J. Dairy Sci. 105:923–939
https://doi.org/10.3168/jds.2021-20293

Invited review: Unknown-parent groups and metafounders in single-step genomic BLUP

Yutaka Masuda,[1]*† Paul M. VanRaden,[2] Shogo Tsuruta,[1] Daniela A. L. Lourenco,[1] and Ignacy Misztal[1]
[1]Department of Animal and Dairy Science, University of Georgia, Athens 30602
[2]Animal Genomics and Improvement Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705

ABSTRACT

Single-step genomic BLUP (ssGBLUP) is a method for genomic prediction that integrates matrices of pedigree (**A**) and genomic (**G**) relationships into a single unified additive relationship matrix whose inverse is incorporated into a set of mixed model equations (MME) to compute genomic predictions. Pedigree information in dairy cattle is often incomplete. Missing pedigree potentially causes biases and inflation in genomic estimated breeding values (GEBV) obtained with ssGBLUP. Three major issues are associated with missing pedigree in ssGBLUP, namely biased predictions by selection, missing inbreeding in pedigree relationships, and incompatibility between **G** and **A** in level and scale. These issues can be solved using a proper model for unknown-parent groups (UPG). The theory behind the use of UPG is well established for pedigree BLUP, but not for ssGBLUP. This study reviews the development of the UPG model in pedigree BLUP, the properties of UPG models in ssGBLUP, and the effect of UPG on genetic trends and genomic predictions. Similarities and differences between UPG and metafounder (MF) models, a generalized UPG model, are also reviewed. A UPG model (QP) derived using a transformation of the MME has a good convergence behavior. However, with insufficient data, the QP model may yield biased genetic trends and may underestimate UPG. The QP model can be altered by removing the genomic relationships linking GEBV and UPG effects from MME. This altered QP model exhibits less bias in genetic trends and less inflation in genomic predictions than the QP model, especially with large data sets. Recently, a new model, which encapsulates the UPG equations into the pedigree relationships for genotyped animals, was proposed in simulated purebred populations. The MF model is a comprehensive solution to the missing pedigree issue. This model can be a choice for multibreed or crossbred evaluations if the data set allows the estimation of a reasonable relationship matrix for MF. Missing pedigree influences genetic trends, but its effect on the predictability of genetic merit for genotyped animals should be negligible when many proven bulls are genotyped. The SNP effects can be back-solved using GEBV from older genotyped animals, and these predicted SNP effects can be used to calculate GEBV for young-genotyped animals with missing parents.
Key words: bias, genomic selection, pedigree, single-step evaluation, relationship matrix

INTRODUCTION

Single-step genomic BLUP (**ssGBLUP**) is a genomic prediction method used to obtain genomic EBV (**GEBV**) for both genotyped and nongenotyped animals (Legarra et al., 2009; Misztal et al., 2009). The ssGBLUP is based on the inverse of a unified additive relationship matrix ($\mathbf{H}^{-1}$) that is a function of the numerator relationship matrix (**A**) for pedigree animals (Henderson, 1976) and the genomic relationship matrix (**G**) for genotyped animals (VanRaden, 2008). The ssGBLUP is routinely used for genomic evaluations in various domestic animal species (Misztal et al., 2020). Several countries tested ssGBLUP using national-level dairy data sets (Koivula et al., 2018; Masuda et al., 2018b; Oliveira et al., 2019), and a few countries implemented this method officially (https://interbull.org/ib/nationalgenoforms). Legarra et al. (2014), Mäntysaari et al. (2020), and Misztal et al. (2020) reviewed the advantages of ssGBLUP over a "multi-step" method involving a sequence of statistical procedures.

There is another class of single-step methods predicting SNP marker effects instead of breeding values for genotyped animals (Gengler et al., 2012; Fernando et al., 2014; Liu et al., 2014). We refer to this type of method as single-step marker effect model (**ssMEM**).

Received February 12, 2021.
Accepted September 26, 2021.
*Corresponding author: yutaka@rakuno.ac.jp
†Current address: Department of Sustainable Agriculture, Rakuno Gakuen University, Hokkaido, Japan.

As ssGBLUP is extended from genomic BLUP (**GBLUP**; VanRaden, 2008) to consider nongenotyped animals in the model, ssMEM is an extension of marker prediction models (Meuwissen et al., 2001; Habier et al., 2011) with SNP genotypes as covariates. In dairy cattle, ssMEM were investigated on computing strategies (Taskinen et al., 2017; Vandenplas et al., 2018) and predictive ability (Konstantinov and Goddard, 2020; Alkhoder and Liu, 2021). Mäntysaari et al. (2020) reviewed ssMEM in terms of modeling, development, and applications to dairy cattle breeding.

Single-step and multi-step genomic evaluations in dairy populations often face biased GEBV because of intensive selection, unbalanced data structure, and a large fraction of missing pedigree (Wiggans et al., 2017; Mäntysaari et al., 2020; Nani et al., 2020). The GEBV for young animals tend to show inflation defined as $b_1 < 1$ in a cross-validation, where b_1 is the slope coefficient of a simple regression of reliable genetic evaluations (e.g., deregressed proof calculated from the full data) on genomic predictions (based on the truncated data). There are additional issues regarding pedigree incompleteness in US dairy populations because cows are categorized as either registered or nonregistered (Wiggans et al., 2012a). Registered animals have an almost complete pedigree, but nonregistered animals often have unidentified dams. Further, genotyped heifers may not have pedigree information. Wiggans et al. (2018) reported that 97% of sires but 39% of dams of over 2 million genotyped animals in the dairy database were validated as of January 2018.

In quantitative genetics, any unknown parent is assumed to be sampled from a base population, where the expectation of a breeding value is zero (Falconer and Mackay, 1996). However, this assumption is inappropriate for animals in a selected population and for animals imported from other populations because it may lead to biased genetic evaluations. Unknown-parent groups (**UPG**), also known as phantom parents, have been applied to pedigree-based animal models in dairy cattle to reduce bias (Graser et al., 1987; Quaas, 1988; Westell et al., 1988). In a way, unknown parents are replaced by a group effect. This model can account for a change in genetic trend that is not accounted for by known pedigree relationships.

Unknown-parent groups have been incorporated into ssGBLUP through a modification of $\mathbf{H}^{-1}$ (Tsuruta et al., 2011; Misztal et al., 2013; Masuda et al., 2021). In this review, the modified $\mathbf{H}^{-1}$ with UPG will be referred to as "H-inverse." Some UPG models have decreased the inflation of GEBV and bias in genetic trends, and increased the convergence stability of iterative solvers for the mixed model equations (**MME**) (Tsuruta et al., 2014; Matilainen et al., 2016, 2018). However, UPG may not completely eliminate bias and inflation. Reasons for this are potential discrepancies between **G** and **A** due to incomplete relationships and inbreeding in **A** when pedigree is missing (Misztal et al., 2013, 2017). The 2 relationship matrices can be "aligned" in several ways (Chen et al., 2011; Vitezica et al., 2011).

Christensen (2012) and Legarra et al. (2015) developed the metafounder (**MF**) model, a generalization of the UPG model in ssGBLUP. The MF model is intended to ensure "compatibility" between **G** and **A**, and it can be applied to genomic evaluation of animals from related base populations such as breeds or lines. The MF model is relatively new; thus, research has only recently started in dairy cattle.

There is a need for alternative strategies to compensate for missing pedigree in ssGBLUP, and there is also a lack of extensive discussion covering all possible issues arising from missing pedigree. A comprehensive solution could, in principle, be achieved in the context of MF. However, although the UPG theory is well defined for pedigree BLUP (Wolak and Reid, 2017), the most reasonable model for ssGBLUP is still unknown (Misztal et al., 2020). For ssMEM, the study on UPG has just begun (Vandenplas et al., 2021).

The primary objective of this review is to clarify the properties of UPG and MF models in ssGBLUP. We first review the development of the UPG model in pedigree-based BLUP. Then, we describe general issues due to missing pedigree in ssGBLUP and compare several models with alternative H-inverses. Metafounders are discussed in terms of similarities to and differences from UPG. We review published studies regarding the effect of missing pedigree on ssGBLUP genomic predictions for dairy cattle. The recent development of UPG models in ssMEM will also be discussed.

GROUPING STRATEGY IN CLASSICAL BLUP

Before the Animal Model Era

The earliest use of grouping is attributed to Henderson (1949), who tried to predict a cow's "producing ability" from records subject to culling (Pollak and Quaas, 1983). The robustness of linear mixed models against selection was unknown. Animal breeders at that time regularly used least-squares methods that gave biased predictions with selected data. Groups were expected to account for environmental and genetic trends. Later, Henderson et al. (1959) formally described a linear mixed model with random unrelated animal effects (u_{ij}) nested within a fixed group (g_j), where producing ability was expressed as $u_{ij} + g_j$.

The grouping strategy used in the sire model to evaluate dairy bulls assumed that all bulls were unrelated (Henderson, 1973; Powell and Freeman, 1974). Bulls were assigned to groups to account for potential genetic differences and genetic trends by bull stud and year. Some scientists expected the importance of grouping bulls to decrease with the use of $\mathbf{A}^{-1}$ for bulls, computed with the newly developed fast algorithm (Henderson, 1975; Thompson, 1979). Subsequently, Pollak and Quaas (1983) showed that the need for grouping decreased as $\mathbf{A}^{-1}$ became complete. They also demonstrated that the group solutions reflected a genetic selection differential representing the maternal side of the pedigree that was not included in the bull pedigree file (Schaeffer, 1991). Hence, although relationships among bulls were considered in the sire model, group effects were still required to reduce bias in bull genetic predictions.

Unknown-Parent Groups in the Animal Model

Theoretical Development. An animal model is expected to account for selection and drift through $\mathbf{A}^{-1}$ for all animals in the pedigree (Kennedy et al., 1988). Westell et al. (1988, p. 1310) stated that "group effects can be thought of as accounting for selection not accounted for by records of relatives. Under this concept, groups would be assigned only if animals were missing genetic relationships." Usually, groups are arbitrarily defined to reflect selection differentials of unknown parents based on sex, time period, region, breed, and country of origin (Westell et al., 1988; Wiggans et al., 1988).

A single-trait animal model can be written as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}, \qquad [1]$$

where $\mathbf{y}$ is a vector of observations; $\mathbf{b}$ is a vector of fixed effects; $\mathbf{u}$ is a vector of random additive genetic effects for all animals in the pedigree; $\mathbf{e}$ is a vector of random residuals; $\mathbf{X}$ is an incidence matrix relating observations to elements of $\mathbf{b}$; and $\mathbf{Z}$ is an incidence matrix relating observations to elements of $\mathbf{u}$. Vectors $\mathbf{u}$ and $\mathbf{e}$ have expected values equal to zero, $\text{var}(\mathbf{u}) = \mathbf{A}\sigma_u^2$, where σ_u^2 is the additive genetic variance, and $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$, where σ_e^2 is the residual variance. Henderson (1976) showed that $\mathbf{A} = \mathbf{TDT}'$, where $\mathbf{T}$ is a lower triangular matrix, and $\mathbf{D}$ is a diagonal matrix which is the covariance matrix of Mendelian-sampling contributions.

Thompson (1979) recognized that the breeding value of an animal was expressed as an accumulation of Mendelian-sampling terms from its ancestors and itself through $\mathbf{T}$. He then suggested that the group effects could also be accumulated in the same manner under a sire model. Robinson (1986) and Westell and Van Vleck (1987) illustrated how this concept could be extended to the animal model.

The current UPG model used in genetic evaluation was developed by Graser et al. (1987) and Westell et al. (1988) and explained in detail by Quaas (1988). A vector of breeding values with a contribution of group effects is

$$\mathbf{u}^* = \mathbf{u} + \mathbf{Qg}, \qquad [2]$$

where $\mathbf{Q}$ is a matrix of expected fractions of genes in the ith individual coming from the jth group, and $\mathbf{g}$ is a vector of fixed means (i.e., UPG effects). Note that the sum of the group compositions (the sum of elements in a row of $\mathbf{Q}$) is 1.

Transformed Equations with UPG. Replacing vector $\mathbf{u}$ with vector $\mathbf{u}^*$ in mixed model [1] yields

$$\begin{aligned}\mathbf{y} &= \mathbf{Xb} + \mathbf{Zu} + \mathbf{ZQg} + \mathbf{e}\\ &= \mathbf{Xb} + \mathbf{Zu}^* + \mathbf{e}.\end{aligned} \qquad [3]$$

The solutions for $\mathbf{u}$ and $\mathbf{g}$ are obtained by solving the MME based on the first model in [3]. To obtain the solution for $\mathbf{u}^*$ directly, the MME can be reformulated using the Quaas-Pollak ($\mathbf{QP}$) transformation (Quaas and Pollak, 1981) as

$$\begin{bmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & 0\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\mathbf{A}^{-1}\lambda & -\mathbf{A}^{-1}\mathbf{Q}\lambda\\ 0 & -\mathbf{Q}'\mathbf{A}^{-1}\lambda & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}\lambda\end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{u}}^*\\ \hat{\mathbf{g}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\\ 0\end{bmatrix}, \qquad [4]$$

where $\lambda = \sigma_e^2/\sigma_u^2$. Using the joint density function, Quaas (1988) derived the "inverse of a covariance matrix" for $\boldsymbol{\theta}' = \begin{bmatrix}\mathbf{u}^{*\prime} & \mathbf{g}'\end{bmatrix}$; that is,

$$\mathbf{A}^* = \begin{bmatrix}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q}\\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}\end{bmatrix}. \qquad [5]$$

This matrix is not full rank, and therefore, it is not the inverse of any particular relationship matrix. Nevertheless, Henderson's rules can still be used to build matrix [5] without computing $\mathbf{Q}$ by treating the UPG as "phantom animals" that do not contribute to the Mendelian-sampling variance (Quaas, 1988; Westell et al., 1988).

The joint relationship matrix for $\mathbf{u}^*$ and $\mathbf{g}$ [i.e., var($\boldsymbol{\theta}$)], is defined by random group effects, assuming

$\mathbf{g} \sim MVN\left(0, \boldsymbol{\Sigma}\sigma_u^2\right)$, where $\boldsymbol{\Sigma}$ is the covariance matrix among UPG. Using block-matrix partitioning rules (Searle, 1982) and omitting σ_u^2, var($\boldsymbol{\theta}$) and its inverse are as follows:

$$\mathbf{A}_{\Sigma}^{*} = \left(\text{var}\begin{bmatrix}\mathbf{u}+\mathbf{Qg}\\ \mathbf{g}\end{bmatrix}\right)^{-1} = \begin{bmatrix}\mathbf{A}+\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}' & \mathbf{Q}\boldsymbol{\Sigma}\\ \boldsymbol{\Sigma}\mathbf{Q}' & \boldsymbol{\Sigma}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q}\\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}+\boldsymbol{\Sigma}^{-1}\end{bmatrix}. \quad [6]$$

The covariance matrix [6] shows that the incomplete $\mathbf{A}$ is extended with $\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}'$ (i.e., the UPG variation), although the inverse does not clearly show this fact. The inverse [6] becomes $\mathbf{A}^{*}$ for fixed UPG effects because $\boldsymbol{\Sigma}^{-1} \to 0$. Note that a form of the matrix [6] was used in a context of random UPG (van der Werf and Thompson, 1992; VanRaden, 1992; Schaeffer, 1994), but it was not formally derived until recently (Masuda et al., 2021).

Fixed Versus Random UPG. The UPG have been modeled as fixed effects because a group effect is thought to represent the average genetic level of a base population (Graser et al., 1987; Quaas, 1988). Under this assumption, $\mathbf{A}^{*}$ is not full rank, and a group effect cannot be uniquely estimated because of dependencies on, and possibly confounding with, other group fixed effects. The predicted genetic value of an animal should be expressed as a difference from the predicted value of other animals.

Several concerns have been raised about fixed UPG. Kennedy (1991) pointed out shortcomings of assuming UPG as fixed from a theoretical point of view, namely unaccountability of relationships among base individuals and inbreeding in the base populations, no change in additive genetic variances in base populations, and no consideration of genetic drift for small UPG. These concerns were discussed in subsequent studies with random UPG, including VanRaden's model (VanRaden, 1992) and selected-base-population models as well as the metafounder model presented in the sections below. Another issue is numerical instability of UPG estimates. Foulley et al. (1992) suggested random UPG as an option when fixed UPG were not estimated well. A model with random UPG appeared useful to avoid confounding between UPG and fixed effects in the model because it removed estimability problems, and helped with convergence in iterative solvers (Schaeffer, 1994; Sullivan and Schaeffer, 1994; Lidauer et al., 2015). Schaeffer (1994) suggested using $\boldsymbol{\Sigma} = \mathbf{I}$ in [6] because of simplicity. A random UPG model is considered as a better option than the fixed UPG in terms of theory and computational stability.

VanRaden's Random UPG with Approximated Inbreeding

Approximated Inbreeding Coefficients. Inbreeding coefficients are used to monitor genetic diversity in livestock and wild populations, assess inbreeding depression, and construct $\mathbf{A}^{-1}$ in genetic evaluation. The inbreeding coefficients can be calculated with the tabular method (Emik and Terrill, 1949) or more efficient algorithms (e.g., Meuwissen and Luo, 1992). However, if the pedigree is incomplete, unknown parents are assumed to be noninbred and unrelated, and inbreeding is underestimated in populations with assortative mating.

VanRaden (1992) developed an iterative method to approximate inbreeding coefficients using pedigree with UPG, assuming that base individuals in a group may be inbred and individuals among groups may be related. Initially, the inbreeding for UPG j (F_j) and the relationship among UPG j and k (a_{jk}) are set to zero. As a first step, the inbreeding coefficients for animals in the pedigree are calculated using the tabular method as if UPG were real animals. In the next step, F_j and a_{jk} are estimated with the mean inbreeding for and the mean relationship among known parents that have the same genetic background as the corresponding UPG. The 2 steps are repeated until the inbreeding coefficients for all animals converge. The approximated inbreeding coefficients using this approach are expected to be equal to or greater than standard inbreeding coefficients because of inbred ancestors. VanRaden (1992) suggested constructing $\boldsymbol{\Sigma}$ [6] as $\Sigma_{jj} = 1 + F_j$ and $\Sigma_{jk} = a_{jk}$ at convergence.

Inbred and Selected Base Populations. The inbreeding coefficients should be considered in $\mathbf{A}^{-1}$ to properly account for selection on the variance of breeding values in the animal model although the inbreeding coefficients are optional to construct $\mathbf{A}^{-1}$ (Henderson, 1976). If the pedigree is incomplete, the approximated inbreeding can be an alternative to use. VanRaden (1992) suggested a random UPG model for $\mathbf{A}_{\Sigma}^{*}$ [6] with approximated inbreeding coefficients for pedigree animals and approximated $\boldsymbol{\Sigma}$ for UPG. Henderson's rules to construct $\mathbf{A}^{-1}$ are the same except for the elements of the inverse of the matrix of Mendelian-sampling variances ($\mathbf{D}^{-1}$). The ith diagonal element of $\mathbf{D}^{-1}$ is

$$d_{ii} = \frac{4}{\left(1+r\delta_s\right)\left(1-\hat{F}_s\right)+\left(1+r\delta_d\right)\left(1-\hat{F}_d\right)}, \quad [7]$$

where $\hat{F}_s\left(\hat{F}_d\right)$ is the approximated inbreeding coefficient of sire s (dam d), δ_s (δ_d) is 0 if sire (dam) is known and 1 otherwise, and r represents the reduction in additive genetic variance, and it is set to 1 when base popula-

tions are not selected. van der Werf and Thompson (1992) and Alfonso and Estany (1999) discussed how r was estimated. When s (d) is unknown, the inbreeding of the corresponding UPG $\left(\hat{F}_j\right)$ is used.

The random UPG model with formula [7] deals with selected base populations. Henderson (1985) assumed that the base populations had been selected, and he tried to alter $\mathbf{A}$ to account for the selected base populations with nonzero means of breeding values. Later, Henderson (1988) simplified his 1985 model and found it to be identical to the standard UPG model (Graser et al., 1987). If base populations are generated as a consequence of selection or drift, ideally, all base animals should be traced back to an unselected hypothetical population, as Schaeffer (1991) explained. van der Werf and Thompson (1992) developed the random UPG model [7] independently of VanRaden (1992), and they argued that r could satisfy $\sigma_u^2 = r\sigma_{u0}^2$, where σ_u^2 is the additive genetic variance after selection, and σ_{u0}^2 is the additive genetic variance before selection. These authors suggested using $\mathbf{\Sigma} = r\mathbf{I}$ and $r = 1 - \overline{F}$, where $\overline{F}$ is the average inbreeding coefficient in the selected base population. Alfonso and Estany (1999) were not aware of VanRaden (1992), and they provided a formal derivation to construct $\mathbf{A}^{-1}$ based on selected base populations with fixed UPG. Their effort resulted in VanRaden's formula [7].

MISSING PEDIGREE ISSUES IN SINGLE-STEP GBLUP

Complete Pedigree with No Selection

Misztal et al. (2009) suggested including genotyped and nongenotyped animals in the standard animal model [1]. The unified additive relationship matrix when only a portion of animals in the pedigree have genotypes, omitting the additive genetic variance, is as follows:

$$\mathbf{H} = \begin{bmatrix} \mathrm{var}(\mathbf{u}_1) & \mathrm{cov}(\mathbf{u}_1, \mathbf{u}_2) \\ \mathrm{cov}(\mathbf{u}_2, \mathbf{u}_1) & \mathrm{var}(\mathbf{u}_2) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix},$$

where subscript 1 is for nongenotyped animals and subscript 2 is for genotyped animals (Legarra et al., 2009). The inverse of the unified additive relationship matrix $\mathbf{H}$, derived by Aguilar et al. (2010) and Christensen and Lund (2010), is as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

According to VanRaden et al. (2009) and Aguilar et al. (2010), $\mathbf{G}$ is expected to partially account for the same additive genetic relationships as $\mathbf{A}$, and therefore, $\mathbf{A}_{22}$ should discount for double-counting. In practice, to ensure positive definiteness, $\mathbf{G}$ is often "blended" with $\mathbf{A}_{22}$ (or $\mathbf{I}$) as $w\mathbf{G} + (1 - w)\mathbf{A}_{22}$, where w is an arbitrary weight (VanRaden, 2008). The weight $1 - w$ can be defined as the fraction of the residual polygenic variance (Liu et al., 2016; Mäntysaari et al., 2017) not accounted for by $\mathbf{G}$. In this review, $\mathbf{G}$ is assumed to be positive definite, thus there is no residual polygenic variance.

Matrix $\mathbf{H}$ shows that $\mathbf{G}$ and $\mathbf{A}_{22}$ need to be "compatible" to satisfy $\mathrm{E}(\mathbf{G}) = \mathbf{A}_{22}$ when marker genotypes are considered as random variables (VanRaden, 2008; Christensen et al., 2012). Compatibility means that $\mathbf{G}$ and $\mathbf{A}_{22}$ have the same average of diagonals and the same average of off-diagonals. If this is not guaranteed, the GEBV are thought to be biased (Chen et al., 2011; Vitezica et al., 2011).

The average elements of $\mathbf{A}_{22}$ (i.e., average inbreeding and average additive genetic relationship) are determined by the number of generations in the pedigree, pedigree completeness, effective population size, selection, drift, and mating system (Falconer and Mackay, 1996). Meanwhile, the average elements of $\mathbf{G}$ are computed using allele frequencies ($\mathbf{AF}$) of genomic markers (VanRaden, 2008). For example, using AF calculated from genomic data, the average of the off-diagonal elements in $\mathbf{G}$ is expected to be 0, and the average of the diagonal elements of $\mathbf{G}$ is expected to be 1 (Chen et al., 2011). In an ideal population (i.e., large, in random mating, and without selection and drift), or when the pedigree is shallow, individuals are most likely to be unrelated and noninbred in $\mathbf{A}_{22}$, and AF are expected to be constant over generations. Thus, $\mathbf{G}$ can be compatible with $\mathbf{A}_{22}$.

Complete Pedigree Under Selection

In a typical breeding program, the population is subject to selection, and genotyped animals are often from recent generations, although the pedigree can contain animals from a large number of generations. Average elements in $\mathbf{A}_{22}$ are greater than in $\mathbf{G}$ when computed with current AF. Average elements of the additive relationship matrix are associated with expected breeding values in selected populations (Chen et al., 2011; Vitezica et al., 2011). The "genomic base" (population with mean GEBV equal to 0 for genotyped

animals) is defined by the population providing the AF. In contrast, the first generation of the pedigree defines a "genetic base" with mean EBV equal to 0. The mean GEBV based on current AF can be equal to 0 for the current genotyped population, even when the mean EBV is nonzero. In other words, $\mathbf{G}$ constructed with current AF cannot account for past selection that occurred before the genomic information was collected.

To make the $\mathbf{A}_{22}$ and $\mathbf{G}$ matrices compatible, or equivalently to make the genomic and genetic bases be at the same level, $\mathbf{G}$ should be calculated using the same AF as in the base population defined by pedigree (VanRaden, 2008; Meuwissen et al., 2011; Meyer et al., 2018). The base AF are unknown, and an accurate estimation of the frequencies is often difficult (Misztal et al., 2020). A simple solution is to align $\mathbf{G}$ to $\mathbf{A}_{22}$ in terms of the averages of their diagonal and off-diagonal elements. This "alignment" has the same effect as adding a constant to the GEBV to shift the genomic base [see Legarra et al. (2014) for a detailed discussion].

Strandén and Christensen (2011) showed that the general mean added to the GEBV accounted for the difference in AF as follows:

$$\mathbf{u}_{2:\text{aligned}} = \mathbf{u}_2 + \mathbf{1}\mu, \qquad [8]$$

where $\mathbf{1}$ is a vector of ones, and μ is the general mean that absorbs the difference between genomic and genetic bases. This adjustment is justified when $\mathbf{G}$ is constructed as $\mathbf{WW}'/s$ and $\mathbf{W} = \mathbf{M} - \mathbf{1p}'$, where $\mathbf{M}$ is a matrix of genomic markers, $\mathbf{p}$ is a vector of known functions of AF, and s is a constant. The genomic relationships of VanRaden's first method (VanRaden, 2008) satisfy this condition. This explicit-mean model was studied in SNP BLUP by Hsu et al. (2017) and in ssGBLUP by Bermann et al. (2021).

The aligned GEBV are obtainable without computing μ using the aligned $\mathbf{G}$ as the variance of [8]:

$$\mathbf{G}_{\text{aligned}} = \beta\mathbf{G} + \alpha\mathbf{11}', \qquad [9]$$

where α accounts for selection and drift in GEBV and β adjusts for the inflation of GEBV due to selection (Vitezica et al., 2011; Christensen et al., 2012; Misztal et al., 2017). The constants α and β are calculated based on average elements of $\mathbf{G}$ and $\mathbf{A}_{22}$ (Chen et al., 2011; Christensen et al., 2012; Gao et al., 2012). Formula [9] clearly assumes that $\mathbf{G}$ is lacking variation due to μ. It should be noted that the individual element of $\mathbf{G}$ does not necessarily match the corresponding element in $\mathbf{A}_{22}$ because the alignment modifies the average and scale of elements in $\mathbf{G}$ as a whole so that the genomic base is adjusted to the genetic base. Meyer et al. (2018) reviewed various alignment methods to estimate unbiased genetic trends under selection.

Vitezica et al. (2011) showed that the alignment formula [8] was equivalent to a UPG model. When μ is explicitly included in the model, $\mathbf{H}^{-1}$ is augmented by an unaligned $\mathbf{G}^{-1}$ as follows:

$$\mathbf{H}^{-1}_{\text{aligned}} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \beta^{-1}\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & \beta^{-1}\mathbf{G}^{-1}1 \\ 0 & \beta^{-1}1'\mathbf{G}^{-1} & \beta^{-1}1'\mathbf{G}^{-1}1 + \alpha^{-1} \end{bmatrix} \qquad [10]$$

for a vector of interest $\boldsymbol{\theta}' = \begin{bmatrix} \mathbf{u}'_{1:\text{aligned}} & \mathbf{u}'_{2:\text{aligned}} & \mu \end{bmatrix}$, where $\mathbf{A}^{-1}$ is padded with 0 for equations due to μ, and $\mathbf{u}'_{1:\text{aligned}}$ is the vector of breeding values aligned with a known function of μ for nongenotyped animals. The aligned breeding values in $\boldsymbol{\theta}$ are expected to be unbiased because selection and drift are properly accounted for.

Incomplete Pedigree Under Selection

Several studies have described compatibility issues in real populations under selection when some animals have missing parents regardless of their genotyping status (Misztal et al., 2013; Tsuruta et al., 2014, 2019). In dairy cattle populations, UPG have often been applied only to pedigree relationships because software packages for ssGBLUP supported pedigree UPG only (Tsuruta et al., 2011; Koivula et al., 2015). An H-inverse with pedigree UPG is as follows:

$$\mathbf{H}^{*}_{\omega} = \mathbf{A}^{*} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad [11]$$

where ω is an arbitrary constant. This constant is required to reduce the inflation of GEBV. Also, it could avoid divergence in MME iterative solvers (Misztal et al., 2013). Martini et al. (2018) presented a detailed discussion on the use of ω to decrease GEBV inflation. A genomic model with this H-inverse will be referred to as the "Omega model" (**Omega-M**).

The Omega-M was "derived" ad hoc; thus, it may not be theoretically justified. Although the Omega-M has been widely used to obtain GEBV with UPG, inflation and bias of GEBV as well as computational problems remained even though UPG, alignment, and ω were considered (Tsuruta et al., 2014). Further, when ω is too small (<0.8), the validation accuracy of GEBV may be reduced (Mäntysaari et al., 2020).

Misztal et al. (2013, 2017) suggested that missing pedigree could cause at least 3 inseparable issues asso-

ciated with bias and inflation of GEBV in ssGBLUP: (1) biased predictions by selection due to a lack of information, (2) underestimation of inbreeding and additive relationships, and (3) poor compatibility between $\mathbf{G}$ and $\mathbf{A}_{22}$. Issues 1 and 2 are associated with incorrect UPG, and issues 2 and 3 are related to an incomplete $\mathbf{A}_{22}$ against a complete $\mathbf{G}$. Underestimation of inbreeding for genotyped animals may create a discrepancy between pedigree and genomic inbreeding and relationships. This issue is attributed to unequal pedigree lengths for genotyped animals when $\mathbf{G}$ is aligned to average elements of $\mathbf{A}_{22}$ with missing pedigree. For an animal with a deeper pedigree, the aligned genomic relationships are expected to be smaller than the pedigree relationships. These facts are problematic in the inverse scale, and the diagonals of $\left(\mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\right)$ can be negative, possibly leading to ill-conditioned MME. In addition, with imperfect alignment, GEBV tend to be biased up with shorter (shallower) pedigree, and biased down with longer (deeper) pedigree (Chen et al., 2011). The arbitrary constant ω can compensate for a portion of the misalignment. When UPG are applied only to $\mathbf{A}^{-1}$, neglecting the other relationship matrices, the UPG estimates are expected to be biased (Misztal et al., 2013).

UPG MODELS IN SINGLE-STEP GBLUP

Misztal et al. (2013) projected the UPG model [3] into the single-step GBLUP model by partitioning $\mathbf{Q}$ into $\mathbf{Q}'=\begin{bmatrix}\mathbf{Q}_1' & \mathbf{Q}_2'\end{bmatrix}$ to be conformable with $\mathbf{H}^{-1}$, and defining $\boldsymbol{\theta}$ as follows:

$$\boldsymbol{\theta}=\begin{bmatrix}\mathbf{u}_1^*\\ \mathbf{u}_2^*\\ \mathbf{g}\end{bmatrix}=\begin{bmatrix}\mathbf{u}_1+\mathbf{Q}_1\mathbf{g}\\ \mathbf{u}_2+\mathbf{Q}_2\mathbf{g}\\ \mathbf{g}\end{bmatrix}.$$

As Matilainen et al. (2018) showed, $\mathbf{A}_\Sigma^*$ [6] can be divided into 3×3 blocks according to $\boldsymbol{\theta}$; thus,

$$\begin{aligned}\mathbf{A}_\Sigma^*&=\begin{bmatrix}\mathbf{A}^{11} & \mathbf{A}^{12} & -\begin{bmatrix}\mathbf{A}^{11} & \mathbf{A}^{12}\end{bmatrix}\mathbf{Q}\\ \mathbf{A}^{21} & \mathbf{A}^{22} & -\begin{bmatrix}\mathbf{A}^{21} & \mathbf{A}^{22}\end{bmatrix}\mathbf{Q}\\ -\mathbf{Q}'\begin{bmatrix}\mathbf{A}^{11}\\ \mathbf{A}^{21}\end{bmatrix} & -\mathbf{Q}'\begin{bmatrix}\mathbf{A}^{12}\\ \mathbf{A}^{22}\end{bmatrix} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}+\boldsymbol{\Sigma}^{-1}\end{bmatrix}\\ &=\begin{bmatrix}\mathbf{A}^{11} & \mathbf{A}^{12} & \mathbf{A}^{13}\\ \mathbf{A}^{21} & \mathbf{A}^{22} & \mathbf{A}^{23}\\ \mathbf{A}^{31} & \mathbf{A}^{32} & \mathbf{A}_\Sigma^{33}\end{bmatrix},\end{aligned} \quad [12]$$

where subscript 3 refers to UPG. The functions of $\mathbf{Q}$ in the third row and column will be referred to as the UPG equations. Matrix $\mathbf{A}_\Sigma^*$ is sparse, and the sparsity is useful to compute $\mathbf{A}_{22}^{-1}$ indirectly (Henderson, 1976) as follows:

$$\mathbf{A}_{22}^{-1}=\mathbf{A}^{22}-\mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}. \quad [13]$$

Notably, $\mathbf{A}^{11}$ contains information for nongenotyped ancestors of genotyped animals only.

Properties of UPG Models

QP Model. Misztal et al. (2013) replaced $\mathbf{A}^{-1}$ with $\mathbf{H}^{-1}$ in the QP-transformed equation [4] and presented the following $\mathbf{H}_{Q\Sigma}^*$ matrix associated with $\boldsymbol{\theta}$:

$$\mathbf{H}_{Q\Sigma}^*=\mathbf{A}_\Sigma^*+\begin{bmatrix}0 & 0 & 0\\ 0 & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} & -\left(\mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\right)\mathbf{Q}_2\\ 0 & -\mathbf{Q}_2'\left(\mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\right) & \mathbf{Q}_2'\left(\mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\right)\mathbf{Q}_2\end{bmatrix}. \quad [14]$$

Formula [14] assumes random UPG although the original QP transformation assumed fixed UPG $\left(\boldsymbol{\Sigma}^{-1}=0\right)$. Regardless of whether UPG are fixed or random, the model corresponding to the $\mathbf{H}_{Q\Sigma}^*$ matrix [14] will be referred to as the "QP model" (**QP-M**).

Matrix $\mathbf{H}_{Q\Sigma}^*$ can be split into several matrices as follows:

$$\begin{aligned}\mathbf{H}_{Q\Sigma}^*&=\mathbf{A}_\Sigma^*+\begin{bmatrix}0 & 0 & 0\\ 0 & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1} & 0\\ 0 & 0 & 0\end{bmatrix}+\\ &\quad\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & -\mathbf{G}^{-1}\mathbf{Q}_2\\ 0 & -\mathbf{Q}_2'\mathbf{G}^{-1} & \mathbf{Q}_2'\mathbf{G}^{-1}\mathbf{Q}_2\end{bmatrix}-\\ &\quad\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & -\mathbf{A}_{22}^{-1}\mathbf{Q}_2\\ 0 & -\mathbf{Q}_2'\mathbf{A}_{22}^{-1} & \mathbf{Q}_2'\mathbf{A}_{22}^{-1}\mathbf{Q}_2\end{bmatrix}\\ &=\mathbf{A}_\Sigma^*+\mathbf{E}_{GA22}+\mathbf{E}_G-\mathbf{E}_{A22},\end{aligned}$$

The matrix $\mathbf{E}_G$ is similar to $\mathbf{1}'\mathbf{G}^{-1}\mathbf{1}$ and $\mathbf{G}^{-1}\mathbf{1}$ in the aligned model [10] and relates UPG effects ($\mathbf{g}$) to the base adjustment (μ). In addition, $\mathbf{E}_G$ contains a direct association between $\mathbf{u}_2^*$ and $\mathbf{g}$, which makes the 2 effects inseparable. In a simulation study, Masuda et al. (2021)

suggested that, if alignment is applied, $\mathbf{u}_2^*$, $\mathbf{g}$, and μ can be confounded, and GEBV become sensitive to UPG solutions. The matrix $\mathbf{E}_{A22}$ is expected to offset $\mathbf{A}_\Sigma^*$. However, $\mathbf{E}_{A22}$ has a structure similar to $\mathbf{E}_G$; thus, $\mathbf{u}_2^*$ and $\mathbf{g}$ may be inseparable in the MME, and the solution for $\mathbf{g}$ may also be biased (Masuda et al., 2021).

Altered QP Model. Masuda et al. (2019a), Bradford et al. (2019), and Tsuruta et al. (2019) thought that UPG should merely compensate for missing pedigree-relationships, not for genomic relationships. These authors arbitrarily removed $\mathbf{G}^{-1}$ from the UPG equations in $\mathbf{H}_{Q\Sigma}^*$ as follows:

$$\mathbf{H}_{A\Sigma}^* = \mathbf{A}_\Sigma^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -\left(-\mathbf{A}_{22}^{-1}\right)\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2'\left(-\mathbf{A}_{22}^{-1}\right) & \mathbf{Q}_2'\left(-\mathbf{A}_{22}^{-1}\right)\mathbf{Q}_2 \end{bmatrix}. \quad [15]$$

The model associated with $\mathbf{H}_{A\Sigma}^*$ matrix [15] that disregards fixed and random UPG will be referred to as the "altered-QP model" (**AltQP-M**). Matrix $\mathbf{H}_{A\Sigma}^*$ [15] does not have $\mathbf{E}_G$, and there is no immediate confounding among $\mathbf{u}_2^*$, $\mathbf{g}$, and μ. However, $\mathbf{E}_{A22}$ is a potential source of bias for genetic trends because it creates the direct link between $\mathbf{u}_2^*$ and $\mathbf{g}$, and these 2 effects may not be inseparable as in QP-M.

Encapsulated-UPG Model. Without UPG, when all animals are genotyped, $\mathbf{A}_{22}$ is equal to $\mathbf{A}$, and $\mathbf{H}^{-1}$ reduces to $\mathbf{G}^{-1}$ (i.e., the standard GBLUP). If the same principle applies to the UPG model, when all animals are genotyped, the H-inverse should become $\mathbf{G}^{-1}$. However, this does not occur with either QP-M or AltQP-M. Thus, Masuda et al. (2021) suggested including the UPG equations in $\mathbf{A}_{22}^{-1}$ as follows:

$$\mathbf{H}_{E\Sigma}^* = \mathbf{A}_\Sigma^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^* & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad [16]$$

where

$$\mathbf{A}_{22}^* = \mathbf{A}^{22} - \begin{bmatrix} \mathbf{A}^{21} & \mathbf{A}^{23} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{13} \\ \mathbf{A}^{31} & \mathbf{A}_\Sigma^{33} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}^{12} \\ \mathbf{A}^{32} \end{bmatrix}. \quad [17]$$

Using [16] with all animals genotyped in MME [4], $\hat{\mathbf{g}}$ is canceled out, and only $\mathbf{G}^{-1}$ remains in MME for $\hat{\mathbf{u}}$ (i.e., reducing to GBLUP). Matrix $\mathbf{A}_{22}^*$ [17] can be derived by absorbing all elements other than $\mathbf{A}^{22}$ into $\mathbf{A}^{22}$ in [12]. Note that [17] has a similar structure to [13]. The model with the $\mathbf{H}_{E\Sigma}^*$ matrix will be referred to as the "encapsulated-UPG model" (**EUPG-M**), which is valid for random UPG because $\begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{13} \\ \mathbf{A}^{31} & \mathbf{A}_\Sigma^{33} \end{bmatrix}$ can have an inverse only when $\mathbf{\Sigma}^{-1} \neq 0$. It should be noted that $\mathbf{A}_{22}^*$ [17], like $\mathbf{A}_{22}^{-1}$ [13], can be constructed using only the ancestors of genotyped animals, but $\begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{13} \\ \mathbf{A}^{31} & \mathbf{A}_\Sigma^{33} \end{bmatrix}$ should be recalculated with the subset pedigree.

Masuda et al. (2021), using simulation, showed that EUPG-M could separate the UPG effects from GEBV and that the predicted genetic trends were sufficiently accurate compared with the true genetic trend. The UPG effects in EUPG-M do not have a direct link with GEBV through $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, and the prediction of $\mathbf{g}$ is more independent from UPG than in QP-M and AltQP-M.

Theoretical Justification

Masuda et al. (2021) presented derivations of QP-M, AltQP-M, and EUPG-M. Each model was derived from a density function after adding genomic information to the original density, $p\left(\mathbf{u}, \mathbf{g} \mid \mathbf{A}, \mathbf{\Sigma}, \sigma_u^2\right)$. Differences among these 3 UPG models stem from the specific genomic relationship matrix for $\mathbf{c} = \begin{bmatrix} \mathbf{u}_2^{*\prime} & \mathbf{g}' \end{bmatrix}$ added to the original density.

Specifically, EUPG-M is constructed under the assumption that $\text{var}\left(\mathbf{u}_2^*\right) = \mathbf{G}\sigma_u^2$, as in the standard ssGBLUP (Aguilar et al., 2010). The AltQP-M is derived assuming that the genomic information gives both $\text{var}\left(\mathbf{u}_2^*\right) = \mathbf{G}\sigma_u^2$ and $\text{var}\left(\mathbf{g}\right) = \mathbf{\Sigma}\sigma_u^2$, but $\text{cov}\left(\mathbf{u}_2^*, \mathbf{g}'\right) = 0$. This assumption is questionable because $\mathbf{\Sigma}$ was already specified in the initial density given the pedigree. The QP-M assumes that

$$\text{var}\begin{bmatrix} \mathbf{u}_2^* + \mathbf{Q}_2\mathbf{g} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{G} + \mathbf{Q}_2\mathbf{\Sigma}\mathbf{Q}_2' & \mathbf{Q}_2\mathbf{\Sigma} \\ \mathbf{\Sigma}\mathbf{Q}_2' & \mathbf{\Sigma} \end{bmatrix}\sigma_u^2, \quad [18]$$

which indicates that $\mathbf{G}$, considering its similarity to $\mathbf{A}$ in [6], contains incomplete relationships due to missing pedigree. This assumption may be inappropriate because $\mathbf{G}$ is usually expected to describe the additive genetic relationships among genotyped animals regardless of pedigree completeness (Christensen, 2012; Misztal et al., 2013; Legarra et al., 2015).

The inverse of [18] is the same as [6] replacing $\mathbf{A}^{-1}$ with $\mathbf{G}^{-1}$, and $\mathbf{Q}$ with $\mathbf{Q}_2$. Plieschke et al. (2015) suggested to use this inverse in a GBLUP model to treat UPG as breeds of origin. They split $\mathbf{G}$ into 2 com-

ponents: variation within breeds ($\mathbf{G}_S$) and segregation among breeds ($\mathbf{G}_A$), equivalent to the approach by García-Cortés and Toro (2006). In this approach, $\mathbf{G}_S$ does not contain variation due to UPG (i.e., breeds), and the between-breed variation is restricted to $\mathbf{G}_A$. Makgahlela et al. (2013) suggested similar methods to estimate these 2 genomic relationship matrices. A potential model that included nongenotyped animals would need to replace $\mathbf{G}$ with $\mathbf{G}_S$ and $\boldsymbol{\Sigma}$ with $\mathbf{G}_A$ in [14].

Comparison Among UPG Models

Simulated Populations. Masuda et al. (2021) compared genetic trends for proven bulls and genotyped animals among QP-M, AltQP-M, and EUPG-M in a simulated purebred population where 50% of genotyped females had unidentified dams. The QP-M and AltQP-M yielded potentially biased genetic trends for proven bulls, and they also underestimated UPG predictions because of confounding with GEBV. The EUPG-M accurately predicted UPG and the genetic trend bias for proven bulls was minimal. The genetic trend for females was accurate with all UPG models. Predictability, inflation, and bias of young genomic predictions were comparable among the 3 UPG models.

Tsuruta et al. (2019) compared GEBV between QP-M and AltQP-M using similar simulated data. These authors found that both models slightly overestimated genetic trends for nongenotyped proven bulls, and underestimated genetic trends for genotyped proven bulls, but the bias was limited. However, QP-M yielded lower GEBV predictability and inflation for young-genotyped animals than AltQP-M. They did not report UPG prediction biases. Tsuruta et al. (2019) used 3 times more phenotypes and genotypes than Masuda et al. (2021). The biases in predictions can be removed with larger data sets.

Bradford et al. (2019) compared QP-M and Omega-M using simulated dairy cattle populations with more missing pedigree for animals with smaller breeding values. The QP-M yielded lower GEBV accuracies and greater bias for young-genotyped animals. They did not observe any improvement of AltQP-M over QP-M. However, they mentioned that the amount of information used to estimate fixed UPG effects may not have been enough.

Dairy Cattle Populations. The QP-M has been considered in several studies as a replacement for Omega-M to have better convergence (Matilainen et al., 2016; Bradford et al., 2019; Mäntysaari et al., 2020). Tsuruta et al. (2019) reported that the MME for QP-M converged in US Holstein, but the MME for Omega-M did not. Similarly, Matilainen et al. (2018) found that the MME for Omega-M diverged in Nordic Red dairy cattle, and indicated that QP-M had better convergence and produced more reasonable genetic trends.

The QP-M is likely sensitive to the amount of genomic data or condition of $\mathbf{G}$. Tsuruta et al. (2019), using US Holstein data until 2010, found that QP-M underestimated genetic trends during the last few years, and that AltQP-M also underestimated genetic trends but to a lesser degree. These 2 models gave the same genetic trends with US Holstein data until 2014 (Tsuruta et al., 2019). Masuda et al. (2018a) reported lower predictability with QP-M than with Omega-M, probably because of inaccurate UPG solutions for production traits obtained with US Holstein data until 2011. Later, Masuda et al. (2018b) obtained genetic trends with QP-M using a US Holstein data set with more than 700 thousand genotypes. until 2015. Koivula et al. (2018) used QP-M to calculate GEBV of protein yield for all genotyped and nongenotyped animals in Nordic Red dairy cattle. The genetic trend for old animals estimated with 30,186 genotypes was almost identical to both trends with 21,416 genotypes and without genotypes (i.e., pedigree BLUP). The amount of data required to estimate stable genetic trends in QP-M depends on unbalancedness of pedigrees and diverse UPG (Mäntysaari et al., 2020).

Few researchers have examined AltQP-M. Tsuruta et al. (2019) and Masuda et al. (2019a) observed a reduction in inflation of GEBV with AltQP-M relative to QP-M for type and production traits in US Holstein. Using more recent data for production traits in the same population, AltQP-M showed greater predictive ability and generally less inflation in GEBV for young bulls and heifers than QP-M (Cesarani et al., 2021). These authors also found the predictive ability of AltQP-M to be more stable than that of QP-M at various pedigree depths. Finally, the AltQP-M required fewer iterations than QP-M to solve the MME.

Inbreeding in ssGBLUP

It is crucial to use inbreeding coefficients in the construction of $\mathbf{A}^{-1}$, regardless of whether they are the standard ones or VanRaden's approximated inbreeding, to improve compatibility between the pedigree and genomic relationship matrices. For pedigree BLUP, inbreeding coefficients are often ignored when forming $\mathbf{A}^{-1}$ because of a computational simplicity and no trouble in convergence of iterative solvers. The EBV biases do not seem to be a concern (Mehrabani-Yeganeh et al., 2000). However, for ssGBLUP, neglecting inbreeding when constructing $\mathbf{A}^{-1}$ causes severe biases in

GEBV with UPG as well as computational problems (Matilainen et al., 2016; Tsuruta et al., 2019). If $\mathbf{A}_{22}^{-1}$ is computed indirectly with [13], either standard or estimated inbreeding coefficients should be considered for $\mathbf{A}^{-1}$. Several studies reported that utilization of standard inbreeding coefficients in $\mathbf{A}^{-1}$ decreased the inflation of GEBV and the convergence behavior of iterative solvers (Matilainen et al., 2016; Garcia-Baccino et al., 2017; Masuda et al., 2018b).

Estimated inbreeding coefficients can improve compatibility between pedigree and genomic relationship matrices. Misztal et al. (2017) showed that estimated inbreeding coefficients removed convergence issues and reduced GEBV inflation compared with standard inbreeding coefficients. Tsuruta et al. (2019) reported a similar outcome when using estimated inbreeding coefficients in QP-M and AltQP-M for 18 type traits in US Holstein.

Missing Pedigree and Predictive Ability

The effect of missing pedigree on realized (or validation) accuracy of GEBV could be limited (Bradford et al., 2019; Tsuruta et al., 2019; Masuda et al., 2021). Provided sufficient availability of genotypes, the GEBV of young-genotyped animals can be expressed as direct genomic predictions (**DGP**) that are functions of $\mathbf{G}^{-1}$ and $\mathbf{u}_2$ without UPG (or $\mathbf{u}_2^*$ with UPG; Lourenco et al., 2015). The DGP are similar to genomic BLUP (GBLUP). Their accuracy depends on the number of genotyped animals with reliable GEBV (VanRaden et al., 2009) and the number of independent chromosome segments, which is a function of the effective population size and the length of the genome (Goddard, 2009). If there are many daughter-proven genotyped bulls like in the US dairy population, DGP should be accurate regardless of missing pedigree and choice of UPG model. The GEBV of young-genotyped animals can be indirectly calculated through SNP effects back-solved from $\mathbf{u}_2^*$, $\mathbf{G}^{-1}$, and the genomic marker matrix ($\mathbf{W}$) (Lourenco et al., 2015; Garcia et al., 2020).

An incomplete pedigree will reduce the accuracy of DGP when the number of proven bulls is limited (Tonussi et al., 2017). This reduction in accuracy is also evident in genotyped populations of small size where GEBV are determined by pedigree relationships in addition to genomic relationships (Lourenco et al., 2015).

METAFOUNDERS

Metafounders are groups that serve as proxies for animals in base populations. Like UPG, it is the responsibility of the user to assign MF to unknown parents. The goal of the MF model is to modify $\mathbf{A}$ so that, in theory, it becomes compatible with $\mathbf{G}_{05}$ ($\mathbf{G}$ with AF equal to 0.5), which does not need alignment. The modified relationship matrix accounting for MF ($\mathbf{\Gamma}$) proposed by Legarra et al. (2015) was the same as in VanRaden's UPG model (VanRaden, 1992). Pedigree inbreeding and relationships are entirely accounted for by $\mathbf{\Gamma}$. The genetic and genomic bases are set to a hypothetical population with AF equal to 0.5. Hence, the MF model was presented as a possible solution for missing pedigree issues in ssGBLUP. The similarities and differences between UPG and MF are described in this section.

MF Model

The numerator relationship matrix with MF is designated as $\mathbf{A}_\Gamma$ and its inverse as $\mathbf{A}_\Gamma^{-1}$ in this review. Vector $\mathbf{g}$ in the MF model contains random MF effects instead of random UPG effects. Vector $\mathbf{g}$ follows a normal distribution, $\mathbf{g} \sim N\left(0, \mathbf{\Gamma}\sigma_u^2\right)$, where $\mathbf{\Gamma}$ is a covariance matrix among MF. The vector of additive genetic effects ($\mathbf{u}_\Gamma$) contains genotyped animals, nongenotyped animals, and MF, i.e., $\mathbf{u}_\Gamma' = \left[\mathbf{u}_{\Gamma 1}' \ \mathbf{u}_{\Gamma 2}' \ \mathbf{g}'\right]$, and its distribution is $\mathbf{u}_\Gamma \sim N\left(0, \mathbf{A}_\Gamma\sigma_u^2\right)$, where

$$\mathbf{A}_\Gamma = \begin{bmatrix} \mathbf{A}_{\Gamma 11} & \mathbf{A}_{\Gamma 12} & \mathbf{A}_{\Gamma 1m} \\ \mathbf{A}_{\Gamma 21} & \mathbf{A}_{\Gamma 22} & \mathbf{A}_{\Gamma 2m} \\ \mathbf{A}_{\Gamma m1} & \mathbf{A}_{\Gamma m2} & \mathbf{\Gamma} \end{bmatrix} \text{ and}$$

$$\mathbf{A}_\Gamma^{-1} = \begin{bmatrix} \mathbf{A}_\Gamma^{11} & \mathbf{A}_\Gamma^{12} & \mathbf{A}_\Gamma^{1m} \\ \mathbf{A}_\Gamma^{21} & \mathbf{A}_\Gamma^{22} & \mathbf{A}_\Gamma^{2m} \\ \mathbf{A}_\Gamma^{m1} & \mathbf{A}_\Gamma^{m2} & \mathbf{A}_\Gamma^{mm} + \mathbf{\Gamma}^{-1} \end{bmatrix},$$

with subscript m for MF. The H-inverse with MF is as follows:

$$\mathbf{H}_\Gamma^{-1} = \mathbf{A}_\Gamma^{-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{05}^{-1} - \mathbf{A}_{\Gamma 22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad [19]$$

Legarra et al. (2015) showed that $\mathbf{A}_\Gamma$ could be computed with either the tabular method or with Colleau's method (Colleau, 2002) by tracing animals back to the MF, and the estimated inbreeding could be calculated using a modification of the method of Meuwissen and Luo (1992). Matrix $\mathbf{A}_\Gamma^{-1}$ can be constructed with VanRaden's approach in a manner similar to $\mathbf{A}_\Sigma^*$ by replacing $\mathbf{\Sigma}$ with $\mathbf{\Gamma}$, and treating MF as real animals.

Simplified H-Inverse

The 3 assumptions of the MF model, namely: (1) $\mathbf{A}_{\Gamma 22}$ is compatible with $\mathbf{G}_{05}$ that includes the variation due to MF, (2) $\mathbf{\Gamma}$ is defined before obtaining $\mathbf{G}_{05}$, and (3) genomic information only updates the distribution of $\mathbf{u}_{\Gamma 2}$, are equivalent to those for EUPG-M [17]. Thus, the derivation of $\mathbf{H}_{\Gamma}^{-1}$ is the same as that for EUPG-M (Masuda et al., 2021), and an alternative H-inverse with MF is as follows:

$$\mathbf{H}_{\Gamma}^{*} = \mathbf{A}_{\Gamma}^{-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{05}^{-1} - \mathbf{A}_{\Gamma 22}^{*} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad [20]$$

where $\mathbf{A}_{\Gamma 22}^{*}$ is identical to $\mathbf{A}_{22}^{*}$ [17] except for the replacement of $\mathbf{A}_{\Sigma}^{*}$ with $\mathbf{A}_{\Gamma}^{-1}$. Similar to EUPG-M, $\mathbf{A}_{\Gamma}^{11}$ can be for ancestors of genotyped animals only as in [17] constructed with the subset pedigree. Masuda et al. (2019b) used formula [20] to apply a MF model to a US Holstein data set with 2.3 million genotyped animals.

Masuda et al. (2021) found that EUPG-M gave numerically identical GEBV to those from the MF model in a simulated purebred population where all animals were related in the pedigree. The 2 models are equivalent in practice, although there are theoretical differences between them.

Modeling of Group Effects

There is a critical difference between UPG and MF when defining group effects. In the UPG model, the vector of group effects ($\mathbf{g}$) can be defined as the vector of means of additive genetic animal effects in a base population. However, in the MF model, vector $\mathbf{g}$ is simply a vector of group effects (i.e., MF effects are not additive genetic effects of base animals).

For simplicity, as in Christensen (2012), consider a single MF and let u_b be the additive genetic effect of a base animal. Then, group effect $\mathbf{g}$ reduces to g and $\mathbf{\Gamma}$ to γ; hence, $\text{var}(g) = \gamma\sigma_u^2$. The self-relationship of the base animal is $\text{var}(u_b) = (1 + \gamma/2)\sigma_u^2$ in the base population (Legarra et al., 2015), and the inbreeding coefficient of the base animal is $\gamma/2$, computed as the correlation between gametes, which can be negative (Wright, 1922). The inbreeding coefficient of the MF is $F_{\Gamma} = \gamma - 1 = 2\text{var}(u_b) - 3$. Conversely, in the UPG model, $\mathbf{\Sigma}$ reduces to $a = \text{var}(u_b)$, and the inbreeding coefficient of the base animal is $F_a = a - 1 = \text{var}(u_b) - 1$. The association between the inbreeding coefficients of the 2 models, obtained by equating the 2 expressions using $\text{var}(u_b)$, is as follows:

$$F_{\Gamma} = 2F_a - 1.$$

When 2 MF (1 and 2) are considered, the relationship between MF is γ_{12}, which is equal to the corresponding additive relationship between UPG, Σ_{12}.

The association between F_{Γ} and F_a explains how the UPG and MF models are similar. Element d_{ii} can be obtained using Henderson's rules with MF as follows:

$$d_{ii} = \frac{4}{2 - F_{\Gamma s} - F_{\Gamma d}},$$

where $F_{\Gamma s}$ ($F_{\Gamma d}$) is the inbreeding coefficient of the sire (dam). This formula assumes that all parents are known regardless of whether they are real animals or MF. When an MF's $F_{\Gamma s}$ ($F_{\Gamma d}$) is replaced with a base animal's $2F_a - 1$, the above expression for d_{ii} becomes identical to VanRaden's formula [7] dealing with the case of missing parents.

A negative F_{Γ} represents an excess of heterozygotes (Legarra et al., 2015). An extreme example is $\gamma = 0$ and $F_{\Gamma} = -1$ (equivalently, $a = 1$ and $F_a = 0$); in this case, the algorithms for $\mathbf{A}_{\Gamma}$ and $\mathbf{A}_{\Gamma}^{-1}$ are reduced to the standard methods for $\mathbf{A}$ and $\mathbf{A}^{-1}$. In the Meuwissen and Luo algorithm (Meuwissen and Luo, 1992), unrelated base individuals (coded as 0 in their program) have a -1 value for inbreeding, and they are treated as known parents.

Scale of the Additive Genetic Variance

Assuming a single MF, matrix $\mathbf{A}_{\Gamma}$ is identical to the following alignment formula:

$$\mathbf{A}_{\Gamma} = (1 - \gamma/2)\mathbf{A} + \gamma\mathbf{1}\mathbf{1}'. \qquad [21]$$

However, with multiple MF, the explicit alignment is an approximation (Legarra et al., 2015). In addition to MF, VanRaden et al. (2011) developed a formula similar to (21) to align $\mathbf{A}$ with $\mathbf{G}_{05}$, or $\mathbf{G}$ with base AF, for multiple purebred populations.

Whereas the additive genetic variance is estimated using the standard $\mathbf{A}$, the actual pedigree relationships used in the mixed model are scaled as in $\mathbf{A}_{\Gamma}$ with [21], resulting in an incorrect scale for the additive genetic variance. Therefore, Legarra et al. (2015) suggested a simple adjustment for the additive genetic variance used for multiple MF:

$$\sigma_u^{2*} \approx \frac{\sigma_u^2}{k} \text{ with } k = 1 + \frac{\overline{\text{diag}(\mathbf{\Gamma})}}{2} - \overline{\mathbf{\Gamma}}, \qquad [22]$$

where σ_u^2 is the additive genetic variance assuming unrelated base animals for $\mathbf{A}$, $\sigma_u^{2^*}$ is the additive genetic variance among related base animals for $\mathbf{A}_\Gamma$, $\text{diag}(\mathbf{\Gamma})$ is the average of the diagonals in $\mathbf{\Gamma}$, and $\overline{\mathbf{\Gamma}}$ is the average of all elements in $\mathbf{\Gamma}$. For a single MF, $k = 1 - \gamma/2$.

Estimation of Gamma Parameters

Estimation methods for $\mathbf{\Gamma}$ have been investigated because this parameter needs to be estimated before genomic prediction with MF. A maximum-likelihood method (Christensen, 2012) and summary statistics for $\mathbf{G}$ and $\mathbf{A}_{22}$ (Legarra et al., 2015) resulted in computational issues and inaccurate estimates. Garcia-Baccino et al. (2017) developed approaches to calculate $\mathbf{\Gamma}$ based on the estimated AF (or equivalently, the mean gene content) of each marker in the base populations using $\mathbf{A}_{22}$, $\mathbf{Q}_2$, and marker genotypes (McPeek et al., 2004; Gengler et al., 2007). The matrix $\mathbf{\Gamma}$ is estimated as $2\mathbf{V}\mathbf{V}' = 8\text{var}(\mathbf{P})$, where $\mathbf{V} = \{v_{ij}\}$ is a matrix of estimated mean gene-contents and $\mathbf{P} = \{p_{ij}\}$ is a matrix of estimated AF for group i and marker j (Garcia-Baccino et al., 2017; Mäntysaari et al., 2020). Diagonal elements of $\mathbf{\Gamma}$ must range between 0 and 2, and off-diagonal elements between −1 and 1. Garcia-Baccino's approaches do not guarantee $\mathbf{\Gamma}$ to be in the parameter space.

A weakness is that the base AF may not be robustly calculated because of extremely low allele frequencies and unbalanced assignment of MF across genotypes (Aldridge et al., 2019; Calus and Vandenplas, 2019; Kudinov et al., 2020). A simple solution is to reduce the number of MF as a trade-off for accurate modeling of missing pedigree and breed origins (Calus and Vandenplas, 2019). Other options include truncated pedigree, redefined base populations, and setting a lower limit for AF (Aldridge et al., 2019; Calus and Vandenplas, 2019; Kudinov et al., 2020). If MF are defined by period, a smooth extrapolation for the elements in $\mathbf{\Gamma}$ can be an option (Calus and Vandenplas, 2019; Kudinov et al., 2020).

Although $\mathbf{\Gamma}$ may not be reliably calculated, $\mathbf{\Sigma}$ can be obtained using the association between UPG and MF, $\gamma_{jj} = 2\left(\Sigma_{jj} - 1\right) = 2\hat{F}_j$ and $\gamma_{jk} = \Sigma_{jk}$ for groups j and k. However, this conversion is not perfect because the resulting matrix may not be positive definite.

Application of MF

Metafounders have been tested in crossbred and multibreed domestic animal populations (Christensen et al., 2015; Xiang et al., 2017; van Grevenhof et al., 2019). Christensen et al. (2015) suggested that all animals in a pedigree could be related across breeds through $\mathbf{\Gamma}$ in the MF model. Combining all available records from separate breeds, MF have an advantage over UPG which assume that all base animals, both within and across breeds, are considered unrelated in the pedigree.

Masuda et al. (2019a) reported that the MF model yielded GEBV with similar or slightly better inflation and bias properties as well as comparable genetic trends to QP-M and AltQP-M in US Holstein. Bradford et al. (2019) obtained similar results for purebred data in a simulation study. Kudinov et al. (2020) applied MF to Nordic Red Dairy cattle consisting of multiple breeds and concluded that MF give almost the same validation results and genetic trends as Omega-M for random UPG. Granado-Tajada et al. (2020) indicated that the advantage of using up to 14 MF to predict GEBV for milk yield over pedigree BLUP in 2 sheep populations was not clear because of a limited number of genotyped animals. Macedo et al. (2020) reported that the MF model produced unbiased GEBV, whereas QP-M and Omega-M yielded biased GEBV for milk yield in dairy sheep.

RELATED ISSUES

Additive Genetic Variance

The additive genetic variance in the MF model is rescaled for consistent prediction [22]. Conversely, the additive genetic variance is not usually rescaled when using UPG in genetic evaluation models. If nonzero inbreeding for UPG is assumed, it may be necessary to use the additive genetic variance of a base population that may have undergone selection. However, it is unclear how to estimate variance components and what variance components should be used for genetic evaluation in a UPG model (van der Werf, 1992; Pieramati and Van Vleck, 1993).

Genetic parameters in dairy cattle are often estimated based on a pedigree model with historical data. It is expected that the additive genetic variance can decrease under genomic selection, but the pedigree-based model tends to overestimate the additive genetic variance in the latest generation (Hidalgo et al., 2020). Thus, these genetic parameters may not be suitable for predicting GEBV for young animals. Cross-validation of GEBVs gives an optimal additive genetic variance for prediction if this variance is considered to be a prediction parameter rather than a genetic parameter as in the MF model. Lowering the additive genetic variance (and eventually heritability) is sufficient to reduce the inflation of GEBV for cows (Wiggans et al., 2012b). Misztal et al. (2017) and Tsuruta et al. (2018) found that a halved additive genetic variance did not change

the accuracy but reduced the inflation of GEBV for type traits in young Holstein bulls.

Simplest Solution to Missing Pedigree Issues

If the purpose is to rank young-genotyped animals and a long-term genetic trend is not the primary interest, using data truncation (i.e., excluding data from old generations) to avoid UPG is an option. Bradford et al. (2019) and Tsuruta et al. (2019) reported that a model without UPG gave the same validation accuracy for young animals as UPG models in simulated and field data sets. Lourenco et al. (2014) found that the use of only phenotypes and pedigree in the last few generations resulted in GEBV of the same accuracy and less inflation than when using the full data set. This is because of sufficient information for genomic prediction and better compatibility among relationship matrices in the current generation. Although data truncation does not change GEBV accuracies with enough data, GEBV inflation and bias may remain. Data truncation is also advantageous for faster convergence and shorter computing times (Pocrnic et al., 2017).

Data truncation can create a case where enough genotypes have been collected, and all relevant animals are genotyped. Under these conditions, H-inverse approaches to $\mathbf{G}^{-1}$ and ssGBLUP is close to GBLUP. This situation will become more realistic in the future with the existence of a large-genotyped population where missing pedigree will not be a severe issue.

Single-Step Marker Effect Models

Two typical ssMEM can be considered here: "hybrid models," Bayesian regression models by Fernando et al. (2014, 2016) and "single-step SNP BLUP" (**ssSNPBLUP**) by Liu et al. (2014, 2016) and Mäntysaari and Strandén (2016). With the equivalence between GBLUP and SNP BLUP under some assumptions (Strandén and Garrick, 2009), ssGBLUP and ssSNPBLUP can be equivalent when the markers account for the additive genetic variation perfectly (Liu et al., 2014) and the marker effects follow the multivariate normal distribution (Fernando et al., 2014). The hybrid models are flexible to assume any a priori distribution of each marker effect as marker regression models (Meuwissen et al., 2001; Habier et al., 2011).

In a selected population, the mean GEBV for genotyped animals should be adjusted by μ [8], the difference with the mean GEBV for nongenotyped animals. Whereas Fernando et al. (2014) explicitly considered μ in the hybrid models (noted as "J" in their study), ssSNPBLUP can be extended to have μ (Hsu et al., 2017; Tribout et al., 2019), as demonstrated by Vandenplas et al. (2020).

The ssSNPBLUP models with UPG can be obtained easily. Vandenplas et al. (2021) applied the QP transformation to Liu's model and derived the QP-M-equivalent formula for ssSNPBLUP. The same result is expected by incorporating the $\mathbf{Q}$ elements into the model as covariates. Additionally, Vandenplas et al. (2021) claimed that it is straightforward to use MF in ssSNPBLUP. This statement implies that EUPG-M is also applicable to ssSNPBLUP because of the equivalence to MF.

Large G

The maximum rank of $\mathbf{G} = \mathbf{ZZ}'/s$ is the smaller of the number of genotyped animals (Ng) and the number of markers (Nm), as implied by VanRaden (2008). Genomic data are highly redundant, and the actual rank of $\mathbf{G}$ is smaller than either Ng or Nm (Macciotta et al., 2010; Pocrnic et al., 2016). Given this fact, $\mathbf{G}^{-1}$ does not exist with many genotyped animals. Misztal et al. (2014) developed the so-called APY (algorithm for proven and young) to build the "inverse" of $\mathbf{G}$, say $\mathbf{G}^{-1}_{APY}$, while keeping the additive genetic variation of the original $\mathbf{G}$ matrix. In addition, Mäntysaari et al. (2017) suggested a single-step method using the Woodbury formula to avoid computing $\mathbf{G}^{-1}$. Because both methods only replace $\mathbf{G}^{-1}$ with an equivalent expression, any UPG or MF model can be applied.

However, a question arises. What role does the alignment play if $E(\mathbf{G}) = \mathbf{A}_{22}$ may not occur when $\mathbf{G}$ is large? Although the original matrix computed as the inverse of an APY inverse (i.e., $\left[\mathbf{G}^{-1}_{APY}\right]^{-1}$) may seem incompatible with $\mathbf{G}$ (Strandén et al., 2017), the GEBV are still reasonable. Masuda et al. (2021) showed that a nonaligned $\mathbf{G}$ gave numerically the same GEBV as an aligned $\mathbf{G}$ under $Ng > Nm$ (i.e., the 2 $\mathbf{G}$ matrices gave different μs, but identical $\mathbf{u}_2$). These authors hypothesized that alignment was not necessarily required when $\mathbf{G}$ was (nearly) singular, whereas the other studies reported the alignment was needed to remove the bias in genomic predictions (Legarra et al., 2014).

FINAL REMARKS

In practice, a set of UPG is assigned according to the selection intensity and expected genetic merit of unknown parents. A complex grouping strategy could lead to confounding among group effects or to imprecise UPG solutions (Quaas, 1988; Fikse, 2009). Although a UPG definition creates no problems in pedigree BLUP, the same definition may cause convergence issues and

questionable predictions in ssGBLUP (Tsuruta et al., 2014; Masuda et al., 2019b; Mäntysaari et al., 2020). An optimal definition of UPG is data specific and generally unknown. Thus, users need to consult UPG assignments and ssGBLUP models to obtain unbiased and stable predictions using cross-validation methods. This is also true for the MF model because the definition of MF also relies on the user.

CONCLUSIONS

This review presented and discussed issues related to missing pedigree in ssGBLUP, properties of several UPG models, and how MF are related to UPG. The QP-M has a good convergence behavior, but may produce biased genetic trends and underestimate UPG effects. The AltQP-M produces less bias in genetic trends than QP-M and less inflation of GEBV for young-genotyped animals, especially in large data sets. The EUPG-M incorporates UPG contributions into pedigree relationships for genotyped animals, and it was proposed for purebred populations. The MF model is a comprehensive solution to missing pedigree issues and it is a choice for multibreed or crossbred evaluations if the data set permits the estimation of a reasonable $\boldsymbol{\Gamma}$. Although missing pedigree influences genetic trends, its effect on predictability for genotyped animals should be negligible when many proven bulls are genotyped. In this situation, the indirect-prediction method is useful to predict GEBV for young-genotyped animals with missing parents.

ACKNOWLEDGMENTS

The authors thank Andrés Legarra (INRAE, Castanet-Tolosan, France) for helpful comments on an earlier version of this manuscript. Two anonymous reviewers are thanked for their valuable feedback to improve the quality of the manuscript. The authors gratefully acknowledge the editing of the manuscript by Mauricio Elzo (University of Florida). This study was partially funded by the Agriculture and Food Research Initiative Competitive Grant No. 2020-67015-31030 from the National Institute of Food and Agriculture of the US Department of Agriculture (Beltsville, MD). The authors have not stated any conflicts of interest.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. https://doi.org/10.3168/jds.2009-2730.

Aldridge, M. N., J. Vandenplas, and M. P. L. Calus. 2019. Efficient and accurate computation of base generation allele frequencies. J. Dairy Sci. 102:1364–1373. https://doi.org/10.3168/jds.2018-15264.

Alfonso, L., and J. Estany. 1999. An expression of mixed animal model equations to account for different means and variances in the base population. Genet. Sel. Evol. 31:105. https://doi.org/10.1186/1297-9686-31-2-105.

Alkhoder, H., and Z. Liu. 2021. Application of a single-step SNP BLUP model to conformation traits of German Holsteins. Interbull Bull. 56:30–40.

Bermann, M., D. A. L. Lourenco, and I. Misztal. 2021. Technical note: Automatic scaling in single-step genomic BLUP. J. Dairy Sci. 104:2027–2031. https://doi.org/10.3168/jds.2020-18969.

Bradford, H. L., Y. Masuda, P. M. VanRaden, A. Legarra, and I. Misztal. 2019. Modeling missing pedigree in single-step genomic BLUP. J. Dairy Sci. 102:2336–2346. https://doi.org/10.3168/jds.2018-15434.

Calus, M., and J. Vandenplas. 2019. Computation of many relationships between metafounders replacing phantom parents. Pages 596–596 in Book of Abstracts of the 70th Annual Meeting of the European Federation of Animal Science. Wageningen Academic Publishers.

Cesarani, A., Y. Masuda, S. Tsuruta, E. Nicolazzi, P. M. VanRaden, D. A. L. Lourenco, and I. Misztal. 2021. Genomic predictions for yield traits in US Holsteins with unknown parent groups. J. Dairy Sci. 104:5843–5853. https://doi.org/10.3168/jds.2020-19789.

Chen, C.-Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale. J. Anim. Sci. 89:2673–2679. https://doi.org/10.2527/jas.2010-3555.

Christensen, O. F. 2012. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. Genet. Sel. Evol. 44:37. https://doi.org/10.1186/1297-9686-44-37.

Christensen, O. F., A. Legarra, M. S. Lund, and G. Su. 2015. Genetic evaluation for three-way crossbreeding. Genet. Sel. Evol. 47:98. https://doi.org/10.1186/s12711-015-0177-6.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571. https://doi.org/10.1017/S1751731112000742.

Colleau, J.-J. 2002. An indirect approach to the extensive calculation of relationship coefficients. Genet. Sel. Evol. 34:409. https://doi.org/10.1186/1297-9686-34-4-409.

Emik, L. O., and C. E. Terrill. 1949. Systematic procedures for calculating inbreeding coefficients. J. Hered. 40:51–55. https://doi.org/10.1093/oxfordjournals.jhered.a105986.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics. Pearson Education Limited.

Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96. https://doi.org/10.1186/s12711-016-0273-2.

Fernando, R. L., J. C. M. Dekkers, and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50. https://doi.org/10.1186/1297-9686-46-50.

Fikse, F. 2009. Fuzzy classification of phantom parent groups in an animal model. Genet. Sel. Evol. 41:42. https://doi.org/10.1186/1297-9686-41-42.

Foulley, J. L., E. Hanocq, and D. Boichard. 1992. A criterion for measuring the degree of connectedness in linear models of genetic evaluation. Genet. Sel. Evol. 24:315. https://doi.org/10.1186/1297-9686-24-4-315.

Gao, H., O. F. Christensen, P. Madsen, U. S. Nielsen, Y. Zhang, M. S. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genet. Sel. Evol. 44:8. https://doi.org/10.1186/1297-9686-44-8.

Garcia, A. L., Y. Masuda, S. Tsuruta, S. Miller, I. Misztal, and D. Lourenco. 2020. Indirect predictions with a large number of genotyped animals using the algorithm for proven and young. J. Anim. Sci. 98:skaa154. https://doi.org/10.1093/jas/skaa154.

Garcia-Baccino, C. A., A. Legarra, O. F. Christensen, I. Misztal, I. Pocrnic, Z. G. Vitezica, and R. J. Cantet. 2017. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. Genet. Sel. Evol. 49:34. https://doi.org/10.1186/s12711-017-0309-2.

García-Cortés, L. A., and M. Á. Toro. 2006. Multibreed analysis by splitting the breeding values. Genet. Sel. Evol. 38:601. https://doi.org/10.1186/1297-9686-38-6-601.

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal 1:21–28. https://doi.org/10.1017/S1751731107392628.

Gengler, N., G. Nieuwhof, K. Konstantinov, and M. E. Goddard. 2012. Alternative single-step type genomic prediction equations. Page 131 in Proc. 63rd Annual Meeting of the European Association for Animal Production, Bratislava, Slovakia. Wageningen Academic Publishers.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136:245–257. https://doi.org/10.1007/s10709-008-9308-0.

Granado-Tajada, I., A. Legarra, and E. Ugarte. 2020. Exploring the inclusion of genomic information and metafounders in Latxa dairy sheep genetic evaluations. J. Dairy Sci. 103:6346–6353. https://doi.org/10.3168/jds.2019-18033.

Graser, H.-U., S. P. Smith, and B. Tier. 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. J. Anim. Sci. 64:1362–1370. https://doi.org/10.2527/jas1987.6451362x.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12:186. https://doi.org/10.1186/1471-2105-12-186.

Henderson, C. 1949. Estimation of changes in herd environment. J. Dairy Sci. 32:706.

Henderson, C. 1975. Use of relationships among sires to increase accuracy of sire evaluation. J. Dairy Sci. 58:1731–1738. https://doi.org/10.3168/jds.S0022-0302(75)84777-1.

Henderson, C. R. 1973. Sire evaluation and genetic trends. J. Anim. Sci. 1973(Symposium):10–41. https://doi.org/10.1093/ansci/1973.Symposium.10.

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. https://doi.org/10.2307/2529339.

Henderson, C. R. 1985. Best linear unbiased prediction using relationship matrices derived from selected base populations. J. Dairy Sci. 68:443–448. https://doi.org/10.3168/jds.S0022-0302(85)80843-2.

Henderson, C. R. 1988. A simple method to account for selected base populations. J. Dairy Sci. 71:3399–3404. https://doi.org/10.3168/jds.S0022-0302(88)79945-2.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. Von Krosigk. 1959. The estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192–218. https://doi.org/10.2307/2527669.

Hidalgo, J., S. Tsuruta, D. Lourenco, Y. Masuda, Y. Huang, K. A. Gray, and I. Misztal. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. J. Anim. Sci. 98:skaa032. https://doi.org/10.1093/jas/skaa032.

Hsu, W. L., D. J. Garrick, and R. L. Fernando. 2017. The accuracy and bias of single-step genomic prediction for populations under selection. G3 (Bethesda) 7:2685–2694. https://doi.org/10.1534/g3.117.043596.

Kennedy, B. W. 1991. C. R. Henderson: The unfinished legacy. J. Dairy Sci. 74:4067–4081. https://doi.org/10.3168/jds.S0022-0302(91)78602-5.

Kennedy, B. W., L. R. Schaeffer, and D. A. Sorensen. 1988. Genetic properties of animal models. J. Dairy Sci. 71:17–26. https://doi.org/10.1016/S0022-0302(88)79975-0.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2018. Reducing bias in the dairy cattle single-step genomic evaluation by ignoring bulls without progeny. J. Anim. Breed. Genet. 135:107–115. https://doi.org/10.1111/jbg.12318.

Koivula, M., I. Strandén, J. Pösö, G. P. Aamand, and E. A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775–2784. https://doi.org/10.3168/jds.2014-8975.

Konstantinov, K. V., and M. E. Goddard. 2020. Application of multivariate single-step SNP best linear unbiased predictor model and revised SNP list for genomic evaluation of dairy cattle in Australia. J. Dairy Sci. 103:8305–8316. https://doi.org/10.3168/jds.2020-18242.

Kudinov, A. A., E. A. Mäntysaari, G. P. Aamand, P. Uimari, and I. Strandén. 2020. Metafounder approach for single-step genomic evaluations of Red Dairy cattle. J. Dairy Sci. 103:6299–6310. https://doi.org/10.3168/jds.2019-17483.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. Livest. Sci. 166:54–65. https://doi.org/10.1016/j.livsci.2014.04.029.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. Genetics 200:455–468. https://doi.org/10.1534/genetics.115.177014.

Lidauer, M. H., J. Pösö, J. Pedersen, J. Lassen, P. Madsen, E. A. Mäntysaari, U. S. Nielsen, J.-Å. Eriksson, K. Johansson, T. Pitkänen, I. Strandén, and G. P. Aamand. 2015. Across-country test-day model evaluations for Holstein, Nordic Red Cattle, and Jersey. J. Dairy Sci. 98:1296–1309. https://doi.org/10.3168/jds.2014-8307.

Liu, Z., M. E. Goddard, B. Hayes, F. Reinhardt, and R. Reents. 2016. Technical note: equivalent genomic models with a residual polygenic effect. J. Dairy Sci. 99:2016–2025. https://doi.org/10.3168/jds.2015-10394.

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. J. Dairy Sci. 97:5833–5850. https://doi.org/10.3168/jds.2014-7924.

Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, T. J. Lawlor, S. Forni, and J. I. Weller. 2014. Are evaluations on young genotyped animals benefiting from the past generations? J. Dairy Sci. 97:3930–3942. https://doi.org/10.3168/jds.2013-7769.

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand, T. S. Amen, L. Wang, D. W. Moser, and I. Misztal. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. J. Anim. Sci. 93:2653–2662. https://doi.org/10.2527/jas.2014-8836.

Macciotta, N. P. P., G. Gaspa, R. Steri, E. L. Nicolazzi, C. Dimauro, C. Pieramati, and A. Cappio-Borlino. 2010. Using eigenvalues as variance priors in the prediction of genomic breeding values by principal component analysis. J. Dairy Sci. 93:2765–2774. https://doi.org/10.3168/jds.2009-3029.

Macedo, F. L., O. F. Christensen, J.-M. Astruc, I. Aguilar, Y. Masuda, and A. Legarra. 2020. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. Genet. Sel. Evol. 52:47. https://doi.org/10.1186/s12711-020-00567-1.

Makgahlela, M. L., I. Strandén, U. S. Nielsen, M. J. Sillanpää, and E. A. Mäntysaari. 2013. The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. J. Dairy Sci. 96:5364–5375. https://doi.org/10.3168/jds.2012-6523.

Mäntysaari, E. A., R. Evans, and I. Strandén. 2017. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. J. Anim. Sci. 95:4728–4737. https://doi.org/10.2527/jas2017.1912.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. J. Dairy Sci. 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Mäntysaari, E. A., and I. Strandén. 2016. Single-step genomic evaluation with many more genotyped animals. Page 336 in Proc. 67th Annual Meeting of the European Association for Animal Production, Belfast, Northern Ireland. Wageningen Academic Publishers.

Martini, J. W. R., M. F. Schrauf, C. A. Garcia-Baccino, E. C. Pimentel, S. Munilla, A. Rogberg-Muñoz, R. J. Cantet, C. Reimer, N. Gao, V. Wimmer, and H. Simianer. 2018. The effect of the H^{-1} scaling factors τ and ω on the structure of H in the single-step procedure. Genet. Sel. Evol. 50:16. https://doi.org/10.1186/s12711-018-0386-x.

Masuda, Y., I. Misztal, P. M. VanRaden, and T. J. Lawlor. 2018a. Genomic predictability of single-step GBLUP for production traits in US Holstein. J. Dairy Sci. 101(Suppl. 2):182.

Masuda, Y., S. Tsuruta, A. Legarra, M. Bermann, P. M. VanRaden, H. L. Bradford, and I. Misztal. 2021. Comparison of models for missing pedigree in single-step genomic prediction. J. Anim. Sci. 99:skab019. https://doi.org/10.1093/jas/skab019.

Masuda, Y., S. Tsuruta, E. Nicolazzi, and I. Misztal. 2019a. Single-step GBLUP including more than 2 million genotypes with missing pedigrees for production traits in US Holstein. Accessed Feb. 12, 2021. https://interbull.org/static/web/10_30_Masuda_final.pdf.

Masuda, Y., S. Tsuruta, E. Nicolazzi, and I. Misztal. 2019b. Genomic prediction with missing pedigrees in single-step GBLUP for production traits in US Holstein. Page 593 in 70th Annual Meeting of the European Federation of Animal Science, Ghent, Belgium. Wageningen Academic Publishers.

Masuda, Y., P. M. VanRaden, I. Misztal, and T. J. Lawlor. 2018b. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. J. Dairy Sci. 101:5194–5206. https://doi.org/10.3168/jds.2017-13310.

Matilainen, K., M. Koivula, I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Managing genetic groups in single-step genomic evaluations applied on female fertility traits in Nordic Red Dairy cattle. Interbull Bull. 50:71–75.

Matilainen, K., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2018. Single step genomic evaluation for female fertility in Nordic Red dairy cattle. J. Anim. Breed. Genet. 135:337–348. https://doi.org/10.1111/jbg.12353.

McPeek, M. S., X. Wu, and C. Ober. 2004. Best linear unbiased allele-frequency estimation in complex pedigrees. Biometrics 60:359–367. https://doi.org/10.1111/j.0006-341X.2004.00180.x.

Mehrabani-Yeganeh, H., J. Gibson, and L. Schaeffer. 2000. Including coefficients of inbreeding in BLUP evaluation and its effect on response to selection. J. Anim. Breed. Genet. 117:145–151. https://doi.org/10.1046/j.1439-0388.2000.00241.x.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Meuwissen, T. H. E., T. Luan, and J. A. Woolliams. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J. Anim. Breed. Genet. 128:429–439. https://doi.org/10.1111/j.1439-0388.2011.00966.x.

Meuwissen, T. H. E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24:305. https://doi.org/10.1186/1297-9686-24-4-305.

Meyer, K., B. Tier, and A. Swan. 2018. Estimates of genetic trend for single-step genomic evaluations. Genet. Sel. Evol. 50:39. https://doi.org/10.1186/s12711-018-0410-1.

Misztal, I., H. L. Bradford, D. A. L. Lourenco, S. Tsuruta, Y. Masuda, A. Legarra, and T. J. Lawlor. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bull. 51:38–42.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. https://doi.org/10.3168/jds.2009-2064.

Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. J. Dairy Sci. 97:3943–3952. https://doi.org/10.3168/jds.2013-7752.

Misztal, I., D. A. L. Lourenco, and A. L. Legarra. 2020. Current status of genomic evaluation. J. Anim. Sci. 98:skaa101. https://doi.org/10.1093/jas/skaa101.

Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet. 130:252–258. https://doi.org/10.1111/jbg.12025.

Nani, J. P., L. R. Bacheller, J. B. Cole, and P. M. VanRaden. 2020. Discovering ancestors and connecting relatives in large genomic databases. J. Dairy Sci. 103:1729–1734. https://doi.org/10.3168/jds.2019-17580.

Oliveira, H., D. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. Brito, F. Silva, and F. Schenkel. 2019. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. J. Dairy Sci. 102:2365–2377. https://doi.org/10.3168/jds.2018-15466.

Pieramati, C., and L. D. Van Vleck. 1993. Effect of genetic groups on estimates of additive genetic variance. J. Anim. Sci. 71:66–70. https://doi.org/10.2527/1993.71166x.

Plieschke, L., C. Edel, E. C. G. Pimentel, R. Emmerling, J. Bennewitz, and K.-U. Götz. 2015. A simple method to separate base population and segregation effects in genomic relationship matrices. Genet. Sel. Evol. 47:53. https://doi.org/10.1186/s12711-015-0130-8.

Pocrnic, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203:573–581. https://doi.org/10.1534/genetics.116.187013.

Pocrnic, I., D. A. L. Lourenco, H. L. Bradford, C. Y. Chen, and I. Misztal. 2017. Impact of pedigree depth on convergence of single-step genomic BLUP in a purebred swine population. J. Anim. Sci. 95:3391–3395. https://doi.org/10.2527/jas2017.1581.

Pollak, E., and R. Quaas. 1983. Definition of group effects in sire evaluation models. J. Dairy Sci. 66:1503–1509. https://doi.org/10.3168/jds.S0022-0302(83)81965-1.

Powell, R., and A. Freeman. 1974. Genetic trend estimators. J. Dairy Sci. 57:1067–1075. https://doi.org/10.3168/jds.S0022-0302(74)85010-1.

Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:91–98. https://doi.org/10.1016/S0022-0302(88)79986-5.

Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. J. Dairy Sci. 64:1868–1872. https://doi.org/10.3168/jds.S0022-0302(81)82778-6.

Robinson, G. K. 1986. Group effects and computing strategies for models for estimating breeding values. J. Dairy Sci. 69:3106–3111. https://doi.org/10.3168/jds.S0022-0302(86)80774-3.

Schaeffer, L. R. 1991. C. R. Henderson: Contributions to predicting genetic merit. J. Dairy Sci. 74:4052–4066. https://doi.org/10.3168/jds.S0022-0302(91)78601-3.

Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77:2671–2678. https://doi.org/10.3168/jds.S0022-0302(94)77209-X.

Searle, S. R. 1982. Matrix Algebra Useful for Statistics. John Wiley and Sons.

Strandén, I., and O. F. Christensen. 2011. Allele coding in genomic evaluation. Genet. Sel. Evol. 43:25. https://doi.org/10.1186/1297-9686-43-25.

Strandén, I., and D. J. Garrick. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. J. Dairy Sci. 92:2971–2975. https://doi.org/10.3168/jds.2008-1929.

Strandén, I., K. Matilainen, G. Aamand, and E. Mäntysaari. 2017. Solving efficiently large single-step genomic best linear unbiased prediction models. J. Anim. Breed. Genet. 134:264–274. https://doi.org/10.1111/jbg.12257.

Sullivan, P. G., and L. R. Schaeffer. 1994. Fixed versus random genetic groups. Pages 483–486 in Proc 6th World Congr. Genet. Appl. to Livest. Prod. Guelph. Accessed Oct. 26, 2021. http://www.wcgalp.org/proceedings/1994/fixed-versus-random-genetic-groups.

Taskinen, M., E. A. Mäntysaari, and I. Strandén. 2017. Single-step SNP-BLUP with on-the-fly imputed genotypes and residual poly-

genic effects. Genet. Sel. Evol. 49:36. https://doi.org/10.1186/s12711-017-0310-9.
Thompson, R. 1979. Sire evaluation. Biometrics 35:339–353. https://doi.org/10.2307/2529955.
Tonussi, R. L., R. M. O. Silva, A. F. B. Magalhães, R. Espigolan, E. Peripolli, B. F. Olivieri, F. L. B. Feitosa, M. V. A. Lemos, M. P. Berton, H. L. J. Chiaia, A. S. C. Pereira, R. B. Lôbo, L. A. F. Bezerra, C. U. Magnabosco, D. A. L. Lourenço, I. Aguilar, and F. Baldi. 2017. Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data. PLoS One 12:e0181752. https://doi.org/10.1371/journal.pone.0181752.
Tribout, T., D. Boichard, V. Ducrocq, and J. Vandenplas. 2019. A fast method to fit the mean of unselected base animals in single-step SNP-BLUP. Page 211 in Proc. 70th Annual Meeting of the European Association for Animal Production, Ghent, Belgium. Wageningen Academic Publishers.
Tsuruta, S., D. Lourenco, I. Misztal, and T. Lawlor. 2018. Possible causes of inflation in genomic evaluations for dairy cattle. Page 342 in Proc. 10th World Congr. Genet. Appl. to Livest. Prod. Auckland. Accessed Oct. 26, 2021. http://www.wcgalp.org/system/files/proceedings/2018/possible-causes-inflation-genomic-evaluations-dairy-cattle.pdf.
Tsuruta, S., D. A. L. Lourenco, Y. Masuda, I. Misztal, and T. J. Lawlor. 2019. Controlling bias in genomic breeding values for young genotyped bulls. J. Dairy Sci. 102:9956–9970. https://doi.org/10.3168/jds.2019-16789.
Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204. https://doi.org/10.3168/jds.2011-4256.
Tsuruta, S., I. Misztal, D. A. L. Lourenco, and T. J. Lawlor. 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. J. Dairy Sci. 97:5814–5821. https://doi.org/10.3168/jds.2013-7821.
van der Werf, J. H. J. 1992. Restricted maximum likelihood estimation of additive genetic variance when selected base animals are considered fixed. J. Anim. Sci. 70:1068–1076. https://doi.org/10.2527/1992.7041068x.
van der Werf, J. H. J., and R. Thompson. 1992. Variance decomposition in the estimation of genetic variance with selected data. J. Anim. Sci. 70:2975–2985. https://doi.org/10.2527/1992.70102975x.
van Grevenhof, E. M., J. Vandenplas, and M. P. Calus. 2019. Genomic prediction for crossbred performance using metafounders. J. Anim. Sci. 97:548–558. https://doi.org/10.1093/jas/sky433.
Vandenplas, J., H. Eding, M. Bosmans, and M. P. Calus. 2020. Computational strategies for the preconditioned conjugate gradient method applied to ssSNPBLUP, with an application to a multivariate maternal model. Genet. Sel. Evol. 52:24. https://doi.org/10.1186/s12711-020-00543-9.
Vandenplas, J., H. Eding, and M. P. Calus. 2021. Genetic groups in single-step single nucleotide polymorphism best linear unbiased predictor. J. Dairy Sci. 104:3298–3303. https://doi.org/10.3168/jds.2020-19460.
Vandenplas, J., H. Eding, M. P. L. Calus, and C. Vuik. 2018. Defated preconditioned conjugate gradient method for solving single-step BLUP models efficiently. Genet. Sel. Evol. 50:51. https://doi.org/10.1186/s12711-018-0429-3.
VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J. Dairy Sci. 75:3136–3144. https://doi.org/10.3168/jds.S0022-0302(92)78077-1.
VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.
VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94:5673–5682. https://doi.org/10.3168/jds.2011-4500.
VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. https://doi.org/10.3168/jds.2008-1514.
Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb.) 93:357–366. https://doi.org/10.1017/S001667231100022X.
Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71:1310–1318. https://doi.org/10.3168/jds.S0022-0302(88)79688-5.
Westell, R. A., and L. D. Van Vleck. 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. J. Dairy Sci. 70:1006–1017. https://doi.org/10.3168/jds.S0022-0302(87)80106-6.
Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic selection in dairy cattle: The USDA experience. Annu. Rev. Anim. Biosci. 5:309–327. https://doi.org/10.1146/annurev-animal-021815-111422.
Wiggans, G. R., T. A. Cooper, P. M. VanRaden, K. M. Olson, and M. E. Tooker. 2012a. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. J. Dairy Sci. 95:1552–1558. https://doi.org/10.3168/jds.2011-4985.
Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. J. Dairy Sci. 71:1319–1329. https://doi.org/10.3168/jds.S0022-0302(88)79689-7.
Wiggans, G. R., P. M. VanRaden, and L. R. Bacheller. 2018. Methods for discovering and validating relationships among genotyped animals. Interbull Bull. 53:27–30.
Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2012b. Adjustment of all cow evaluations for yield traits to be comparable with bull evaluations. J. Dairy Sci. 95:3444–3447. https://doi.org/10.3168/jds.2011-5000.
Wolak, M. E., and J. M. Reid. 2017. Accounting for genetic differences among unknown parents in microevolutionary studies: How to include genetic groups in quantitative genetic animal models. J. Anim. Ecol. 86:7–20. https://doi.org/10.1111/1365-2656.12597.
Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338. https://doi.org/10.1086/279872.
Xiang, T., O. F. Christensen, and A. Legarra. 2017. Genomic evaluation for crossbred performance in a single-step approach with metafounders. J. Anim. Sci. 95:1472–1480. https://doi.org/10.2527/jas.2016.1155.

ORCIDS

Yutaka Masuda https://orcid.org/0000-0002-3428-6284
Paul M. VanRaden https://orcid.org/0000-0002-9123-7278
Shogo Tsuruta https://orcid.org/0000-0002-6897-6363
Daniela A. L. Lourenco https://orcid.org/0000-0003-3140-1002
Ignacy Misztal https://orcid.org/0000-0002-0382-1897