Genet. Sel. Evol. (1989) 21, 467-477

Original article

Marker assisted selection using best linear unbiased prediction

R.L. Fernando* and M. Grossman

University of Illinois, Department of Animal Sciences, 1207 West Gregory Drive Urbana, IL 61801, USA

(received 12 May 1989; accepted 28 August 1989)

Summary – Best linear unbiased prediction (BLUP) is applied to a mixed linear model with additive effects for alleles at a market quantitative trait locus (MQTL) and additive effects for alleles at the remaining quantitative trait loci (QTL). A recursive algorithm is developed to obtain the covariance matrix of the effects of MQTL alleles. A simple method is presented to obtain its inverse. This approach allows simultaneous evaluation of fixed effects, effects of MQTL alleles, and effects of alleles at the remaining QTLs, using known relationships and phenotypic and marker information. The approach is sufficiently general to accommodate individuals with partial or no marker information. Extension of the approach to BLUP with multiple markers is discussed.

marker-assisted selection – best linear unbasied prediction – genetic marker

Résumé – Sélection assistée par un marqueur: utilisation du meilleur prédicteur linéaire sans biais (BLUP). *La méthode du BLUP (meilleure prédiction linéaire sans biais) est appliquée à un modèle linéaire mixte comprenant des effets additifs associé aux allèles d'un locus quantitatif flanqué d'un gène marqueur, et d'effets additifs pour les autres locus quantitatifs. Un algorithme récursif permet d'obtenir la matrice de covariances associée aux effets des allèles du locus marqué. Une méthode simple est aussi proposée pour calculer l'inverse de cette matrice. Cette approche permet d'évaluer simultanément les effets fixés, les effets des allèles du locus marqué, et les effets génétiques additifs de l'ensemble des autres locus, d'après les relations de parenté, les données phénotypiques et l'information sur les marqueurs. Cette approche est assez générale pour tenir compte de données incomplètes chez certains individus. On discute l'extension à un BLUP avec plusieurs marqueurs.*

sélection assistée par un marqueur – meilleure prédiction linéaire sans biais – marqueur génétique

INTRODUCTION

Genetic engineering techniques have produced a variety of molecular genetic markers with the potential to identify a large number of genetic polymorphisms (Soller

*Author to whom correspondence should be addressed.

and Beckmann, 1982; Smith and Simpson, 1986; Schumm *et al.*, 1988). Marker-assisted selection is one application of these techniques to animal and plant breeding. Information on marker loci that are linked to quantitative trait loci, together with phenotypic information, could be used to increase genetic progress by increasing accuracy of selection and by reducing generation interval (Soller, 1978; Smith and Simpson, 1986).

Geldermann (1975) proposed a least-squares procedure to estimate effects of marker alleles on quantitative traits. Based on selection index principles, Soller (1978) combined marker information and phenotypic information to obtain genetic evaluations. This method has been used to study the additional genetic progress expected from marker-assisted selection (Soller, 1978; Soller and Beckmann, 1983, Smith and Simpson, 1986). Because of the complex nature of animal breeding data, however, these methods may not be applicable directly to marker-assisted selection with field data.

Data from field-recorded populations are affected by non-genetic nuisance factors, such as age of animal, age of dam, management system, season of birth and herb. Also, non-random mating, selection and overlapping generations contribute to the complexity of the data. Best linear unbasied prediction (BLUP; Henderson, 1973, 1975, 1982) deals with these complications when predicting breeding values from phenotypic data. The objective of this paper is to present methodology for the application of BLUP to marker-assisted selection in animal breeding. Each methodological development is illustrated with a numerical example using a single hypothetical pedigree

METHODOLOGY

Consider a single polymorphic marker locus (ML), closely linked to a quantitative trait locus (QTL). Let M_i^p and M_i^m denote alleles at the ML that individual i inherited from its paternal (p) and its maternal (m) parent, and let Q_i^p and Q_i^m denote alleles at the market QTL (MQTL) linked to M_i^p and M_i^m, as shown below:

M_i^p	Q_i^p

M_i^m	Q_i^m

Let v_i^p and v_i^m be the additive effects of Q_i^p and Q_i^m. Additive effects of alleles at the remaining QTLs, unlinked to the ML, will be denoted by the residual additive effect u_i. Now, the additive effect for individual i, a_i, can be written as

$$a_i = v_i^p + v_i^m + u_i \tag{1}$$

The usual model to obtain BLUP if additive effects, given phenotypic information, is

$$y_i = \mathbf{x}_i'\beta + a_i + e_i \tag{2}$$

where y_i is the phenotypic value of individual i, $\mathbf{x}_i'$ is a vector of known constants, β is a vector of unknown fixed effects, and e_i is a random error. Using equ.(2), BLUP allows information from relatives to contribute to the predictor of a_i through the

covariance matrix of a_i values. Note that this covariance matrix depends on the type of genetic information available. When only relationship information ($\boldsymbol{r}$) is available, the covariance of a_i values is

$$\mathrm{Cov}(\mathbf{a}|\boldsymbol{r}) = \boldsymbol{G}_{a|r}$$

which is proportional to the numerator relationship matrix (*e.g.*, Henderson, 1976). When marker information ($\boldsymbol{m}$) is also available, the covariance matrix a_i values is

$$\mathrm{Cov}(\mathbf{a}|\boldsymbol{r},\boldsymbol{m}) = \boldsymbol{G}_{a|r,m}$$

It can be shown that $\boldsymbol{G}_{a|r} \neq \boldsymbol{G}_{a|r,m}$, in general. For example, the covariance between half-sibs that receive the same ML allele from their common parent is higher than the covariance between half-sibs that receive different ML alleles. This is because half-sibs receiving the same ML allele also receive the same MQTL allele with greater frequency than half-sibs receiving different ML alleles.

A. Marker model

To obtain BLUP with phenotypic and marker information, it is convenient to use

$$y_i = \mathbf{x}_i'\beta + v_i^p + v_i^m + u_i + e_i \tag{3}$$

which is equivalent to equ.(2). The covariance matrix of v_i values ($\boldsymbol{G}_v$) depends on relationship and marker information. The covariance matrix of u_i values ($\boldsymbol{G}_u$) depends only on relationship information and is proportional to the numerator relationship matrix (*e.g.*, Henderson, 1976). Given the covariance matrices $\boldsymbol{G}_v$ and $\boldsymbol{G}_u$, BLUPs of v_i and u_i values can be obtained using the mixed model equations (Henderson, 1973). The inverse of $\boldsymbol{G}_u$, which is required on the mixed model equations, usually is obtained using an algorithm given by Henderson (1976). A recursive algorithm to construct $\boldsymbol{G}_v$ is given in section B, and an algorithm to obtain its inverse is in section C.

B. Covariance matrix of MQTL effects

1. Theory. To construct $\boldsymbol{G}_v$, consider the covariance between additive effects of MQTL alleles. Without loss of generality, consider only paternal MQTL alleles. Suppose arbitrary individuals o and o' have sires s and s'. The MQTL alleles inherited by o and o' from their sires are Q_o^p and Q_o^p, having additive effects v_o^p and $v_{o'}^p$. For paternal MQTL alleles in o and o', the covariance between their additive effects v_o^p and $v_{o'}^p$ is

$$\begin{aligned}\mathrm{Cov}(v_o^p, v_{o'}^p) &= \mathrm{Cov}(v_o^p, v_{o'}^p | Q_o^p \equiv Q_{o'}^p) \cdot P(Q_o^p \equiv Q_{o'}^p) \\ &= \mathrm{Var}(v_o^p) \cdot P(Q_o^p \equiv Q_{o'}^p)\end{aligned} \tag{4}$$

where $\mathrm{Var}(v_o^p) = \sigma_v^2$ is the additive variance of an MQTL allele and $P(Q_o^p \equiv Q_{o'}^p)$ is the probability that Q_o^p is identical by descent to $Q_{o'}^p$. For an arbitrary pair of individuals, one is not a direct descendent of other. If o is not a direct descendant of the o', Q_o^p can be identical by descent to $Q_{o'}^p$ in 2 mutually exclusive ways:

1) Q_o^p is identical by descent to the maternal MQTL allele of the sire of o' ($Q_{s'}^p$) and o' inherits $Q_{s'}^p$ or
2) Q_o^p is identical by descent to the paternal MQTL allele of the sire of o' ($Q_{s'}^m$) and o' inherits $Q_{s'}^m$.

If marker information is available, the conditional probability that o' inherits $Q^p_{s'}$, given that o' inherits $M^p_{s'}$, is $(1-r)$, where r is the recombination rate between the ML and the MQTL. Thus if o' inherits $M^p_{s'}$, the probability in equ.(4) can be calculated recursively as

$$P(Q^p_o \equiv Q^p_{o'}) = P(Q^p_o \equiv Q^p_{s'}) \cdot (1-r) + P(Q^p_o \equiv Q^m_{s'}) \cdot r \tag{5}$$

Similarly, if o' inherits $M^m_{s'}$

$$P(Q^p_o \equiv Q^p_{o'}) = P(Q^p_o \equiv Q^p_{s'}) \cdot r + P(Q^p_o \equiv Q^m_{s'}) \cdot (1-r) \tag{6}$$

If marker information is not available, so that it is not known whether o' inherits $M^p_{s'}$ or $M^m_{s'}$, 0.5 replaces r in equs.(5) and (6). This is because, in the absence of marker information, $Q^p_{s'}$ and $Q^m_{s'}$ have equal probability of being transmitted to o'.

The above development leads to a tabular method to construct $\boldsymbol{G}_v$, which is similar to the method used to construct the numerator relationship matrix (*e.g.* Henderson, 1976). Note that $\boldsymbol{G}_v$ has twice as many rows as individuals because each individual has 2 effects: 1 for the paternal and 1 for the maternal MQTL allele. The rows and columns of $\boldsymbol{G}_v$ should be ordered so that those corresponding to progeny follow those for their parents. Let the row indices of $\boldsymbol{G}_v$, corresponding to the effects of MQTL alleles of individual $o(v^p_o, v^m_o)$, be i^p_o, i^m_o; of its sire $s(v^p_s, v^m_s)$, be i^p_s, i^m_s; and of its dam $d(v^p_d, v^m_d)$, be i^p_d, i^m_d. Also, let element ij of $\boldsymbol{G}_v$ be g_{ij}. Then from equs.(4), (5) and (6), the elements of row i^p_o, below the diagonal, are obtained as

$$g_{i^p_o, j} = (1-\rho^p_o) g_{i^p_s, j} + \rho^p_o g_{i^m_s, j} \tag{7a}$$

for $j = 1 \ldots i^p_o - 1$, where $\rho^p_o = r$ if o inherits M^p_s or $\rho^p_o = (1-r)$ if o inherits M^m_s. Elements of column i^p_o, above the diagonal, are obtained from the corresponding row elements because $\boldsymbol{G}_v$ is symmetric. Similarly, elements of row i^m_o, below the diagonal, are obtained as

$$g_{i^m_o, j} = (1-\rho^m_o) g_{i^p_d, j} + \rho^m_o g_{i^m_d, j} \tag{7b}$$

for $j = 1 \ldots i^m_o - 1$, where $\rho^m_o = r$ if o inherits M^p_d and $\rho^m_o = (1-r)$ if o inherits M^m_d. Elements of column i^m_o, above the diagonal, are obtained from the corresponding row elements.

From equ.(4), the diagonal elements of $\boldsymbol{G}_v$ are equal to σ^2_v. If marker information cannot be used to determine which of the 2 marker alleles o are inherited from its sire or its dam, then 0.5 replaces ρ^p_o in equ.(7a) or ρ^m_o in (7b).

2. Numerical example. Consider the pedigree in Table I. To construct $\boldsymbol{G}_v$, rows and columns are arranged by individual and by paternal and maternal MQTL alleles within individual (Table II). For convenience, we will assume that $\sigma^2_v = 1$ and that $r = 0.1$. The first two individuals are assumed to be unrelated; thus the upper left 4×4 submatrix of $\boldsymbol{G}_v$ is the identity matrix. Elements on the diagonal are equal to $\sigma^2_v = 1$. Now, row elements below the diagonal can be obtained from equs.(7a) and (7b); column elements above the diagonal are obtained by symmetry. Each row element for v^p_3 is equal to $(1-r) = 0.9$ times the corresponding row element for v^p_1 plus $r = 0.1$ times the corresponding row element for v^m_1. Each row element for v^m_3 is equal to $r = 0.1$ times the corresponding row element for v^p_2 plus $(1-r) = 0.9$ times the corresponding row element for v^m_2. The ML allele inherited by 4 from

its sire is unknown. Thus, each row element for v_4^p is the mean ($r = 0.5$) of the corresponding row element for v_1^p and for v_1^m. Marker information is available for v_4^m, so that each row element for v_4^m is $(1 - r) = 0.9$ times the corresponding row element for v_3^p plus $r = 0.1$ times the corresponding row element for v_3^m.

Table I. Pedigree with marker information

Animal	*Sire*	*Dam*	*Marker inherited from*	
			Sire	*Dam*
1	–	–	–	–
2	–	–	–	–
3	1	2	M_1^p	M_2^m
4	1	3	–	M_3^p

Table II. Covariance matrix of MQTL effects: $\boldsymbol{G}_v$

	v_1^p	v_1^m	v_2^p	v_2^m	v_3^p	v_3^m	v_4^p	v_4^m
v_1^p	1.0	0.0	0.0	0.0	0.9	0.0	0.5	0.81
v_1^m	0.0	1.0	0.0	0.0	0.1	0.0	0.5	0.09
v_2^p	0.0	0.0	1.0	0.0	0.0	0.1	0.0	0.01
v_2^m	0.0	0.0	0.0	1.0	0.0	0.9	0.0	0.09
v_3^p	0.9	0.1	0.0	0.0	1.0	0.0	0.5	0.9
v_3^m	0.0	0.0	0.1	0.9	0.0	1.0	0.0	0.1
v_4^p	0.5	0.5	0.0	0.0	0.5	0.0	1.0	0.45
v_4^m	0.81	0.09	0.01	0.09	0.9	0.1	0.45	1.0

C. Algorithm for inverting $\boldsymbol{G}_v$

1. Theory. The approach taken here follows that by Quaas *et al.* (1984) and Quaas (1988) to invert the matrix of additive relationships. We define a linear model to relate the effect of the paternal MQTL allele of an individual (o) to effects of paternal and maternal MQTL alleles of its sire (s)

$$v_o^p = (1 - \rho_o^p)v_s^p + \rho_o^p v_s^m + \varepsilon_o^p \tag{8a}$$

where ε_o^p is a residual effect. Similarly, a linear model for effect of the maternal MQTL allele of o is

$$v_o^m = (1 - \rho_o^m)v_d^p + \rho_o^m v_d^m + \varepsilon_o^m \tag{8b}$$

It can be shown that the residuals ε_o^p in equ.(8a) and ε_o^m in (8b) have a diagonal covariance matrix ($\boldsymbol{G}_\varepsilon$; see Appendix). Now, the vector of effects of MQTL alleles ($\mathbf{v}$) can be written as

$$\mathbf{v} = \boldsymbol{P}\mathbf{v} + \boldsymbol{\varepsilon} \tag{9}$$

where $\boldsymbol{P}$ is a matrix with each row containing only two non-zero elements, if the parent is known or containing only zeros, if the parent is unknown; and where $\boldsymbol{\varepsilon}$ is a vector of residuals. For example, row i_o^p will have $(1-\rho_o^p)$ in column i_s^p and ρ_o^p in column i_s^m, if the sire of i is known. Similarly, row i_o^m will have $(1-\rho_o^m)$ in the column i_d^p and ρ_o^m in column i_d^m, if the dam of i is known.

To proceed, we need the diagonal elements of $\boldsymbol{G}_\varepsilon$. Consider, for example, the variance of ε_o^p. From equ.(8a), if the sire of o is known

$$\begin{aligned}\mathrm{Var}(v_o^p) = (1-\rho_o^p)^2 \cdot \mathrm{Var}(v_s^p) + (\rho_s^p)^2 \cdot \mathrm{Var}(v_s^m) \\ +2(1-\rho_o^p)\rho_o^p \cdot \mathrm{Cov}(v_s^p, v_s^m) + \mathrm{Var}(\varepsilon_o^p)\end{aligned}$$

because effects of MQTL alleles of sire s are uncorrelated with residuals of its offspring o (see Appendix). Hence

$$\begin{aligned}\mathrm{Var}(\varepsilon_o^p) = \mathrm{Var}(v_o^p) - (1-\rho_o^p)^2 \cdot \mathrm{Var}(v_s^p) - (\rho_o^p)^2 \cdot \mathrm{Var}(v_s^m) \\ -2(1-\rho_o^p)\rho_o^p \cdot \mathrm{Cov}(v_s^p, v_s^m)\end{aligned} \tag{10}$$

The covariance between the effects of paternal and maternal MQTL alleles can be written as

$$\mathrm{Cov}(v_s^p, v_s^m) = \mathrm{Var}(v_s^p) \cdot P(Q_s^p \equiv Q_s^m) = \mathrm{Var}(v_s^p) \cdot F_s \tag{11}$$

where F_s is the inbreeding of sire s. Now, equ.(10) can be written as

$$\mathrm{Var}(\varepsilon_o^p) = 2\sigma_v^2(1-\rho_o^p)\rho_o^p(1-F_s) \tag{12a}$$

because $\mathrm{Var}(v_o^p) = \mathrm{Var}(v_s^p) = \mathrm{Var}(v_s^m) = \sigma_v^2$, and where $(1-\rho_o^p)\rho_o^p = (1-r)r$ for ρ_o^p or for $\rho_o^p = (1-r)$. When the sire is not inbred: $\mathrm{Var}(\varepsilon_o^p) = 2\sigma_v^2(1-r)r$, if marker information is available; or $\mathrm{Var}(\varepsilon_o^p) = \sigma_v^2/2$, if marker information is not available. If the sire is not known, $\mathrm{Var}(\varepsilon_o^p) = \sigma_v^2$.

Similarly, if dam of o is known, the variance of ε_o^m is

$$\mathrm{Var}(\varepsilon_o^m) = 2\sigma_v^2(1-\rho_o^m)\rho_o^m(1-F_d) \tag{12b}$$

where $(1-\rho_o^m)\rho_o^m = (1-r)r$ for $\rho_o^m = r$ or for $\rho_o^m = (1-r)$ and where F_d is the inbreeding of dam d. When the dam is not inbred: $\mathrm{Var}(\varepsilon_o^m) = 2\sigma_v^2(1-r)r$, if marker information is available; or $\mathrm{Var}(\varepsilon_o^m) = \sigma_v^2/2$, if marker information is not available. If the dam is not known, $\mathrm{Var}(\varepsilon_o^m) = \sigma_v^2$.

Rearranging (9), $\boldsymbol{v}$ can be written as

$$\boldsymbol{v} = (\boldsymbol{I}-\boldsymbol{P})^{-1}\boldsymbol{\varepsilon} \tag{13}$$

for non-singular $(\boldsymbol{I}-\boldsymbol{P})$, and thus $\boldsymbol{G}_v$ can be written as

$$\boldsymbol{G}_v = (\boldsymbol{I}-\boldsymbol{P})^{-1}\boldsymbol{G}_\varepsilon(\boldsymbol{I}-\boldsymbol{P}')^{-1} \tag{14}$$

From equ.(14), it is clear that $\boldsymbol{G}_v^{-1}$ can be written as

$$\boldsymbol{G}_v^{-1} = (\boldsymbol{I}-\boldsymbol{P}')\boldsymbol{G}_\varepsilon^{-1}(\boldsymbol{I}-\boldsymbol{P}) \tag{15}$$

As shown earlier, $\boldsymbol{P}$ has a simple structure, with each row containing at most 2 non-zero elements, and $\boldsymbol{G}_\varepsilon^{-1}$ is diagonal.

To obtain the rules for inverting $\boldsymbol{G}_v^{-1}$, equ.(15) is written as

$$\boldsymbol{G}_v^{-1} = \boldsymbol{Q}\boldsymbol{G}_\varepsilon^{-1}\boldsymbol{Q}' \tag{16}$$

where $\boldsymbol{Q} = (\boldsymbol{I}-\boldsymbol{P}')$. Because $\boldsymbol{G}_\varepsilon$ is diagonal, equ.(16) can be written as

$$\boldsymbol{G}_v^{-1} = \sum_{j=1}^{2n} \boldsymbol{q}_j \, \boldsymbol{q}_j' d_j \tag{17}$$

where n is number of individuals in the pedigree, $\boldsymbol{q}_j$ is column j of $\boldsymbol{Q}$, and d_j is diagonal element j of $\boldsymbol{G}_\varepsilon^{-1}$. By definition of $\boldsymbol{Q}$, element j of $\boldsymbol{q}_j$ is unity. Further, $\boldsymbol{q}_j$ will have, at most, only 2 other non-zero elements; for $j = i_o^p$, element i_s^p equals $-(1-\rho_o^p)$ and element i_s^m equals $-\rho_o^p$, if the sire of o is known. Similarly, for $j = i_o^m$, element i_d^p equals $-(1-\rho_o^m)$ and element i_d^m equals $-\rho_o^m$, if the dam of o is known. Thus, given parent and marker information of an individual, the contributions to $\boldsymbol{G}_v^{-1}$, corresponding to effects of paternal and maternal MQTL alleles of the individual, are easily obtained.

Now, to obtain the inverse of $\boldsymbol{G}_v$: 1) calculate diagonals of $\boldsymbol{G}_\varepsilon$: when the parent is known, the diagonal is given by equ.(12a) or (12b), and when the parent is unknown, the diagonal is σ_v^2; 2) set $\boldsymbol{G}_v^{-1}$ to the null matrix; 3) for each offspring o, with sire s and dam d, add the following to the indicated elements of $\boldsymbol{G}_v^{-1}$:

if sire is known, add $(1-\rho_o^p)^2 d_{i_o^p}$ to diagonal element i_s^p, i_s^p;
$(1-\rho_o^p)\rho_o^p) d_{i_o^p}$ to elements i_s^p, i_s^m and i_s^m, i_s^p;
$-(1-\rho_o^p) d_{i_o^p}$ to elements i_s^p, i_o^p and i_o^p, i_s^p;
$(\rho_o^p)^2 d_{i_o^p}$ to diagonal element i_s^m, i_s^m;
and $-\rho_o^p d_{i_o^p}$ to elements i_s^m, i_o^p and i_o^p, i_s^m;
if dam is known, add $(1-\rho_o^m)^2 d_{i_o^m}$ to diagonal element i_d^p, i_d^p;
$(1-\rho_o^m)\rho_o^m d_{i_o^m}$ to elements i_d^p, i_d^m and i_d^m, i_d^p;
$(1-\rho_o^m) d_{i_o^m}$ to elements i_d^p, i_o^m and i_o^m, i_d^p;
$(\rho_o^m)^2 d_{i_o^m}$ to diagonal element i_d^m, i_d^m;
and $-\rho_o^m d_{i_o^m}$ to elements i_d^m, i_o^m and i_o^m, i_d^m
and always add $d_{i_o^p}$ to element i_o^p, i_o^p;
and $d_{i_o^m}$ to element i_o^m, i_o^m

2. Numerical example. Consider the pedigree in Table I. To construct $\boldsymbol{G}_\varepsilon$, we again take $\sigma_v^2 = 1$ and $r = 0.1$. Because the parents of individuals 1 and 2 are not known, the first 4 elements on the diagonal of $\boldsymbol{G}_\varepsilon$ are $\sigma_v^2 = 1$. For individual 3, each parent is known and marker information is available. Thus, from equs.(12a) and (12b), the two diagonals of $\boldsymbol{G}_\varepsilon$ corresponding to effects of paternal and maternal MQTL alleles of individual 3 are $2(1-r)r = 0.18$. Each parent of individual 4 is also known, but the marker inherited from the sire is not known. Therefore, the diagonal of $\boldsymbol{G}_\varepsilon$ corresponding to v_4^p is 0.5, and that corresponding to v_4^m is $2(1-r)r = 0.18$.

The $\boldsymbol{P}$ matrix for this example is given in Table III. The first 4 rows of $\boldsymbol{P}$ are null because parents of the first 2 individuals are not known. The sire of individual 3 is 1, and M_1^p was transmitted to 3. Thus, the row corresponding to v_3^p has $(1-r) = 0.9$ in the column corresponding to v_1^p and $r = 0.1$ in the column corresponding to v_1^m. Similarly, the dam of individual 3 is 2, and M_2^m was transmitted to 3. Thus, the row corresponding to v_3^m has $r = 0.1$ in the column corresponding to v_2^p and $(1-r) = 0.9$ in the column corresponding to v_2^m. The sire of individual 4 is 1, but marker information is not available. Thus, the row corresponding to v_4^p has 0.5 in the columns corresponding to v_1^p and v_1^m. The dam of individual 4 is 3, and M_3^p

was transmitted to 4. Thus, the row corresponding to v_4^m has $(1-r)=0.9$ in the column corresponding to v_3^p and $r=0.1$ in the column corresponding to v_3^m.

The matrix $\boldsymbol{Q}=(\boldsymbol{I}-\boldsymbol{P}')$ is given in Table IV. The product $\boldsymbol{Q}\boldsymbol{G}_\varepsilon^{-1}\boldsymbol{Q}'$ is given in Table V. It can be verified that this is identical to the inverse of the matrix $\boldsymbol{G}_v$ in Table II.

Table III. $\boldsymbol{P}$ matrix

	v_1^p	v_1^m	v_2^p	v_2^m	v_3^p	v_3^m	v_4^p	v_4^m
v_1^p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
v_1^m	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
v_2^p	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
v_2^m	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
v_3^p	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0
v_3^m	0.0	0.0	0.1	0.9	0.0	0.0	0.0	0.0
v_4^p	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0
v_4^m	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0

Table IV. $\boldsymbol{Q}$ matrix

	v_1^p	v_1^m	v_2^p	v_2^m	v_3^p	v_3^m	v_4^p	v_4^m
v_1^p	1.0	0.0	0.0	0.0	−0.9	0.0	−0.5	0.0
v_1^m	0.0	1.0	0.0	0.0	−0.1	0.0	−0.5	0.0
v_2^p	0.0	0.0	1.0	0.0	0.0	−0.1	0.0	0.0
v_2^m	0.0	0.0	0.0	1.0	0.0	−0.9	0.0	0.0
v_3^p	0.0	0.0	0.0	0.0	1.0	0.0	0.0	−0.9
v_3^m	0.0	0.0	0.0	0.0	0.0	1.0	0.0	−0.1
v_4^p	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
v_4^m	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Table V. $\boldsymbol{Q}\boldsymbol{G}_\varepsilon^{-1}\boldsymbol{Q}'$

	v_1^p	v_1^m	v_2^p	v_2^m	v_3^p	v_3^m	v_4^p	v_4^m
v_1^p	6.0	1.0	0.0	0.0	−5.0	0.0	−1.0	0.0
v_1^m	1.0	1.556	0.0	0.0	−0.556	0.0	−1.0	0.0
v_2^p	0.0	0.0	1.056	0.5	0.0	−0.556	0.0	0.0
v_2^m	0.0	0.0	0.5	5.5	0.0	−5.0	0.0	0.0
v_3^p	−5.0	−0.556	0.0	0.0	10.056	0.5	0.0	−5.0
v_3^m	0.0	0.0	−0.556	−5.0	0.5	5.611	0.0	−0.556
v_4^p	−1.0	−1.0	0.0	0.0	0.0	0.0	2.0	0.0
v_4^m	0.0	0.0	0.0	0.0	−5.0	−0.556	0.0	5.556

D. BLUP with multiple markers

If information on another marker locus linked to a QTL is available, the model can be expanded to include effects of alleles of this MQTL. This approach, however, results in $2n$ additional equations for each marker introduced into the analysis. Thus, for a large number of individuals (n) and a large number of MQTLs, solving the mixed model equations may not be feasible. An alternative would be to use equ.(2), with

$$a_i = \sum_k (v_{ki}^p + v_{ki}^m) + u_i \tag{18}$$

where v_{ki}^p and v_{ki}^m are effects of paternal and maternal alleles of the k^{th} MQTL. The covariance matrix of effects of MQTL alleles at each locus ($\boldsymbol{G}_{v_k}$) can be constructed using the tabular method described in Section II.B. Then, assuming gametic equilibrium, the covariance of matrix a_i values ($\boldsymbol{G}_{a|r,m}$) can be obtained as

$$\boldsymbol{G}_{a|r,m} = \boldsymbol{Z}(\sum_k \boldsymbol{G}_{v_k})\boldsymbol{Z}' + \boldsymbol{G}_u \tag{19}$$

where $\boldsymbol{Z}$ is a $n \times 2n$ matrix with elements for row i containing a 1 corresponding to each of the paternal and maternal MQTL effects of individual i and zeros for the remaining elements. The problem with this approach, however, is that it could not be applied to large systems, unless a simple algorithm to invert $\boldsymbol{G}_{a|r,m}$ is available.

DISCUSSION

Results presented here are an application of BLUP to marker-assisted selection. This is a generalization of the method presented by Soller (1978) and Soller and Beckmann (1983). This generalization allows simultaneous evaluation of fixed effects, MQTL effects and the residual QTL effects, using known relationships and phenotypic and marker information. It is sufficiently general to accommodate individuals with partial or no marker information.

Several authors have calculated the additional genetic progress expected from marker-assisted selection (Soller, 1978, Soller and Beckmann, 1983; Smith and Simpson, 1986). Because the method presented here is a generalization of the method considered by these authors, their results give an indication of the advantage expected by using marker-assisted BLUP.

Application of this procedure requires knowledge of the recombination rate (r) between the marker and the MQTL and the variance of the additive effect of the MQTL alleles (σ_v^2). Assuming that effects of MQTL alleles are normally distributed, the model presented here could be used to estimate r and σ_v^2 by restricted maximum likelihood (REML; Patterson and Thompson, 1971). The robustness of REML estimation, with respect to the distribution of effects of MQTL alleles, needs to be examined.

REFERENCES

Geldermann H. (1975) Investigation on inheritance of quantitative characters in animals by gene markers. 1. Methods. *Theor. Appl. Genet.* 46, 319-330

Henderson C.R. (1973) Sire evaluation and genetic trends. *In: Animal Breeding and Genetics Symposium in Honor of Dr Jay L. Lush*, American Society of Animal Science and American Dairy Science Association, Champaign, IL, pp. 10-41

Henderson C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-439

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

Henderson C.R. (1982) Best linear unbiased prediction in populations that have undergone selection. *In: World Congress on Sheep and Beef Cattle Breeding*, (Barton R.A. & Smith W.C. ed.) vol. 1, Dunsmore Press, Palmerston North, NZ, pp. 191-200

Henderson C.R. (1988) Use of an average numerator relationship matrix for multiple-sire joining. *J. Anim. Sci.* 66, 1614-1621

Patterson H.D. & Thompson R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554

Quaas R.L. (1988) Additive genetic model with groups and relationships. *J. Dairy Sci.* 71, 1338-1345

Quaas R.L., Anderson R.D. & Gilmour A.R. (1984) *BLUP School Handbook; Use of Mixed Models for Prediction and for Estimation of (Co)variance Components.* Animal Genetics and Breeding Unit, University of New England, N.S.W. 2351, Australia.

Schumm J.W., Knowlton R.G., Braman J.C., Barker D.F., Botstein D., Akots G., Brown V.A., Gravius T.C., Helms C., Hsiao K., Rediker K., Thurston J.G. & Donis-Keller H. (1988) Identification of more that 500 RFLPs by screening random genomic clones. *Am. J. Hum. Genet.* 42, 143-159

Smith C. & Simpson S.P. (1986) The use of genetic polymorphisms in livestock improvement. *J. Anim. Breeding. Genet.* 103, 205-217

Soller M. (1978) The use of loci associated with quantitative traits in dairy cattle improvement. *Anim. Prod.* 27, 133-139

Soller M., Beckmann J.S. (1982) Restriction fragment length polymorphisms and genetic improvement. *In: World Congress on Genetics Applied to Livestock Production*, Madrid, vol. 6, pp. 396-404

Soller M., Beckmann J.S. (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theor. Appl. Genet.* 67, 25-33

APPENDIX

Proof that G_ϵ is diagonal

Let o be an individual that is not a direct descendant of o'. From equs.(8a) and (8b), the additive effects of the MQTL alleles of o and o' are

$$v_o^z = (1 - \rho_o^z)v_\xi^p + \rho_o^z v_\xi^m + \varepsilon_o^z \tag{A1}$$

and

$$v_{o'}^{z'} = (1 - \rho_{o'}^{z'})v_{\xi'}^p + \rho_{o'}^{z'} v_{\xi'}^m + \varepsilon_{o'}^{z'} \tag{A2}$$

where z can take values p or m, $\xi = s$ when $z = p$, or $\xi = d$ when $z = m$. Similarly, z' can take values p or m, $\xi' = s'$ when $z' = p$, or $\xi' = d$ when $z' = m$. Note that for an arbitrary pair of individuals, ones is not direct descendant of the other. Therefore, to prove that $\boldsymbol{G}_\epsilon$ is diagonal, it is sufficient to show thar the covariance between ε_o^z and $\varepsilon_{o'}^{z'}$ is null.

From equs.(A1) and (A2), the covariance between additive effects of MQTL alleles v_o^z and $v_{o'}^{z'}$ can be written as

$$\begin{aligned}\mathrm{Cov}(v_o^z, v_{o'}^{z'}) &= \mathrm{Cov}[v_o^z, (1 - \rho_{o'}^{z'})v_{\xi'}^p + \rho_{o'}^{z'} v_{\xi'}^m + \varepsilon_{o'}^{z'}] \\ &= (1 - \rho_{o'}^{z'})\mathrm{Cov}(v_o^z, v_{\xi'}^p) + \rho_{o'}^{z'}\mathrm{Cov}(v_o^z, v_{\xi'}^m) \\ &\quad + \mathrm{Cov}(v_o^z, \varepsilon_{o'}^{z'})\end{aligned} \tag{A3}$$

But, from equs.(7a) and (7b)

$$\mathrm{Cov}(v_o^z, v_{o'}^{z'}) = (1 - \rho_{o'}^{z'})\mathrm{Cov}(v_o^z, v_{\xi'}^p) + \rho_{o'}^{z'}\mathrm{Cov}(v_o^z, v_{\xi'}^m) \tag{A4}$$

Thus, for equ.(A3) to equal (A4), the third term in equ.(A3), $\mathrm{Cov}(v_o^z, \varepsilon_{o'}^{z'})$, must be zero. The same reasoning can be used to show that $\mathrm{Cov}(v_\xi^p, \varepsilon_{o'}^{z'})$ and $\mathrm{Cov}(v_\xi^m \varepsilon_{o'}^{z'})$ are zero. Therefore, given that $\mathrm{Cov}(v_o^z, \varepsilon_{o'}^{z'})$, $\mathrm{Cov}(v_\xi^p, \varepsilon_{o'}^{z'})$ and $\mathrm{Cov}(v_\xi^m, \varepsilon_{o'}^{z'})$ are zero, $\mathrm{Cov}(\varepsilon_o^z, \varepsilon_{o'}^{z'})$ must be zero.

Further, taking o to be the a parent of o', the residual $(\varepsilon_{o'}^{z'})$ in equ.(A2) is uncorrelated with $v_{\xi'}^p$ and with $v_{\xi'}^m$ in equ.(A2), because $\mathrm{Cov}(v_o^z, \varepsilon_{o'}^{z'}) = 0$, as shown above. The result that be effect of each MQTL allele of a parent is uncorrelated with the residual $(\varepsilon_{o'}^{z'})$ of its offspring was used to obtain equ.(10)