



Modeling missing pedigree in single-step genomic BLUP

H. L. Bradford,^{1,2*} Y. Masuda,³ P. M. VanRaden,² A. Legarra,⁴ and I. Misztal³

¹Department of Animal and Poultry Science, Virginia Tech, Blacksburg 24061

²Animal Genomics and Improvement Laboratory, Agriculture Research Service, USDA, Beltsville, MD 20705-2350

³Department of Animal and Dairy Science, University of Georgia, Athens 30605

⁴GenPhySE, INRA, INPT, ENVT, Université de Toulouse, Castanet-Tolosan 31520, France

ABSTRACT

The objective was to compare methods of modeling missing pedigree in single-step genomic BLUP (ssGBLUP). Options for modeling missing pedigree included ignoring the missing pedigree, unknown parent groups (UPG) based on **A** (the numerator relationship matrix) or **H** (the unified pedigree and genomic relationship matrix), and metafounders. The assumptions for the distribution of estimated breeding values changed with the different models. We simulated data with heritabilities of 0.3 and 0.1 for dairy cattle populations that had more missing pedigrees for animals of lesser genetic merit. Predictions for the youngest generation and UPG solutions were compared with the true values for validation. For both traits, ssGBLUP with metafounders provided accurate and unbiased predictions for young animals while also appropriately accounting for genetic trend. Accuracy was least and bias was greatest for ssGBLUP with UPG for **H** for the trait with heritability of 0.3 and with UPG for **A** for the trait with heritability of 0.1. For the trait with heritability of 0.1 and UPG for **H**, the UPG accuracy (SD) was -0.49 (0.12), suggesting poor estimates of genetic trend despite having little bias for validations on young, genotyped animals. Problems with UPG estimates were likely caused by the lesser amount of information available for the lower heritability trait. Hence, UPG need to be defined differently based on the trait and amount of information. More research is needed to investigate accounting for UPG in **A**₂₂ to better account for missing pedigrees for genotyped animals.

Key words: genetic group, metafounder, single-step genomic BLUP (ssGBLUP), unknown parent group

INTRODUCTION

Unknown parent groups (UPG) were developed to account for missing pedigrees in a population (Thompson, 1979; Quaas, 1988). Ignoring these missing pedigrees in genetic evaluations required assuming all the missing parents shared the same average breeding value. Hence, UPG were developed to model differences in genetic merit across categories of missing parents such as sex, birth year, selection path, country, and so on (Quaas, 1988; Westell et al., 1988). Although UPG were originally developed to account for missing parents of one population, UPG can also account for breed differences (Legarra et al., 2007), and appropriate UPG definitions become even more important in multi-breed evaluations (VanRaden et al., 2007). With the implementation of genomic selection, UPG need to be incorporated into single-step genomic models.

The first genomic implementation simply used the original pedigree-based UPG in single-step genomic BLUP (ssGBLUP; Tsuruta et al., 2014). The theory was then extended to include UPG with the combined pedigree and genomic relationship matrix (Misztal et al., 2013). These approaches had different assumptions for UPG effects, and the practical implications of the different approaches were undocumented. Each model has been implemented individually, as UPG for **A** (the numerator relationship matrix; Tsuruta et al., 2014) and **H** (the unified pedigree and genomic relationship matrix; Johnston et al., 2018; Masuda et al., 2018) with limited comparisons between models. Legarra et al. (2015) suggested another approach to model missing pedigrees called “metafounders” that used genotypes to relax the common assumptions for UPG.

Researchers have applied ssGBLUP to the Nordic dairy cattle evaluations for many traits. Although analyses with UPG for **A** worked well for medium-heritability production traits (Koivula et al., 2015), models with UPG for **A** converged extremely slowly for 11 low-heritability fertility traits, whereas models with UPG for **H** converged well and produced reasonable predictions (Matilainen et al., 2016). The Nordic

Received July 24, 2018.

Accepted November 12, 2018.

*Corresponding author: hbradford@vt.edu

dairy data included different breeds and relatively few incomplete pedigrees. Misztal et al. (2017) looked at optimization of ssGBLUP for type traits, which have nearly complete pedigrees. The greatest reliability and least bias were obtained with UPG for **H**, while also accounting for inbreeding for unknown parents (VanRaden, 1992) in the construction of **H**. Masuda et al. (2018) reported lower reliability for Holstein production data using UPG for **H** than for **A** and for no UPG. Their study used 764,000 genotyped animals with many animals with incomplete pedigrees.

Thus, our understanding of how to model UPG in a genomic evaluation context is based on very few studies that are incomplete and sometimes contradictory. The objectives of this study were to understand the benefits and limitations of different approaches for modeling missing pedigree in ssGBLUP for a single breed. Simulation was used to compare predictions from different models to true values for UPG effects and validation animals.

MATERIALS AND METHODS

Animal care and use committee approval was not needed as all data were simulated.

Models

We modeled data as follows:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{Q}\mathbf{s} + \mathbf{e},$$

where **y** was the vector of simulated phenotypes, μ was the fixed mean, **Z** was an incidence matrix relating animals to observations, **a** was the vector of random additive genetic effects, **Q** was the matrix containing UPG compositions for all individuals, **s** was the vector of UPG effects, and **e** was the vector of random residuals. Residuals were assumed distributed as $N(\mathbf{0}, \mathbf{I}\sigma_e^2)$. Assumptions for the distribution of **a** changed with the different models. The **Qs** term was not included in models without UPG, and EBV were equivalent to **a** in these models. For models with UPG, the EBV were

$$\mathbf{u} = \mathbf{Q}\mathbf{s} + \mathbf{a},$$

where **u** was the total EBV including UPG effects. All analyses included inbreeding in the calculation of numerator relationship matrices (VanRaden, 1992). Data were modeled with and without UPG, with and without genomics, and with different assumptions regarding the genomic implementation for UPG. All data were

analyzed in BLUPF90 (Misztal et al., 2018) and our own software for the metafounders analysis.

BLUP

We implemented BLUP with and without UPG. The mixed model equations with no UPG were

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{1} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix},$$

where $\hat{\mu}$ was the estimated overall mean, **A** was the numerator relationship matrix, and λ was the ratio of residual to genetic variance. The variance components were assumed the same as those used to simulate the data. In this model, EBV were computed directly as $\hat{\mathbf{u}}$. The mixed model equations for BLUP with QP-transformed UPG were (Quaas, 1988)

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{1} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} & -\lambda\mathbf{A}^{-1}\mathbf{Q} \\ \mathbf{0} & -\lambda\mathbf{Q}'\mathbf{A}^{-1} & \lambda\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix},$$

where $\hat{\mathbf{s}}$ = estimates for UPG solutions. The inverse covariance matrix,

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix},$$

can be built directly using the rules in Quaas (1988).

ssGBLUP

The first genomic model was the classic implementation of ssGBLUP (Aguilar et al., 2010, Christensen and Lund, 2010). With this model, no UPG were modeled, and the mixed model equations were

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{1} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

with

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

where \mathbf{u}_1 was the vector of breeding values ordered nongenotyped followed by genotyped animals and

hereafter defined as \mathbf{u} ; \mathbf{G} was the genomic relationship matrix (VanRaden, 2008); and \mathbf{A}_{22} was the partition of the numerator relationship matrix for genotyped animals. \mathbf{G} was blended as $\mathbf{G} = 0.99\mathbf{G}_{\text{original}} + 0.01\mathbf{I}$, where $\mathbf{G}_{\text{original}}$ was computed by the first method of VanRaden (2008) assuming 0.5 allelic frequencies. No tuning was used, meaning \mathbf{G} was not scaled to match \mathbf{A}_{22} (Chen et al., 2011; Forni et al., 2011; Vitezica et al., 2011; Christensen et al., 2012). These assumptions were made to construct \mathbf{G} in a similar manner as metafounders. Results with default blending ($\mathbf{G} = 0.95\mathbf{G}_{\text{original}} + 0.05\mathbf{A}_{22}$) and default tuning (as in Christensen et al., 2012) in BLUPF90 (Misztal et al., 2018) were very similar (unpublished data).

ssGBLUP with UPG for \mathbf{A}^{-1}

The next development in modeling UPG was to incorporate the original pedigree-based UPG into the ssGBLUP mixed model equations. The left-hand-side of the QP-transformed coefficient matrix of the mixed model equations, ignoring fixed effects, was

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 & \mathbf{Z}'_1\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & -\mathbf{A}^1\mathbf{Q} \\ \mathbf{A}^{21} & \mathbf{A}^{22} & -\mathbf{A}^2\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & -\mathbf{Q}'\mathbf{A}^{-2} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where

$$\begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & -\mathbf{A}^1\mathbf{Q} \\ \mathbf{A}^{21} & \mathbf{A}^{22} & -\mathbf{A}^2\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & -\mathbf{Q}'\mathbf{A}^{-2} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix} = \mathbf{A}^*$$

above, 1 denotes nongenotyped animals, and 2 denotes genotyped animals. The UPG solutions in the last row were functions of pedigree relationships but not corrected for genomic or pedigree relationships among genotyped animals.

ssGBLUP with UPG for \mathbf{H}^{-1}

The next model was proposed by Misztal et al. (2013) by extending the Quaas (1988) model for UPG to ssGBLUP. With this model, the covariance matrix \mathbf{A} was replaced by \mathbf{H} for UPG solutions in ssGBLUP.

The left-hand-side of the QP-transformed coefficient matrix, ignoring fixed effects, was

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 & \mathbf{Z}'_1\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & -\mathbf{A}^1\mathbf{Q} \\ \mathbf{A}^{21} & \mathbf{A}^{22} & -\mathbf{A}^2\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & -\mathbf{Q}'\mathbf{A}^{-2} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & -\mathbf{Q}'_2(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & \mathbf{Q}'_2(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix},$$

where \mathbf{Q}_2 was the matrix containing UPG compositions for genotyped animals. In this model, UPG solutions depend on \mathbf{A}^{-1} , \mathbf{A}_{22}^{-1} , and \mathbf{G}^{-1} in a similar manner as EBV solutions. This model was equivalent to explicitly modeling UPG as covariates in ssGBLUP. One potential issue with this implementation is the construction of \mathbf{A}_{22}^{-1} by ignoring UPG.

ssGBLUP with Metafounders

An alternative approach was to model missing pedigrees with metafounders (Legarra et al., 2015). Additional random founder animals (metafounders) modeled the missing pedigrees, and these metafounders could have nonzero inbreeding and relationships to each other. The relationship matrix among metafounders, $\mathbf{\Gamma}$, was a function of genetic similarity across populations [$\Gamma_{ij} = 8\text{cov}(p^i, p^j)$, where p^i represents the allelic frequencies in metafounder population i] and was estimated using generalized least squares, which was previously shown to be an accurate estimator (Garcia-Baccino et al., 2017). The $\mathbf{\Gamma}$ was used to calculate $\mathbf{A}(\mathbf{\Gamma})^{-1}$ which is the inverse of the numerator relationship matrix in the context of metafounders, as described in Legarra et al. (2015). First, inbreeding for all animals was computed starting from $\mathbf{\Gamma}$. This approach resulted in inbreeding greater than or equal to the inbreeding without metafounders because the relationships between metafounders were positive and metafounders had nonzero inbreeding. The $\mathbf{A}(\mathbf{\Gamma})^{-1}$ was initialized by adding $\mathbf{\Gamma}^{-1}$ to the metafounder elements. Then, the rest of $\mathbf{A}(\mathbf{\Gamma})^{-1}$ was created by Henderson (1976) rules. Finally, $\mathbf{A}(\mathbf{\Gamma})^{-1}$ was multiplied by $1 - \frac{\text{diag}(\mathbf{\Gamma})}{2} + \bar{\Gamma}$, where $\text{diag}(\mathbf{\Gamma})$ was the mean diagonal element of $\mathbf{\Gamma}$, and $\bar{\Gamma}$ was the mean element in $\mathbf{\Gamma}$, to fit the genetic variance correctly. The left-hand-side of the mixed model equations was

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{1} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{H}(\Gamma)^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{m}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix},$$

where $\hat{\mathbf{m}}$ were the metafounder solutions, \mathbf{Z} was enlarged to include 3 metafounders, and

$$\mathbf{H}(\Gamma)^{-1} = \begin{bmatrix} \mathbf{A}(\Gamma)^{11} & \mathbf{A}(\Gamma)^{12} & \mathbf{A}(\Gamma)^{13} \\ \mathbf{A}(\Gamma)^{21} & \mathbf{A}(\Gamma)^{22} & \mathbf{A}(\Gamma)^{23} \\ \mathbf{A}(\Gamma)^{31} & \mathbf{A}(\Gamma)^{32} & \mathbf{A}(\Gamma)^{33} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}(\Gamma)_{22}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where 1 denotes nongenotyped individuals, 2 denotes genotyped individuals, and 3 denotes metafounders.

Simulation Population. Data were simulated using QMSim v1 (Sargolzaei and Schenkel, 2009) for 20 replicates, and the parameter file is available in Supplemental File S1 (<https://doi.org/10.3168/jds.2018-15434>). We simulated 2 sex-limited traits in separate simulations, one with heritability of 0.3 and one with heritability of 0.1. The historical population began with 5,000 individuals, decreased gradually to 250 at generation 1,000, and increased gradually to 30,000 at generation 1,100. Fluctuations in population size were used to generate linkage disequilibrium consistent with a dairy cattle population. The last historical generation had 350 males and 29,650 females. From this population, 50 males and 14,950 females were randomly selected to be founders of the recent population. The replacement rate was 30% with random mating for 10 generations. Each female had 1 offspring per generation, with an equal probability of male or female offspring. Individuals were selected based on EBV calculated from animal-model BLUP within QMSim (Sargolzaei and Schenkel, 2009). All females were phenotyped except the last generation, which was used for validation. The numbers of animals and phenotypes were reported in Table 1. For the trait with heritability of 0.3, genetic trend increased by 4.42 (0.25) genetic SD (SD of replicates) over 10

generations, and for the trait with heritability of 0.1, genetic trend increased by 3.59 (0.33) genetic SD (SD of replicates).

Genome. The genome mimicked that of dairy cattle with 29 autosomes with a genome length of 2,319 cM. The traits were entirely controlled by 500 biallelic QTL that were randomly spaced throughout the genome. The QTL effects were drawn from a gamma distribution (shape = 0.4, scaled internally for genetic variance of 0.3 or 0.1). There were 50,000 biallelic SNP equally spaced throughout the genome. Both SNP and QTL had 0.5 allelic frequencies to start the historical population, 2.5×10^{-5} mutations per meiosis per locus, and 1 crossover per Morgan per meiosis. The SNP had mean (SD) pooled squared correlation coefficient per chromosome of 0.30 (0.02) in the last generation based on internal calculations in QMSim.

All individuals in the most recent 5 generations had the opportunity to be genotyped. Within each generation, individuals were ranked in the top or bottom half of each sex based on BLUP EBV. The top half of males (females) had a 0.35 (0.35) probability of being genotyped, and the bottom half of males (females) had a 0.10 (0.20) probability of being genotyped. Genotyped individuals were randomly selected with the constraint that all sires within those generations were genotyped. In total, 18,685 individuals were genotyped in each replicate.

Missing Pedigrees. After simulating the entire data in QMSim, we generated missing pedigrees based on sex, genotype status, and genetic merit. Genetic merit was evaluated by identifying the top half and bottom half of males by comparing the males' EBV to the average male EBV in a given generation. The EBV were calculated from animal-model BLUP because animals were selected based on BLUP EBV using full data and pedigree. The same process was done for females independently of the males. The simulated probabilities for missing dams were presented in Table 2, and the numbers of animals with unknown dams were presented in Table 3.

Table 1. Descriptive statistics for simulated data

Item	Males		Females		Total	
	n	SD	n	SD	n	SD
$h^2 = 0.3$						
Pedigree	89,718	190	74,782	190	164,500	0
Genotyped	10,282	44	8,404	36	18,686	8
Phenotypes	0	0	82,247	159	82,247	159
$h^2 = 0.1$						
Pedigree	89,659	237	74,841	237	164,500	0
Genotyped	10,272	41	8,412	35	18,684	8
Phenotypes	0	0	82,177	212	82,177	212

Table 2. Simulated probabilities for missing dams based on sex, genotype status, and genetic merit (top vs. bottom half of a sex within each generation based on EBV)

Type	Males	Females
Top, genotyped	0.00	0.05
Top, nongenotyped	0.00	0.05
Bottom, genotyped	0.10	0.10
Bottom, nongenotyped	0.30	0.30

Validations

EBV. The last generation of the recent population was used for validation. These animals had no phenotypes and some had genotypes. All genotyped young animals were used in validations ($5,042 \pm 10$). All EBV were standardized so that the founders of the most recent population had a mean of 0. Accuracy was measured as the correlation between true breeding value (TBV) and EBV. The dispersion was measured as the regression coefficient from regressing TBV on EBV. Bias was evaluated for all genotyped young animals with a missing dam (157 ± 2) as TBV minus EBV and scaled by the genetic SD. Bias was assessed this way to directly observe the impact of different models for missing pedigree on EBV predictions. The number of rounds to convergence in preconditioned conjugate gradient was reported based on a convergence criterion of 10^{-14} (Tsuruta et al., 2001). Analyses were compared based on the heritability, type of model for UPG, and complete versus missing pedigrees.

UPG. Unknown parent groups were defined as generations 0 to 4 (**UPG₁**), 5 to 7 (**UPG₂**), and 8 to 10 (**UPG₃**). The numbers of times each UPG was present in the pedigree and was the parent of a phenotyped animal is presented in Table 4. True UPG effects were calculated as the weighted average TBV for the missing parents within each UPG definition. The accuracy of UPG solutions was the correlation between true and estimated values. Bias was assessed for estimable functions (contrasts) of UPG, defined as $UPG_2 - UPG_1$, $UPG_3 - UPG_1$, and $UPG_3 - UPG_2$. Bias was the true

value of the estimable function minus estimated and scaled by the genetic SD.

RESULTS AND DISCUSSION

The $\mathbf{\Gamma}$ estimates for the trait with heritabilities of 0.3 and 0.1 were, respectively,

$$\mathbf{\Gamma} = \begin{bmatrix} 0.54 & 0.55 & 0.55 \\ 0.55 & 0.59 & 0.61 \\ 0.55 & 0.61 & 0.66 \end{bmatrix}$$

and

$$\mathbf{\Gamma} = \begin{bmatrix} 0.54 & 0.54 & 0.54 \\ 0.54 & 0.59 & 0.60 \\ 0.54 & 0.60 & 0.65 \end{bmatrix},$$

and were estimated based on the first replicate and used for all replicates. The $\mathbf{\Gamma}$ were relationship matrices among the 3 metafounders, defined the same way as the 3 UPG. The $\mathbf{\Gamma}$ were nearly identical between the heritabilities because $\mathbf{\Gamma}$ was dependent not on phenotype but on variation in pedigrees and genotypes. The $\mathbf{\Gamma}$ indicated strong correlations between metafounders, which made sense because metafounders were proxies for animals that actually belong to the same population. These correlations were ignored when using UPG to model missing pedigree and were much stronger than those estimated between swine breeds (Xiang et al., 2017) but similar to those between dairy breeds (Legarra et al., 2015). Had $\mathbf{\Gamma}$ been a 0 matrix, the model would be equivalent to having UPG for \mathbf{A} . Had $\mathbf{\Gamma}$ been an identity matrix, there would be no equivalence with other models because metafounders would still create greater inbreeding and different variances in the Mendelian samplings used in the setup of \mathbf{A}^{-1} .

We calculated the true value of $\mathbf{\Gamma}$ based on the allelic frequencies by generation for the 50,000 SNP using all simulated animals. For each metafounder, the allelic

Table 3. Descriptive statistics for animals with simulated missing dams based on sex, genotype status, and genetic merit (top vs. bottom half of a sex within each generation based on EBV)

Type	$h^2 = 0.3$		$h^2 = 0.1$	
	n	SD	n	SD
Top female, nongenotyped	368	18	378	8
Bottom female, genotyped	10,089	90	10,066	75
Bottom female, nongenotyped	183	11	184	13
Bottom male, genotyped	10,667	91	10,688	78
Bottom male, nongenotyped	1,548	42	1,541	32

Table 4. Numbers (SD) of animals with unknown parent groups (UPG) as parents for the entire pedigree and for phenotyped animals

UPG ¹	Heritability = 0.3		Heritability = 0.1	
	Pedigree	Phenotype	Pedigree	Phenotype
1	39,723 (66)	20,169 (51)	39,721 (80)	20,181 (63)
2	6,873 (66)	3,677 (64)	6,911 (73)	3,684 (62)
3	6,260 (67)	2,151 (53)	6,225 (59)	2,126 (35)

¹UPG were defined as follows: 1 = generations 0 to 4, 2 = generations 5 to 7, 3 = generations 8 to 10.

frequencies were calculated based on the mean allelic frequency for the generations included in a given meta-founder. The $\mathbf{\Gamma}$ was 8 times the covariance of allelic frequencies between 2 metafounders. The matrices corresponding to the true $\mathbf{\Gamma}$ for the trait with heritabilities of 0.3 and 0.1, respectively, were

$$\begin{bmatrix} 0.54 & 0.56 & 0.55 \\ 0.56 & 0.63 & 0.65 \\ 0.55 & 0.65 & 0.71 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.54 & 0.55 & 0.54 \\ 0.55 & 0.61 & 0.63 \\ 0.54 & 0.63 & 0.70 \end{bmatrix}.$$

The mean matrices were all within 0.03 of those for the first replicates. The SD for these true values were all less than 0.03, with more variability for the 2 more recent UPG. The estimated $\mathbf{\Gamma}$ closely aligned with the true $\mathbf{\Gamma}$, especially for the relationships with the first meta-founder. Because genotypes were used to calculate $\mathbf{\Gamma}$ and genotyping did not start until generation 6, the estimates for the first meta-founder were based on genotyped animals from later generations. This result indicated the method was robust to estimating metafounders that were recently related to but not included in the genotyped population.

Table 5 shows EBV validation results for different combinations of models. For the trait with heritability of 0.3, the best models were ssGBLUP with no UPG, with UPG for **A**, and with metafounders. The ssGBLUP with metafounders model was especially similar to the ssGBLUP analysis of the complete pedigree. Single-step genomic BLUP with UPG for **H** had more biased EBV than ssGBLUP without UPG. Bias was greater for nongenotyped animals missing a dam but followed a similar trend to the presented results. This increased bias was likely caused by biased UPG solutions for ssGBLUP with UPG for **H** (Table 6). Overall genetic

trend was underestimated by more than 1 genetic SD over 10 generations (Table 7). In addition, the BLUP analyses were unbiased with the complete pedigree but severely biased when pedigrees were missing. Hence, the issues with modeling UPG may stem from the basic assumptions for UPG effects.

For the trait with heritability of 0.1, the best models were ssGBLUP with metafounders and with UPG for **H** because the other models had biased EBV for young animals with missing dams. Again, the metafounders method was most similar to ssGBLUP with complete pedigree. Including UPG caused a slight decrease in accuracy and an increase in dispersion, especially for UPG for **A**, although the difference is not likely to be significant. The decrease in accuracy and increase in dispersion were also observed when including UPG in BLUP. One concern about ssGBLUP with UPG for **A** was the 20% increase in the number of rounds to convergence, indicating greater numeric instability (Table 5). Although all simulations converged in a reasonable number of rounds, this behavior is problematic in industry data (Matilainen et al., 2016).

We further investigated EBV by summarizing bias for genotyped animals based on the amount of pedigree information in one replicate of the trait with heritability of 0.1 (Figure 1). For ssGBLUP with UPG for **A**, bias was evident in overestimation of EBV for genotyped animals missing dams. For ssGBLUP with UPG for **H**, bias occurred for individuals missing dams or maternal granddams, indicating issues with the final UPG solution. For ssGBLUP with metafounders, no systematic bias existed based on pedigree completeness. The EBV from ssGBLUP with UPG for **H** and with metafounders are plotted in Figure 2. The EBV were shifted slightly to the left of the graph for animals with missing pedigree, indicating a discrepancy between the models. The EBV differed more between models when the missing pedigree was more recent. This difference occurred because EBV for young animals depend on the UPG contribution from $(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2\hat{\mathbf{s}}$ in ssGBLUP with UPG for **H**. Animals had larger UPG contributions when recent pedigrees were missing because

Table 5. Validation results for BLUP and single-step genomic BLUP (ssGBLUP) models including or excluding unknown parent groups (UPG)

Model	UPG ¹	Accuracy ²	Accuracy SD	b ₁ ³	b ₁ SD	Bias ⁴	Bias SD	Rounds (SD) ⁵
h² = 0.3								
BLUP	Complete	0.45	0.04	1.03	0.07	0.08	0.04	86 (2)
BLUP	None	0.36	0.03	0.59	0.06	1.57	0.14	85 (2)
BLUP	A	0.42	0.03	0.83	0.06	1.65	0.19	81 (4)
ssGBLUP	Complete	0.77	0.02	1.07	0.03	0.04	0.03	181 (8)
ssGBLUP	None	0.76	0.02	1.10	0.04	-0.14	0.05	180 (8)
ssGBLUP	A	0.76	0.02	1.06	0.04	-0.21	0.05	221 (8)
ssGBLUP	H	0.73	0.03	1.05	0.05	-0.83	0.10	170 (8)
ssGBLUP	Meta	0.77	0.02	1.05	0.04	-0.03	0.04	122 (6)
h² = 0.1								
BLUP	Complete	0.38	0.06	0.96	0.12	0.06	0.06	113 (5)
BLUP	None	0.28	0.05	0.49	0.08	1.23	0.14	113 (6)
BLUP	A	0.26	0.05	0.40	0.08	2.88	0.18	110 (5)
ssGBLUP	Complete	0.65	0.03	1.02	0.06	0.01	0.06	263 (8)
ssGBLUP	None	0.63	0.04	1.08	0.06	-0.31	0.07	261 (8)
ssGBLUP	A	0.61	0.03	0.90	0.06	-0.37	0.09	294 (8)
ssGBLUP	H	0.62	0.04	1.02	0.06	0.11	0.14	238 (8)
ssGBLUP	Meta	0.64	0.04	0.94	0.05	-0.01	0.07	161 (8)

¹Complete: full data were used with no missing pedigrees to model; None: did not model missing pedigrees, A: UPG formed for **A**⁻¹, H: UPG formed for **H**⁻¹, and Meta: metafounders.

²SE ≤ 0.01.

³SE ≤ 0.03.

⁴Bias = (TBV - EBV)/σ_u, 0.01 ≤ SE ≤ 0.04, where TBV = true breeding value and σ_u = additive genetic SD.

⁵Metafounders were analyzed with different software to convergence criteria of 10⁻¹² and were not directly comparable.

A₂₂⁻¹ accounted for only known pedigree, whereas **G**⁻¹ accounted for all genetic relationships even if pedigree was missing. Based on these results, the metafounder solutions were expected to be less biased than UPG solutions for either model. The models with UPG in ssGBLUP did not enable fair comparisons of individuals with different pedigree completeness because of the biased EBV for individuals with missing pedigrees. The

EBV bias likely resulted from biased estimates of UPG effects.

In general, UPG solutions were underestimated (Table 6). The BLUP with UPG model underestimated genetic trend by more than 1 genetic SD for the trait with heritability of 0.3 and by more than 3 genetic SD for the trait with heritability of 0.1. The ssGBLUP with UPG for **H** solutions were similar to the

Table 6. Validation results (SD) for unknown parent group (UPG) solutions from BLUP and single-step genomic BLUP (ssGBLUP) models

Model	UPG ¹	Accuracy	UPG bias ²		
			2 - 1 ³	3 - 1 ⁴	3 - 2 ⁵
h² = 0.3					
BLUP	A	0.97 (0.02)	1.13 (0.10)	1.33 (0.13)	0.20 (0.12)
ssGBLUP	A	0.98 (0.01)	0.83 (0.10)	0.49 (0.12)	-0.34 (0.09)
ssGBLUP	H	0.95 (0.03)	1.35 (0.09)	1.64 (0.13)	0.30 (0.12)
ssGBLUP	Meta	0.98 (0.01)	0.76 (0.09)	0.76 (0.12)	0.00 (0.08)
h² = 0.1					
BLUP	A	-0.38 (0.15)	2.80 (0.09)	3.10 (0.12)	0.30 (0.14)
ssGBLUP	A	0.88 (0.03)	1.30 (0.09)	0.20 (0.12)	-1.10 (0.13)
ssGBLUP	H	-0.49 (0.12)	3.01 (0.09)	3.36 (0.12)	0.35 (0.15)
ssGBLUP	Meta	0.93 (0.03)	0.89 (0.06)	0.69 (0.06)	-0.20 (0.08)

¹Models based on the numerator relationship matrix (**A**), the combined pedigree and genomic relationship matrix (**H**), or metafounders (Meta).

²Bias = (TBV - EBV)/σ_u, where TBV = true breeding value and σ_u = additive genetic SD. UPG defined as generations 0 to 4 (1), 5 to 7 (2), and 8 to 10 (3).

³Estimable function defined as UPG 2 - UPG 1.

⁴Estimable function defined as UPG 3 - UPG 1.

⁵Estimable function defined as UPG 3 - UPG 2.

Table 7. Estimable functions of unknown parent group (UPG) solutions from BLUP and single-step genomic BLUP (ssGBLUP) models

Model	UPG ¹	2 - 1 ²	SD	3 - 1 ³	SD	3 - 2 ⁴	SD
$h^2 = 0.3$							
BLUP	A	0.51	0.19	1.74	0.20	1.22	0.10
ssGBLUP	A	0.81	0.19	2.58	0.22	1.77	0.09
ssGBLUP	H	0.29	0.17	1.42	0.19	1.13	0.10
ssGBLUP	Meta	0.88	0.18	2.31	0.20	1.43	0.07
Simulated	True	1.37	0.11	2.85	0.20	1.48	0.11
$h^2 = 0.1$							
BLUP	A	-1.46	0.18	-0.59	0.26	0.87	0.16
ssGBLUP	A	0.04	0.20	2.31	0.29	2.27	0.19
ssGBLUP	H	-1.67	0.16	-0.85	0.24	0.82	0.17
ssGBLUP	Meta	0.26	0.15	1.68	0.22	1.42	0.11
Simulated	True	1.16	0.13	2.37	0.22	1.22	0.12

¹UPG defined as generations 0 to 4 (1), 5 to 7 (2), and 8 to 10 (3); models based on the numerator relationship matrix (**A**), the combined pedigree and genomic relationship matrix (**H**), or metafounders (Meta).

²Estimable function defined as $(UPG\ 2 - UPG\ 1)/\sigma_u$, where σ_u = additive genetic SD.

³Estimable function defined as $(UPG\ 3 - UPG\ 1)/\sigma_u$.

⁴Estimable function defined as $(UPG\ 3 - UPG\ 2)/\sigma_u$.

BLUP with UPG solutions. This result coincides with the more similar genetic trend and EBV predictions for BLUP with UPG and for ssGBLUP with UPG for **H** than for BLUP with UPG and ssGBLUP with UPG for **A** (Matilainen et al., 2016). For most estimable functions, ssGBLUP with metafounders was the least biased model, especially for the more problematic trait with heritability of 0.1. Interestingly, the metafounders approach was often best at predicting group differences for both heritabilities (Table 7). For ssGBLUP with UPG for **A** and the trait with heritability of 0.1, the UPG solutions were essentially the same for the first 2 UPG (Table 7). For ssGBLUP with UPG for **H** and the

trait with heritability of 0.1, the first UPG solution was the largest, resulting in an unexpected trend for UPG solutions. Because the population underwent selection, these solutions were not reasonable, because the first UPG should have the smallest solution. The inconsistency in the direction of genetic trend affected the accuracy of the UPG solutions (Table 6). It is worth noting that the issues with UPG were not restricted to ssGBLUP because BLUP also had problems with estimating genetic trend. The UPG solutions should be checked for unexpected behavior, as the issue was not immediately evident in EBV validations for young animals. In addition, care is needed to ensure enough data in the

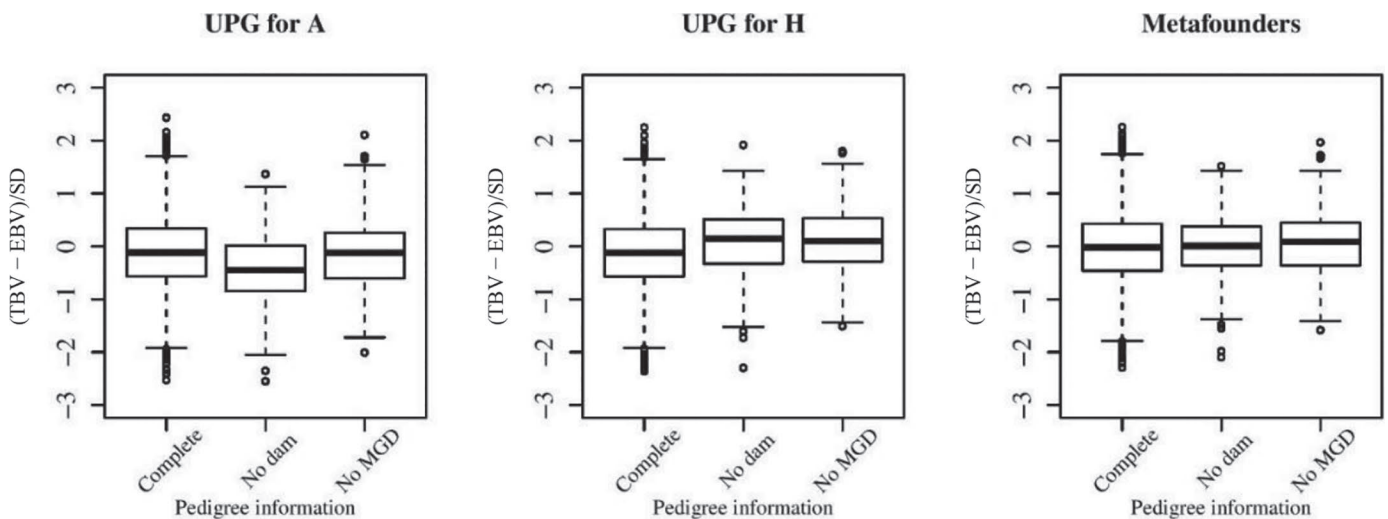


Figure 1. Boxplots of bias $(TBV - EBV)/SD$, where TBV = true breeding value) for genotyped animals with 2 generations of complete pedigree, missing dam, and missing maternal granddam (MGD) in single-step genomic BLUP with unknown parent groups (UPG) defined for **A**, UPG for **H**, or metafounders. The box represents the first and third quartiles, the line is the median, the whiskers extend to the furthest point up to 1.5 times the interquartile range, and circles are for outliers beyond that distance.

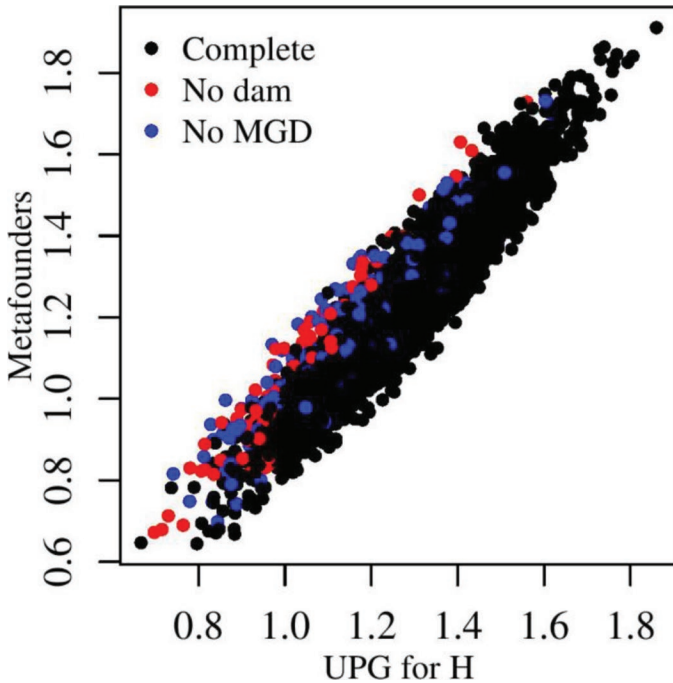


Figure 2. Estimated breeding values from single-step genomic BLUP with unknown parent groups (UPG) for **H** and with metafounders for genotyped animals with 2 generations of complete pedigree, missing dam, and missing maternal granddam (MGD).

most recent UPG to prevent sudden changes in genetic trend, especially when validating with reduced data (Tsuruta et al., 2014). These results were likely caused by inappropriate UPG definitions given the amount of information to estimate the UPG effects, particularly for the low heritability simulation. Heritability was the main difference between the simulations, and the lower heritability means that less overall information was available from the same number of phenotypes. This lack of information seemed to be diminished by the combined use of genotypes for animals missing pedigree together with metafounders.

When there was not enough information to estimate UPG solutions well, a feasible option was to not include UPG and just consider the pedigree as missing to have accurate predictions for young animals. This situation was noticeable for the low-heritability scenario, in which no UPG was more accurate and unbiased than including UPG for BLUP and ssGBLUP models, but ssGBLUP with metafounders was about the same as not including UPG (Table 5). Hence, in populations with limited genetic trend for low-heritability traits, modeling UPG could negatively affect genetic predictions. Attempts to model more UPG also caused erratic results in the trait with heritability of 0.3, again caused by a lack of information to estimate the UPG solutions. Care is needed in defining UPG over time; the same

UPG definitions were not ideal across traits with different heritability.

For ssGBLUP with UPG for **H**, one issue is the UPG solutions depend on **G**. Theoretically, the relationships in **G** are based on identity by state and are based on SNP distributed throughout the genome. Thus, relationships in **G** were complete and unaffected by missing pedigree relationships. The contribution from **G** could be ignored in the UPG solutions with the left-hand-side of the mixed model equations:

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 & \mathbf{Z}'_1\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & -\mathbf{A}^1\mathbf{Q} \\ \mathbf{A}^{21} & \mathbf{A}^{22} & -\mathbf{A}^2\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & -\mathbf{Q}'\mathbf{A}^{-2} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & \mathbf{A}_{22}^{-1}\mathbf{Q}_2 \\ \mathbf{0} & \mathbf{Q}'_2\mathbf{A}_{22}^{-1} & -\mathbf{Q}'_2\mathbf{A}_{22}^{-1}\mathbf{Q}_2 \end{bmatrix}.$$

This model was tested but preliminary results did not improve over ssGBLUP with UPG for **H** (unpublished data). In ssGBLUP with UPG for **H**, UPG solutions and EBV still depend on $\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$, meaning that discrepancies between these matrices affected predictions. Even if the matrices were scaled to look similar, \mathbf{A}_{22} still had missing pedigree relationships; \mathbf{A}_{22} was created without accounting for any missing pedigrees. Care is needed to ensure that \mathbf{A}_{22} and **G** are on the same base (Vitezica et al., 2011). This requirement was automatically fulfilled by using metafounders, which forced $\mathbf{A}(\mathbf{T})_{22}$ to be on the same base as **G**. Research is needed to include UPG in \mathbf{A}_{22} to better match **G** and \mathbf{A}_{22} and to potentially have more stable UPG solutions. This approach may be more similar to using metafounders.

It is unclear why UPG for **A** worked well in some cases and UPG for **H** worked well in other cases. In general, **G** is not affected by pedigree but is affected by line or breed differences (Plieschke et al., 2015) and \mathbf{A}_{22} is affected by both missing pedigree (and subsequently unaccounted for inbreeding) and line or breed differences. Therefore, formulas for UPG in ssGBLUP do not need to include **G** for purebred populations, and they do not need to include \mathbf{A}_{22} if all genotyped animals have complete pedigree. In the Nordic population where UPG for **H** worked well, few pedigrees were missing, and accounting for the country (line) differences were important (Matilainen et al., 2016). The study by Misztal et al. (2017) used only genotypes for animals born up to 2014 with nearly complete pedigrees. In contrast, Masuda et al. (2018) used genotyped animals born up to 2015 with many incomplete pedigrees. Al-

though the optimal method for UPG seems to be based on metafounders, a simpler method is currently under investigation.

The best approach to model UPG will likely be situation-dependent because what works best in dairy with missing pedigrees across time might not work best in swine with breed- or line-specific missing pedigrees. This simulation contained many simplifying assumptions with only dams missing and all females being phenotyped. Dairy populations have more complicated missing pedigree structures in addition to imported genetics that may have limited available pedigree and phenotypic information. When both parents were missing and UPG were defined by sex, the UPG effects could be confounded. The UPG definitions needed to balance genetic trend with the amount of phenotypic information to estimate these effects.

CONCLUSIONS

Much information was needed to estimate UPG solutions well. In our simulation, one UPG definition worked well with a moderately heritable trait but not with a lowly heritable trait. The best solution was to include metafounders to model missing pedigrees and to provide reasonable predictions of the genetic trend. If metafounders were not used and limited information was available to estimate UPG effects, the best approach was to not model the missing pedigree to obtain accurate predictions for young animals. In some cases with poor modeling of UPG, biases were 3 genetic SD for UPG solutions. More research is needed on the application of ssGBLUP to different types of populations with missing pedigrees.

ACKNOWLEDGMENTS

This research was supported in part by an appointment to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE; Oak Ridge, TN) through an interagency agreement between the US Department of Energy (DOE) and the US Department of Agriculture (USDA). ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-SC0014664. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of USDA, ARS, DOE, or ORAU/ORISE. Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. A. Legarra is grateful for FEDER (European Regional Development Fund) funds through Poctefa (Programme Interreg

V-A Espagne-France-Andorre, Huesca, Spain) project ARDI (Research, Development, Innovation in Sheep). We appreciated the helpful comments by 2 anonymous reviewers.

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>.
- Chen, C.-Y., I. Misztal, I. Aguilar, A. Legarra, and W. 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., 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>.
- Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* 43:1. <https://doi.org/10.1186/1297-9686-43-1>.
- 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>.
- 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.
- Johnston, D. J., M. H. Ferdosi, N. K. Connors, V. Boerner, J. Cook, C. J. Girard, A. A. Swan, and B. Tier. 2018. Implementation of single-step genomic BREEDPLAN evaluations in Australian beef cattle. Page 269 in *Proc. World Congr. Genet. Appl. Livest. Prod.*, Auckland, New Zealand. Accessed Dec. 19, 2018. <http://www.wcgalp.org/system/files/proceedings/2018/implementation-single-step-genomic-breedplan-evaluations-australian-beef-cattle.pdf>.
- Koivula, M., I. Strandén, J. Pösö, G. Aamand, and E. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. *J. Dairy Sci.* 98:2775. <https://doi.org/10.3168/jds.2014-8975>.
- Legarra, A., J. Bertrاند, T. Strabel, R. Sapp, J. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286–295. <https://doi.org/10.1111/j.1439-0388.2007.00671.x>.
- 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>.
- Masuda, Y., I. Misztal, P. M. VanRaden, and T. J. Lawlor. 2018. Genomic predictability of single-step GBLUP for production traits in US Holstein. *J. Dairy Sci.* 101(Suppl. 2):182. (Abstr.)
- 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.
- 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., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 family of programs. Vol. 2018. Accessed Dec. 17, 2018. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.
- Misztal, I., Z.-G. Vitezica, A. Legarra, I. Aguilar, and 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>.
- Plieschke, L., C. Edel, E. C. 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>.
- Quaas, R. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:1338–1345. [https://doi.org/10.3168/jds.S0022-0302\(88\)79691-5](https://doi.org/10.3168/jds.S0022-0302(88)79691-5).
- Sargolzaei, M., and F. S. Schenkel. 2009. QMSim: A large-scale genome simulator for livestock. *Bioinformatics* 25:680–681. <https://doi.org/10.1093/bioinformatics/btp045>.
- Thompson, R. 1979. Sire evaluation. *Biometrics* 35:339–353. <https://doi.org/10.2307/2529955>.
- Tsuruta, S., I. Misztal, D. Lourenco, and T. 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>.
- Tsuruta, S., I. Misztal, and I. Strandén. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. *J. Anim. Sci.* 79:1166–1172. <https://doi.org/10.2527/2001.7951166x>.
- VanRaden, P. 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](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., M. Tooker, J. Cole, G. Wiggans, and J. Megonigal Jr.. 2007. Genetic evaluations for mixed-breed populations. *J. Dairy Sci.* 90:2434–2441. <https://doi.org/10.3168/jds.2006-704>.
- 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., R. 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](https://doi.org/10.3168/jds.S0022-0302(88)79688-5).
- 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>.