standard deviation from 30 replicates for selection by marker-assisted selection (MAS), and combination (COMB) of MAS and BLUP selection; for different heritability (h^2), with a Normal distribution of QTL effects, base population with 20 generations of random mating (F_{20})

	Generation	MAS	COMB
		Mean SD	Mean SD
$h^2=0.1$	1	0.00±0.05	0.00±0.04
	2	0.33±0.12	1.01±0.11
	3	0.43±0.13	1.58±0.13
	4	0.52±0.15	2.09±0.14
	5	0.57±0.15	2.43±0.15
	6	0.63±0.15	2.72±0.14
	7	0.66±0.16	2.97±0.16
	8	0.69±0.18	3.21±0.16
	9	0.72±0.18	3.43±0.16
	10	0.74±0.18	3.62±0.16
$h^2=0.25$	1	0.00±0.09	0.00±0.10
	2	0.53±0.16	1.11±0.18
	3	0.67±0.20	1.92±0.21
	4	0.78±0.20	2.50±0.22
	5	0.86±0.22	2.95±0.23
	6	0.92±0.22	3.32±0.23
	7	0.98±0.22	3.61±0.23
	8	1.04±0.23	3.81±0.23
	9	1.08±0.25	3.97±0.25
	10	1.13±0.29	4.16±0.27
$h^2=0.50$	1	0.00±0.12	0.00±0.11
	2	0.65±0.25	1.22±0.18
	3	0.92±0.30	2.19±0.20
	4	1.10±0.34	2.78±0.24
	5	1.24±0.34	3.26±0.25
	6	1.32±0.34	3.62±0.28
	7	1.43±0.35	3.95±0.28
	8	1.49±0.36	4.23±0.28
	9	1.52±0.43	4.51±0.28
	10	1.57±0.46	4.75±0.29

increase rates of genetic response. With outbreeding populations of livestock that have been largely closed for many generations, linkage disequilibrium will be limited with loosely linked markers, corresponding more to the F_{10} and F_{20} than the F_2 and F_5 base populations.

A feature of the selection was the continuous reduction in response over the ten generations of selection. This was due to the reduction in additive genetic variance as gene frequencies moved from the initial average value of 0.5, and some of the larger QTLs reached extreme frequencies, and the favorable alleles, or their markers, became fixed. If QTLs with moderate to large effects exist, past selection would have used them, and they would be at high frequency or fixed. The exception would be a large QTL currently in transition to a high frequency. Thus, a priori, it would not be expected that there would be many genes of large effect segregating. In fact,

Table 9. Average gene frequency (mean and SD on 30 replicates) of 20 QTLs with the largest effects at the tenth generation of selection by marker-assisted selection (MAS), best linear unbiased prediction of breeding value (BLUP), and a combined index (COMB); for different heritability (h^2) with a Normal distribution of QTL effects. All selections initiated at gene frequencies of 0.5 but at different levels of linkage disequilibrium[a]

	MAS	BLUP	COMB
	mean SD	mean SD	mean SD
$h^2=0.1$			
F_2	0.84±0.04	0.79±0.05	0.82±0.05
F_5	0.59±0.10	0.79±0.04	0.81±0.05
F_{10}	0.56±0.10	0.78±0.04	0.80±0.05
F_{20}	0.54±0.08	0.78±0.04	0.80±0.04
$h^2=0.25$			
F_2	0.86±0.05	0.80±0.05	0.84±0.04
F_5	0.69±0.09	0.79±0.05	0.83±0.05
F_{10}	0.60±0.09	0.80±0.04	0.82±0.05
F_{20}	0.56±0.08	0.79±0.04	0.81±0.04
$h^2=0.5$			
F_2	0.88±0.05	0.81±0.04	0.85±0.03
F_5	0.80±0.06	0.80±0.04	0.83±0.04
F_{10}	0.70±0.08	0.80±0.04	0.82±0.03
F_{20}	0.65±0.07	0.79±0.04	0.82±0.03

[a] Shown in Table 4

Table 10. Correlation (r) between the true (A) and estimated breeding values ($\hat{A}$) for a Gamma distribution of QTL gene effects

Top QTL loci (%)	r	Top MQTL loci (%)	Base population	
			F_2 r	F_5 r
1	0.86	1	0.67	0.47
2	0.92	2	0.73	0.57
5	0.95	5	0.82	0.64
10	0.98	10	0.81	0.64
20	0.99	20	0.82	0.63
40	1.00	40	0.80	0.64
60	1.00	60	0.79	0.63
80	1.00	80	0.81	0.65
100	1.00	100	0.80	0.64

$V_A=0.25$

from the continuous responses achieved in many selected lines and populations in practice, it would seem that there are many QTLs with small effects, rather than a few QTLs with large effects. This makes identification and use of MQTL associations more difficult and less useful.

The proportion of the additive genetic variation explained by the markers depends on the linkage disequilibrium The disequilibrium will be partial, rather than complete. The procedure of double sampling, proposed

Table 11. Average genetic mean (in genetic standard deviation units) and standard deviation from 30 replicates for selection by marker-assisted selection (MAS), best linear unbiased prediction of breeding value (BLUP), and combination (COMB) of two methods; for different heritability (h^2), with a Gamma distribution of QTL effects, base population with five generations random mating (F_5)

	Generation	MAS	BLUP	COMB
		Mean SD	Mean SD	Mean SD
$h^2=0.1$	1	0.00±0.11	0.00±0.05	0.00±0.07
	2	1.56±0.10	1.39±0.06	1.51±0.17
	3	1.84±0.10	1.67±0.08	1.74±0.15
	4	1.87±0.10	1.74±0.18	1.99±0.14
	5	1.90±0.09	1.82±0.20	2.15±0.10
	6	1.96±0.11	1.87±0.19	2.21±0.12
	7	1.99±0.12	1.89±0.23	2.23±0.09
	8	1.99±0.13	1.92±0.23	2.25±0.09
	9	1.99±0.14	1.95±0.23	2.25±0.09
	10	1.99±0.16	1.97±0.23	2.25±0.09
$h^2=0.25$	1	0.00±0.11	0.00±0.07	0.00±0.11
	2	1.62±0.09	1.51±0.18	1.60±0.23
	3	1.90±0.11	1.68±0.31	1.86±0.16
	4	2.04±0.10	1.79±0.31	2.24±0.15
	5	2.14±0.12	1.91±0.31	2.40±0.13
	6	2.20±0.13	2.00±0.32	2.46±0.14
	7	2.22±0.14	2.07±0.31	2.52±0.15
	8	2.24±0.17	2.12±0.32	2.54±0.15
	9	2.24±0.19	2.16±0.32	2.54±0.15
	10	2.24±0.19	2.20±0.32	2.54±0.16
$h^2=0.50$	1	0.00±0.16	0.00±0.10	0.00±0.15
	2	1.67±0.14	1.54±0.17	1.62±0.30
	3	2.03±0.16	1.84±0.32	1.94±0.17
	4	2.23±0.17	2.12±0.32	2.35±0.17
	5	2.35±0.21	2.26±0.34	2.54±0.17
	6	2.39±0.23	2.33±0.34	2.62±0.19
	7	2.43±0.24	2.38±0.35	2.69±0.21
	8	2.46±0.25	2.43±0.34	2.73±0.22
	9	2.48±0.25	2.47±0.36	2.74±0.22
	10	2.50±0.25	2.51±0.34	2.74±0.24

Table 12. Average gene frequency (mean and SD on 30 replicates) of ten QTLs with the largest effects at the tenth generation of selection by marker-assisted selection (MAS), best linear unbiased prediction of breeding value (BLUP), and combined index (COMB); for different heritability (h^2) with a Gamma distribution of QTL effects. All selections initiated at a gene frequency of 0.5 at the same level of linkage disequilibrium (F_5)[a]

	MAS	BLUP	COMB
	mean SD	mean SD	mean SD
$h^2=0.1$	0.77±0.04	0.80±0.05	0.93±0.07
$h^2=0.25$	0.79±0.05	0.81±0.05	0.93±0.07
$h^2=0.5$	0.81±0.05	0.82±0.04	0.94±0.06

[a] See Table 13

Table 13. Average linkage disequilibrium of the top ten QTL and nearest marker at the start and after ten generations of selection on marker-assisted selection (MAS) and combined index (COMB) for a Gamma distribution QTL gene effects (F_5 base population)

	Before selection	MAS After selection	COMB After selection
$h^2=0.1$	0.104	0.035	0.047
$h^2=0.25$	0.104	0.030	0.046
$h^2=0.50$	0.104	0.028	0.041

by Lande and Thompson (1990) to get unbiased MQTL effects, gives higher sampling errors of the effects for the same total number of animals tested. Sampling errors of the effects reduce the effectiveness of the MQTL information in selection, and large test numbers are needed. The information is also likely to be breed and line specific, rather than being general across populations as for other genetic parameters. It will usually not be worthwhile to re-estimate the MQTL effects until a number (n) of generations have passed, unless the linkages are quite loose, >0.1 Morgan, since the disequilibrium falls as $(1-r)^n$, where r is the recombination frequency. The methods tend to estimate the MQTL effects individually because extreme QTL genotypes are at low frequency in the base populations. The MQTL effects may interact with each other, so that their cumulative effect may be less than the sum of the individual effects, giving negative epistasis. The relevant trait in selection is economic merit. Some loci may have positive effects for some traits but have unfavorable effects for others. It is their net effect on economic merit which is important in selection.

Marker-assisted selection may be useful in the selection of sex limited traits, in early selection of animals for further testing, and to allow selection of juveniles or embryos (Georges and Massey 1991). With linkage equilibrium, marker-assisted selection must be within families, requiring large amounts of data to determine within-family linkage phase and zygosity at the QTL, and with loss of information in tracking the markers and the associations over generations. Linkage disequilibrium allows the determination and use of MQTL effects across a population. However, only the genetic variation associated with the disequilibrium is exploited by MAS, and this gets used up as the disequilibrium is exploited. Lande and Thompson (1990) propose hybridization among selected lines to regenerate the disequilibrium. Selecting in an elite stock for economic merit would be done in replicate lines which would be hybridized later. Keeping two, or more, lines with the same testing facilities as for one large line will reduce the selection intensity and so also the genetic response compared with that made in a single line. Also,

the regeneration of the disequilibrium in the hybrid or line cross will depend on the different genetic drift within the replicate lines. It was shown theoretically by Lande and Thompson (1990), and supported by studies of disequilibrium on crossing replicate lines after selection in this simulation, that genetic drift is not effective in creating new disequilibrium. In fact, with selection by MAS for the same MQTL associations in replicate lines, genetic drift of the large MQTL is restricted by selection, and recombination may cancel or offset the disequilibrium generated by crossing replicate lines. Thus, the opportunities of continued selection response by cycles of selection and hybridization of restricted lines are likely to be quite limited.

A more powerful way to increase the useful disequilibrium will be to find markers even closer to the QTLs. With 100 markers and 100 QTLs in this study, the average map distance of a QTL to the nearest marker was 0.147 Morgans. This could be steadily reduced as more random markers are found. Alternatively, a specific search for DNA markers closer to the individual QTL could be made by "chromosome walking" and similar methods. Eventually, with identification and choice of very close unique polymorphic markers, each QTL allele (identical by descent) will be uniquely associated with its own DNA marker, or marker haplotype. At this level the disequilibrium will be complete, and selection can be effective on the individual QTL alleles themselves. This corresponds to direct selection on QTL alleles (Smith 1967).

Acknowledgements. This research was funded by the Natural Sciences and Engineering Research Council of Canada. We thank Dr. J. C. M. Dekkers and Dr. J. P. Gibson for useful comments.

References

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor Appl Genet 67:35–43

Dekkers JCM, Dentine MR (1991) Quantitative genetic variation associated with chromosomal markers in segregating populations. Theor Appl Genet 81:212–220

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman, Harlow, UK, p 19

Fernando RL (1990) Statistical problems in marker assisted selection for QTL. In: Proc 4th World Congr Genet Appl Livestock Prod 13:433–436

Fernando RL, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Evol 21: 467–477

Georges M, Massey JM (1991) Velogenetics, or the synergistic use of marker assisted selection and germ line manipulation. Theriogenology 35:151–159

Goddard ME (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878–886

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

Kashi Y, Hallerman E, Soller M (1990) Marker assisted selection of candidate bulls for progeny testing programmes. Anim Prod 51:63–74

Kennedy BW, Sorensen DA (1988) Properties of mixed model methods for prediction of genetic merit. In: Weir BS, Eisen EJ, Goodman MM, Namkoong G (eds) Quantitative genetics. Sinauer, MA, pp 91–103

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

Quaas RL, Pollak EJ (1980) Mixed model methodology for farm and ranch beef cattle testing programs. J Anim Sci 51:1277–1287

Smith C (1967) Improvement of metric traits through specific genetic loci. Anim Prod 9:349–358

Smith C, Simpson SP (1986) The use of polymorphisms in livestock improvement. J Anim Breed Genet 103:205–217

Soller M, Beckmann JS (1983) Genetic polymorphisms in varietal identification and genetic improvement. Theor Appl Genet 67:25–33