Genet. Sel. Evol. 31 (1999) 375–394

Original article

Marker assisted estimation of breeding values when marker information is missing on many animals

Theo H.E. Meuwissen [a*], Mike E. Goddard [b]

[a] Research Institute of Animal Science and Health, Box 65, 8200 AB Lelystad, the Netherlands
[b] Institute of Land and Food Resources, University of Melbourne, Parkville, Victoria 3052, Australia

(Received 15 December 1998; accepted 4 June 1999)

Abstract – Two methods are presented that use information from a large population of commercial animals, which have not been genotyped for genetic markers, to calculate marker assisted estimates of breeding value (MA-EBV) for nucleus animals, where the commercial animals are descendants of the marker genotyped nucleus animals. The first method reduced the number of mixed model equations per commercial animal to one, instead of one plus twice the number of marked quantitative trait loci in conventional MA-EBV equations. Without this reduction, the time taken to solve the mixed model equations including markers could be very large especially if the number of commercial animals and the number of markers is large. The solutions of the reduced set of equations were exact and did not require more iterations than the conventional set of equations. A second method was developed for the situation where the records of the commercial animals were not directly available to the nucleus breeding programme but conventional non-MA-EBVs and their accuracies were available for nucleus animals from a large scale (e.g. national) breeding value evaluation, which uses nucleus and commercial information. Using these non-MA-EBV, the MA-EBV of the nucleus animals were approximated. In an example, the approximated MA-EBV were very close to the exact MA-EBV. © Inra/Elsevier, Paris

marker assisted selection / breeding value estimation / quantitative trait loci / DNA markers

Résumé – Évaluation génétique assistée par marqueurs quand l'information sur les marqueurs est rare. On présente deux méthodes d'utilisation de l'information provenant d'une grande population d'animaux commerciaux, non typés pour des marqueurs, en vue de l'évaluation génétique d'animaux typés dans les noyaux de

* Correspondence and reprints
E-mail: t.h.e.meuwissen@id.dlo.nl

sélection qui sont à l'origine des populations commerciales. La première méthode limite à une seule équation du modèle mixte pour chaque animal commercial au lieu de une plus deux fois, le nombre de loci marqués, quand on utilise les équations classiques du BLUP assisté par marqueurs. Ceci permet de réduire substantiellement le temps de calcul quand le nombre d'animaux commerciaux et le nombre de marqueurs sont grands. Les solutions de ce système réduit sont exactes et ne demandent pas plus d'itérations que le système classique d'équations. La seconde méthode est proposée quand les données des animaux commerciaux ne sont pas directement accessibles aux sélectionneurs du noyau de sélection alors que leurs évaluations classiques (non assistées par marqueurs) le sont. Ces évaluations tiennent alors compte des données des animaux du noyau et hors noyau. Dans ce cas, la méthode est approchée. Sur un exemple, cette approximation a été trouvée très proche de l'évaluation exacte assistée par marqueurs.

sélection assistée par marqueurs / évaluation génétique / loci à caractères quantitatifs / marqueur à ADN

1. INTRODUCTION

Fernando and Grossman [3] presented a method to calculate the best linear unbiased predicted-estimates of breeding values (BLUP-EBV) using the information that DNA markers are linked to a quantitative trait locus (QTL). Goddard [4] extended the method to the use of flanking marker information. Although, these methods are relatively easy to use, the number of equations rapidly becomes large when there are many animals. Even with only one marked QTL, there are three equations per animal: two estimating both gametic effects at the QTL and one for the polygenic effect (the joint effect of the background genes). Every extra marked QTL increases the number of equations per animal by two. Moreover, when the flanking markers are close to the QTL, the probabilities of double cross-overs become small and the equations close to singular, and thus difficult to solve [13]. Meuwissen and Goddard [8] avoided these singularity problems by assuming a negligible probability of double recombinations within the flanking markers.

As genetic markers become more frequently used in commercial breeding programmes, the situation will commonly arise where only a small fraction of the animals have been genotyped. The phenotypes of non-genotyped animals may, however, be vital to the calculation of the effects of marked QTL as, for instance, in a granddaughter design where only bulls are genotyped but only cows are phenotyped. Calculation of two QTL effects for each marker for many non-genotyped animals is wasteful and may inhibit the implementation of marker assisted selection. Hoeschele [7] greatly reduced the number of equations in very general population structures, but this method is complicated and therefore difficult to apply in practice, mainly because it eliminates as many equations as possible. A more simple breeding structure such as a genotyped nucleus and non-genotyped commercial population structure can greatly simplify the elimination of equations. In some situations the organisation controlling the nucleus breeding programme may not have access to the records on commercial animals but may still need to include this information in the calculation of marker assisted EBVs (MA-EBVs) on nucleus animals.

The aim of this paper is to present a method that reduces the number of marker assisted breeding value estimation equations in a population where the nucleus animals are marker genotyped and the commercial animals are not genotyped. The reduction mainly eliminates the equations of non-genotyped animals. Furthermore, an approximate method of calculating MA-EBVs on nucleus animals is presented, which uses only the conventional non-MA-EBVs of nucleus animals from a national genetic evaluation to represent the data from commercial animals.

2. METHODS

2.1. Reducing the number of equations

The population was split into nucleus and commercial animals. Here, the definition of a commercial animal is: an animal that is not marker genotyped and has no descendants that are genotyped. The nucleus animals are all marker genotyped animals plus their ancestors. The method will still work if a commercial animal is erroneously considered as a nucleus animal, although the number of equations will not be reduced for such an animal. The method will fail, however, if a nucleus animal is erroneously considered as a commercial animal. For simplicity we ignored fixed effect equations, but including them is straightforward. Similarly, we assumed here only one marked QTL, since the inclusion of more marked QTL is straightforward. Partitioning the population into nucleus and commercial animals, the model can be written as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_3 & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \quad (1)$$

where $\mathbf{y}_1$ ($\mathbf{y}_2$) is the vector of phenotypic records of nucleus (commercial) animals; $\mathbf{a}_1$ ($\mathbf{a}_2$) is the vector of polygenic effects of nucleus (commercial) animals; $\mathbf{q}_1$ is the vector of marked QTL effects of the nucleus animals; $\mathbf{q}_2$ ($\mathbf{q}_3$) is the vector of paternally (maternally) derived QTL effects of the commercial animals; $\mathbf{e}_1$ ($\mathbf{e}_2$) is the vector of environmental effects of nucleus (commercial) animals; $\mathbf{Z}_1$ is the incidence matrix of polygenic effects of nucleus animals; $\mathbf{Z}_2$ is the incidence matrix of QTL effects of the nucleus animals; and $\mathbf{Z}_3$ is the incidence matrix of polygenic effects of the commercial animals. Note that $\mathbf{Z}_3$ is also used as the incidence matrix of the paternally and of the maternally derived QTL effects of the commercial animals, because these effects have the same incidence matrix as the polygenic effects of the commercial animals. The $\mathbf{Z}_2$ matrix can differ substantially from $\mathbf{Z}_1$ when the inheritance of QTL effects is traced from parent to offspring by the markers [8]. In order to solve the BLUP equations, we need the inverses of the (co)variance matrix of $[\mathbf{a}_1' \ \mathbf{a}_2']$ and of $[\mathbf{q}_1' \ \mathbf{q}_2' \ \mathbf{q}_3']$, which are obtained using the methods of Quaas [10, 11] and Fernando and Grossman [3], respectively.

In order to reduce the number of equations of the commercial animals, the 'reduced animal model' approach of Quaas and Pollak [12] was adopted. This approach was also used by Cantet and Smith [2] and Bink et al. [1] to absorb

the equations of non-parents in a model with QTL and polygenic effects. We re-write equation (1) as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{q}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \tag{2}$$

where $\mathbf{u}_2 = \mathbf{a}_2 + \mathbf{q}_2 + \mathbf{q}_3$. For the mixed model equations that follow from equations (2), we need the inverse of the (co)variance matrix of $[\mathbf{a}_1' \ \mathbf{q}_1' \ \mathbf{u}_2']$. Following Quaas [10, 11], we will assume that the animals within the nucleus and within the commercial are sorted from old to young. Next, we write every element of $[\mathbf{a}_1' \ \mathbf{q}_1' \ \mathbf{u}_2']$ in terms of its 'parental' elements plus an independent deviation from the 'parental' elements, where 'parental' elements denote the $\mathbf{a}_1$, $\mathbf{q}_1$ or $\mathbf{u}_2$ elements of the parents of the current animal:

$$\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{q}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} \\ \mathbf{R} & \mathbf{S} & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{q}_1 \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \tag{3}$$

where $\mathbf{P}$ is an indicator matrix of the parents of $\mathbf{a}_1$, such that $P_{ij} = 0.5$ if animal j is a parent of animal i, and otherwise $P_{ij} = 0$; $Q_{ij} = \theta_{ij}$ if QTL_i is with probability θ_{ij} a direct copy of QTL_j, where QTL_j was one of the two 'parental' QTL alleles of QTL_i, with 'parental' denoting that QTL_j was involved in the Mendelian sampling process that resulted in QTL_i, and for all other i and j: $Q_{ij} = 0$; $R_{ij} = 0.5$ if nucleus animal j is a parent of commercial animal i, and otherwise $R_{ij} = 0$; $S_{ij} = 0.5$ if one of the two QTL of commercial animal i is a direct copy of the nucleus gamete j with a probability of 0.5 (the probability is always 0.5 because commercial animals are not marker genotyped), otherwise $S_{ij} = 0$; $T_{ij} = 0.5$ when commercial animal j is a parent of i, otherwise $T_{ij} = 0$.

The elements of ε_1, ε_2 and ε_3 are all independent, unless the markers are not completely informative, i.e. it is not always possible to trace which marker is inherited from the sire and which from the dam. In the latter case, the elements of ε_2 may be correlated and the method of Wang et al. [14] can be used to set up (the inverse of) the (co)variance matrix of the QTL effects of the nucleus animals. The calculation of the (co)variance matrix of the QTL effects of the nucleus animals becomes even more complex when ancestors of nucleus animals have missing marker genotypes; however, for this situation, Wang et al. provide an approximate method to set up the (co)variance matrix of QTL effects. We will ignore these complications of obtaining the inverse of the (co)variance matrix of the QTL effects of the nucleus animals here, because the method that is used to obtain the inverse of this (co)variance matrix does not affect the setting up of the inverse of the (co)variance matrix of the $\mathbf{u}_2$ equations. This is because the situation of uninformative marker information and ungenotyped ancestors of genotyped animals did not occur within the group of commercial animals, since none of the commercial animals were genotyped.

Let the variance of the polygenic effects be denoted by σ_a^2 and the variance of the QTL effect of one gamete be denoted by σ_q^2, then their variances are:

$$\text{Var}(\varepsilon_1) = \mathbf{D}_1, \ \text{Var}(\varepsilon_2) = \mathbf{D}_2, \ \text{ and } \ \text{Var}(\varepsilon_3) = \mathbf{D}_3$$

where $\mathbf{D}_1$ is a diagonal matrix with $\mathbf{D}_{1ii}$ equal to σ_a^2, $0.75\sigma_a^2$ or $0.5\sigma_a^2$ when no, one or both parents are known of nucleus animal i, respectively; $\mathbf{D}_2$ is a diagonal with $\mathbf{D}_{2ii}$ equal to σ_q^2 or $2\theta_{ij}(1-\theta_{ij})\sigma_q^2$ when gamete i is a founder gamete or is derived from gamete j with probability θ_{ij} [3], respectively; and $\mathbf{D}_3$ is a diagonal with $\mathbf{D}_{3ii}$ equal to σ_u^2, $0.75\sigma_u^2$ or $0.5\sigma_u^2$, when no, one or both parents of commercial animal i are known, respectively, where $\sigma_u^2 = \sigma_a^2 + 2\sigma_q^2$. Next we solve equation (3) for $\mathbf{v}' = [\mathbf{a}_1' \ \mathbf{q}_1' \ \mathbf{u}_2']$ to obtain:

$$\mathbf{v} = \mathbf{V}^{-1}\,\boldsymbol{\varepsilon}$$

where

$$\mathbf{V} = \begin{bmatrix} \mathbf{I}-\mathbf{P} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}-\mathbf{Q} & \mathbf{0} \\ -\mathbf{R} & -\mathbf{S} & \mathbf{I}-\mathbf{T} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

Taking variances on both sides yields,

$$\mathrm{Var}(\mathbf{v}) = \mathbf{V}^{-1}\,\mathbf{D}\,\mathbf{V}'^{-1} \tag{4}$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_3 \end{bmatrix}$$

Finally the inverse of $\mathrm{Var}(\mathbf{v})$ is $\mathbf{G}^{-1}$ which is obtained as:

$$\mathbf{G}^{-1} = \mathbf{V}'\,\mathbf{D}^{-1}\,\mathbf{V} \tag{5}$$

Similar to Quaas [10, 11], the following rules can be found to set up $\mathbf{G}^{-1}$.

1) For the polygenic effects of the nucleus animals part of $\mathbf{G}^{-1}$: follow Quaas' rules (multiply by $1/\sigma_a^2$ to account for the different variances in different parts of $\mathbf{G}^{-1}$).

2) For the QTL effects of the nucleus animals part of $\mathbf{G}^{-1}$: follow the rules of Fernando and Grossman [3] (multiply by $1/\sigma_q^2$).

3) For the genetic effects, $\mathbf{u}_2$, of commercial animal i:

– if both parents are unknown: add $1/\sigma_u^2$ to position (i,i);

– if one parent s is known with QTL alleles a_1 and a_2 add to the indicated positions:

$$\begin{array}{c} \\ a_1 \\ a_2 \\ s \\ i \end{array} \begin{array}{c} \begin{array}{cccc} a_1 & a_2 & s & i \end{array} \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 & -2/3 \\ 1/3 & 1/3 & 1/3 & -2/3 \\ 1/3 & 1/3 & 1/3 & -2/3 \\ -2/3 & -2/3 & -2/3 & 4/3 \end{pmatrix} \end{array} \Big/ \sigma_u^2 \tag{6}$$

If there are no equations for the QTL alleles a_1 and a_2, i.e. s was a commercial animal, the additions to their positions are cancelled, and the additions simplify to the original rule of Quaas [10, 11];

– if both parent s and d of animal i are known with QTL alleles a_1 and a_2 of s and alleles a_3 and a_4 of d, add to the indicated positions:

$$\begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ s \\ d \\ i \end{array} \begin{array}{c} \begin{array}{ccccccc} a_1 & a_2 & a_3 & a_4 & s & d & i \end{array} \\ \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & -1 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & -1 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & -1 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & -1 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & -1 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & 1/2 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 2 \end{pmatrix} \end{array} \Big/ \sigma_u^2 \tag{7}$$

If there are no equations for the QTL alleles a_1, a_2, a_3 and/or a_4 the additions to their positions are cancelled. When all alleles a_1, a_2, a_3 and/or a_4 have no equations, the additions simplify to the original rule of Quaas [10, 11].

As can be seen from the above additions, the commercial animals add the same values as in Quaas' rules to the elements of their parents, but if the parents are nucleus animals these values are added to their polygenic and QTL effects.

After setting up the $\mathbf{G}^{-1}$ matrix, we can set up and solve Henderson's [6] mixed model equations:

$$(\mathbf{W}'\mathbf{W} + \mathbf{G}^{-1}\,\sigma_e^2)\mathbf{v} = \mathbf{W}'\mathbf{y} \tag{8}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_3 \end{bmatrix}$$

and σ_e^2 is the environmental variance.

These equations will yield exact solutions of the estimates of polygenic ($\mathbf{a}_1$) and QTL effects ($\mathbf{q}_1$) of the nucleus animals, and of the sum of the polygenic and QTL effects of the commercial animals ($\mathbf{u}_2$) (unless approximations have to be applied for setting up the (co)variance matrix of the QTL effects of the nucleus animals owing to missing marker genotypes of ancestors of nucleus animals). A small example of the calculation of the $\mathbf{G}^{-1}$ matrix is given in Appendix A.

2.2. The use of conventional EBV to predict MA-EBV

In the case of cattle breeding schemes especially, the commercial animals may not be owned by the breeding organisation and this organisation may not have access to the phenotypic information of the commercial animals. However, BLUP breeding value estimates and their accuracies may be available from a

national breeding value evaluation. We would like to use this information to improve the accuracy of the marker assisted breeding value estimates in the nucleus. This problem is similar to that of incorporating AI sire evaluations into intraherd breeding value predictions by Henderson [5], and our approach will therefore also be similar to that of Henderson.

The first step is to absorb the commercial animal equations into the nucleus equations, which will reveal which information from the commercial animals is needed. The full mixed model equations are [writing out equations (8) and (5)] see (8bis) in the following page.

Absorption of the commercial animal equations ($\mathbf{u}_2$) yields equation (9), shown in the following page, where $\mathbf{B} = \mathbf{D}_3^{-1} - \mathbf{D}_3^{-1}(\mathbf{I} - \mathbf{T})[\mathbf{Z}_3'\mathbf{Z}_3 + (\mathbf{I} - \mathbf{T}')\mathbf{D}_3^{-1}(\mathbf{I} - \mathbf{T})]^{-1}(\mathbf{I} - \mathbf{T}')\mathbf{D}_3^{-1}$, and $\mathbf{b} = \mathbf{D}_3^{-1}(\mathbf{I} - \mathbf{T})[\mathbf{Z}_3'\mathbf{Z}_3 + (\mathbf{I} - \mathbf{T}')\mathbf{D}_3^{-1}(\mathbf{I} - \mathbf{T})]^{-1}\mathbf{Z}_3'\mathbf{y}_2$. Note that equation (9) reduces to the MA-EBV equations of the nucleus animals without accounting for any information of commercial animals, if $\mathbf{B}$ and $\mathbf{b}$ are set to zero. The term $\mathbf{R}'\mathbf{B}\mathbf{R}$ leads to additions to the equations of the nucleus parents of the commercial animals. Similarly, $\mathbf{S}'\mathbf{B}\mathbf{S}$ leads to additions to the equations of the QTL that are carried by the nucleus parents of the commercial animals. Further, $\mathbf{R}'\mathbf{B}\mathbf{S}$ leads to additions to the animal $*$ QTL block of the equation (9) of the nucleus parents (of commercial animals) and their QTL effects. The terms $\mathbf{R}'\mathbf{b}$ and $\mathbf{S}'\mathbf{b}$ result in additions to the right hand side of the equations pertaining to the parents of nucleus animals and their QTL effects, respectively. We will approximate these terms $\mathbf{R}'\mathbf{B}\mathbf{R}$, $\mathbf{S}'\mathbf{B}\mathbf{S}$, $\mathbf{R}'\mathbf{B}\mathbf{S}$, $\mathbf{R}'\mathbf{b}$ and $\mathbf{S}'\mathbf{b}$ using the results from a conventional national evaluation of breeding values.

The solutions of EBV of nucleus animals of the conventional national evaluation should equal the solutions from the equations of the nucleus animals after absorption of the commercial animals. The conventional equations for nucleus animals after absorption of commercial animals are:

$$(\mathbf{M} + \mathbf{R}'\mathbf{B}\mathbf{R})\ \mathbf{EBV} = \mathbf{Z}_1'\mathbf{y}_1 + \mathbf{R}'\mathbf{b} \tag{10}$$

where $\mathbf{EBV}$ is a vector of conventional EBV of nucleus animals (known from national evaluation), $\mathbf{M} = [\mathbf{Z}_1'\mathbf{Z}_1 + (\mathbf{I} - \mathbf{P})'\mathbf{D}_1^{-1}(\mathbf{I} - \mathbf{P})\sigma_e^2\ \sigma_a^2/\sigma_u^2]$, which is the coefficient matrix of the conventional mixed model equations when only information from nucleus animals is used (note that $(\mathbf{I} - \mathbf{P})'\mathbf{D}_1^{-1}(\mathbf{I} - \mathbf{P})/\sigma_a^2$ equals the inverse of the relationship matrix of the nucleus animals). Note also that the additions $\mathbf{R}'\mathbf{B}\mathbf{R}$ and $\mathbf{R}'\mathbf{b}$ are the same as those in the MA-EBV equation (9). Hence, if we obtain approximations for $\mathbf{R}'\mathbf{B}\mathbf{R}$ and $\mathbf{R}'\mathbf{b}$ in equation (10) we can approximate equation (9). We know the EBV and their accuracies, r_i, which result from equation (10). Let the matrix $\mathbf{C} = (\mathbf{M} + \mathbf{R}'\mathbf{B}\mathbf{R})^{-1}$, then the diagonal elements of $\mathbf{C}$ are:

$$C_{ii} = (1 - r_i^2)/\lambda$$

where $\lambda = \sigma_e^2/\sigma_u^2$. Now it is assumed that $\mathbf{R}'\mathbf{B}\mathbf{R}$ can be approximated by a diagonal matrix Δ, i.e. we find a diagonal matrix Δ such that:

$$(\mathbf{M} + \Delta)^{-1} \approx \mathbf{C}$$

$$\begin{bmatrix} \mathbf{Z}'\mathbf{Z} + \left[(\mathbf{I}-\mathbf{P}')\mathbf{D}_1^{-1}(\mathbf{I}-\mathbf{P}) + \mathbf{R}'\mathbf{D}_3^{-1}\mathbf{R}\right]\sigma_e^2 & \mathbf{Z}_1'\mathbf{Z}_2 + \mathbf{R}'\mathbf{D}_3^{-1}\mathbf{S}\sigma_e^2 & \mathbf{R}'\mathbf{D}_3^{-1}(\mathbf{I}-\mathbf{T})\sigma_e^2 \\ \mathbf{Z}_2'\mathbf{Z}_1 + \mathbf{S}'\mathbf{D}_3^{-1}\mathbf{R}\sigma_e^2 & \mathbf{Z}_2'\mathbf{Z}_2 + \left[(\mathbf{I}-\mathbf{Q}')\mathbf{D}_2^{-1}(\mathbf{I}-\mathbf{Q}) + \mathbf{S}'\mathbf{D}_3^{-1}\mathbf{S}\right]\sigma_e^2 & \mathbf{S}'\mathbf{D}_3^{-1}(\mathbf{I}-\mathbf{T})\sigma_e^2 \\ (\mathbf{I}-\mathbf{T}')\mathbf{D}_3^{-1}\mathbf{R}\sigma_e^2 & (\mathbf{I}-\mathbf{T}')\mathbf{D}_3^{-1}\mathbf{S}\sigma_e^2 & \mathbf{Z}_3'\mathbf{Z}_3 + (\mathbf{I}-\mathbf{T}')\mathbf{D}_3^{-1}(\mathbf{I}-\mathbf{T})\sigma_e^2 \end{bmatrix} * \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{q}_1 \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1'\mathbf{y}_1 \\ \mathbf{Z}_2'\mathbf{y}_1 \\ \mathbf{Z}_3'\mathbf{y}_2 \end{bmatrix} \tag{8bis}$$

$$\begin{bmatrix} \mathbf{Z}_1'\mathbf{Z}_1 + \left[(\mathbf{I}-\mathbf{P}')\mathbf{D}_1^{-1}(\mathbf{I}-\mathbf{P}) + \mathbf{R}'\mathbf{B}\mathbf{R}\right]\sigma_e^2 & \mathbf{Z}_1'\mathbf{Z}_2 + \mathbf{R}'\mathbf{B}\mathbf{S}\sigma_e^2 \\ \mathbf{Z}_2'\mathbf{Z}_1 + \mathbf{S}'\mathbf{B}\mathbf{R}\sigma_e^2 & \mathbf{Z}_2'\mathbf{Z}_2 + \left[(\mathbf{I}-\mathbf{Q}')\mathbf{D}_2^{-1}(\mathbf{I}-\mathbf{Q}) + \mathbf{S}'\mathbf{B}\mathbf{S}\right]\sigma_e^2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1'\mathbf{y}_1 + \mathbf{R}'\mathbf{b} \\ \mathbf{Z}_2'\mathbf{y}_1 + \mathbf{S}'\mathbf{b} \end{bmatrix} \tag{9}$$

where only the diagonal elements of $\mathbf{C}$ are known. The diagonal elements of Δ, Δ_{ii}, yield the effective number of records that should be added to a nucleus animal i, such that the accuracy of its EBV is equal to the accuracy when the commercial animals were included. A similar effective number of records was derived by Henderson [5], but in his situation the animals within the herd did not contribute significantly to the EBV of the sire. Here, we used the following iteration scheme to disentangle the information that came from the nucleus animals, which is represented by the matrix $\mathbf{M}$, and the information that comes from the commercial animals, which is represented by the matrix Δ.

Newton's iteration algorithm was used to calculate the diagonal matrix Δ such that $\text{diag}((\mathbf{M}+\Delta)^{-1}) = \text{diag}(\mathbf{C})$, where $\text{diag}(\mathbf{X})$ denotes a vector containing the diagonal elements of the matrix $\mathbf{X}$. Let the vector $\delta = \text{diag}(\Delta)$. The iteration scheme estimates δ by:

step 1: a first approximation $\Delta_{[0]}$ or, equivalently, $\delta_{[0]}$ is obtained from:

$$\delta_{[0]} = \text{diag}(\mathbf{C}^{-1}) - \text{diag}(\mathbf{M})$$

step 2: improve δ by Newton-Raphson iteration:

$$\delta_{[p+1]} = \delta_{[p]} - \mathbf{H}_{[p]}^{-1}[\text{diag}((\mathbf{M}+\Delta_{[p]})^{-1}) - \text{diag}(\mathbf{C})]$$

where $[p]$ denotes the pth iteration; and $\mathbf{H}$ is a matrix of derivatives of $\text{diag}((\mathbf{M}+\mathbf{D})^{-1})$ with respect to δ, which can be shown to equal $-(\mathbf{M}+\Delta)^{-1} * (\mathbf{M}+\Delta)^{-1}$, where $*$ denotes element by element multiplication.

Given the approximated mixed model coefficient matrix of the nucleus animals after absorption of the commercial animals, $\mathbf{M}+\Delta$, an approximation of the right hand side of equation (10), is obtained from:

$$(\mathbf{M}+\Delta)\mathbf{EBV} = \mathbf{Z}_1'\mathbf{y}_1 + \Delta\mathbf{RHS}$$

where $\Delta\mathbf{RHS}$ is an approximation of the term $\mathbf{R}'\mathbf{b}$ in equation (10). Since, $\mathbf{EBV}$ and $\mathbf{Z}_1'\mathbf{y}_1$ are known, $\Delta\mathbf{RHS}$ can be calculated from the above equation.

Next we will calculate the absorbed coefficient matrix of the marker assisted mixed model equation (9), and their right hand side. From the previous section we concluded that we could approximate $\mathbf{R}_i'\mathbf{B}\mathbf{R}_i$ by Δ_{ii}, where $\mathbf{R}_i$ is the ith column of $\mathbf{R}$. The vector $\mathbf{R}_i$ indicates which commercial animal is an offspring of nucleus animal i by containing a $1/2$ if the commercial animal is an offspring of i or a 0 otherwise. If a_1 is one of the QTL alleles of nucleus animal i, the a_1th column of $\mathbf{S}$, $\mathbf{S}_{a1}$, contains a $1/2$ if the commercial animal is an offspring of animal i. If every nucleus animal has two unique QTL alleles, as in the model of Fernando and Grossman [3], it follows that $\mathbf{R}_i = \mathbf{S}_{a1} = \mathbf{S}_{a2}$, with a_1 and a_2 denoting the QTL alleles of animal i. Hence:

$$\mathbf{R}_i'\mathbf{B}\mathbf{R}_i = \mathbf{R}_i'\mathbf{B}\mathbf{S}_{ax} = \mathbf{S}_{ax}'\mathbf{B}\mathbf{S}_{ax} \approx \Delta_{ii}$$

and, similarly,

$$\mathbf{R}_i'\mathbf{b} = \mathbf{S}_{ax}'\mathbf{b} \approx \Delta\mathbf{RHS}_i$$

where a_x denotes a_1 or a_2. Thus, the addition Δ_{ii} to the diagonal of the polygenic equation of the nucleus animal i should also be added to the off-diagonal of the polygenic equation i and QTL allele equation a_1 and a_2; to the diagonal of both QTL equations a_1 and a_2; and to the off-diagonal elements of a_1 and a_2. And the term $\Delta\mathbf{RHS}_i$ should be added to the right hand side of the equation of animal i, and of the QTL equations a_1 and a_2. In conclusion, to account for the information of commercial animals, for every nucleus animal i we add to the coefficient matrix of the MA equations of the nucleus animals that ignores information of commercial animals:

$$\begin{array}{c} \\ a_1 \\ a_2 \\ i \end{array} \begin{array}{c} \begin{array}{ccc} a_1 & a_2 & i \end{array} \\ \begin{pmatrix} \Delta_{ii} & \Delta_{ii} & \Delta_{ii} \\ \Delta_{ii} & \Delta_{ii} & \Delta_{ii} \\ \Delta_{ii} & \Delta_{ii} & \Delta_{ii} \end{pmatrix} \end{array} \tag{11}$$

where a_1 and a_2 denote the equations for the QTL effects of animal i; and we add to the right hand side of these nucleus equations for every nucleus animal i:

$$\begin{array}{c} a_1 \\ a_2 \\ i \end{array} \begin{pmatrix} \Delta\mathbf{RHS}_i \\ \Delta\mathbf{RHS}_i \\ \Delta\mathbf{RHS}_i \end{pmatrix} \tag{12}$$

Thus, the additions (11) and (12) result in an approximation of the marker assisted nucleus equations (9) using only the EBV and accuracies to account for the information of commercial animals.

The equality of $\mathbf{R}_i$ to $\mathbf{S}_{ax}$ requires that the QTL allele a_x is only present in one animal i. However, in the model of Meuwissen and Goddard [8], QTL alleles might be traced from parent to offspring with certainty, because flanking markers were used and double recombinations were ignored. In this model different animals may carry the same QTL allele a_x, and $\mathbf{S}_{ax} = \Sigma_{i \in Ax} \mathbf{R}_i$, where the summation is over all animals i that carry QTL allele a_x. This complication of $\mathbf{S}_{ax}$ being the sum of several $\mathbf{R}_i$ terms does not affect the additions in equations (11) and (12) which are due to terms that are linear in $\mathbf{S}_{ax}$, because the correct additions are still performed as all the animals contributing to $\mathbf{S}_{ax}$ are evaluated. However, the additions to the QTL allele * QTL allele block of equation (11), are due to second order terms of $\mathbf{S}_{ax}$, which implies that more off diagonal terms of the absorption matrix $\mathbf{B}$ have to be added. We will ignore these extra off diagonal terms of $\mathbf{B}$, which are due to the second order terms of $\mathbf{S}_{ax}$, and perform the additions as described in equation (12), which adds another level of approximation to this method.

In the above, the fixed effect structure of the nucleus animal data was ignored, but can be accounted for by absorbing the fixed effect equations into the equations of the nucleus animals, i.e. the matrix $\mathbf{M}$ would be the conventional mixed model coefficient matrix after absorption of fixed effects. Alternatively, if absorption of fixed effects is computationally too demanding, the following steps can be undertaken to account for fixed effects:

step 1: approximate Δ_{ii} as in the forementioned Newton algorithm, except that

$(\mathbf{M} + \Delta_{[p]})^{-1}$ is replaced by the animal * animal block of:

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_1 \\ \mathbf{Z}_1'\mathbf{X} & \mathbf{M} + \Delta_{[p]} \end{pmatrix}^{-1}$$

where $\mathbf{X}$ is the design matrix of the fixed effect structure of the nucleus data;

step 2: if the fixed effect solutions are not available from the national breeding value evaluation, solve for the fixed effect solutions, β, using:

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}_1 - \mathbf{X}'\mathbf{Z}_1\,\mathbf{EBV}$$

step 3: calculate $\Delta\mathbf{RHS}$ using:

$$\mathbf{Z}_1'\mathbf{X}\beta + (\mathbf{M} + \Delta)\mathbf{EBV} = \mathbf{Z}_1'\mathbf{y}_1 + \Delta\mathbf{RHS}$$

The above methods that account for fixed effects assume that different fixed effects are estimated in the nucleus than in the commercial animals, which will be the case in most situations. A brief example of the use of non-MA-EBV in the estimation of MA-EBV of nucleus animals is given in Appendix A.

2.3. Simulation

A data set was simulated to test whether the reduced number of marker assisted mixed model equations indeed yielded the same solutions as the original full set of equations, to compare the number of iterations needed to solve the reduced set of equations and the original equations (which might be more diagonally dominant), and to test the approximate absorption of all equations for commercial animals. The data set resulted from five generations of simulation of a nucleus and a female commercial population, where the nucleus animals are selected on conventional BLUP-EBV, and the unselected commercial females are mated to the selected sires of the nucleus. The parameters of the simulated data set are presented in *table I.* MA-EBVs were calculated in a manner similar to that of Meuwissen and Goddard [8] in which it is assumed that if markers cannot trace the inheritance of QTL alleles from parent to offspring, then the QTL allele inherited is treated as equally likely to be either of the two alleles in the parent (the possibility that a segregation analysis of the marker data might improve this prediction, was ignored). The probability that the markers could not trace the inheritance of the QTL was assumed to be 0.1, which occurred in 158 instances in the nucleus. In these instances a new QTL effect was postulated and estimated. Including the 400 founder QTL effects (= 2 * 200 founder nucleus animals), the number of QTL effects of nucleus animals was 558. The commercial animals were not marker genotyped and so no QTL effects could be traced. If no equations were eliminated, this would result in 10 000 (= 2 * 5 000 commercial animals) commercial QTL effects.

The equations were solved by Gauss-Seidel iteration. The convergence criterion was:

$$SS/SST < 10^{-10}$$

Table I. Parameters of the simulated data set.

No. of discrete generations		5
Total no. of nucleus animals		1 000
Structure of nucleus generation:	no. of sires	5
	no. of dams per sire	5
	no. of offspring per dam	8
	total	200
Total no. of commercial females		5 000
Structure of commercial generation:	no. of sires from nucleus	5
	no. of dams from commercial	1 000[1]
	total	1 000
Total no of animals (nucleus plus commercial)		6 000
No. of marked QTL		1
Variances: genetic variance at QTL		0.1
polygenic		0.25
environmental		1
Probability that flanking markers could not trace QTL		0.1
Position of QTL in marker bracket:		half way
Selection of sires and dams on:		conventional BLUP–EBV

[1] All commercial dams produced one offspring with a randomly sampled nucleus sire to replace the commercial generation.

Table II. Results of the calculations of the MA-EBV.

Method	No. of equations	No. of iterations[1]
No reduction in no. of equations	16 558	416
One equation per commercial animal	6 558	369

[1] Gauss-Seidel iterations were performed with a convergence criterion of 10^{-10}, as indicated in the text. MA-EBV = marker assisted estimates of breeding value.

where SS is the sum of squares of the deviations of the left hand side from the right hand side of the equations; SST is the sum of squares of the solutions. The number of iterations needed to reach this convergence criterion is a (imperfect) measure of how easily the equations could be solved. This measure is not perfect because the solution vectors of both methods are not the same, and SS may be small while the solution vector is still far from the exact solution.

3. RESULTS

Table II shows the results of the EBV calculations. Without any reduction in the number of equations, the total number of equations is: 16 558 (6 000 animal and 10 558 QTL effects). When the QTL equations of the commercial animals were eliminated, the number of equations was reduced by 10 000. In practice, this figure will often be much larger, because the commercial population will be much larger than in the simulation. Furthermore, the number of iterations that was needed to reach the convergence criterion, was smaller with the reduced set of equations. This suggests that the reduced set of equations was not any harder to solve than the original large set of equations. The solutions to both sets of equations were virtually identical (result not shown), except that the reduced set did not yield estimates of individual QTL effects of commercial animals. If the estimates of the QTL effects of the commercial animals were required, they could be obtained by back solving (see Appendix B). *Table III* shows the results of the approximate method using conventional EBV of nucleus animals to estimate the MA-EBV of the nucleus animals. When we want to predict the total breeding value of the animals, u_i, the use of conventional EBV of nucleus animals, instead of the phenotypes of commercial animals, leads to very accurate predictions of MA-EBV in the simulated data set. The predictions of the individual QTL effects, q_i, were also accurate, i.e. the correlation and regression is close to 1. In this approximate method the number of equations was only 1 558.

Table III. Comparison of approximate MA-EBV ($\widehat{u}_a$) and estimates of QTL effects ($\widehat{q}_a$), which were obtained by using conventional EBV of commercial animals and their accuracies, to the original MA-EBV ($\widehat{u}$) and QTL effect estimates ($\widehat{q}$), which were obtained by using the original records of the commercial animals.

Coefficient	Value
Regression of $\widehat{u}_a$ on $\widehat{u}$	0.9991
Correlation of $\widehat{u}_a$ and $\widehat{u}$	0.9999
Regression of $\widehat{q}_a$ on $\widehat{q}$	1.0029
Correlation of $\widehat{q}_a$ and $\widehat{q}$	1.0000

(MA-EBV = marker assisted estimates of breeding value)

4. DISCUSSION

4.1. Reduction in the number of equations

A simple method was presented that reduced the number of equations of non-marker genotyped commercial animals from three to one, where the latter equation estimates the total breeding value (polygenic plus QTL) of the commercial animals. The reduction in the number of equations was large when the number of commercial animals was large, and the reduced set of equations was not any harder to solve (*table II*).

An alternative to this method is the use of a reduced animal model [1, 2], which would eliminate equations for commercial animals only if they are not parents. Thus the method presented eliminates more equations than the reduced animal model approach but less than Hoeschele [7]. The importance of using the data on commercial animals when calculating MA-EBVs for the nucleus animals is illustrated by the case of a granddaughter design where all phenotypic data come from the commercial granddaughters. The method proposed could be used in a national genetic evaluation where the number of additional equations would be proportional to the size of the nucleus.

Extension of the method to more marked QTL is straightforward. Let the parental QTL alleles of the first marked QTL be denoted by a_1, a_2 (a_3, a_4), and those of the second marked QTL by b_1, b_2 (b_3, b_4), where the elements between the brackets are needed when a second nucleus parent is known. Now extra rows and columns for b_1, b_2 (b_3, b_4) are augmented to the additions in equations (6) and (7), where the values in the augmented rows and columns are the same as those in the rows and columns of a_1, a_2 (a_3, a_4). Also, the off diagonal elements between aj and bk are the same as those between a_1 and a_2 for all j and k. Hence, with n marked QTL, the number of equations of commercial animals reduces from $2n + 1$ to just one. In many marker assisted selection schemes, there may also be many nucleus animals that are not genotyped, namely the old ancestors of the nucleus that were born before marker genotyping started. Some cryo-conserved semen of old sires may still be available for genotyping, but many old ancestors will remain non-marker genotyped. In situations, where the old ancestors result in a computationally unmanageable number of equations, the approach of Hoeschele [7] eliminates all QTL equations of non-genotyped ancestors that are not on a genetic pathway between two marker genotyped animals. Although more difficult to apply, this method can result in a substantial reduction in the number of QTL equations of non-genotyped ancestors. In the present simulated data set, all nucleus animals were genotyped and a further reduction in the number of equations was not obtained by using the approach of Hoeschele [7].

The rules presented for setting up the $\mathbf{G}^{-1}$ matrix, did not account for the inbreeding of the animals. If inbreeding is not negligible, Quaas' [10, 11] rules can be used to account for inbreeding in the nucleus animal $*$ nucleus animal part of the $\mathbf{G}^{-1}$ matrix, which results in reducing the elements of the $\mathbf{D}_1$ matrix by a fraction equal to the average inbreeding coefficient of the parents. Also, Wang's [14] method accounts for inbreeding, where the inbreeding coefficient is calculated at the QTL given the marker information. In the equations of the commercial animals, the average inbreeding coefficients can be accounted for by reducing the σ_u^2 term in additions (6) and (7) by a fraction equal to the average inbreeding coefficients of the parents. The latter will be slightly biased because the average inbreeding coefficient of the parents will be different at the QTL. This bias can be corrected by using a weighted average inbreeding coefficient, where the conventional inbreeding coefficient averaged over the parents, the inbreeding at the QTL of the sire, and that at the QTL of the dam, are weighed in proportion to their variances, i.e. σ_a^2, σ_q^2 and σ_q^2, respectively.

4.2. Use of conventional EBV to predict MA-EBV

A method was developed that uses the information of conventional EBV of nucleus animals and their accuracies instead of the data on commercial animals to increase the accuracy of MA-EBV of nucleus animals. In the simulated data set, the approximate MA-EBV, which used conventional EBVs, were very close to the original MA-EBV based on the full set of equations which used the phenotypic records from the commercial animals. The method was also tested in a granddaughter design [15], where a genotyped grand sire has genotyped sons which have conventional EBVs based on daughter records. In this situation, the prediction of the QTL effects of the sons and the grand sires was identical to when the original records of the daughters had been used (result not shown). This method could be used by a breeding organisation controlling the nucleus breeding programme without including any information on commercial animals. The method can be compared to the use of daughter yield deviations to represent data on the (commercial) daughters of a nucleus bull. However, this method uses all commercial descendents of a nuclear animal (via its conventional EBV) and avoids double counting of the information from descendents in the nucleus. The method could be extended for QTL detection studies based on REML estimation of the variance due to a QTL, bracketed by the markers.

The approximate method to incorporate EBVs from the commercial animals into the nucleus MA-EBV is similar to the use of foreign EBV in the national evaluation of a country. Except that the foreign EBV calculation does (almost) not use local information, such that the foreign EBV yield independent extra information. Hence, the accuracy of the foreign EBV can be directly converted into an effective number of records (or daughters) that is added to the diagonals of the coefficient matrix, and into deregressed proofs, i.e. extra records (or daughters) are invented that contain the information of the foreign EBV. Here, the situation was more complicated because the conventional EBV of the nucleus animal already contained the information of the commercial animals, which made it more difficult to determine the extra information.

The calculation of the MA-EBV using conventional EBV of commercial animals relies strongly on the accuracy of the conventional EBV. In the present simulation study, the accuracies were calculated by inversion of the conventional mixed model matrix, i.e. exact accuracies were used. In the case of a national evaluation of EBV, the number of equations is too large for direct inversion and the accuracies have been approximated. These approximations are often good (e.g. [9]). However, poor approximations of the accuracies of the conventional EBV will probably reduce the accuracies of the MA-EBV substantially. In any case, the method presented here seems to make as much use as possible of the conventional EBV of national evaluations to improve the accuracies of the MA-EBV of the nucleus animals.

REFERENCES

[1] Bink M.C.A.M., Quaas R.L., Van Arendonk J.A.M., Bayesian estimation of dispersion parameters with a reduced animal model including polygenic and QTL effects, Genet. Sel. Evol. 30 (1998) 103–125.

[2] Cantet R.J.C., Smith C., Reduced animal model for marker assisted selection using best linear unbiased prediction, Genet. Sel. Evol. 23 (1991) 221–233

[3] Fernando R.L., Grossman M., Marker-assisted selection using best linear unbiased prediction, Genet. Sel. Evol. 21 (1989) 467–477.

[4] Goddard M.E., A mixed model for analyses of data on multiple genetic markers, Theor. Appl. Genet. 83 (1992) 878–886.

[5] Henderson C.R., Use of AI relatives in intraherd prediction of breeding values and producing abilities, J. Dairy Sci. 58 (1975) 1910–1916.

[6] Henderson C.R., Applications of linear models in animal breeding, Can. Catal. Publ. Data, Univ. Guelph, Guelph, 1984.

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

[8] Meuwissen T.H.E., Goddard M.E., The use of marker haplotypes in animal breeding schemes, Genet. Sel. Evol. 28 (1996) 161–176.

[9] Misztal I., Wiggans G.R., Approximation of prediction error variances in large-scale animal models, J. Dairy Sci. 71 (Suppl. 2) (1988) 27–32.

[10] Quaas R.L., Computing the diagonal elements of a large numerator relationship matrix, Biometrics 32 (1976) 949–953.

[11] Quaas R.L., Additive genetic model with groups and relationships, J. Dairy Sci. 71 (1988) 1338–1345.

[12] Quaas R.L., Pollak E.J., Mixed model methodology for farm an ranch beef cattle testing programs, J. Anim. Sci. 51 (1980) 1277–1287.

[13] Ruane J., Colleau J.J., Marker-assisted selection for a sex-limited character in a nucleus breeding population, J. Dairy Sci. 79 (1996) 1666–1678.

[14] Wang T., Fernando R.L., Van der Beek S., Grossman M., Van Arendonk J.A.M., Covariance between relatives for a marked quantitative trait locus, Genet. Sel. Evol. 27 (1995) 251–274.

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

APPENDIX A

An example illustrating the calculation of the inverse of the genetic covariance matrix, G, for the reduced set of equations

The pedigree and marker genotypes of five animals are given in *table AI*. The marker genotype of animal 5 is unknown, and this animal has no descendants with known marker genotypes. Hence, given the definition in the main text, animal 5 is a commercial animal, and its QTL allele equations will be absorbed. Let the polygenic and QTL allelic variance be equal to one, i.e. $\sigma_a^2 = \sigma_q^2 = 1$, and $\sigma_u^2 = \sigma_a^2 + 2\sigma_q^2 = 3$. The first step is to set up the inverse of the polygenic–nucleus animal part of the genetic covariance matrix, $\mathbf{G}^{-1}$, where the nucleus animals are 1, 2, 3 and 4. This part of the $\mathbf{G}^{-1}$ matrix is obtained by using Quaas' rules [10, 11]:

$$\begin{bmatrix} 2 & 1 & -1 & -1 \\ 1 & 2 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

The second step is to set up the QTL effects part of the $\mathbf{G}^{-1}$ matrix. For simplicity, we will assume here that the recombination rate between the marker and the QTL is zero. (This situation is similar to the situation where flanking markers are used to trace the QTL alleles, and the double recombination rate is assumed negligible.) Because the recombination rate is zero, the inheritance of the QTL alleles is identical to that of the marker alleles, i.e. there are four QTL alleles q_1, q_2, q_3 and q_4, and the presence of marker allele A_x also indicates the presence of QTL allele q_x, where $x = 1$, 2, 3 or 4. The QTL allele part of the $\mathbf{G}^{-1}$ matrix is calculated using the rules of Fernando and Grossman [3] or Wang et al. [14], and yields an identity matrix here because the QTL alleles q_1, q_2, q_3 and q_4 are all founder alleles. After this step, the polygenic and QTL allele part of the nucleus animals of the $\mathbf{G}^{-1}$ matrix is:

$$\begin{bmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad [A1]$$

The third step is to add the equations for the commercial animals to nucleus animal equations using additions (6) or (7) of the main text. Here, only animal 5 is a commercial animal and it has both parents known (namely animals 3 and 4) such that addition (7) applies. Note that $1/\sigma_u^2$ is 1/3, such that the complete $\mathbf{G}^{-1}$ matrix for the reduced set of equations is:

$$\mathbf{G}^{-1} = \begin{bmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\ -1 & -1 & \frac{1}{6} & 2\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & 1\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 1\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 1\frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 1\frac{1}{6} & -\frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Use of conventional EBV to predict MA-EBV

Let us consider again the example of *table AI*, and make use of the conventional non-MA-EBV that are calculated for all animals, but are only available on the nucleus animals 1–4 together with their accuracies. These EBVs and their accuracies are calculated assuming an error variance of $\sigma_e^2 = 3$. The first step is to obtain a first approximation of the extra information due to the commercial animals by calculating $\delta_{[0]} = \text{diag}(\mathbf{C}^{-1}) - \text{diag}(\mathbf{M})$, where the ith diagonal element of $\mathbf{C}^{-1}$ is $\lambda/(1 - r_i^2)$, with $\lambda = \sigma_e^2/\sigma_u^2 = 1$; and $\mathbf{M}$ is the coefficient matrix of the conventional animal model equations for the nucleus animals:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 & -1 & -1 \\ 1 & 2 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

Hence, $\delta_{[0]} = \text{diag}(\mathbf{C}^{-1}) - \text{diag}(\mathbf{M}) = [-0.6326 \;\; -0.6326 \;\; -0.2222 \;\; -0.2222]'$.

In step 2 we first set up the $\mathbf{H}$ – matrix of derivatives of $\text{diag}((\mathbf{M} + \Delta)^{-1})$ with respect to δ:

$$\mathbf{H} = -(\mathbf{M} + \Delta_{[0]})^{-1} * (\mathbf{M} + \Delta_{[0]})^{-1} = -\begin{bmatrix} 0.3907 & 0.0113 & 0.0349 & 0.0349 \\ 0.0113 & 0.3907 & 0.0349 & 0.0349 \\ 0.0349 & 0.0349 & 0.2455 & 0.0181 \\ 0.0349 & 0.0349 & 0.0181 & 0.2445 \end{bmatrix} \quad \text{[A2]}$$

where $\Delta_{[0]}$ = a diagonal matrix with the elements of $\delta_{[0]}$ on the diagonals. Next we calculate:

$$\text{diag}((\mathbf{M} + \Delta_{[0]})^{-1}) - \text{diag}(\mathbf{C}) = [0.2027 \;\; 0.2027 \;\; 0.1345 \;\; 0.1345]'$$

The update of δ is now obtained from:

$$\delta_{[1]} = \delta_{[0]} - \mathbf{H}^{-1}[\text{diag}((\mathbf{M} + \Delta_{[0]})^{-1}) - \text{diag}(\mathbf{C})]$$
$$= [-0.1972 \;\; -0.1972 \;\; 0.1742 \;\; 0.1742]' \quad \text{[A3]}$$

After three more updates of δ by equations (A2) and (A3) the values of δ converged to $[0.03 \;\; 0.03 \;\; 0.36 \;\; 0.36]'$, which are equal to the Δ_{ii} in addition (11).

Next we set up the marker assisted mixed model equations of the nucleus animals, without accounting for the commercial animals, upon which the additions (11) will be performed. The inverse of the genetic (co)variance matrix

of these equations is obtained from (A1), such that the coefficient matrix of these equations is:

$$\begin{bmatrix} \mathbf{Z}_1'\mathbf{Z}_1 & \mathbf{Z}_1'\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 \end{bmatrix} + 3 * [A1] = \begin{bmatrix} 7 & 3 & -3 & -3 & 1 & 1 & 0 & 0 \\ 3 & 7 & -3 & -3 & 0 & 0 & 1 & 1 \\ -3 & -3 & 7 & 0 & 1 & 0 & 1 & 0 \\ -3 & -3 & 0 & 7 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 5 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 5 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 5 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 5 \end{bmatrix}$$

where $\mathbf{Z}_1$ is the design matrix of polygenic effects, i.e. a $(4 * 4)$ identity matrix; $\mathbf{Z}_2$ is the design matrix of the QTL effects, i.e. $\mathbf{Z}_2 = [1\ 1\ 0\ 0;\ 0\ 0\ 1\ 1;\ 1\ 0\ 1\ 0;\ 0\ 1\ 0\ 1]$; and the factor 3 is due to $\sigma_e^2 = 3$. Using the estimates of δ or, equivalently, Δ_{ii}, we next perform the additions in equation (11) to obtain the coefficient matrix of the mixed model equations that does account for the information of commercial animal 5:

$$\begin{bmatrix} 7.03 & 3 & -3 & -3 & 1.03 & 1.03 & 0 & 0 \\ 3 & 7.03 & -3 & -3 & 0 & 0 & 1.03 & 1.03 \\ -3 & -3 & 7.36 & 0 & 1.36 & 0 & 1.36 & 0 \\ -3 & -3 & 0 & 7.36 & 0 & 1.36 & 0 & 1.36 \\ 1.03 & 0 & 1.36 & 0 & 5.39 & 1.03 & 1.36 & 0 \\ 1.03 & 0 & 0 & 1.36 & 1.03 & 5.39 & 0 & 1.36 \\ 0 & 1.03 & 1.36 & 0 & 1.36 & 0 & 5.39 & 1.03 \\ 0 & 1.03 & 0 & 1.36 & 0 & 1.36 & 1.03 & 5.39 \end{bmatrix} \quad [A4]$$

Next, the polygenic * polygenic part of these equations, i.e. the (1:4, 1:4) block, should be multiplied by the non-MAS-EBV (see *table A1*) to obtain the new right hand side. This new right hand side deviates from the original hand side, i.e. $\mathbf{Z}_1'\mathbf{y}$, by $\Delta\mathbf{RHS} = [0.03\ -4.03\ 0\ 4.36]'$. This $\Delta\mathbf{RHS}$ is used to perform additions (12) to obtain the right hand side of the MA-mixed model equations of the nucleus animals:

$$\begin{bmatrix} \mathbf{Z}_1'\mathbf{y} \\ \mathbf{Z}_2'\mathbf{y} \end{bmatrix} + \text{additions[12]}$$

$$= [1.03\ \ -7.03\ \ 0\ \ 7.36\ \ -1.97\ \ 7.39\ \ -3.03\ \ 0.33]' \quad [A5]$$

The solutions from the coefficient matrix (A4) and the right hand side (A5) are $[0.887\ \ -0.887\ \ 0.177\ \ 0.823\ \ -0.759\ \ 1.20\ \ -0.20\ \ -0.241]'$, which are the estimates of the polygenic and QTL effects of the nucleus animals 1–4, using the information of the commercial animal 5.

Table AI. An example of a pedigree of five animals with their marker genotypes, records (y), non-MA-EBV, and their accuracies (r) (zeros indicate unknown parents).

Individual	Sire	Dam	Genotype	y	non-MA-EBV	r
1	0	0	A_1A_2	1	1	0.76
2	0	0	A_3A_4	−3	−1	0.76
3	1	2	A_1A_3	0	0	0.80
4	1	2	A_2A_4	3	1	0.80
5	3	4	unknown	–	–	–

APPENDIX B

Back-solving for the QTL effects of the commercial animals in the reduced set of equations

The reduced set of equations yields estimates of $v' = [a_1' \; q_1' \; u_2']$ (see text). Given the estimate $\widehat{\mathbf{V}}$, we can solve for the Mendelian sampling components: $\widehat{\varepsilon} = \mathbf{V}\hat{\mathbf{v}}$, where $\mathbf{V}$ and ε are defined in equation (4). The equations of the commercial animals yield no information to separate the Mendelian sampling components of the $\mathbf{u}_2$ effects, ε_3, into the components due to QTL effects and due to polygenes. Hence, the splitting of these components is proportional to their variance components, i.e.

$$\widehat{\varepsilon}_{31} = \widehat{\varepsilon}_3 \, \sigma_a^2/\sigma_u^2,$$
$$\widehat{\varepsilon}_{32} = \widehat{\varepsilon}_3 \otimes 1(1/2)\sigma_q^2/\sigma_u^2$$

where $\widehat{\varepsilon}_{31}$ and $\widehat{\varepsilon}_{32}$ are the Mendelian sampling components of $\mathbf{a}_2$ and $\mathbf{q}_2$ effects of the commercial animals, with $\mathbf{q}_2$ denoting the QTL effects sorted such that the paternal QTL effect of an animal is always followed by its maternal QTL effect. Next the estimates of $\mathbf{a}_2$ and $\mathbf{q}_2$ are obtained by solving:

$$\hat{\mathbf{a}}_2 = \mathbf{R}\hat{\mathbf{a}}_1 + \mathbf{T}\hat{\mathbf{a}}_2 + \widehat{\varepsilon}_{31}$$
$$\hat{\mathbf{q}}_2 = \mathbf{S}^*\hat{\mathbf{q}}_1 + \mathbf{T}^*\hat{\mathbf{q}}_2 + \widehat{\varepsilon}_{32}$$

where $\mathbf{S}^*(\mathbf{T}^*)$ is matrix with element (i, j) equal to half, if nucleus (commercial) QTL_j is a 'parental' QTL of commercial QTL_i, and zero otherwise; $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{q}}_1$ are known from the MA-EBV evaluation of the reduced set of equations; the elements of $\hat{\mathbf{a}}_2$ and $\hat{\mathbf{q}}_2$, which are needed in the right hand side of the above equations, have been calculated before they are needed when the animals are sorted from old to young within these vectors.