Genet. Res., Camb. (1995), **66**, *pp.* 71–83

Marker assisted selection for genetic improvement of animal populations when a single QTL is marked

J. RUANE AND J. J. COLLEAU*

Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, Centre de Jouy-en-Josas, F-78352 Jouy-en-Josas, France

(*Received 6 July 1994 and in revised form 5 January 1995*)

Summary

A Monte Carlo simulation study to evaluate the benefits of marker assisted selection (MAS) in small populations with one marked bi-allelic quantitative trait locus (QTL) is described. In the base generation, linkage phase equilibrium between the markers, QTL and polygenes was assumed and frequencies of 0·5 for the two QTL alleles were used. Six discrete generations of selection for a single character measured on both sexes followed. An additive genetic model was used with the QTL positioned midway between two highly polymorphic markers. Schemes were simulated with a distance of 10 cM between the QTL and either of the two markers and with the QTL explaining 1/8 of the total genetic variance in the base generation. Values of 0·5, 0·25 or 0·1 were assumed for the heritability. Eight males and 16, 32 or 64 females were selected each generation with each dam producing four sons and four daughters as candidates for the next generation. Animals were evaluated with a conventional BLUP animal model or with a model using marker information. MAS resulted in substantially higher QTL responses (4–54%), especially with low heritabilities, than conventional BLUP but lower polygenic responses (up to 4%) so that the overall effect on the total genetic response, although in the majority of cases favourable, was relatively small. With QTLs of larger size (explaining 25% of the genetic variance) comparable results were found. When the distance between the QTL and the markers was reduced to 2 cM, genetic responses were increased very slightly with a heritability of 0·5 whereas with a heritability of 0·1 responses were increased by up to 10%, compared with conventional BLUP. Results emphasize that MAS should be most useful for lowly heritable traits and that once QTLs for such traits have been identified the search for closely linked polymorphic markers should be intensified.

1. Introduction

The rapid progress made in developing genetic linkage maps of polymorphic molecular markers in domestic animal species (see e.g. Bishop *et al.* 1994; Rohrer *et al.* 1994) has made it possible for the first time to begin the systematic search for individual loci affecting quantitative traits (QTLs) of economic importance. To summarize, this is done by recording animals for the character of interest; typing them for polymorphic markers of known chromosomal location; testing for statistical associations between marker alleles and phenotypic score and, if associations are found, inferring the presence of polymorphic QTLs adjacent to the marker loci (e.g. Paterson *et al.* 1988; Andersson *et al.* 1994).

The physical linkage between polymorphic markers and QTLs and the potential linkage disequilibrium (LD) between marker and QTL alleles that this generates can be used for Marker Assisted Selection (MAS) in three different situations. Firstly, where the aim is to transfer only a single gene from a 'donor' to a 'recipient' population, MAS can be used to accelerate recovery of the recipient genome (Hospital *et al.* 1992). In the second situation, more relevant to plant than animal breeding, the crossing of two divergent populations or inbred lines creates LD in the hybrid population which, at least in the early generations, can be successfully exploited for genetic improvement (Lande & Thompson, 1990; Zhang & Smith, 1992; Gimelfarb & Lande, 1994). In the third situation, which is examined here, the population of interest is outbred and LD created within families is exploited despite the fact that, at a population level,

* Corresponding author.

the markers and QTLs are in equilibrium (Soller, 1978). Thus, a certain marker allele may co-segregate with a favourable QTL allele in one family and with an unfavourable allele in a different family.

The aim of the present study is to examine the potential impact of MAS, under conditions which we consider should be favourable for the success of MAS, in a nucleus breeding scheme with several generations of directional single-trait selection. A single QTL is flanked by two highly polymorphic markers and both sexes are recorded. Each selected female produces eight offspring for selection in the next generation. In practical terms, the schemes may represent beef cattle (with the use of embryo transfer) or pig improvement programmes.

The value of MAS is evaluated in relatively small closed populations. Since the costs of applying molecular genetics are relatively expensive, it is considered that full implementation of MAS, i.e. with all selection candidates of both sexes typed, is likely to occur within centralized nucleus rather than dispersed field testing breeding programmes.

The genetic model assumes a single marked QTL and a large number of unmarked polygenes. This could correspond to the practical situation where statistical detection is considered certain for only a single QTL or where, due to the slow uptake of new technologies, only the largest QTL is considered in the breeding programme. Its simplicity as a basic model to investigate MAS is also appealing.

Recent studies of MAS in animal breeding have tended to focus on the use of multiple regression of phenotype on markers as a global method to identify markers linked to QTLs; to estimate marker effects and to evaluate candidates for selection based on their marker genotypes (Meuwissen & Van Arendonk, 1992; Zhang & Smith, 1992; Gimelfarb & Lande, 1994). The number of marker loci typed and the number of QTLs detected and used for selection purposes can thus be very large. Here, the perspective is quite different in that we assume that a single QTL of known variance and mode of action has been identified, which is flanked by two polymorphic markers and that the recombination rate between the QTL and the markers is known.

The merits of MAS are measured by evaluating candidates for selection over several generations using the model of Fernando & Grossman (1989). This allows marker information to be used in an animal model with Best Linear Unbiased Prediction (BLUP) of breeding values. Comparison is made with a classical BLUP animal model ignoring marker data. Following Goddard (1992), we extend the single marker model of Fernando & Grossman (1989) to the situation of two flanking markers but we allow for the further possibility of double crossing over within the marker bracket. Calculation of the probability of origin of QTL alleles, a key element of the model, is extended, using a relatively simple approximation, to the situation where uncertainty exists concerning the transmission of marker alleles from parents to offspring.

2. Materials and methods

Six discrete generations of single character selection in a closed population are simulated. Founder animals are chosen at random from a base population in linkage equilibrium at generation 0. A hierarchical mating design is used to produce selection candidates at generation 1 that are evaluated with either the BLUP animal model or the model of Fernando & Grossman (1989). Animals are then selected and mated at random to produce the candidates for selection at generation 2. This continues until the last cycle of selection, at generation 6. The number of sires and dams selected each generation is constant, including the base generation. Family sizes are fixed and each selected female has four sons and four daughters.

The QTL simulated is bi-allelic and both markers are highly polymorphic, each marker allele being present only once among the founders.

(i) *Founders*

B unrelated animals are chosen at random to be founders from a base population in which the markers, QTL and polygenes are in linkage phase equilibrium and in which the QTL alleles are present at an initial frequency of 0·5. The variances due to a single QTL allele (σ_v^2) and to the QTL ($2\sigma_v^2$) in the base generation are thus 0·25 α^2 and 0·5 α^2 respectively, where α represents the average effect of the gene substitution (Falconer, 1989). The values of the favourable A allele and unfavourable B allele are $\alpha/2$ and $-\alpha/2$ respectively.

An additive genetic model is assumed. The genotype (g) of a founder i is simulated by

$$g_i = v_{i1} + v_{i2} + u_i,$$

where v_{i1} and v_{i2} are the QTL allelic effects, chosen at random with a 50% probability of an allele being A or B, and where u_i represents the polygenic effect, drawn at random from a normal distribution with mean 0 and variance σ_u^2. Phenotypes (p) are generated by

$$p_i = \mu + g_i + e_i,$$

where e_i represents an environmental effect drawn at random from a normal distribution with mean 0 and variance σ_e^2. For simplicity, the only fixed effect considered was the mean.

For the B founders it is assumed that there are 2B unique marker alleles each at the M and N marker loci flanking the QTL and that the phase of the markers is known. Thus, at the marker loci all founders are double heterozygotes and have maximum poly-

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

morphism. To measure inbreeding at the QTL, founder QTL alleles are each given a unique identification number.

(ii) *Non-founders*

The genotype of a non-founder i bred by sire s and dam d is simulated by

$$g_i = v_i^p + v_i^m + u_s/2 + u_d/2 + r_i,$$

where v_i^p and v_i^m represent the values of the QTL alleles received respectively from the sire and dam (i.e. of paternal and maternal origin); u_s and u_d represent the polygenic values of the parents and r_i represents the Mendelian sampling term specific to each individual, drawn from a normal distribution with mean 0 and variance of $(\sigma_u^2/2)\ (1-(F_s+F_d)/2))$, where F_s and F_d represent the inbreeding coefficients of the sire and dam. Phenotypes are generated as for the founders.

Marker and QTL alleles were transmitted from parents to offspring in classical Mendelian fashion, allowing for recombination. Recombination is modelled using Haldane's (1919) function which converts map distances to recombination rates on the assumption that crossing over between M and the QTL and between the QTL and N occur independently. The QTL is assumed to be located midway between the markers. All individuals are typed.

(iii) *Evaluation models*

For both models, animals of both sexes and of all generations produce phenotypes. Equations to calculate the random and fixed effects of each model are solved iteratively following Schaeffer & Kennedy (1986).

(*a*) *Conventional BLUP animal model*

$$y = 1\mu + Wg + e,$$

where the vectors y, g and e represent the phenotype, genotype and environmental effects respectively, 1 is a vector of one's and μ represents the mean. The incidence matrix W is the identity matrix I. Marker information is not used.

(*b*) *Fernando & Grossman (1989) model*

$$y = 1\mu + Zv + Wu + e,$$

where v and u represent the QTL and polygenic effects, both treated as random uncorrelated variables and where Z and W (again equal to the identity matrix) are incidence matrices. The mixed model equations are

$$\begin{bmatrix} 1'1 & 1'Z & 1'W \\ Z'1 & Z'Z+\lambda_1 G_v^{-1} & Z'W \\ W'1 & W'Z & W'W+\lambda_2 A^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{v} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'y \\ Z'y \\ W'y \end{bmatrix},$$

where λ_1 and λ_2 and σ_e^2/σ_v^2 and σ_e^2/σ_u^2 and A is the additive relationship matrix. Fernando & Grossman (1989) provide relatively simple rules for calculating the elements of $\sigma_v^{-2}G_v^{-1}$, the inverse of the variance covariance matrix of QTL allelic effects. To apply them we need first to calculate the probability of origin of QTL alleles for each animal.

(iv) *Calculation of the probability of origin of QTL alleles*

At the QTL each individual has two alleles. Assuming the QTL alleles are not subject to mutation, one of these two is transmitted to the next generation. Based on the marker alleles transmitted by the parents to the offspring, we can calculate the probability that the offspring received either one or the other QTL allele. The values of the QTL alleles received by the offspring i from the sire (v_i^p) and from the dam (v_i^m) are then defined by Fernando & Grossman (1989) as

$$v_i^p = \theta_i^p v_s^m + (1-\theta_i^p) v_s^p + \epsilon_i^p$$
$$v_i^m = \theta_i^m v_d^m + (1-\theta_i^m) v_d^p + \epsilon_i^m,$$

where θ_i^p is the probability that the QTL allele received by the offspring i from its sire s (Q_i^p) is the same as that of its paternal grandmother (i.e. Q_s^m); $1-\theta_i^p$ is the probability that it is the same as that of its paternal grandfather (i.e. Q_s^p) and θ_i^m that the QTL allele received from its dam d (Q_i^m) is the same as that of its maternal grandmother (i.e. Q_d^m). v_s^m and v_s^p represent the values of the two QTL alleles of the sire (again, of maternal and paternal origin) and v_d^m and v_d^p represent those of the dam. ϵ_i^p and ϵ_i^m are residual effects.

The values of θ_i^p and θ_i^m are calculated for the situation where two markers are used. The chromosomes of a parent (a sire s in this case) with the Q QTL and the flanking marker loci M and N are represented diagramatically as

$$\begin{array}{ccc} \mathrm{M}_1 & Q_s^p & \mathrm{N}_1 \\ \text{----------} & & \\ \text{----------} & & \end{array}$$

$$\mathrm{M}_2 \leftarrow x \rightarrow Q_s^m \leftarrow x \rightarrow \mathrm{N}_2$$

The arrangement is denoted H_s and is written $\mathrm{M}_1\mathrm{N}_1/\mathrm{M}_2\mathrm{N}_2$ to indicate that the marker haplotypes $\mathrm{M}_1\mathrm{N}_1$ and $\mathrm{M}_2\mathrm{N}_2$ came from the sire and dam of s respectively.

Following Haldane (1919), the recombination rate (r) between the QTL and either of the two markers, as a function of the map distance (x) between them is derived by

$$r = 0{\cdot}5(1-e^{-2x}).$$

The values of θ_i^p depend on (*a*) whether the sire s has heterozygous or homozygous marker loci and (*b*)

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

Table 1. *Calculation of probability of origin of QTL alleles*

Sire marker loci	Markers transmitted from sire to offspring	θ_i^p
Heterozygous at M Heterozygous at N	$M_1 N_1$ $M_2 N_2$ $M_1 N_2$ $M_2 N_1$	$r^2/(1-2r+2r^2)$ $(1-r)^2/(1-2r+2r^2)$ 0·5 0·5
Homozygous at M Heterozygous at N	$M_1 N_1$ $M_1 N_2$	r $1-r$
Heterozygous at M Homozygous at N	$M_1 N_1$ $M_2 N_1$	r $1-r$
Double Homozygous	$M_1 N_1$	0·5

θ_i^p gives the probability that the offspring i has received Q_s^m from its sire. The QTL is situated midway between the two markers and r represents the recombination rate between the QTL and either of the two marker loci M and N.

the marker alleles transmitted to the offspring i. This is demonstrated in Table 1. Although to avoid repetition we consider only θ_i^p and the sire–offspring pathway, it must be remembered that values of θ_i^m for the dam–offspring pathway are calculated in a similar fashion.

Calculation of θ_i^p and θ_i^m is complicated by the fact that even though all animals may be typed, in certain situations there may be uncertainty concerning the marker alleles transmitted by the sire and dam to the offspring. With a single marker locus this may only arise if the sire, dam and offspring are heterozygous and have the same two marker alleles. With flanking markers uncertainty may arise in three situations.

(*a*) If the sire, dam and offspring are double heterozygotes and have the same four alleles. This is the only situation in which the linkage phase of markers for the offspring is also unknown.

(*b*) If the sire and dam are double heterozygotes with four marker alleles in common and the offspring is homozygous at one marker locus and heterozygous at the other.

(*c*) If the sire and dam have three different alleles in common and each of the offspring's four alleles are the same as one of these three and if the offspring is homozygous at one locus and heterozygous at the other.

For a single marker locus Bink & Van Arendonk (1994) suggest that, due to situations of uncertainty, the method of Wang *et al.* (1991), although more complicated, is preferred. Whereas Fernando & Grossman (1989) label the two QTL alleles according to parental origin, Wang *et al.* (1991) consider that the first QTL allele is linked to one marker allele and that the second QTL allele is linked to the second marker allele. Extension of Wang *et al.* (1991) to flanking markers should be quite straightforward for situations where the marker linkage phase is known (i.e. (*b*) and (*c*)). However, there are problems in dealing with situation (*a*), where the offspring has four possible marker haplotypes.

Instead, we maintain the distinction between QTL alleles based on parental origin and suggest an approximation for the calculation of probability of origin of QTL alleles in situations of uncertainty. The general algorithm can be presented in a Bayesian setting, with the repeated use of conditional probabilities. The approximation works as follows. In all situations, we know that an individual i receives Q_i^p from its sire and Q_i^m from its dam. In situations of uncertainty, we consider all the possible pathways by which i could have received its markers from its sire and dam. For each pathway, θ_i^p and θ_i^m are calculated (following the rules of Table 1) and then weighted by he probability of occurrence of each pathway to give the final values.

If i then becomes a parent in the next generation there is a probability c_i that Q_i^p is associated with the first marker haplotype and Q_i^m with the second marker haplotype and a probability $1-c_i$ that Q_i^p is associated with the second marker haplotype and Q_i^m with the first. If i is the offspring of sire s and c_s has a value 0·5 (i.e. there is an equal probability that Q_s^p could be on either one of the sire's chromosomes) then θ_i^p is 0·5, regardless of the markers received. For an offspring i in situation (*a*) it is not necessary to go through these calculations as θ_i^p, θ_i^m and c_i are all 0·5. Expressing the algorithm in algebraic form, we say that for an individual i, T_i represents the raw marker data, h_i represents any one of the four possible arrangements of these markers and H_i is the true arrangement. Then,

$$\theta_i^p = \sum_{h_i}\sum_{h_s}\sum_{h_d}(\theta_i^p|H_i = h_i, H_s = h_s, H_d = h_d)$$
$$\mathbf{Pr}(H_i = h_i, H_s = h_s, H_d = h_d|T_i, T_s, T_d, T_s^*, T_d^*)$$

and the second term simplifies to

$$\mathbf{Pr}(H_i = h_i|T_i, h_s, h_d)\,\mathbf{Pr}(H_s = h_s|T_s, T_s^*)$$
$$\mathbf{Pr}(H_d = h_d|T_d, T_d^*),$$

where T_s^* and T_d^* represent marker data from the parents and grandparents of the sire and dam

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

Table 2. *Cumulative selection responses from conventional BLUP when eight sires and 16, 32 or 64 dams are selected. Each dam has eight offspring*

	16 Dams Generation						32 dams Generation						64 dams Generation					
h^2	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
ΔG																		
0·5	0·77	1·37	1·93	2·46	2·96	3·45	0·86	1·53	2·15	2·74	3·31	3·86	0·94	1·67	2·34	2·99	3·61	4·21
0·25	0·64	1·14	1·60	2·03	2·44	2·82	0·71	1·27	1·79	2·28	2·74	3·18	0·78	1·40	1·96	2·50	3·02	3·50
0·1	0·49	0·88	1·24	1·57	1·88	2·17	0·54	1·00	1·40	1·78	2·14	2·47	0·62	1·12	1·57	2·00	2·40	2·79
ΔU																		
0·5	0·68	1·21	1·71	2·20	2·67	3·13	0·76	1·35	1·91	2·46	3·00	3·53	0·82	1·47	2·09	2·70	3·30	3·88
0·25	0·56	1·01	1·42	1·81	2·18	2·54	0·63	1·12	1·58	2·04	2·47	2·88	0·69	1·23	1·74	2·24	2·72	3·19
0·1	0·43	0·78	1·10	1·40	1·67	1·93	0·48	0·88	1·24	1·58	1·91	2·22	0·54	0·98	1·38	1·77	2·14	2·50
ΔV																		
0·5	0·09	0·16	0·22	0·26	0·29	0·31	0·10	0·18	0·24	0·28	0·31	0·32	0·12	0·19	0·25	0·29	0·32	0·33
0·25	0·08	0·14	0·19	0·23	0·26	0·28	0·09	0·15	0·20	0·24	0·27	0·30	0·10	0·17	0·22	0·26	0·29	0·31
0·1	0·06	0·10	0·14	0·18	0·21	0·24	0·06	0·11	0·16	0·20	0·23	0·25	0·08	0·14	0·19	0·23	0·26	0·29

Three levels of heritability (h^2) in the base population are simulated. In all cases, the genetic standard deviation in the base generation equals 0·707. The QTL is responsible for 12·5% of the total genetic variance. The map distance between the QTL and either of the two flanking markers is 10 cM. The genetic response (ΔG) is the sum of response at the polygenes (ΔU) and at the QTL (ΔV). Standard errors, expressed as a percentage of the cumulative responses, range from 0·1 to 1·2% for ΔG, 0·2–1·3% for ΔU and 0·4–3·3% for ΔV.

respectively. To calculate θ_i^m replace θ_i^p with θ_i^m in the above equations. An example is given in the Appendix which shows that these elements are calculated by first checking for uncertain transmission of marker haplotypes, by calculating the probability and θ_i^p value associated with each possible pathway of transmission (if uncertainty exists) and by then weighting the θ_i^p values by their respective probabilities.

The values of θ_i^p and θ_i^m, once calculated, are then used to construct the non-zero elements of the G_v^{-1} matrix, following the rules of Fernando & Grossman (1989) except that, as an approximation, inbreeding coefficients of parents based on pedigree information are used in the equations. These coefficients are calculated anyway and we thus avoid building up the G_v matrix.

(v) *Schemes simulated*

Three population sizes with three different heritability values are simulated. Eight sires and 16, 32 or 64 dams are selected. Each dam produces four sons and four daughters so that the total number of candidates per generation is 128, 256 or 512 and the proportion of males selected is 1/8, 1/16 or 1/32 respectively. The proportion of females selected is 1/4 for all schemes. As the number of dams is raised from 16 to 64, the number of half sibs per candidate increases from 8 to 56.

The QTL, polygenic and genetic variances are 0·0625, 0·4375 and 0·5 respectively. The environmental variance has a value of 0·5, 1·5 or 4·5 giving corresponding heritability values of 0·5, 0·25 or 0·1. The map distance (x) between the QTL and either of the marker loci is 10 cM. This value is used because low resolution linkage maps with markers spaced about 20 cM apart exist already for cattle and pigs and will soon be available for sheep and chickens (Beattie, 1994).

In secondary simulations, the implications of altering the map distance between the QTL and the markers or the size of the QTL effect were examined. For the first factor, x was reduced to 2 cM. For the second factor, while keeping the genetic variance constant at 0·5 and x at 10 cM, the QTL variance was increased to 0·125 and the polygenic variance reduced to 0·375. For both factors, schemes with eight sires and 16, 32 and 64 dams were simulated.

Schemes with MAS were replicated 70–950 times (average 220) and with conventional BLUP 500–4000 times (average 1830). Genetic evaluation was more time-consuming with MAS, so fewer replicates were possible. For both models, bigger schemes were replicated less often.

3. Results

(i) *Main simulations*

Table 2 presents cumulative selection responses from conventional BLUP for each of the nine basic situations (three heritability values × three population sizes) and for each of the genetic components (QTL, polygenes and overall breeding values).

Results show, as expected, that greater responses are achieved with larger population sizes and higher heritability values. The maximum QTL response possible is 0·354 units and between 40–70% and 50–80% of this is achieved after three or four rounds

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

Table 3. *Relative response to selection from MAS when eight sires and 16, 32 or 64 dams are selected*

	16 Dams Generation						32 dams Generation						64 dams Generation					
h^2	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
ΔG																		
0·5	102·1	102·1	101·4	101·2	100·6	100·5	100·6	101·0	100·7	100·1	99·8	99·7	100·0	100·3	100·4	100·0	99·9	99·7
0·25	100·8	101·3	101·4	101·5	101·4	101·0	100·1	101·2	101·8	100·9	100·5	100·3	100·4	102·2	101·5	101·5	101·0	101·0
0·1	98·3	99·4	100·3	100·7	100·7	100·2	100·9	103·5	104·2	103·6	102·3	101·1	100·8	104·1	104·3	103·4	102·5	101·3
ΔU																		
0·5	100·8	100·3	99·8	99·7	99·7	99·8	98·5	98·3	98·5	98·4	98·6	98·9	98·0	97·8	98·4	98·7	99·3	99·4
0·25	100·1	98·7	98·8	99·0	99·3	99·4	97·3	97·4	98·0	97·8	98·3	98·6	98·1	98·1	98·0	99·1	99·5	99·9
0·1	97·6	96·2	96·0	95·7	96·0	96·2	98·0	97·6	97·8	97·7	97·6	97·2	99·4	99·2	98·7	98·7	99·4	99·1
ΔV																		
0·5	111·2	115·0	114·0	113·4	109·1	107·6	115·9	121·1	118·1	115·0	111·3	107·6	114·8	119·8	116·8	111·3	106·7	104·1
0·25	106·2	120·7	121·5	121·3	118·9	115·8	121·0	129·7	131·4	126·8	120·9	116·1	116·9	132·6	129·6	122·9	115·2	110·2
0·1	103·5	123·4	133·5	140·0	138·5	133·5	122·2	149·0	154·0	150·2	141·6	135·0	110·0	138·0	145·0	139·0	128·4	120·7

The cumulative selection responses – genetic (ΔG), polygenic (ΔU) and QTL (ΔV) – are expressed, in percentage terms, relative to conventional BLUP i.e. (MAS response / conventional BLUP response) × 100%.

of selection respectively. As a rough guide, QTL responses of 0·07, 0·14, 0·21, 0·28 and 0·35 units correspond to frequencies of 0·6, 0·7, 0·8, 0·9 and 1·0 of the favourable allele. Response per generation is greatest early on and declines in later generations as the genetic variance in the population is progressively eroded.

Table 3 shows the relative efficiency of MAS compared to conventional BLUP. For all situations examined, MAS yields greater response at the QTL. This superiority is especially pronounced when heritabilities are low – from 4 to 54% with h^2 of 0·1 (compared with 4–21% superiority with h^2 of 0·5). On the other hand, in almost all situations and all generations the polygenic response is lower with MAS (up to 4%). Putting these two effects together, MAS, in general, achieves higher rates of genetic progress than conventional BLUP, especially when heritabilities are low. However, in most cases the superiority is relatively small.

Examination of the accuracies of selection (defined as the correlation between true and estimated breeding values) in Table 4 shows that, in general, MAS is not more efficient at evaluating and ranking the selection candidates than conventional BLUP at generation one; that it tends to be most efficient at generations two and three (1–9% more accurate) and that this superiority tends to decline in later generations. With low heritabilities, the relative accuracy of selection from MAS is greatest.

Inbreeding coefficients were calculated in two different ways – from the proportion of individuals with two QTL alleles that are identical by descent or from pedigree information and the relationship matrix. The latter expresses the probability that an individual contains two alleles that are identical by descent at a locus neutral with respect to the character under selection. Inbreeding rates were calculated with $\Delta F = (F_t - F_{t-1})/(1 - F_t)$ (Falconer, 1989), where F_t is the inbreeding coefficient at generation t. Results are presented in Table 5.

The table shows that as schemes get bigger and h^2 increases, inbreeding rates are progressively reduced. Secondly, in a selected population the rate of inbreeding is higher for loci affecting the selected trait than for neutral loci. Thirdly, compared to conventional BLUP, MAS has no effect on inbreeding at neutral loci but it substantially increases (by 10–40%) inbreeding rates at the QTL, especially with low heritability values.

(ii) *Secondary simulations*

Table 6 shows that the impact of reducing the distance between the QTL and the markers depends on the heritability of the trait of interest. With a heritability of 0·5, QTL response is increased by up to 17% when the markers are 2 cM from the QTL instead of 10 cM. However, the increase in QTL response is accom-

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

Table 4. *Accuracies of selection (defined as the correlation between true and estimated breeding values for selection candidates) for MAS with eight sires and 16, 32 or 64 dams (D) selected*

		Generation						
	h^2	1	2	3	4	5	6	Av.
16 D	0·5	0·77	0·70	0·68	0·67	0·65	0·64	0·69
		(100·1%)	(101·5)	(100·5)	(100·9)	(100·5)	(100·0)	(100·6)
	0·25	0·64	0·56	0·55	0·53	0·52	0·50	0·55
		(99·9)	(102·0)	(102·3)	(102·0)	(102·1)	(101·2)	(101·5)
	0·1	0·47	0·43	0·42	0·41	0·39	0·37	0·42
		(99·2)	(102·3)	(104·6)	(105·1)	(104·9)	(101·7)	(102·8)
32 D	0·5	0·78	0·70	0·68	0·67	0·66	0·65	0·69
		(100·8)	(101·0)	(101·0)	(100·3)	(100·1)	(100·0)	(100·5)
	0·25	0·65	0·58	0·57	0·54	0·52	0·53	0·57
		(100·9)	(103·4)	(104·1)	(99·8)	(99·9)	(102·6)	(101·8)
	0·1	0·51	0·46	0·46	0·42	0·40	0·37	0·44
		(101·5)	(105·4)	(109·3)	(103·4)	(99·3)	(96·9)	(102·7)
64 D	0·5	0·78	0·71	0·69	0·67	0·67	0·66	0·70
		(100·4)	(101·3)	(101·1)	(100·1)	(100·2)	(100·0)	(100·5)
	0·25	0·66	0·59	0·58	0·55	0·53	0·54	0·57
		(100·2)	(103·0)	(104·4)	(101·5)	(99·7)	(102·1)	(101·8)
	0·1	0·51	0·48	0·46	0·44	0·43	0·40	0·45
		(99·1)	(105·9)	(106·0)	(100·7)	(103·1)	(98·4)	(102·2)

Results are presented by generation and for the average (Av.) of the six generations of selection. Results are also expressed in % terms relative to conventional BLUP and are given in parentheses – values above 100% indicate superiority over conventional BLUP.

Table 5. *Average inbreeding rates per generation (from generations 2 to 7) with MAS, calculated for loci neutral with respect to the selected trait (ΔF_p) or for the QTL (ΔF_{qtl})*

	ΔF_{qtl}			ΔF_p		
h^2	16 D	32 D	64 D	16 D	32 D	64 D
0·50	7·11	5·93	5·54	5·18	4·32	4·06
	(118)	(110)	(111)	(101)	(98)	(102)
0·25	9·29	7·97	8·46	6·55	5·65	5·25
	(124)	(118)	(136)	(99)	(98)	(101)
0·10	11·72	11·34	10·40	7·69	6·95	6·38
	(137)	(138)	(140)	(98)	(100)	(99)

Heritability values (h^2) are 0·5, 0·25 and 0·1 and eight sires and 16, 32 or 64 dams (D) are selected. Inbreeding rates relative to conventional BLUP, expressed as a percentage, are given in parentheses – values above 100% indicate superiority over conventional BLUP.

panied by a slight reduction in polygenic response so that overall, genetic response is increased by only 0–1·7% (16 dams), by 0·4–1·4% (32 dams) and has no positive effect with 64 dams selected. On the other hand, with a heritability of 0·1, the reduction in marker-QTL distance from 10 to 2 cm yields a far greater increase in QTL response (up to 95%) which, despite the fact that the polygenic response is again reduced (by 0–9%), results in the genetic response being raised by up to 8%. Compared with conventional BLUP, this represents a 0–10% improvement in genetic gain.

Table 7 shows that when the size of QTL effect is increased, the QTL response from MAS is still higher than that of conventional BLUP, but the superiority is less pronounced – thus suggesting that conventional BLUP is a relatively efficient tool for fixing QTL alleles of large effect. Polygenic responses are, however, about 5% lower with MAS when the QTL accounts for 25% of the genetic variance so that the

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

Table 6. *Effect of distance between QTL and markers on cumulative selection responses – QTL (ΔV) and genetic (ΔG) – for MAS*

		Generation					
	h^2	1	2	3	4	5	6
ΔV							
16 D	0·5	111·8	116·6	114·4	108·5	106·3	103·0
	0·1	153·1	194·9	169·2	138·7	122·4	112·0
32 D	0·5	112·2	110·1	109·7	105·6	102·5	101·4
	0·1	174·1	168·0	137·0	118·0	109·3	104·4
64 D	0·5	115·0	110·5	106·3	103·4	101·5	99·9
	0·1	171·0	149·3	123·3	109·4	104·0	101·8
ΔG							
16 D	0·5	101·0	101·7	101·1	100·5	100·5	100·0
	0·1	103·8	107·9	105·4	101·6	100·8	100·0
32 D	0·5	100·8	101·4	101·3	100·7	100·7	100·4
	0·1	105·7	104·0	101·0	99·2	100·6	100·6
64 D	0·5	98·4	100·7	99·9	100·1	100·6	100·3
	0·1	106·6	105·3	102·2	100·1	99·6	99·3

Eight sires and 16, 32 or 64 dams (D) are selected and the heritability is 0·5 or 0·1. The map distance between the QTL and either of the two flanking markers is 2 cM. Selection responses are expressed in % terms relative to the standard MAS schemes (of the same heritability value) presented in Table 3 i.e. where the marker – QTL distance is 10 cM.

Table 7. *Effect of size of QTL effect on cumulative selection responses – QTL (ΔV) and genetic (ΔG) – for MAS*

		Generation					
	QTL Var.	1	2	3	4	5	6
ΔV							
16 D	12·5	111·2	115·0	114·0	113·4	109·1	107·6
	25·0	109·2	113·7	114·0	108·6	105·4	103·5
32 D	12·5	115·9	121·1	118·1	115·0	111·3	107·6
	25·0	116·4	111·6	107·6	103·7	101·3	99·9
64 D	12·5	114·8	119·8	116·8	111·3	106·7	104·1
	25·0	117·6	115·9	106·5	102·5	100·1	98·8
ΔG							
16 D	12·5	102·1	102·1	101·4	101·2	100·6	100·5
	25·0	101·3	101·5	101·1	99·5	99·1	99·4
32 D	12·5	100·6	101·0	100·7	100·1	99·8	99·7
	25·0	102·4	101·0	100·7	99·3	99·3	99·2
64 D	12·5	100·0	100·3	100·4	100·0	99·9	99·7
	25·0	99·8	100·4	99·8	98·9	98·8	98·9

Eight sires and 16, 32 or 64 dams (D) are selected and the heritability is 0·5. The map distance between the QTL and either of its flanking markers is 10 cM and the QTL is responsible for 12·5 % (i.e. as in Table 3) or 25 % of the total genetic variance. Selection responses are expressed in % terms relative to conventional BLUP.

total genetic responses are little different from those of conventional BLUP. The genetic response of MAS compared with conventional BLUP, expressed as a percentage, thus tends to drop slightly as the amount of genetic variance explained by the QTL increases. Similar results (not shown) are also found with heritabilities of 0·1.

4. Discussion

Results presented here show that, with respect to response at the QTL, MAS works since it is far more efficient than conventional BLUP. For all schemes simulated and for all generations, the cumulative QTL response was superior. This was particularly the case when the heritability was low. In this situation, the contribution of the candidate's own phenotype to its estimated breeding value is reduced and marker information has a greater relative importance, as suggested by Lande & Thompson (1990).

If, for reasons such as line breeding, fixation of the favourable allele was a primary goal of the selection programme, MAS would be preferable to conventional BLUP. In addition, fixation could be accelerated by putting more weight on the estimated QTL effects and less on the estimated polygenic effects. Another point of interest is that the frequency of the favourable allele in the base generation was 0·5. Both MAS and conventional BLUP succeeded in fixing the allele and the differences in QTL responses reflected differences in fixation rates. If the favourable allele was present at a lower frequency, MAS might be even more efficient since, with the use of markers, the probability of losing the favourable allele due to drift should be lower than with conventional BLUP.

On the other hand, it is worth recalling that in many ways conditions in this study were highly favourable for the success of MAS. All animals were typed for two very highly polymorphic marker loci flanking the QTL while the linkage phase of markers for all founder animals was assumed known as well as the recombination rate between the QTL and the markers. The proportion of the genetic variance explained by the QTL and the additive nature of the genes controlling the trait of interest were also known. If any of these assumptions are modified, the QTL response could be considerably reduced.

In the simulated populations considered here, some assumptions of the Fernando & Grossman (1989) model are violated. Firstly, it assumes a covariance of zero between the QTL and polygenic effects which, although true at generation one, is later invalid due to gametic phase disequilibrium induced by selection. Secondly, to develop their rules for calculating the elements of the inverse of the variance covariance matrix of QTL allelic effects, they assume that the QTL allelic variance is constant from one generation to the next (whereas it changes due to selection and inbreeding). Thirdly, it assumes a normal distribution of QTL effects, thus indicating that the number of QTL alleles is large whereas in this study a biallelic QTL was considered. Simulated results show, however, that the model can successfully identify animals

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

Table 8. *Illustration of gametic phase disequilibrium when eight sires and 64 dams are selected at the first round selection (i.e. generation 1) with MAS or conventional BLUP*

	Conventional BLUP			Marker assisted selection (MAS)		
Genotype	Freq. (%)	QTL Value	Polygenic Value	Freq. (%)	QTL Value	Polygenic Value
Sires						
AA	49·1	0·354	0·951	53·4	0·354	0·929
AB	42·6	0·000	1·108	40·3	0·000	1·109
BB	8·3	−0·354	1·279	6·3	−0·354	1·294
Weighted means		0·145	1·045		0·167	1·025
Dams						
AA	38·1	0·354	0·495	41·0	0·354	0·470
AB	48·3	0·000	0·639	47·1	0·000	0·641
BB	13·6	−0·354	0·799	11·9	−0·354	0·820
Weighted means		0·087	0·606		0·103	0·592

For both models the proportion of males and females selected is 3.13 and 25% respectively and the heritability is 0·5. The frequencies, average polygenic values and QTL values of the AA, AB and BB genotypes (A represents the favourable allele) as well as the weighted polygenic and QTL means are given.

Table 9. *Accuracy of evaluation of polygenic and QTL effects for MAS with eight sires and 16 or 64 dams (D) selected*

		r_u Generation			r_v Generation		
	h^2	1	2	3	1	2	3
16 D	0·5	0·72	0·63	0·62	0·29	0·26	0·23
		(99·4)	(98·5)	(97·5)	(106·8)	(121·0)	(122·4)
	0·25	0·59	0·50	0·49	0·24	0·24	0·23
		(99·4)	(97·3)	(98·2)	(104·8)	(137·4)	(136·2)
	0·1	0·44	0·38	0·36	0·18	0·20	0·21
		(98·7)	(95·6)	(95·3)	(103·7)	(146·2)	(172·2)
64 D	0·5	0·72	0·62	0·63	0·32	0·28	0·21
		(98·6)	(96·9)	(99·1)	(114·5)	(133·0)	(119·6)
	0·25	0·60	0·50	0·51	0·27	0·29	0·25
		(98·1)	(95·2)	(99·4)	(117·5)	(160·3)	(157·1)
	0·1	0·47	0·40	0·39	0·20	0·27	0·25
		(98·1)	(96·0)	(97·0)	(105·1)	(175·4)	(176·1)

Results are presented for the first three generations. r_u represents the correlation between polygenic effects and estimated breeding values and r_v represents the correlation between QTL effects and estimated breeding values. Results are also expressed in % terms relative to conventional BLUP and are given in parentheses – values above 100% indicate superiority over conventional BLUP.

with superior QTL values and that, despite apparent weaknesses of the model, it is relatively robust.

The increased QTL response achieved by MAS compared to conventional BLUP is however accompanied by a reduced polygenic response so that (apart from results in Table 6) the effect of MAS on the total genetic response is relatively small (especially considering the cost and effort involved in typing the animals). This is observed right from the first generation of selection. How can we explain that the polygenic response is lower with MAS? One factor which certainly plays a role is the negative covariances generated between QTL and polygenic effects as a consequence of selection (Bulmer, 1971). This gametic phase disequilibrium is demonstrated in Table 8 for a specific example. For both MAS and conventional BLUP, selected animals of genotype AA tend to have low polygenic values while BB animals tend to have high polygenic values (otherwise they would not have been selected).

We might envisage then that MAS, which leads to increased QTL responses, might also yield lower polygenic responses than conventional BLUP as a direct result of gametic phase disequilibrium. How-

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

ever, this does not seem to be the case. Detailed examination of the genotype frequencies and within-genotype polygenic means of selected animals indicates that the intensities or accuracies of selection of polygenic effects must be lower with MAS than with conventional BLUP. Confirming this is not straightforward since conventional BLUP provides estimates of breeding values only and not of QTL or polygenic effects individually. Nevertheless, we can measure the correlations between QTL or polygenic effects with estimated breeding values (EBVs) for both models (Table 9) and these show very clearly that QTL effects are more accurately evaluated and that polygenic effects are less accurately measured with MAS than with conventional BLUP. In passing, it is also interesting to note that whereas the accuracy of selection of polygenic effects falls substantially from generations one to two it changes relatively little for QTL effects, with the result that the precision with which QTL effects are estimated (and hence the resulting QTL response achieved – Table 3) is much higher for MAS compared with conventional BLUP at generations two and three than at generation one.

Thus, in conclusion it seems that the principal reason for MAS yielding lower polygenic responses is that the polygenic effects are evaluated less accurately than with conventional BLUP. The reduced precision of polygenic effects is not due to gametic phase disequilibrium since it is found even when evaluating the candidates for selection at generation one, which are bred by founders chosen at random. Instead, by examining the three components of the polygenic accuracy of selection, the reduction seems to be due to reduced covariances between polygenic effects and EBVs, which probably stem from statistical considerations such as higher error variances of estimated polygenic effects with MAS.

Our findings were validated using a simple deterministic prediction model in which each of the unrelated candidates for selection is evaluated using performance data of the candidate, of its parents, of the mates of its sire (a hierarchical mating design is used) and of the average performance of full and half sib groups. These groups are further subdivided into four subgroups based on the pairwise values of θ_t^p (which may be 0·99 or 0·01) and θ_t^m (which may be 0·99 or 0·01) of the animals. Weights were calculated according to selection index theory assuming either a conventional polygenic model or a model including a single QTL with normally distributed effects (Lande, 1975), an assumption implicit in the model of Fernando & Grossman (1989). Responses were calculated assuming the QTL model was true.

Results for generation one from the deterministic model with a population of eight sires and 16 dams show that the variances of EBVs are 0·1 and 5·2% higher with MAS compared with conventional BLUP with heritabilities of 0·5 and 0·1 respectively. This increase is due to greater covariances between EBVs and QTL values (59 and 58% higher respectively) and not due to covariances between EBVs and polygenic values (which in fact decrease by 4·1 and 1·1% respectively). Therefore, the predicted polygenic response, which is proportional to the covariance between EBVs and polygenic values divided by the standard deviation of the distribution of EBVs, is reduced. The agreement between deterministic and simulated results suggests that the findings from simulation concerning polygenic responses are not linked to the fact that a non-normal distribution of QTL effects was used in the simulations. Indeed, using normally distributed QTL effects for the deterministic model accentuated our findings – MAS, compared with conventional selection, was even more efficient for QTL response and even less efficient for polygenic response.

The relative selection responses (Table 3) and accuracies of selection (Table 4) from MAS compared with conventional BLUP are not constant over time. In general, they reach a maximum at generations 2 and 3 and tend to decline thereafter. This can be explained by the fact that, with each cycle of selection, the favourable allele moves closer to fixation. With MAS, the relative accuracy of selection declines over time because errors in estimating the probability of origin of QTL alleles accumulate and because marker information is used to estimate QTL differences which, eventually, do not actually exist. This is best illustrated by the situation where the QTL accounts for 25% of the total genetic variance (Table 7). Here, the favourable allele reaches a frequency of about 0·95 after three generations of selection and consequently, the selection response from generation four onwards is actually inferior to conventional BLUP. These results concerning the effect of time on selection responses echo to a certain degree those of Gibson (1994) who, in the situation where the QTL genotype is known with certainty, found that, compared to ignoring QTL information, genetic gain using the genotypes was greater in the short term but lower in the long term.

Simulated results presented here confirm the conclusion of Lande & Thompson (1990) that MAS may be most useful when the heritability of the trait of interest is low. Results in Table 3 for the situation where the distance between flanking markers is 20 cM suggest that this is true. Results in Table 6 demonstrate this even more clearly by showing that by reducing the distance between flanking markers to 4 cM there is only a minor effect on genetic response when the heritability is 0·5 whereas it increases genetic response substantially when the heritability is 0·1. These findings emphasize that MAS may be of considerable use for lowly heritable traits and that, once QTLs for such characters have been identified the search for closely linked marked should be intensified, as proposed by Smith & Smith (1993).

In developing the methodology for evaluating candidates for MAS it was assumed that the calculation of the probability of origin of QTL alleles was conditiond only on marker information. This reduces the efficiency of MAS. Furthermore, when inferences are made using marker data alone, the probabilities of origin of QTL alleles for the animals selected are no longer correct. Repeated errors of this kind over successive generations affect the system of equations for the Fernando & Grossman (1989) model and thus reduce the efficiency of MAS. Accounting for performance would improve evaluation but the relevant theory remains to be worked out and is beyond the scope of this paper.

An approximation for the calculation of the probability of origin of QTL alleles under conditions of uncertain transmission of marker alleles was used in this study. What effect did it have on our results? Because the number of unique marker alleles is twice the number of founder animals, the problem of uncertainty is greatest when 16 dams are selected. Since each founder has unique marker alleles, the first cases of uncertainty arise in generation two. From simulation, the proportion of 'uncertain' animals was calculated and was found to represent on average about 3 and 4% of animals per generation from generations two to six with heritabilities of 0·5 and 0·1 respectively. The effect on genetic response was investigated by assuming that no uncertainty exists and by correctly assigning the markers received from the sire and from the dam for 'uncertain' animals. Response per generation was found to be changed by no more than 0·3% so we can safely conclude that uncertain transmission of marker haplotypes had no effect on the results presented here.

It should be emphasized that, based on this study, no inferences may be made about the efficiency of MAS if the QTL shows dominance. BLUP animal models accounting for marker information in such situations have yet to be developed and, to evaluate their efficiency once developed, results should be compared with conventional BLUP animal models which include dominance effects (e.g. De Boer & Hoeschele, 1993).

In a deterministic comparison of dairy cattle breeding schemes, Meuwissen & Van Arendonk (1992) found that the use of MAS did not alter the standard deviation of genetic gain and thus concluded that inbreeding rates are not much affected by MAS. In this study the standard deviation of response for MAS (results not shown) was similar to that of conventional BLUP. Table 5 shows, however, that the conclusion of Meuwissen & Van Arendonk (1992) is valid only when loci independent of the character under selection are considered. Inbreeding rates at neutral loci remained relatively constant but, at the QTL, they were 10–40% higher with MAS than with conventional BLUP. Inbreeding rates for conventional BLUP were 9–24% higher at the QTL than for neutral loci. For MAS, since more selection pressure was applied to the QTL, inbreeding rates were 36–63% higher at the QTL than for neutral loci.

This study shows that the application of MAS is not as easy or straightforward as we might have expected. However, in the schemes examined, all animals of both sexes were recorded for the character of interest. Further investigations should concentrate on those situations where MAS is likely to provide additional advantages, as proposed by Zhang & Smith (1993), such as the evaluation of animals before they express the character of interest or, in the case of sex-limited traits, the evaluation of animals that cannot be recorded. Preliminary results from the simple deterministic model mentioned earlier, for the situation where candidates have no performance data, support this conclusion.

The authors are grateful to two anonymous referees for their constructive comments and careful reading of the manuscript. J.R. acknowledges receipt of an INRA postdoctoral fellowship.

References

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., Andersson-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.

Beattie, C. W. (1994). Livestock genome maps. *Trends in Genetics* **10**, 334–338.

Bink, M. C. A. M. & Van Arendonk, J. A. M. (1994). Marker-assisted prediction of breeding values in dairy cattle populations. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Guelph, **21**, 233–236.

Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S. L. F., Hawkins, G. A., Toldo, S. S., Fries, R., Grosz, M. D., Yoo, J. & Beattie, C. (1994). A genetic linkage map for cattle. *Genetics* **136**, 619–639.

Bulmer, M. G. (1971). The effect of selection on genetic variability. *American Naturalist* **105**, 201–211.

De Boer, I. J. M. & Hoeschele, I. (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theoretical and Applied Genetics* **86**, 245–258.

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*, 3rd edn. Harlow, UK: Longman.

Fernando, R. L. & Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genetics, Selection and Evolution* **21**, 467–477.

Gibson, J. P. (1994). Short-term gain at the expense of long-term response with selection of identified loci. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Guelph, **21**, 201–204.

Gimelfarb, A. & Lande, R. (1994). Simulation of marker assisted selection in hybrid populations. *Genetical Research* **63**, 39–47.

Goddard, M. E. (1992). A mixed model for analyses of data on multiple genetic markers. *Theoretical and Applied Genetics* **83**, 878–886.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between loci of linked factors. *Journal of Genetics* **8**, 299–309.

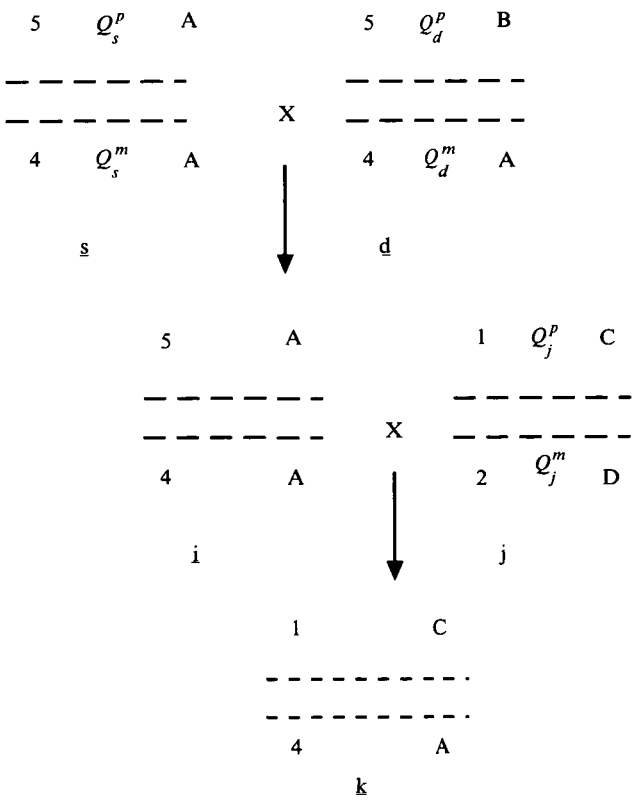
Hospital, F., Chevalet, C. & Mulsant, P. (1992). Using

markers in gene introgression breeding programs. *Genetics* **132**, 1199–1210.
Lande, R. (1975). The maintenance of genetic variability by mutation in a quantitative character with linked loci. *Genetical Research* **26**, 221–235.
Lande, R. & Thompson, R. (1990). Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743–756.
Meuwissen, T. H. E. & Van Arendonk, J. A. M. (1992). Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science* **75**, 1651–1659.
Paterson, A. H., Lander, E. S., Hewitt, J. D., Petersen, S., Lincoln, S. E. & Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721–726.
Rohrer, G. A., Alexander, L. J., Keele, J. W., Smith, T. P. & Beattie, C. (1994). A microsatellite linkage map of the porcine genome. *Genetics* **136**, 234–245.
Schaeffer, L. R. & Kennedy, B. W. (1986). Computing strategies for solving mixed model equations. *Journal of Dairy Science* **69**, 575–579.
Smith, C. & Smith, D. B. (1993). The need for close linkages in marker-assisted selection for economic merit in livestock. *Animal Breeding Abstracts* **61**, 197–204.
Soller, M. (1978). The use of loci associated with quantitative traits in dairy cattle improvement. *Animal Production* **27**, 133–139.
Wang, T., Van der Beek, S., Fernando, R. L. & Grossman, M. (1991). Covariance between effects of marked QTL alleles. *Journal of Animal Science* **69**, 202 (supplement).
Zhang, W. & Smith, C. (1992). Computer simulation of marker-assisted selection utilizing linkage disequilibrium. *Theoretical and Applied Genetics* **83**, 813–820.
Zhang, W. & Smith, C. (1993). Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. *Theoretical and Applied Genetics* **86**, 492–496.

Appendix

Approximation to calculate the probabilities of origin of QTL alleles

To illustrate the method, θ_i^p, θ_k^p and θ_k^m are calculated for the following pedigree. The phase and origin of the marker haplotypes are assumed known for s, d and j so that H_s = 5A/4A (indicating that the 5A and 4A marker haplotypes came from the sire and dam of s respectively), H_d = 5B/4A and H_j = 1C/2D respectively. Note that 1, 2, 4 and 5 represent marker alleles at the M locus and A, B, C and D are alleles at the N locus.



(*a*) *Calculation of* θ_i^p

(i) Which markers did i receive from s and d? i.e. Calculate $\mathbf{Pr}(H_i = h_i \mid T_i, H_s, H_d)$. In most cases this will be obvious. If not construct a gametic table, thus α and β represent the probabilities that the gametes are non-recombinant or recombinant and have values of 0·5 $(1-2r+2r^2)$ and $(r-r^2)$ respectively, and r is the recombination rate between the QTL and either of the two markers. α and β sum to 0·5.

Gametes of s	Probability	Gametes of d	Probability
S_1 = 5 A	α	D_1 = 5 B	α
S_2 = 4 A	α	D_2 = 4 A	α
S_3 = 5 A	β	D_3 = 5 A	β
S_4 = 4 A	β	D_4 = 4 B	β

The individual i could have received its markers by four different pathways – $S_1 D_2$, $S_3 D_2$, $S_2 D_3$ or $S_4 D_3$. The probabilities of these four events are $\alpha^2/(\alpha+\beta)^2$; $\alpha\beta/(\alpha+\beta)^2$; $\alpha\beta/(\alpha+\beta)^2$ and $\beta^2/(\alpha+\beta)^2$ respectively. With r equal to 0·1 they are 0·6724, 0·1476, 0·1476 and 0·0324 respectively.

(ii) *What is the value of* θ_i^p? i.e. calculate $(\theta_i^p \mid H_i = h_i, H_s = h_s, H_d = h_d)$ for the four pathways. From Table 1 the values of θ_i^p for the four pathways are $r^2/(1-2r+2r^2)$; 0·5; $(1-r)^2/(1-2r+2r^2)$ and 0·5 respectively. With r equal to 0·1 these have values 0·0122, 0·5, 0·9878 and 0·5.

(iii) *Weighted* θ_i^p. Now, weight θ_i^p for each path by the probability of occurrence of that path i.e. θ_i^p = (0·6724) (0·0122) + (0·1476) (0·5) + (0·1476) (0·9878) + (0·0324) (0·5) = 0·244. Note that similar computations must also be done for θ_i^m.

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press

(*b*) *Calculation of* θ_k^p *and* θ_k^m

In this situation, there is certainty that k received the haplotype 4A from its sire i and 1C from its dam j. Thus $H_k = 4A/1C$. The phase of the markers is known with certainty for its dam but not for its sire. Its sire i is 5A/4A if $S_1 D_2$ or $S_3 D_2$ occurred (probability = 0·6724 + 0·1476 = 0·82) and 4A/5A if $S_2 D_3$ or $S_4 D_3$ occurred (probability = 0·18). The value of θ_k^p is thus 0·82 $(1-r)$ + 0·18 (r) = 0·756.

There is no uncertainty regarding the marker phase of j so θ_k^m, from Table 1, has a value $r^2/(1-2r+2r^2)$.

https://doi.org/10.1017/S0016672300034406 Published online by Cambridge University Press