Genet Sel Evol (1996) 28, 161-176

Original article

The use of marker haplotypes in animal breeding schemes

THE Meuwissen [1], ME Goddard

Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia

(Received 2 May 1995; accepted 13 December 1995)

Summary – Information on marker haplotypes was used to increase rates of genetic gain in closed nucleus breeding schemes. The schemes were simulated for ten discrete generations: firstly five generations of conventional (non-MAS) and then five generations of marker-assisted selection (MAS). The inheritance of quantitative trait loci (QTL) alleles was traced by marker haplotypes with probability $1 - r$. Emphasis was on extra genetic gains during the early generations of MAS, because it was assumed that new QTL were detected continuously. In the first generation of MAS, genetic gain was increased by 8.8 and 38%, when selection was, respectively, after the recording of the trait (eg, selection for growth rate) or before (eg, fertility). The marked QTL explained 33% of the genetic variance, and $r = 0.1$. The extra genetic gain decreased with the number of generations of MAS as the variance of the QTL became more and more exploited. The extra response rates due to MAS increased more than proportionally to the variance of the QTL and they increased with decreasing heritabilities. When r increased from 0.05 to 0.2, the genetic gain from MAS decreased by only 7.7% (selection before recording). MAS was approximately equally efficient for sex-limited and non-sex-limited traits. In the case of a carcass trait, which is measured after slaughtering, extra response rates were up to 64%. If recording was after selection, additional genetic gains increased markedly with increasing numbers of offspring per dam, because markers rendered within-family selection feasible in this situation. It was concluded that the extra rates of gain from MAS can be large when there is a continuous detection of new QTL, and when selection is before the recording of the trait.

molecular genetic marker / quantitative trait locus / marker assisted selection / animal breeding scheme

Résumé – L'utilisation d'haplotypes marqueurs dans les schémas de sélection animaux. *L'information sur des haplotypes de marqueurs a été utilisée pour augmenter les gains génétiques dans des schémas de sélection à noyaux fermés. De tels schémas ont*

[1] On leave from: DLO – Institute for Animal Science and Health, Box 65, 8200 AB Lelystad, The Netherlands

été simulés sur dix générations séparées : d'abord cinq générations de sélection classique (non assistée par marqueur) puis cinq générations de sélection assistée par marqueur (SAM). La transmission des allèles au locus quantitatif (QTL) était suivie par des haplotypes marqueurs avec une probabilité $1 - r$. *L'accent était mis sur les gains génétiques supplémentaires obtenus lors des premières générations de SAM, puisque l'on a supposé que de nouveaux QTL étaient continuellement détectés. Dans la première génération de SAM, le gain génétique était accru de 8,8 et 38 %, selon que le contrôle de la performance intervenait avant la mise à la reproduction (par exemple une sélection sur la vitesse de croissance) ou après (par exemple la fertilité), et sous l'hypothèse d'un QTL marqué expliquant 33 % de la variance génétique et avec* $r = 0, 1$. *Le gain génétique supplémentaire diminuait avec le nombre de générations de SAM puisque la variance du QTL était de plus en plus exploitée. Les réponses supplémentaires dues à SAM augmentaient plus que proportionnellement à la variance du QTL et augmentaient à mesure que l'héritabilité décroissait. Quand* r *augmentait de 0,05 à 0,2, le gain génétique de SAM ne diminuait que de 7,7 % (avec un contrôle après la mise à la reproduction). La SAM était à peu près également efficace pour des caractères exprimés dans un seul sexe que pour des caractères exprimés dans les deux sexes. Dans le cas d'un caractère de carcasse, mesuré après abattage, les gains de réponse atteignaient 64 %. Pour un caractère mesuré après la mise à la reproduction, les gains génétiques additionnels augmentaient notablement avec le nombre de descendants par mère, parce que les marqueurs rendaient alors possible une sélection intrafamille. On conclut que les gains dus à MAS peuvent être importants quand il y a détection continue de nouveaux QTL et que le contrôle de performance se fait après la mise à la reproduction.*

marqueur moléculaire / locus de caractère quantitatif / sélection assistée par marqueur / schéma de sélection

INTRODUCTION

In recent years, genetic maps of DNA markers have become available for several species of livestock (Barendse et al, 1994; Bishop et al, 1994; Rohrer et al, 1994) and more marker maps are under construction (Haley et al, 1990). In the near future, it is expected that maps with approximate distances between adjacent markers of 10–20 cM will cover most of the genome (see Visscher and Haley, 1995, for a review). In regions where quantitative trait loci (QTL) are found, higher map densities may be achieved.

Some experiments to map quantitative trait loci (QTL) on the marker map have been conducted (Anderson et al, 1994; Georges et al, 1994). More QTL mapping experiments will probably follow and the approximate position and effect of the largest QTL will be assessed. It will be difficult to distinguish whether an effect is due to one or several closely linked QTL, but regions where the QTL for the economically most important traits map can and will be located.

In previous studies, associations between single markers and QTL were based on daughter or granddaughter designs (Kashi et al, 1990; Weller et al, 1990; Meuwissen and Van Arendonk, 1992) and identified QTL had to be traced for two or more generations away from the sire in which they were identified before being used for selection. When marker haplotypes, that surround a QTL, do not recombine, the QTL can be traced with certainty (neglecting double recombinants) and BLUP

(Best linear unbiased prediction) methods can estimate QTL effects from previous generations (Goddard, 1992), such that no daughter or granddaughter design is needed. Furthermore, in contrast to previous studies (eg, Gibson, 1994), emphasis will be on the selection response during early generations of selection, since this is economically most important, and new QTL will be continuously detected in ongoing MAS schemes. This paper will describe the use of marker haplotypes in animal breeding schemes and will identify situations where MAS is particularly useful.

MODEL

Genetic model

It will be assumed here that regions where QTL are present have been identified by QTL mapping experiments. In such a region the presence of one QTL with many alleles is assumed, since the actual number of alleles is unknown. Assuming many alleles minimizes the change in allele frequencies due to selection, which makes the extra response from MAS last longer. This is not important here, where emphasis is on early generation response rates, but with a finite number of alleles and possibly extreme allele frequencies, the extra response from MAS will be reduced during later generations of selection. Also, the assumption of many alleles reflects the, perhaps realistic, situation where the assumed QTL effect is actually due to a cluster of closely linked QTL: the effect of each cluster is then represented by an allele.

A number of markers are scattered around the QTL, together forming a marker haplotype. In the absence of recombination within the haplotype, the inheritance of the haplotype is followed by DNA marker analysis. Double recombinations between two adjacent markers within the haplotype are neglected, which is reasonable even for haplotypes that cover a large distance as long as the distance between the two adjacent markers remains small. Hence, the inheritance of the QTL follows that of the marker haplotype.

When recombination occurs, it is assumed that the inheritance of the QTL is not traceable. Probability statements about the inheritance of the QTL could be made, but this is not attempted here since they require accurate estimates of the position of the QTL, which are generally not available (Haley and Knott, 1994). In its simplest form, the marker haplotype may be formed by two markers bracketing the QTL. When the markers are non-informative with respect to their inheritance, ie, from marker analysis it cannot be deduced whether the markers were inherited from the dam or from the sire, a situation similar to recombination occurs; the inheritance of the QTL effect could not be followed.

The QTL alleles of base generation animals were obtained by sampling from the distribution $\mathbf{N}(0, 1/2V_{\mathrm{QTL}_i})$, where V_{QTL_i} = variance due to the ith QTL. The factor 1/2 is due to the fact that an animal has a paternal and a maternal QTL allele, each of which contributes half of the total variance due to the QTL. Effects of QTL of descendants of base generation animals were obtained by Mendelian sampling from their parental effects. The probability that the marker haplotype recombined (at least once), and the Mendelian sampling of the QTL alleles could not be followed by the marker haplotypes, was r. The actual marker haplotypes

were not simulated: only recombination or no recombination within a haplotype was simulated. This procedure was replicated for all marked QTL.

A polygenic effect, g_i, was simulated to reflect the non-marked genes. In the base generation, polygenic effects were sampled from $\mathbf{N}(0, V_a)$, where V_a = additive variance of polygenic effects. In later generations, it was sampled from $\mathbf{N}(1/2g_s + 1/2g_d; 1/2V_a)$, where s and d denote the sire and dam respectively. Phenotypic records, y_i, were obtained by adding an environmental effect to the sum of the polygenic and QTL effects. The environmental effect was sampled from $\mathbf{N}(0, V_e)$.

Breeding value estimation

Estimation of breeding values with marker brackets or haplotypes follows Goddard (1992). Records were analyzed by the model:

$$\mathbf{y} = \mathbf{Zu} + \Sigma_i \mathbf{ZQ}_i\mathbf{q}_i + \mathbf{e},$$

where $\mathbf{y}$ = vector of records, $\mathbf{u}$ = vector of polygenic effects, $\mathbf{Z}$ = incidence matrix linking animals to records, $\mathbf{q}_i$ = vector of allelic effects for the ith marked QTL, $\mathbf{Q}_i$ = incidence matrix linking QTL alleles to animals (every animal has two QTL alleles, hence every row of $\mathbf{Q}_i$ has two elements equal to 1 and the remaining elements are 0), and $\mathbf{e}$ = vector of environmental effects.

As an example, consider two base generation animals, s and d, and one offspring, o. The alleles of the base generation animals are all considered as different base population alleles: q_{sp}, q_{sm}, q_{dp}, and q_{dm}, where p and m denote the paternal and maternal allele respectively. Now, suppose that the offspring o received the maternal marker haplotype of its sire s (actually, since s is a base animal, one of the haplotypes is arbitrarily denoted as maternal) and a recombined marker haplotype of its dam d. Hence, the paternal QTL allele of o is a copy of q_{sm} and the maternal allele is either q_{dp} or q_{dm}. For the maternal allele of o a new QTL allele is postulated and included in the vector $\mathbf{q}$, with a mean value of $1/2q_{dp} + 1/2q_{dm}$ and a variance around this mean of $E(1/2[1/2(q_{dp} - q_{dm})]^2 + 1/2[1/2(q_{dp} - q_{dm})]^4 = 1/4V_{QTL}$. Hence the total variance of q_{om} is $V(1/2q_{dp} + 1/2q_{dm}) + 1/4V_{QTL} = 1/4(1/2V_{QTL} + 1/2V_{QTL}) + 1/4V_{QTL} = 1/2V_{QTL}$ (which equals that of the other QTL alleles, eg, q_{sp}), and $\mathrm{Cov}(q_{dm}, q_{om}) = \mathrm{Cov}(q_{dm}, 1/2q_{dm}) = 1/4V_{QTL}$. It follows that:

$$\mathbf{Qq} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} q_{sp} \\ q_{sm} \\ q_{dp} \\ q_{dm} \\ q_{om} \end{bmatrix},$$

and:

$$\mathbf{V(q)} = \mathbf{G} = 1/2V_{QTL} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 1 \end{bmatrix} \quad [1]$$

Note that $\mathbf{G}$ has the same structure as a numerator relationship matrix (Henderson, 1976), where q_{dp} and q_{dm} are the parents of q_{om}. Hence, a pedigree of QTL alleles is formed, and $\mathbf{G^{-1}}$ is obtained from Henderson's rules. Also, $\mathrm{Var}(\mathbf{u}) = \mathbf{A}V_{\mathrm{a}}$, and $\mathbf{A^{-1}}$ follows from Henderson (1976).

Estimates for $\mathbf{u}$ and $\mathbf{q}_1$ are obtained by solving Henderson's (1984) mixed model equations (in the case of one QTL):

$$\begin{bmatrix} \mathbf{Z'Z} + \mathbf{A^{-1}}\lambda & \mathbf{Z'ZQ}_1 \\ \mathbf{Q}_1'\mathbf{Z'Z} & \mathbf{Q}_1'\mathbf{Z'ZQ}_1 + \mathbf{G}_1^{-1}\mathbf{V}_e \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{u}} \\ \widehat{\mathbf{q}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{Z'y} \\ \mathbf{Q}_1'\mathbf{Z'y} \end{bmatrix}, \qquad [2]$$

where $\lambda = V_{\mathrm{e}}/V_{\mathrm{a}}$ and the variance components, V_{QTL_1} (needed for $\mathbf{G}$), V_{a}, and V_{e}, were assumed to be known. Extension to more QTL is straightforward. V_{QTL_i} requires knowledge about the size of the QTL effects and their allele frequencies, which may, at least approximately, be obtained from the QTL mapping experiment. Otherwise, they could be obtained by an REML analysis (Fernando and Grossman, 1989; Goddard, 1992).

In situations, where marker information is not available, the equations:

$$[\mathbf{Z'Z} + \mathbf{A}^{-1}\beta]\widehat{\mathbf{a}} = \mathbf{Z'y}, \qquad [3]$$

are solved to obtain breeding value estimates $\widehat{\mathbf{a}}$, where $\beta = V_{\mathrm{e}}/(V_{\mathrm{a}} + \Sigma V_{\mathrm{QTL}_i})$. Both $\widehat{\mathbf{u}} + \Sigma\widehat{\mathbf{q}}_i$ from equations [2] and $\widehat{\mathbf{a}}$ from equations [3] are estimates of the total breeding value $\mathbf{u} + \Sigma\mathbf{q}_i$, which includes the QTL and the polygenes.

Breeding schemes

The analysis of DNA markers for vast numbers of commercial animals was considered too expensive. Hence only nucleus animals were analyzed, although in some instances the effect of having marker information on commercial offspring of selected sires was assessed. Only closed nucleus breeding schemes were studied, because these are most common across species. In species with low female reproductive rates, alternative breeding schemes occur (mainly open nucleus schemes), but due to the availability of modern reproductive techniques these schemes tend more and more towards closed nucleus schemes (Nicholas and Smith, 1983; Meuwissen, 1991a). Because marker information will be mainly available on nucleus animals, genetic markers will increase this tendency towards closed nucleus schemes. The parameters of the closed nucleus scheme are summarized in table I.

Because a QTL mapping experiment precedes the selection on marker information, it was assumed that marker information was available on five generations of animals prior to the start of MAS, which is in generation 0. Also, in an ongoing MAS scheme where a new QTL is detected, marker information becomes available on previous generations of animals. Breeding schemes were simulated for five generations prior to MAS (generation 0) with selection on $\widehat{\mathbf{a}}$ from equations [3]. Marker information accumulated during these five generations. After these five initial generations of selection, five generations of MAS followed with selection for $\widehat{\mathbf{u}} + \Sigma\widehat{\mathbf{q}}_i$ (from equations [2]). Alternatively, selection on $\widehat{\mathbf{a}}$ from equations [3] continued for another five generations, which was denoted by non-MAS.

Table I. Parameters of the closed nucleus breeding schemes.

No of animals per generation	100 ♂ + 100 ♀
No of sires selected per generation	10
No of dams selected per generation	50
(in the case of table IX	20)
Polygenic variance, V_a	0.25
Environmental variance, V_e	1 or 3
Variance of QTL, if there is only one QTL: V_{QTL_1}	0.125, 0.0625 or 0.03125
if there are three QTL: V_{QTL_1}, V_{QTL_2}, and V_{QTL_3}	0.125, 0.0625 and 0.03125
Probability that marker haplotypes recombine, r	0.05, 0.1, 0.2, 0.4
No of commercial progeny with marker information and records	0 or 1 000

RESULTS

Records available before selection

Table II compares genetic gains after one, two, three and five generations of MAS to the analogous gains with non-MAS. When records were available before selection (eg, growth rate, feed efficiency), extra rates of gain due to MAS were moderate and varied from 8.8% in generation 1 to 2% over five generations. The decline in the extra response is because the variance of the QTL effect decreases as the beneficial QTL alleles increase in frequency. The latter occurs more rapidly with MAS and thus the genetic gains with MAS decrease more rapidly than those with non-MAS. Also, non-MAS puts more selection pressure on the polygenic effects **u** than MAS. Hence, the genetic gain in **u** with non-MAS exceeds that with MAS, which reduces the difference in total selection response. Therefore, non-MAS tends to catch up with MAS as the number of generations increases (see Gibson, 1994).

Eventually both MAS and non-MAS exploit all the variance in the QTL, the advantage of MAS being that it exploits the QTL variance faster. However, if a new QTL is found every ith generation, a stable extra genetic gain is achieved equal to that indicated in table II after i generations of MAS, ignoring the gain from continued use of marked QTL after generation i.

Records available after selection

When records become available after selection (eg, with selection for fertility or longevity), the extra response due to MAS is increased and ranged from 38 to 15% over one to five generations. In this situation, conventional selection is for the average EBV (Estimated breeding value) of the parents and within-family variation is not used by selection. MAS uses the within-family variance associated with the QTL, which results in the large increases in response rates.

Effects of heritability

Extra response rates due to MAS are larger, with lower heritabilities (table II). With decreasing heritabilities the accuracy of selection decreases, but QTL effects

Table II. Genetic gains with and without MAS when records are available before or after selection; there is one QTL ($V_{QTL_1} = 0.125$), the probability that marker haplotypes recombine is 0.1, and the heritability is varied by varying the environmental variance, V_a.[a]

Generation number	*Records before selection*		*Records after selection*	
	non-MAS	*MAS*	*non-MAS*	*MAS*
	h^2 at start is 0.27 ($V_e = 1$)			
1	0.367	0.399	0.239	0.329
2	0.724	0.763	0.481	0.627
3	1.066	1.112	0.713	0.890
5	1.718	1.758	1.152	1.328
	h^2 at start is 0.11 ($V_e = 3$)			
1	0.289	0.350	0.193	0.279
2	0.568	0.662	0.382	0.521
3	0.845	0.958	0.548	0.732
5	1.363	1.449	0.888	1.093

[a] Results are based on 500 replicated simulations. Standard errors vary from 0.005–0.009, 0.007–0.011, 0.009–0.013 and 0.011–0.017 in generations 1, 2, 3 and 5 respectively.

are still fairly accurately estimated. This is because the tracing of copies of the QTL alleles by markers leads to the availability of multiple records on the QTL alleles. The accuracy of estimation with multiple records still decreases with decreasing heritability, but less so than with single records. Hence, the superiority of MAS increases with decreasing heritability.

Size of QTL effects

In situations with one marked QTL and recording before selection, the first generation extra genetic gain due to MAS is 1.3, 4.0, and 8.8% with V_{QTL_1} values of 0.03125, 0.0625, and 0.125 respectively (table III). These figures are 6, 16 and 38%, respectively, when recording is after selection. Hence, the extra gain is more than proportional to V_{QTL_1}. The accuracy of selection increases from σ_I/σ_g to $\sqrt{(\sigma_I^2 + \sigma_{IQTL}^2)}/\sigma_g$, where σ_I^2 = variance of estimated breeding values with non-MAS; σ_{IQTL}^2 = extra variance explained at the QTL by MAS; and σ_g^2 = total genetic variance. Since $\sqrt{(\sigma_I^2 + \sigma_{IQTL}^2)}/\sigma_g$ equals approximately $(1 + 1/2\sigma_{IQTL}^2/\sigma_I^2)\sigma_I/\sigma_g$, the increase in accuracy of selection is approximately proportional to σ_{IQTL}^2. Further, $\sigma_{IQTL}^2 = V_{QTL}r_{QTL}^2 = V_{QTL}[V_{QTL}/(V_{QTL} + \sigma_e^2/n)]$ where r_{QTL} = accuracy of estimation of the QTL effect, σ_e^2 = error variance (after accounting for estimation errors of all other effects in the model), and n = the number of copies of a QTL that are traced by the markers. Hence, σ_{IQTL}^2 increases more than proportionally to V_{QTL}.

In particular, where selection precedes the recording/expression of the trait, genetic gains increase more than the aforementioned proportion due to decreased

Table III. Genetic gains with and without MAS when the size of a single QTL is varied ($V_e = 1$); the probability that marker haplotypes recombine is 0.1.[a]

Generation number	V_{QTL_1} = *0.03125*		V_{QTL_1} = *0.0625*		V_{QTL_1} = *0.125*	
	non-MAS	*MAS*	*non-MAS*	*MAS*	*non-MAS*	*MAS*
	recording before selection					
1	0.316	0.320	0.334	0.348	0.367	0.399
2	0.627	0.632	0.655	0.689	0.724	0.763
3	0.935	0.947	0.985	1.028	1.066	1.112
5	1.541	1.561	1.625	1.663	1.718	1.758
	recording after selection					
1	0.204	0.217	0.221	0.257	0.239	0.329
2	0.394	0.436	0.435	0.514	0.481	0.627
3	0.586	0.663	0.648	0.755	0.713	0.890
5	0.981	1.064	1.065	1.178	1.152	1.328

[a] Results are based on 500 replicated simulations. Standard errors vary from 0.003–0.007, 0.005–0.010, 0.007–0.013 and 0.010–0.017 in generations 1, 2, 3 and 5 respectively.

intra-class correlations between EBVs of relatives (because markers explain within-family variance). This results in increased selection intensities (Hill, 1976; Meuwissen, 1991b).

Effects of additional QTL

Table IV shows the effect of considering one or no QTL, when there were three QTL with V_{QTL_i} of 0.125, 0.0625 and 0.03125. Recording was after selection. The extra genetic gains due to MAS were 28.3, 13.4 and 4.6%, when the largest, the intermediate, or the smallest QTL, respectively, was traced by markers.

Table IV. Genetic gains with MAS when there are three QTL with variance 0.125, 0.062 5, and 0.031 25 ($r = 0.1$ for all marked QTL); records were available after selection[a].

Generation number	*QTL-marked*[b]				
	none	*1*	*2*	*3*	*all*
	h^2 at start is 0.32 ($V_e = 1$)				
1	0.277	0.356	0.314	0.290	0.407
2	0.554	0.684	0.618	0.581	0.776
3	0.826	0.980	0.909	0.860	1.101
5	1.346	1.505	1.439	1.398	1.688

[a] Results are based on 500 replicated simulations. Standard errors vary from 0.006–0.008, 0.009–0.011, 0.011–0.013 and 0.014–0.017 in generations 1, 2, 3 and 5 respectively.

[b] 1, 2 and 3 denote that only the largest ($V_{QTL_1} = 0.125$), the intermediate ($V_{QTL_2} = 0.0625$) or the smallest QTL ($V_{QTL_3} = 0.03125$) respectively, was traced by markers.

When all three QTL were traced by markers, the extra genetic gain was 46.8%, which is very close to the sum of the effects of tracing the individual QTL, ie 46.3% (= 28.3 + 13.4 + 4.6). Hence, the extra response rate from including more QTL effects in a MAS scheme seems close to additive.

Recombination rates

Obviously, highest response rates are achieved when the probability that marker haplotypes recombine, r, is lowest, ie, $r = 0.05$ in table V. When r was 0.05, 0.1, 0.2, and 0.4, and V_{QTL_1} was 0.125, first generation extra genetic gains due to MAS were 38, 38, 28 and 22% respectively. With $V_{QTL_1} = 0.031\ 25$, these figures are 6.3, 6.3, 6.3 and 2.8% respectively. Hence, the extra gain due to MAS decreases only moderately with increasing recombination between marker haplotypes. Because a smaller V_{QTL_1} decreases the accuracy of selection, the number of traced copies of the QTL alleles needed for a sufficiently high accuracy of selection will be higher. Hence, with $r = 0.4$, the extra response was relatively more reduced with $V_{QTL_1} = 0.031\ 25$ than with $V_{QTL_1} = 0.125$.

Table V. Genetic gains with MAS when the probability that QTL alleles cannot be traced, r, and the variance of the QTL alleles (V_{QTL_1}) are varied (records are available after selection and $V_e = 1$)[a].

Generation number	r				
	0.05	*0.1*	*0.2*	*0.4*	*0.4*[b]
			$V_{QTL_1} = 0.125$		
1	0.330	0.329	0.305	0.292	0.255
2	0.636	0.627	0.592	0.557	0.502
3	0.900	0.890	0.858	0.802	0.745
5	1.354	1.328	1.304	1.252	1.184
			$V_{QTL_1} = 0.031\ 25$		
1	0.217	0.217	0.217	0.210	0.203
2	0.434	0.436	0.431	0.412	0.413
3	0.647	0.663	0.646	0.616	0.615
5	1.066	1.064	1.040	1.007	1.002

[a] Results are based on 500 replicated simulations. Standard errors vary between 0.005–0.007, 0.007–0.010, 0.009–0.012 and 0.012–0.015, in generations 1, 2, 3 and 5 respectively.
[b] In the case where double recombinations occur and are not detected within a marker bracket with one QTL in the middle of the bracket.

Double recombinations were ignored in this study for several reasons: r may be high due to non-informative markers and not due to a large distance between the markers; the extreme markers of a haplotype may be far apart, but individual marker brackets may be small without knowing to which bracket the QTL maps, hence, double recombinations will be detected and treated as single recombinants; and the probability of double recombinations is small except for large r. Table V also shows rates of gain in the case of a marker bracket with a QTL in the middle (M_1–QTL_1–M_2), a recombination rate between the markers of 0.4, and

when accounting for double recombinants. A realistic mapping function is obtained from Kosambi (1944), which is used in table V. The distance between M_1 and M_2 is then 0.55 M and the probability of a double recombination between M_1 and QTL_1 and between QTL_1 and M_2 is 0.05. Genetic gains were substantially reduced by double recombinants: 13–5% in generations 1–5 with $V_{QTL_1} = 0.125$. With $r = 0.2$, the probability of double recombinants is only 0.004, which does not yield a real reduction in genetic gain. Hence, a marker bracket of 55 cM is too large and genetic gains are substantially increased by having an additional marker within the bracket, even when this does not increase the precision of the estimate of the QTL site.

A simulation was also conducted, where the markers of the bracket were so far apart that the recombination rate between them and the QTL was 0.5. Hence, the markers yielded no information. In this case, genetic gain was 8% lower than with non-MAS (result not shown), because of the high frequency of double recombinations (25%) that resulted in erroneous tracing of QTL alleles.

Information from commercial offspring

In previous studies on the use of MAS, elite sires (or grandsires) were assumed to have progeny test information on many commercial offspring in order to obtain accurate estimates of effects of QTL alleles (Kashi et al, 1990; Meuwissen and Van Arendonk, 1992). First generation response rates increased by 44% due to MAS, when marker and performance information on 1 000 commercial progeny was available (table VI). Without this progeny information, this figure was 38% (table II). Hence, when all available information on QTL alleles is used, as in equations [2], the availability of marker information on many commercial offspring yielded only moderately increased rates of genetic gains.

Table VI. Genetic gains with and without MAS when selected sires have 1 000 progeny of which the QTL alleles could be traced; records are available after selection and progeny records are available after selection of the nucleus progeny of the sires ($V_e = 1$, $V_{QTL_1} = 0.125$, and $r = 0.1$)[a].

Generation number	*non-MAS*	*MAS*
1	0.254	0.366
2	0.501	0.672
3	0.747	0.924
5	1.224	1.304

[a] Results are based on 500 replicated simulations. Average standard errors are 0.006, 0.009, 0.011 and 0.014 in generations 1, 2, 3 and 5 respectively.

Sex-limited traits

When records are available after selection and only on females, eg, in the case of juvenile MOET (multiple ovulation and embryo transfer) schemes for dairy cattle (Nicholas and Smith, 1983), genetic gains were increased by 38 and 21% after one and five generations of MAS respectively (table VII). The former figure is similar

Table VII. Genetic gains with and without MAS when records are only available on females and after selection ($V_e = 1$, $V_{QTL_1} = 0.125$, and $r = 0.1$)[a].

Generation number	*non-MAS*	*MAS*
1	0.217	0.299
2	0.419	0.575
3	0.622	0.815
5	1.012	1.220

[a] Results are based on 500 replicated simulations. Average standard errors are 0.008, 0.010, 0.012 and 0.015 in generations 1, 2, 3 and 5 respectively.

to that with non-sex-limited traits. The latter is probably increased due to the less efficient MAS on sex-limited traits leading to less reduction of variance at the QTL.

Carcass traits

Table VIII considers selection for a carcass trait, which was measured by slaughtering at random half of the animals of each full sib family. The slaughtered animals were not eligible for selection. Non-slaughtered animals were selected after the information on their slaughtered sibs was recorded. Conventional selection yielded much lower response rates than those in table II because of the reduced selection differential, ie, half of the selection candidates were slaughtered, and the limited information. MAS increased rates of gain by 24% when the same breeding structure was used. In addition, all animals could be selected on marker information before the slaughtering, and the non-selected animals could be slaughtered to provide information for the next generation of selection. This increased rates of gain by 64% in the first generation of selection.

Table VIII. Genetic gains with and without MAS for a carcass trait, which is measured by slaughtering at random half of the full sibs before selection or the non-selected animals after MAS ($V_e = 1$; $V_{QTL_1} = 0.125$, and $r = 0.1$)[a].

Generation number	*recording before selection*		*after selection*
	non-MAS	*MAS*	*MAS*
1	0.168	0.208	0.275
2	0.328	0.404	0.530
3	0.479	0.598	0.742
5	0.780	0.948	1.082

[a] Results are based on 500 replicated simulations. Standard errors vary between 0.005–0.07, 0.007–0.010, 0.009–0.013 and 0.012–0.016, in generations 1, 2, 3, and 5 respectively.

Increased numbers of offspring per dam

In the previous situations, the number of offspring per dam was limited to four (= 200/50). In pigs and in cattle, when modern reproductive techniques are

used, the number of offspring can be much larger. With ten offspring per dam and recording before selection, response rates increased by 13 and 2% after one and five generations of MAS respectively (table IX). When recording was after selection, these figures were 45 and 17% respectively. In comparison to table II, first generations response rates were increased due to the higher female reproductive rates, but in later generations the advantage diminished due to the faster reduction in variances.

Table IX. Genetic gains with and without MAS when records are available before or after selection and females produce ten offspring per generation; there is one QTL ($V_{QTL_1} = 0.125$, $V_e = 1$, and $r = 0.1$)[a].

Generation number	*Records before selection*		*Records after selection*	
	non-MAS	*MAS*	*non-MAS*	*MAS*
1	0.431	0.486	0.250	0.361
2	0.844	0.920	0.470	0.646
3	1.240	1.309	0.700	0.875
5	2.014	2.046	1.100	1.282

[a] Results are based on 500 replicated simulations. Standard errors vary between 0.007–0.011, 0.010–0.013, 0.011–0.016 and 0.014–0.021 in generations 1, 2, 3 and 5 respectively.

When comparing tables II and IX, it may be noted that in the case of selection before recording, non-MAS genetic gain is higher with fewer offspring per dam. This is because estimated breeding values of all full sibs are equal and selection is between full sib families. With ten offspring per dam there are fewer full sib families (20 instead of 50), hence genetic gains are larger with four offspring per dam.

In the previous tables it was assumed that marker information was available on generations of animals prior to the MAS. This occurs in situations where a QTL detection experiment precedes the MAS, and/or in a continuous MAS scheme where a new QTL is found. In other situations there may be little marker information on the generations of animals preceding the MAS, eg, when the QTL detection experiment involved a small subset of the population and there are no DNA samples available on the other animals.

Limited marker information on previous generations

When there is only marker information on the animals in generation 0 and their parents and grandparents, the extra genetic gains due to MAS are only 2.8 and 6% in the first generations of selection, with recording before and after selection respectively (compare tables II and X), because information on the QTL effects has to accumulate during these generations.

The response in the first generation of MAS comes from the estimates of the paternal and maternal QTL alleles of the animals in generation -1, which are based on the EBV of their sires and dams respectively, and are traced by the markers to generation 0 animals. When selection is after recording, the records in generation 0 will also provide information on the effects of the paternal and maternal alleles

Table X. Genetic gains with and without MAS when records are available before or after selection, when there is no marker information on generations preceding generation 0 except that the parents and grandparents of generation 0 are genotyped for the markers; there is one QTL ($V_{QTL_1} = 0.125$, $V_e = 1$, and $r = 0.1$)[a].

Generation number	*recording before selection*	
	Yes	*No*
1	0.377	0.253
2	0.744	0.521
3	1.099	0.785
5	1.751	1.255

[a] Results are based on 500 replicated simulations. Average standard errors are 0.007, 0.009, 0.011 and 0.014, in generations 1, 2, 3 and 5 respectively.

of animals in generation −1. First generation extra response rates are therefore further reduced due to the absence of previous generations of marker information when selection is prior to recording compared to after recording.

After five generations of MAS, extra genetic gains are 2 and 9% respectively, in the absence of marker information on previous generations (table X). This compares to increases of 2 and 15%, respectively, in table II. Hence, the differences in genetic gain become smaller with generation number, but it also takes longer before V_{QTL_1} is exploited and before MAS can capitalize on a new QTL.

DISCUSSION

Assumptions: the genetic model

The true genetic model, ie, number of QTL, distribution of QTL effects, and number of alleles per QTL, is unknown and thus impossible to simulate. Fortunately, the short term results, which were required here, are not too sensitive to the genetic model. This is tested by a simulation with only two additive alleles with initial frequencies of 0.25 of the positive allele. This yielded after the five initial generations of non-MAS a frequency of approximately 0.5 and $V_{QTL} = 0.125$. Rates of gain over one, three and five generations of MAS were 0.340, 0.869, and 1.255 respectively, which compare to the figures of 0.329, 0.890, and 1.328 in table II. Hence, with two alleles initial rates of gain from MAS are as high as with many alleles and equal V_{QTL}, but rates of gain decrease faster due to the larger reduction of the variance at the QTL. In the fifth generation of MAS, V_{QTL} was 0.011 or 0.002 when many alleles or two alleles, respectively, were assumed. In conclusion, short term (up till three generations) predictions of rate of gain from the many alleles model are reasonably accurate, but the longer term rates of gain depend on the number of alleles, the distribution of allele effects, and the initial allele frequencies.

For simplicity, r was defined as the probability that marker haplotypes recombined. More precisely, r is the probability that the Mendelian sampling of QTL alleles could not be traced by markers. This could be due to recombination within the

marker haplotype but also due to markers being non-informative, or the haplotype not being known with certainty. In particular, when there is little recombination within marker haplotypes, the frequencies of certain haplotypes will be increased by selection, which decreases the information content of the marker haplotypes. However, additional markers could be typed within the marker haplotype as generations progressed to maintain the informativeness of the haplotype.

Within the first generation of marker haplotyping, the linkage phase between markers is unknown. This is no problem, because QTL effects are also not estimable from a single generation of marker data. In the next generation the linkage phase is known, since the inheritance of the markers can be traced. This is unless both parents and offspring are heterozygous for the same alleles at all marker loci, which is unlikely with reasonably informative markers; if it occurs, QTL effects will be assumed untraceable (as q_{om} in [1]).

Assumptions: the model of analysis

The selection decreases the variance among QTL alleles, which leads to reduced Mendelian sampling variances. This was not accounted for by either breeding value estimation method, MAS or non-MAS. With MAS, the Mendelian sampling variance of q_{om} from the alleles q_{dp} and q_{dm} is reduced because selection makes q_{dp} and q_{dm} more alike. This reduces the variance of q_{om} in formula [1]. With non-MAS the Mendelian sampling variance is also reduced, because part of this variance is due to V_{QTL}. Ignoring these variance reductions probably did not affect genetic gains much since genetic gains are not very sensitive to errors in variance estimates (Sales and Hill, 1976). However here the model of simulation and data analysis differed slightly as well.

It was assumed that the initial variance due to the QTL was known without error. In practice, this will not be the case. Consider an extreme case where the true $V_{QTL_1} = 0$ but it is assumed to be 0.125, ie, a false QTL is assumed. With recording after selection and $r = 0.1$, this case yields first generation genetic gains of 0.158 and 0.183 with MAS and non-MAS respectively. Hence, a substantial reduction in genetic gain of 14% is incurred when a false QTL is assumed.

Rates of genetic gain

MAS will only yield permanent increased rates of gain when there is a continuous input of newly identified QTL. The extra response rates due to MAS decreased very rapidly with increasing number of generations of selection for the same QTL (eg, table II). The rate at which new QTL will be discovered is difficult to predict. Beneficial QTL alleles, that are generated by mutations, may stay at a very low initial allele frequency for a long time and changes in allele frequency will be mainly due to drift (Falconer, 1989). As long as a QTL allele is at very low frequency, it contributes little to the genetic variance despite a possibly large QTL effect, and hence it is difficult to estimate how many of these QTL are present in a population. Once the allele drifts towards an 'intermediate' (still rather low) frequency, (marker assisted) selection increases the frequency rapidly. Hence, the rate at which new QTL are detected depends on the rate at which mutations drift to 'intermediate' frequencies and the amount of effort going into the detection of new QTL.

The inclusion of new traits in the breeding objective may lead to new QTL becoming relevant. Also, QTL may code for enzymes at rate-limiting steps in metabolic pathays. When the flux through such a step is increased by MAS, another step will become rate limiting and, hence, another QTL will occur.

Marker information increased rates of gain here more than in other studies (eg, Ruane and Colleau, 1994; Zhang and Smith, 1993), because of: 1) the use of marker haplotypes that trace QTL alleles with considerable probability instead of using one marker that is linked to a QTL; 2) the inclusion of marker information from previous generations; 3) the consideration of early response rates instead of longer term response rates (assuming that detection of new QTL continues); 4) the emphasis on traits that are recorded after selection, such that the markers increase the accuracy of selection substantially and increase the intensity of selection; and 5) the assumption that the variance associated with marked QTL was known. With non-MAS, selection may be performed only after recording, whereas MAS may lead to schemes with selection before recording because of the increased selection accuracy of young animals. This reduction of generation intervals may further increase rates of gain due to MAS (Meuwissen and Van Arendonk, 1992). In conclusion, the extra rates of gain from MAS can be large, in particular when there is a continuous detection of new QTL alleles and when traits are measured after selection.

ACKNOWLEDGMENT

This work was supported in part by the Meat Research Corporation of Australia.

REFERENCES

Anderson L, Haley CS, Ellegren H, Knott SA, Johansson M, Anderson K, Anderson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, Hakansson J, Lundstrom K (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* 263, 1771-1774

Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N, Grosse WM, Weiss J, Creighton P, McCarthy F, Ron M, Teale AJ, Fries R, McGraw RA, Moore SS, Georges M, Soller M, Womack JE, Hetzel DJS (1994) A genetic map of the bovine genome. *Nature Genetics* 6, 227-235

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo J, Beattie CW (1994) A genetic linkage map for cattle. *Genetics* 136, 619-639

Falconer DS (1989) *Introduction to Quantitative Genetics.* Longman, London

Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. *Gen Sel Evol* 21, 467-477

Georges M, Nielsen D, MacKinnin M, Mishra A, Okimoto R, Pasquito AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I (1994) Using a complete microsatellite map and the granddaughter design to locate polygenes controlling milk production. *5th World Congress Genet Appl Livest Prod* 21, 81-85

Gibson JP (1994) Short-term gain at the expense of long-term response with selection of identified loci. *5th World Congress Genet Appl Livest Prod* 21, 201-204

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

Haley CS, Knott SA (1994) Interval mapping. *5th World Congress Genet Appl Livest Prod* 21, 25-32

Haley CS, Archibald A, Anderson L, Bosma AA, Davies W, Fredholm M, Geldermann H, Groenen M, Gustavsson I, Ollivier L, Tucker EM, Van de Weghe A (1990) The pig gene mapping project–PiGMaP. *4th World Congress Genet Appl Livest Prod* 13, 67-70

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* 32, 69-83

Henderson CR (1984) *Applications of Linear Models in Animal Breeding.* Can Catal Publ Data, Univ Guelph, Guelph

Hill WG (1976) Order statistics of correlated variables and implications in genetic selection programmes. *Biometrics* 32, 889-902

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

Kosambi DD (1944) The estimation of map distance from recombination values. *Ann Eugenics* 12, 172-175

Meuwissen THE (1991a) The use of increased female reproductive rate in dairy cattle breeding schemes. *Anim Prod* 52, 21-31

Meuwissen THE (1991b) Reduction of selection differentials in finite populations with a nested full-half sib family structure. *Biometrics* 47, 195-204

Meuwissen THE, Van Arendonk JAM (1992) Potential improvements in rate of genetic gain from marker-assisted-selection in dairy cattle breeding schemes. *J Dairy Sci* 75, 1651-1659

Nicholas FR, Smith C (1983) Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim Prod* 36, 341-353

Rohrer GA, Leeson JA, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. *Genetics* 136, 231-245

Ruane J, Colleau JJ (1994) The value of marker assisted selection when one QTL is marked. *45th EAAP Meeting, session G1, 5–8 September, Edinburgh*

Sales J, Hill WG (1976) Effect of sampling errors on efficiency of selection indices. 1. Use of information from relatives for single trait improvement. *Anim Prod* 23, 1-17

Visscher PM, Haley CS (1995) Utilizing genetic markers in pig breeding programmes. *Anim Breed Abstr* 63, 1-8

Weller JI, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* 73, 2525-2537

Zhang W, Smith C (1993) Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. *Theor Appl Genet* 86, 491-496