Backsolving in combined-merit models for marker-assisted best linear unbiased prediction of total additive genetic merit

S Saito, H Iwaisaki

Note

Backsolving in combined-merit models for marker-assisted best linear unbiased prediction of total additive genetic merit

S Saito[1], H Iwaisaki[1,2*]

[1] *Graduate School of Science and Technology;*
[2] *Department of Animal Science, Faculty of Agriculture,*
Niigata University, Niigata 950-21, Japan

Summary – The procedures for backsolving are described for combined-merit models for marker-assisted best linear unbiased prediction, or for the animal and the reduced animal models which contain fixed effects and random effects of total additive genetic merits and residuals. Using the best linear unbiased predictors (BLUP) of the total additive genetic merits and the residuals, with the present procedures, the BLUP of additive genetic effects due to quantitative trait loci (QTLs) unlinked to the marker locus and additive effects due to the marked QTL are also obtained. These backsolutions are identical to the solutions in the Fernando and Grossman animal model.

best linear unbiased prediction / marker-assisted selection / combined-merit model / backsolving / additive effect of marked QTL alleles

Résumé – **Restitution des solutions pour la valeur génétique additive totale en cas de prédiction BLUP utilisant des marqueurs.** *On décrit la procédure de restitution des solutions complètes pour la valeur génétique totale à partir des solutions d'un modèle animal réduit. On peut obtenir également des solutions complètes pour les effets génétiques additifs liés à un QTL marqué et les effets liés aux autres gènes. Ces solutions sont identiques à celles du modèle animal de Fernando et Grossman.*

meilleure prédiction linéaire non biaisée / sélection assistée par marqueur / restitution des solutions / QTL marqués

INTRODUCTION

In recent years, a large number of genetic polymorphisms, for example, restricted fragment length polymorphisms (eg, Botstein et al, 1980), variable numbers of tandem repeats (eg, Jeffreys et al, 1985; Nakamura et al, 1987) and random

---

* Correspondence and reprints

amplified polymorphic DNA (eg, Williams et al, 1990), are being detected by molecular techniques. If these are linked to quantitative trait loci (QTLs) affecting quantitative economic traits and are useful as the genetic markers, then marker-assisted prediction of breeding values may be conducted as discussed by Fernando and Grossman (1989). These authors first presented an animal model (AM) procedure to incorporate marker information in a best linear unbiased prediction (Henderson, 1973, 1975, 1984). Following the work of these authors, various models and procedures for the marker-assisted best linear unbiased prediction have been described further (eg, Cantet and Smith, 1991; Goddard, 1992; Hoeschele, 1993; van Arendonk et al, 1994; Togashi et al, 1996; Saito and Iwaisaki, 1996, 1997b).

Van Arendonk et al (1994) presented a combined-merit model, or the AM model combining the additive effects due to marked QTLs (MQTLs) and the effects of alleles at the remaining QTLs into the total additive genetic merit. A reduced animal model (RAM) version of the combined-merit model is also available (Saito and Iwaisaki, 1997b). With these models, the number of systems of equations to be solved is relatively reduced; however, the best linear unbiased predictors (BLUP) of the additive effects of the MQTL alleles and those of the remaining QTLs are not given directly, even if one wishes to know the values for certain animals.

The objective of this paper is to describe the procedures for computing the backsolving of the MQTL- and the remaining QTL-effects in the cases of the combined-merit AM and RAM.

THEORY

Backsolving in the combined-merit AM

Assuming a MQTL and one observation per animal for simplicity, the AM discussed by Fernando and Grossman (1989) is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{Z}(\mathbf{I}_q \otimes \mathbf{1}')\mathbf{v} + \mathbf{e} \qquad [1]$$

In contrast, the combined-merit AM of van Arendonk et al (1994) is expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \qquad [2]$$

with $\mathbf{a} = \mathbf{u} + (\mathbf{I}_q \otimes \mathbf{1}')\mathbf{v}$, where $\mathbf{y}$ is the $n \times 1$ vector of observations, $\boldsymbol{\beta}$ is the $f \times 1$ vector of fixed effects, $\mathbf{u}$ is the $q \times 1$ random vector of additive genetic effects due to alleles at the QTLs not linked to the marker locus, $\mathbf{v}$ is the $2q \times 1$ random vector of additive effects of the MQTL alleles, $\mathbf{a}$ is the $q \times 1$ random vector of the total additive genetic merits or breeding values, $\mathbf{e}$ is the $n \times 1$ vector of random residuals, $\mathbf{X}$ and $\mathbf{Z}$ are $n \times f$ and $n \times q$ known incidence matrices, respectively, $\mathbf{I}_q$ is an identity matrix whose dimension is q, $\mathbf{1}$ is the column vector $(1 \quad 1)'$, and $\otimes$ stands for the direct product operator. For model [2], the expectation and dispersion matrices for the random effects are assumed to be

$$\mathrm{E}\begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and } \mathrm{Var}\begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

with $\mathbf{G} = \mathbf{A_u}\sigma_u^2 + (\mathbf{I}_q \otimes \mathbf{1'})\mathbf{A_v}(\mathbf{I}_q \otimes \mathbf{1})\sigma_v^2$ and $\mathbf{R} = \mathbf{I}_n\sigma_e^2$, where $\mathbf{A_u}$ is the numerator relationship matrix for the QTLs not linked to the marker locus, $\mathbf{A_v}$ is the gametic relationship matrix for the MQTL, $\mathbf{I}_n$ is an identity matrix whose dimension is n, and σ_u^2, σ_v^2 and σ_e^2 are the variance components for the additive effects due to alleles at the QTLs unlinked to the marker locus, for the additive effects of the MQTL alleles and for the residuals, respectively.

The BLUP of the total additive genetic merits, hence, are obtained by solving the following mixed model equations (MME)

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{G}^{-1}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^\circ \\ \widehat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \qquad [3]$$

Then, in the case of the AM, denoting $\mathrm{Cov}([\mathbf{u'}\ \mathbf{v'}]', \mathbf{a'})[\mathrm{Var}(\mathbf{a})]^{-1}$ by $\mathbf{H'}$, the BLUP of additive genetic effects due to QTLs unlinked to the marker locus and additive effects due to the MQTL are further given by

$$\begin{bmatrix} \widehat{\mathbf{u}} \\ \widehat{\mathbf{v}} \end{bmatrix} = \mathbf{H'}\widehat{\mathbf{a}} \qquad [4]$$

with

$$\mathbf{H} = \mathbf{G}^{-1}\begin{bmatrix} \mathbf{A_u}\sigma_u^2 & \vdots & (\mathbf{I}_q \otimes \mathbf{1'})\mathbf{A_v}\sigma_v^2 \end{bmatrix} \qquad [5]$$

Backsolving in the combined-merit RAM

The RAM (Saito and Iwasaki, 1997b) is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{a_p} + \boldsymbol{\theta} \qquad [6]$$

where $\mathbf{y}$, $\mathbf{X}$ and $\boldsymbol{\beta}$ are the same as in equations [1] and [2], $\mathbf{a_p}$ is the appropriate subvector of $\mathbf{a}$ and the subscript p refers to animals with progeny, $\boldsymbol{\theta}$ is the $n \times 1$ residual effects, and $\mathbf{W}$ is the incidence matrix.

With model [6], the assumptions for expectation and dispersion parameters of the random effects are

$$\mathrm{E}\begin{bmatrix} \mathbf{a_p} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and } \mathrm{Var}\begin{bmatrix} \mathbf{a_p} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{G_p} & \mathbf{0} \\ \mathbf{0} & \mathbf{R_r} \end{bmatrix}$$

where $\mathbf{G_p}$ is the appropriate submatrix of $\mathbf{G}$, and $\mathbf{R_r}$ is further expressed as equation [13] of Saito and Iwasaki (1997b).

The BLUP of the total additive genetic merits for parent animals are then obtained by solving the following MME

$$\begin{bmatrix} \mathbf{X'R_r^{-1}X} & \mathbf{X'R_r^{-1}W} \\ \mathbf{W'R_r^{-1}X} & \mathbf{W'R_r^{-1}W} + \mathbf{G_p}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^\circ \\ \widehat{\mathbf{a_p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R_r^{-1}y} \\ \mathbf{W'R_r^{-1}y} \end{bmatrix} \qquad [7]$$

In the case of the RAM, the BLUP of additive genetic effects due to QTLs unlinked to the marker locus and additive effects due to MQTL as obtained by

solving the MME for the full model, or equations [1], are given by the two steps for backsolving for $\widehat{\mathbf{u}_p}$ and $\widehat{\mathbf{v}_p}$ and then for $\widehat{\mathbf{u}_o}$ and $\widehat{\mathbf{v}_o}$, where the subscript o refers to animals without progeny. That is, considering $\mathrm{Cov}([\mathbf{u_p}'\ \mathbf{v_p}']',\ \mathbf{a_p}')[\mathrm{Var}(\mathbf{a_p})]^{-1}$ and $\mathrm{Cov}([\mathbf{u_p}'\ \mathbf{v_p}']',\ \boldsymbol{\theta}')[\mathrm{Var}(\boldsymbol{\theta})]^{-1}$, the BLUP of $\mathbf{u_p}$ and $\mathbf{v_p}$ are first computed as

$$\begin{bmatrix} \widehat{\mathbf{u}_p} \\ \widehat{\mathbf{v}_p} \end{bmatrix} = \mathbf{H}_1'\widehat{\mathbf{a}_p} + \mathbf{H}_2'\widehat{\boldsymbol{\theta}} \qquad [8]$$

with

$$\mathbf{H}_1 = \mathbf{G}_p^{-1}\left[\mathbf{A}_{\mathbf{u}_p}\sigma_u^2 \ \vdots \ (\mathbf{I}_p \otimes \mathbf{1}')\mathbf{A}_{\mathbf{v}_p}\sigma_v^2 \right] \qquad [9]$$

and

$$\mathbf{H}_2 = \mathbf{R}_o^{-1}\left[[\mathbf{T} - \mathbf{K}]\mathbf{A}_{\mathbf{u}_p}\sigma_u^2 \ \vdots \ [(\mathbf{I}_o \otimes \mathbf{1}')\mathbf{B} - \mathbf{K}(\mathbf{I}_p \otimes \mathbf{1}')]\mathbf{A}_{\mathbf{v}_p}\sigma_v^2 \right] \qquad [10]$$

where $\widehat{\boldsymbol{\theta}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}° - \mathbf{W}\widehat{\mathbf{a}_p}$, $\mathbf{A}_{\mathbf{u}_p}$, $\mathbf{A}_{\mathbf{v}_p}$ and $\mathbf{R}_o$ are the appropriate submatrices of $\mathbf{A}_\mathbf{u}$, $\mathbf{A}_\mathbf{v}$ and $\mathbf{R}_r$, respectively, $\mathbf{K}$ is a matrix relating $\mathbf{a}_o$ to $\mathbf{a}_p$, $\mathbf{T}$ has zero elements except for 0.5 in the column pertaining to a known parent of animal i, and $\mathbf{B}$ is a matrix relating the additive MQTL effects of the animals to those of the parents and contains zero elements except for at most four non-zero elements in each row, which are the conditional probabilities for the MQTL (Wang et al, 1995). For details, see Saito and Iwaisaki (1997b).

Then, with $\widehat{\mathbf{u}_p}$ and $\widehat{\mathbf{v}_p}$ provided, the BLUP of $\mathbf{u}_o$ and $\mathbf{v}_o$ are further obtained as

$$\begin{bmatrix} \widehat{\mathbf{u}_o} \\ \widehat{\mathbf{v}_o} \end{bmatrix} = \begin{bmatrix} \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{u}_p} \\ \widehat{\mathbf{v}_p} \end{bmatrix} + \begin{bmatrix} \widehat{\mathbf{m}} \\ \widehat{\boldsymbol{\varepsilon}} \end{bmatrix} \qquad [11]$$

where $\widehat{\mathbf{m}}$ and $\widehat{\boldsymbol{\varepsilon}}$ represent the vectors of the Mendelian sampling effects and the segregation residuals predicted, respectively, which are given as

$$\begin{bmatrix} \widehat{\mathbf{m}} \\ \widehat{\boldsymbol{\varepsilon}} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_o + \mathbf{D}^{-1}\alpha_u & \mathbf{I}_o \otimes \mathbf{1}' \\ \mathbf{I}_o \otimes \mathbf{1} & (\mathbf{I}_o \otimes \mathbf{11}') + \mathbf{G}_\varepsilon^{-1}\alpha_v \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{S} \\ (\mathbf{I}_o \otimes \mathbf{1})\mathbf{S} \end{bmatrix} \qquad [12]$$

where $\alpha_u = \sigma_e^2/\sigma_u^2$, $\alpha_v = \sigma_e^2/\sigma_v^2$, $\mathbf{S} = \mathbf{y_o} - \mathbf{X_o}\boldsymbol{\beta}° - \mathbf{T}\widehat{\mathbf{u}_p} - (\mathbf{I}_o \otimes \mathbf{1}')\mathbf{B}\widehat{\mathbf{v}_p}$, $\mathbf{D}$ is the diagonal matrix whose diagonal elements equal $0.5 - 0.25(F_s + F_d)$ with the inbreeding coefficients of the sire and the dam, F_s and F_d, and $\mathbf{G}_\varepsilon$ is the block-diagonal matrix (Saito and Iwaisaki, 1997a), in which each block is calculated as

$$\begin{bmatrix} 1 & f_i \\ f_i & 1 \end{bmatrix} - \mathbf{B}_{(i)}\mathbf{A}_{\mathbf{v}_{(i)}}\mathbf{B}_{(i)}'$$

where $\mathbf{A}_{\mathbf{v}_{(i)}}$ and $\mathbf{B}_{(i)}$ are appropriate submatrices of $\mathbf{A}_\mathbf{v}$ and $\mathbf{B}$, respectively, which correspond to the parents of animal i, and f_i is the inbreeding coefficient for the MQTL (Wang et al, 1995).

DISCUSSION

The systems of equations in the combined-merit model approach may be compact, relative to that for the AM of Fernando and Grossman (1989), even if the number of MQTLs is high. Compared with the combined-merit AM, the RAM version, applied to species where the fraction of non-parents is high, would lead to a further reduction of the size of the system of equations, although the sparseness in the coefficient matrix of the MME would be adversely affected.

With these models, the inverse covariance matrix of the total additive genetic merits for individual animals or for parent animals in the pedigree file is needed, and moreover the RAM version requires $\mathbf{R_r}$ to be inverted before it can be introduced into equations [7]. For these calculations, certain computing algorithms are available, as discussed by van Arendonk et al (1994) and Saito and Iwaisaki (1997b). Rapid development in computing power may make applications of this type of approach attractive, especially when a large number of markers are considered.

The most relevant information in selecting animals would be the predictors of the total additive genetic merits, which are given directly by the combined-merit model approach. When the models are applied, and one further wishes to compute BLUP of additive genetic effects due to QTLs not linked to the marker locus and/or additive effects due to the MQTL for all or a part of animals, this can be done by using the procedures for backsolving, as just demonstrated in this paper. The backsolutions derived are equivalent to the solutions for the Fernando and Grossman AM. However, the backsolving obviously requires additional computations. Hence, examination of the most efficient numerical techniques would definitely be needed. As an approach, the use of certain transformation techniques might be useful. For the situation where one absolutely needs the solutions in the full model, further research would also be necessary to determine the relative efficiencies of the combined-merit models for computing as compared to the model of Fernando and Grossman (1989) for both cases, single or multiple markers.

REFERENCES

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32, 314-331

Cantet RJC, Smith C (1991) Reduced animal model for marker assisted selection using best linear unbiased prediction. *Genet Sel Evol* 23, 221-233

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

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

Henderson CR (1973) Sire evaluation and genetic trend. 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, USA, 10-41

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423-447

Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, Ontario, Canada

Hoeschele I (1993) Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. *J Dairy Sci* 76, 1693-1713

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314, 67-73

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235, 1616-1622

Saito S, Iwaisaki H (1996) A reduced animal model with elimination of quantitative trait loci equations for marker-assisted selection. *Genet Sel Evol* 28, 465-477

Saito S, Iwaisaki H (1997a) The covariance structure of residual effects in a reduced animal model for marker-assisted selection. *Anim Sci Technol (Jpn)* 68, 1-6

Saito S, Iwaisaki H (1997b) A reduced animal model approach to predicting total additive genetic merit for marker-assisted selection. *Genet Sel Evol* 29, 25-34

Togashi K, Rege JEO, Yamamoto N, Sasaki O, Takeda H (1996) Marker-QTL-association analysis incorporating diversification of QTL variance and its application. *Anim Sci Technol (Jpn)* 67, 923-929

van Arendonk JAM, Tier B, Kinghorn BP (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319-329

Wang T, Fernando RL, van der Beek S, Grossman M, van Arendonk JAM (1995) Covariance between relatives for a marked quantitative trait locus. *Genet Sel Evol* 27, 251-274

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl Acids Res* 18, 6531-6535