



**HAL**  
open science

## **Multibreed genomic evaluation for production traits of dairy cattle in the United States using single-step genomic best linear unbiased predictor**

A. Cesarani, D. Lourenco, S. Tsuruta, A. Legarra, E.L. Nicolazzi, P.M. Vanraden, I. Misztal

### ► To cite this version:

A. Cesarani, D. Lourenco, S. Tsuruta, A. Legarra, E.L. Nicolazzi, et al.. Multibreed genomic evaluation for production traits of dairy cattle in the United States using single-step genomic best linear unbiased predictor. *Journal of Dairy Science*, 2022, 105 (6), pp.5141-5152. <10.3168/jds.2021-21505>. <hal-03778879>

**HAL Id: hal-03778879**

**<https://hal.inrae.fr/hal-03778879v1>**

Submitted on 16 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



## Multibreed genomic evaluation for production traits of dairy cattle in the United States using single-step genomic best linear unbiased predictor

A. Cesarani,<sup>1\*</sup> D. Lourenco,<sup>1</sup> S. Tsuruta,<sup>1</sup> A. Legarra,<sup>2</sup> E. L. Nicolazzi,<sup>3</sup> P. M. VanRaden,<sup>4</sup> and I. Misztal<sup>1</sup>

<sup>1</sup>Department of Animal and Dairy Science, University of Georgia, Athens 30602

<sup>2</sup>INRA, UMR1388 GenPhySE, Castanet-Tolosan, France, 31320

<sup>3</sup>Council on Dairy Cattle Breeding, Bowie, MD 20716

<sup>4</sup>Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705

### ABSTRACT

Official multibreed genomic evaluations for dairy cattle in the United States are based on multibreed BLUP evaluation followed by single-breed estimation of SNP effects. Single-step genomic BLUP (ssGBLUP) allows the straight computation of genomic (G)EBV in a multibreed context. This work aimed to develop ssGBLUP multibreed genomic predictions for US dairy cattle using the algorithm for proven and young (APY) to compute the inverse of the genomic relationship matrix. Only purebred Ayrshire (AY), Brown Swiss (BS), Guernsey (GU), Holstein (HO), and Jersey (JE) animals were considered. A 3-trait model with milk (MY), fat (FY), and protein (PY) yields was applied using about 45 million phenotypes recorded from January 2000 to June 2020. The whole data set included about 29.5 million animals, of which almost 4 million were genotyped. All the effects in the model were breed specific, and breed was also considered as fixed unknown parent groups. Evaluations were done for (1) each single breed separately (single); (2) HO and JE together (HO\_JE); (3) AY, BS, and GU together (AY\_BS\_GU); (4) all the 5 breeds together (5\_BREEDS). Initially, 15k core animals were used in APY for AY\_BS\_GU and 5\_BREEDS, but larger core sets with more animals from the least represented breeds were also tested. The HO\_JE evaluation had a fixed set of 30k core animals, with an equal representation of the 2 breeds, whereas HO and JE single-breed analysis involved 15k core animals. Validation for cows was based on correlations between adjusted phenotypes and (G)EBV, whereas for bulls on the regression of daughter yield deviations on (G)EBV. Because breed was correctly considered in the model, BLUP results for single and

multibreed analyses were the same. Under ssGBLUP, predictability and reliability for AY, BS, and GU were on average 7% and 2% lower in 5\_BREEDS compared with single-breed evaluations, respectively. However, validation parameters for these 3 breeds became better than in the single-breed evaluations when 45k animals were included in the core set for 5\_BREEDS. Evaluations for Holsteins were more stable across scenarios because of the greatest number of genotyped animals and amount of data. Combining AY, BS, and GU into one evaluation resulted in predictions similar to the ones from single breed, especially when using about 30k core animals in APY. The results showed that single-step large-scale multibreed evaluations are computationally feasible, but fine tuning is needed to avoid a reduction in reliability when numerically dominant breeds are combined. Having evaluations for AY, BS, and GU separated from HO and JE may reduce inflation of GEBV for the first 3 breeds.

**Key words:** all-breed model, across-breed predictions, unknown parent group, genomic selection

### INTRODUCTION

Official multibreed genomic evaluations for dairy cattle in the United States are based on multibreed BLUP evaluations followed by single-breed estimation of SNP effects. This is the so-called multistep method (VanRaden, 2008; VanRaden et al., 2009), where genomic PTA is obtained as the combination of direct genomic value based on markers and parent average. The effectiveness of genomic evaluations relies on the availability of a large reference population where the marker effects can be estimated (Goddard, 2009). The accuracy of this estimation strongly depends on the size of the reference population. VanRaden et al. (2009) demonstrated that, when the genotyping was performed almost exclusively on elite bulls, there was a linear increase in estimation reliability with the number

Received October 30, 2021.

Accepted January 27, 2022.

\*Corresponding author: [acesarani@uiss.it](mailto:acesarani@uiss.it)

of bulls added to the reference population. Nowadays, the availability of genotypes drastically increased and the number of animals with genomic information recently hit 5 million only in the United States (<https://queries.uscdcb.com/Genotype/counts.html>).

Single-step genomic BLUP (**ssGBLUP**) has been adopted in lieu of the multistep method for genetic evaluations in several farm animal species over the years. The **ssGBLUP** involves the inverse of a realized relationship matrix ( $\mathbf{H}^{-1}$ ; Aguilar et al., 2010), which combines the pedigree relationship matrix ( $\mathbf{A}$ ) and the genomic relationship matrix ( $\mathbf{G}$ ). Despite past limits due to problems with unknown parent groups (**UPG**) and computational cost, recent studies demonstrated the validity of this method for evaluating breeding values in several different livestock species including dairy (Himmelbauer et al., 2021; Liu and Alkhoder, 2021; Pimentel et al., 2021) and beef cattle (Lourenco et al., 2015), buffalo (Aspilcueta-Borquis et al., 2015; Cesarani et al., 2021a), goats (Teissier et al., 2018), and sheep (Cesarani et al., 2019; Macedo et al., 2020). Recently, better prediction features of **ssGBLUP** compared with **BLUP** in US Holstein were reported (Cesarani et al., 2021b). Those authors used about 860k genotyped animals and demonstrated that old phenotypic records (i.e., before 2000) can be removed without reducing prediction accuracy of young selection candidates. Moreover, they concluded that the best way to consider **UPG** in **ssGBLUP** is to include them in both  $\mathbf{A}$  and the relationship matrix among genotyped animals ( $\mathbf{A}_{22}$ ), as also described by Tsuruta et al. (2019) and Masuda et al. (2021). Another advantage of **ssGBLUP** is the consideration of genomic preselection (Masuda et al., 2018), that is, animals selected based on their **GEBV** before they have phenotypes available (Petry and Ducrocq, 2011; Jibrila et al., 2020).

Some studies have already investigated multibreed genomic evaluation models. Winkelman et al. (2015) applied the multibreed evaluation, using different blended or hybrid genomic models, to Holstein (**HO**), Jersey (**JE**), and their crosses from New Zealand; these authors reported high reliabilities for pure- and crossbreeds, with the latter showing higher accuracy for some traits. Different types of SNP panels and genomic models (i.e., **GBLUP** and Bayesian methods) were tested in crossbreeds, and the highest accuracy for crossbreeds was achieved when their data were considered (Khansefid et al., 2020). Creating a genomic relationship matrix among all genotyped animals belonging to different breeds in **ssGBLUP** can be challenging; however, it may help overcome the limitation due to the small reference population size in some breeds. This research aimed to develop **ssGBLUP** multibreed genomic predictions for

US dairy cattle using purebred Ayrshire (**AY**), Brown Swiss (**BS**), Guernsey (**GU**), **HO**, and **JE** animals.

## MATERIALS AND METHODS

### Data

No animals were used in this study, and ethical approval for the use of animals was thus deemed unnecessary.

Data used in the official multibreed genomic evaluations for US dairy cattle breeds were provided by the Council on Dairy Cattle Breeding (Bowie, MD). Only purebred **AY**, **BS**, **GU**, **HO**, and **JE** animals were considered. It is important to note that animals are considered purebred if they have more than 90% pedigree-based ancestry from one breed. The data sets were analyzed in 4 different scenarios: (1) each single breed separately (single); (2) **HO** and **JE** together (**HO\_JE**); (3) **AY**, **BS**, and **GU** together (**AY\_BS\_GU**); and (4) all the 5 breeds together (**5\_BREEDS**). Number of phenotypic lactation records (and cows), genotyped animals, and total animals in the pedigree (traced back 3 generations) are shown in Table 1. Milk (**MY**), fat (**FY**), and protein (**PY**) yields recorded from January 2000 to June 2020 were considered. The 305-d yields for the first 5 lactations included projected records for the final lactation and for lactations not yet completed by June 2020. The projected records were expanded to have the same genetic variance as completed records but given less weight to account for their larger error variance (VanRaden, 1997). According to the real lactation length, different weights are applied to project the records to 305 d: the shorter the lactation, the lower the weight. Animals were genotyped with 48 different arrays ranging from <3,000 to >600,000 usable SNPs. Genotypes were imputed, within each breed, to a common set of 79,294 selected SNPs using Findhap version 3 (VanRaden, 2016).

Single-step **GBLUP** was used in all the scenarios. The algorithm for proven and young (**APY**; Misztal et al., 2014) was used when the number of genotyped animals exceeded 50k (i.e., in single for Holstein and Jersey evaluations with 15k random core animals, in **HO\_JE** with 30k random core following an equal representation of the 2 breeds, in **AY\_BS\_GU**, and in **5\_BREEDS** with 15k random core animals). According to Pocrnic et al. (2016), the number of largest eigenvalues explaining 98% of the variance of  $\mathbf{G}$  (**EIGEN98**), which represents the number of core animals in **APY**, in the US cattle populations varies from 11k to 19k; however, this number can be higher in multibreed populations if there is less overlap among breeds (Pocrnic et al., 2019).

**Table 1.** Number of records, cows, genotypes, and total animals in 5 single breeds and in the combined scenarios<sup>1</sup>

Scenario	Phenotypes		Animals	
	N	Cows	Genotypes <sup>2</sup>	Total
AY	116,674	47,174	9,202	94,500
BS	328,811	138,418	47,309	292,923
GU	129,422	58,554	5,032	100,643
HO	40,298,113	17,484,436	3,407,476	26,586,363
JE	4,134,973	1,704,641	427,286	2,467,946
AY_BS_GU	574,907	244,146	61,543	488,066
5_BREEDS	45,007,993	19,433,223	3,896,305	29,542,375
HO_JE	44,433,086	19,189,077	3,834,762	29,054,309

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

<sup>2</sup>Genotypes were used in genomic model only (i.e., single-step genomic BLUP).

Because the EIGEN98 when combining AY, BS, and GU was 19,184, being 14,909 for BS alone, we tested one sub-scenario under AY\_BS\_GU with 15k BS and all AY and GU (AY\_BS\_GU\_30k). Additionally, we compared AY\_BS\_GU and AY\_BS\_GU\_30k against one analysis without APY, meaning  $\mathbf{G}$  with almost 61k animals was directly inverted (AY\_BS\_GU\_direct). The 5\_BREEDS scenario had a sub-scenario where 45k core animals were used, being composed by 5k AY, 5k BS, 5k GU, 15k HO, and 15k JE (5\_BREEDS\_45k).

### Model and Analysis

Two evaluation methods were considered: (1) traditional BLUP (BLUP) with UPG; (2) ssGBLUP with UPG for  $\mathbf{A}$  and  $\mathbf{A}_{22}$ , implemented as in Tsuruta et al. (2019) and Cesarani et al. (2021b). The UPG were defined based on breed, sex, and year of birth. The latter consisted of different groups for animals born before 2000, from 2001 to 2005, from 2006 to 2010, and from 2011 to 2020. Therefore, a total of 8 UPG were assigned within each breed.

A 3-trait repeatability BLUP animal model was applied in the single scenario according to Cesarani et al. (2021b), whereas the model used in the present study for ssGBLUP is described as SS\_UPG2 in the same paper. In particular, the following 3-trait repeatability animal model was applied for the traditional BLUP:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_a\mathbf{Q}_a\mathbf{g}_a + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_p\mathbf{p} + \mathbf{e},$$

where  $\mathbf{y}$  was the vector of the phenotypic records (i.e., milk, fat, and protein lactation yields);  $\mathbf{b}$  was the vector of the considered fixed effects of herd-management, age-parity, inbreeding coefficient covariates, and heterosis;  $\mathbf{h}$  was the vector of the random effect of the herd-sire combination;  $\mathbf{g}$  was the vector of the fixed effect of UPG;  $\mathbf{a}$  was the vector of animal additive genetic

effect;  $\mathbf{p}$  was the vector of permanent environmental effect; and  $\mathbf{e}$  was the vector of residuals. Heterosis is included because purebred animals can have up to 10% of other breeds (VanRaden et al., 2007). A single heterosis regression coefficient per animal was computed, as  $1 - \sum s_i d_i$ , where  $s_i$  and  $d_i$  are fractions of the sire's and dam's genes from breed  $i$  (VanRaden, 1992); this results in 0 for a 100% purebred animal, but is different from 0 if, say, the bull was 100% Holstein and the dam was 95% Holstein and 5% Jersey.  $\mathbf{X}$  was the incidence matrix relating each phenotypic record to the fixed effects in  $\mathbf{b}$ ;  $\mathbf{Q}_a$  was an incidence matrix relating animals in vector  $\mathbf{a}$  to UPG in vector  $\mathbf{g}_a$ ;  $\mathbf{Z}_h$ ,  $\mathbf{Z}_a$ , and  $\mathbf{Z}_p$  were incidence matrices relating phenotypic records in  $\mathbf{y}$  to herd-sire, animal, and permanent environment effects, respectively. The  $\mathbf{h}$  vector had mean zero and variance  $\mathbf{I} \otimes \mathbf{V}_h$ , where  $\mathbf{I}$  was an identity matrix and  $\mathbf{V}_h$  was a  $3 \times 3$  matrix of herd-sire variances and covariances, and  $\otimes$  was the Kronecker product. The vector  $\mathbf{a}$  had mean zero and variance  $\mathbf{A} \otimes \mathbf{V}_a$ , where  $\mathbf{A}$  was the numerator relationship matrix and  $\mathbf{V}_a$  was a  $3 \times 3$  matrix of additive genetic variances and covariances. Finally, the vector  $\mathbf{p}$  had mean zero and variance  $\mathbf{I} \otimes \mathbf{V}_p$ , where  $\mathbf{I}$  was an identity matrix and  $\mathbf{V}_p$  was a  $3 \times 3$  matrix of permanent environment variances and covariances. The ssGBLUP model had the same effect structure of BLUP but the covariance structure for  $\mathbf{Z}_a\mathbf{a}$  and  $\mathbf{Z}_a\mathbf{Q}_a\mathbf{g}_a$  also contained genomic relationships. In particular, the covariance matrix for the additive genetic effect in ssGBLUP was given by

$$\mathbf{H}^* = \mathbf{A}^* + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{G}_{\text{APY}}^{-1} - \mathbf{A}_{22}^{-1} & \mathbf{A}_{22}^{-1}\mathbf{Q}_2 \\ 0 & -\mathbf{Q}_2' \mathbf{A}_{22}^{-1} & \mathbf{Q}_2' \mathbf{A}_{22}^{-1}\mathbf{Q}_2 \end{bmatrix},$$

where  $\mathbf{H}^*$  was the inverse of the realized relationship matrix with UPG added to the pedigree relationship

matrix  $\mathbf{A}$  and to the matrix of pedigree relationships among genotyped animals  $\mathbf{A}_{22}$ ;  $\mathbf{A}^*$  was the inverse of the pedigree relationship matrix with UPG, modified with the QP transformation (Quaas, 1988);  $\mathbf{G}_{\text{APY}}^{-1}$  was the inverse of the genomic relationship matrix ( $\mathbf{G}$ ) built with or without the APY according to the scenario (see above). The  $\mathbf{G}$  matrix was constructed according to the method 1 of VanRaden, 2008 (i.e., centered and scaled by current allele frequencies from all genotyped animals). The genomic matrix was combined with 5% of  $\mathbf{A}_{22}$  to avoid singularity; the latter matrix was built based on the algorithm proposed by Colleau (2002). Moreover,  $\mathbf{G}$  was scaled based on  $\mathbf{A}_{22}$  so the 2 matrices had same average values for both diagonal and off-diagonal elements.

Variance components for MY, FY, and PY were obtained from VanRaden et al. (2014): heritability was 0.20 for all the 3 traits; genetic correlations were 0.45 (MY vs. FY), 0.81 (MY vs. PY), and 0.60 (FY vs. PY). The heterogeneous herd variance was adjusted according to Wiggans and VanRaden (1991). The pre-adjustments for herd-year variance also adjust all breeds to have equal genetic variance (VanRaden et al., 2007). Higher heritability in different breeds is accounted for in the model by higher weights obtained from the ratios of genetic to error variance. The same pre-adjustments and variance ratios are applied to all the 3 traits.

In AY\_BS\_GU, HO\_JE, and 5\_BREEDS scenarios, fixed effects were made breed-specific by concatenating a breed indicator to the effect levels and nesting covariables for inbreeding and heterosis within breed.

The mixed model equations (**MME**) were solved using iteration on data and a block preconditioning conjugate gradient with the software BLUP90IOD2 (ver. 3.113; Tsuruta et al., 2001; Tsuruta and Misztal, 2008). If the MME are  $\mathbf{Ax} = \mathbf{b}$ , the convergence criterion was  $\mathbf{b} - \mathbf{Ax}^2/\mathbf{b}^2$ , with a termination criterion of  $10^{-15}$ . The computations were carried out on a Linux server (x86\_64) equipped with Intel Xeon E5-2683 v4 2.10 GHz processor with 32 cores.

### Validation of Breeding Values

The (G)EBV were validated within each tested scenario. Two data sets were considered: (1) complete, with phenotypes for milk, fat, and protein recorded in cows born from 1992 to 2018; (2) reduced, with phenotypic records from cows born up to 2014. Genotyped cows born between 2015 and 2018 with phenotypes in the complete but not in the reduced data set were in the validation set. The numbers of validation cows were 181 for AY, 2,423 for BS, 750 for GU, 577,340 for HO, and 90,666 for JE. Genotyped bulls with no daughters with phenotypes in the reduced but at least 10 daughters

in the complete data set were in the validation group for AY ( $n = 17$ ), BS ( $n = 107$ ), and GU ( $n = 28$ ), whereas 50 daughters were requested for HO ( $n = 3,278$ ) and JE ( $n = 471$ ) validation bulls.

In this study, theoretical accuracies were not analyzed, and models were compared based on validation reliabilities. The validation for bulls was carried out using daughter yield deviations (**DYD**), estimated in the complete data sets for all bulls following the method by Liu et al. (2004) and the algorithm by Mrode and Swanson (2004). The validation reliability of predictions for bulls was measured using the coefficient of determination ( $R^2$ ) from the regression of DYD on (G)EBV. The validation for cows was based on the estimates of phenotypes adjusted for fixed effects ( $Y_{\text{adj}}$ ), computed based on the complete data set using PRE-DICTF90 (ver. 1.5; Misztal et al., 2018). Note that the  $R^2$  cannot be, strictly speaking, interpreted as a reliability (squared correlation between true and estimated BV) unless the number of daughters in the DYD is very high. The predictive ability [i.e., the correlation between  $Y_{\text{adj}}$  and (G)EBV] was used as a measure of accuracy of predictions for validation cows. Note that this correlation cannot be interpreted as an accuracy (correlation between true and estimated BV). Finally, the inflation of predictions was measured through the regression coefficients ( $b_1$ ) of the regression of DYD or  $Y_{\text{PRED}}$  on (G)EBV.

There are several reasons to use one criterion for bulls and another for cows. First, in this manner, we can compare each of these criteria with preceding studies for bulls (e.g., VanRaden et al., 2009) and for cows (e.g., Bengtsson et al., 2020; VanRaden et al., 2020). Second, measures of goodness of fit for bulls and cows are not comparable, as bulls have been strongly selected, which results in a decrease in cross-validation accuracy (Bijma, 2012). Third, it is not easy to develop a simple statistic that can be used for bulls and cows simultaneously.

## RESULTS AND DISCUSSION

Table 2 reports the comparison between (G)EBV estimated in single and 5\_BREEDS\_45k scenarios. Since animals were not connected by the pedigree and the breed was correctly accounted for by the model, BLUP results for single and all multibreed scenarios were the same: correlations between raw EBVs estimated in BLUP single and BLUP multibreed were 1.00 for all breeds and traits. In the genomic models, correlations between raw GEBVs estimated in ssGBLUP single and ssGBLUP 5\_BREEDS\_45k ranged from 0.95 (MY for AY and BS) to 1. The latter was the case for all correlations involving HO and JE, probably because of

**Table 2.** Comparison between genomic EBV estimated in single and 5\_BREEDS\_45k scenarios<sup>1</sup>

Correlation	BLUP			ssGBLUP <sup>2</sup>		
	MY	FY	PY	MY	FY	PY
AY	1.00	1.00	1.00	0.95	0.96	0.96
BS	1.00	1.00	1.00	0.95	0.96	0.97
GU	1.00	1.00	1.00	0.97	0.98	0.98
HO	1.00	1.00	1.00	1.00	1.00	1.00
JE	1.00	1.00	1.00	1.00	1.00	1.00
Slope <sup>3</sup>						
AY	1.00	1.00	1.00	1.02	0.98	0.99
BS	1.00	1.00	1.00	0.93	0.96	0.93
GU	1.00	1.00	1.00	0.95	0.95	0.95
HO	1.00	1.00	1.00	1.00	1.00	1.00
JE	1.00	1.00	1.00	0.99	0.98	0.98

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

<sup>2</sup>ssGBLUP = single-step genomic BLUP.

<sup>3</sup>The slope of the regression of 5\_BREEDS\_45k on single.

the greatest number of genotypes. Moreover, we also computed the slope of the regression of (G)EBV in 5\_BREEDS on (G)EBV in single models. As expected, slopes for BLUP were always 1, confirming that breed was correctly considered, and animals were not connected by the pedigree. Acceptable slope values were also found for ssGBLUP as values ranged from 0.93 (MY and PY for BS) to 1.02 (MY for AY).

The coefficients of determination ( $R^2$ ) for bulls in single-breed and multibreed analyses are shown in Table 3. As expected, the highest values for BLUP were observed for the numerically dominant breeds (i.e., Jersey and Holstein). All the 7 tested genomic scenarios had greater values compared with BLUP; however, some differences were observed among the genomic scenarios. Within the single scenario, ssGBLUP showed values that were on average ( $\pm$ SD) 0.23 ( $\pm$ 0.14) points greater than the ones from BLUP in all considered breeds. The greatest increases moving from BLUP to ssGBLUP were observed for the breeds with larger numbers of genotyped animals (i.e., BS, JE, and HO). Looking at the 5\_BREEDS scenario, there was a considerable reduction in  $R^2$  for BS compared with the single scenario. Contrasting, there was an overall increase of 16 points in  $R^2$  for AY when moving from single to 5\_BREEDS. A small drop in  $R^2$  was observed for fat and protein in GU. The validation results for AY, BS, and GU should be taken with caution because of the limited numbers of validation bulls, which were 17, 107, and 28, respectively. When the number of validation bulls was large enough (i.e., for HO and JE),  $R^2$  was similar between single and 5\_BREEDS scenarios. A successful transition from single-breed to multibreed evaluations would produce  $R^2$  and dispersion in the latter that are at least as good as in the former.

A factor contributing to the drop in  $R^2$  in our study was the breed representation in the core group. When randomly selecting 15k animals among the 5 breeds to be in the core set, the breeds with less genotyped animals were not well represented. While there were 13,049 Holsteins in the core, there were only 1,720 JE, 182 BS, 32 AY, and 17 GU. After increasing the core size to 45k (5\_BREEDS\_45k), which included 15k HO, 15k JE, 5k AY, 5k BS, and 5k GU,  $R^2$  increased to similar or larger levels as in the single scenario.

Multibreed analyses for the 3 breeds with fewer data were performed to investigate whether the genomic information from HO and JE was responsible for the loss in prediction accuracy and validation reliability. Issues in multibreed evaluations are common when some breeds have many more genotyped animals than others. In such a case, SNP effects and frequencies are driven by the numerically dominant breeds, which causes a drop in validation reliability and an increase in inflation of GEBV for smaller breeds (van den Berg et al., 2020). Validation reliabilities for the AY\_BS\_GU scenario were not homogeneous across breeds when comparing with the single scenario. We observed larger  $R^2$  for AY, slightly larger for BS, and slightly lower for GU. Comparing to AY\_BS\_GU\_30k, which had 15k BS and all AY and GU in the core, results were similar for BS and GU. When the AY\_BS\_GU evaluation involved the direct inverse of  $\mathbf{G}$  (AY\_BS\_GU\_direct), the average  $R^2$  across traits remained 0.40 for BS and 0.26 for GU. The large oscillation in  $R^2$  for AY confirms the number of validation bulls is too small to draw any conclusions.

Having a multibreed evaluation with combined data from the breeds with fewer data is a possibility; therefore, investigating an evaluation with data from the numerically dominant breeds becomes interesting. The

**Table 3.** Coefficients of determination ( $R^2$ ) between bull daughter yield deviations (DYD) and EBV from BLUP and between DYD and genomic EBV from single-step genomic BLUP (ssGBLUP) in the 5 considered breeds<sup>1</sup>

Breed	Trait	N	BLUP	Single	ssGBLUP					HO_JE
					5_BREEDS	5_BREEDS_45k	AY_BS_GU	AY_BS_GU_30k	AY_BS_GU_direct	
AY	Milk	17	0.20	0.27	0.36	0.73	0.43	0.34	0.34	—
	Fat		0.20	0.31	0.58	0.74	0.44	0.34	0.34	—
BS	Protein	107	0.30	0.43	0.56	0.86	0.57	0.48	0.48	—
	Milk		0.15	0.40	0.22	0.37	0.40	0.40	0.40	—
	Fat		0.14	0.30	0.15	0.30	0.34	0.34	0.34	—
	Protein		0.21	0.45	0.21	0.42	0.45	0.45	0.45	—
GU	Milk	28	0.18	0.29	0.33	0.41	0.25	0.26	0.25	—
	Fat		0.20	0.31	0.27	0.40	0.29	0.28	0.28	—
	Protein		0.18	0.25	0.22	0.28	0.24	0.23	0.24	—
	Milk		0.28	0.77	0.77	0.77	—	—	—	0.77
HO	Fat	3,278	0.31	0.77	0.76	0.76	—	—	—	0.77
	Protein		0.30	0.74	0.74	0.74	—	—	—	0.74
	Milk		0.43	0.71	0.69	0.75	—	—	—	0.76
	Fat		0.36	0.64	0.63	0.67	—	—	—	0.68
JE	Protein	471	0.45	0.69	0.67	0.74	—	—	—	0.74
	Milk		0.45	0.69	0.67	0.74	—	—	—	0.74

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

evaluation combining HO and JE produced a similar  $R^2$  for HO bulls across all traits and a slight increase in  $R^2$  for milk in JE when compared with 5\_BREEDS\_45k. Small changes of 0.01 or 0.02 for some traits when changing the model to account for different breeds can be because of the scaling of  $\mathbf{G}$  to match  $\mathbf{A}_{22}$ . In such a case, diagonals and off-diagonals of  $\mathbf{G}$  are set to have the same average as in  $\mathbf{A}_{22}$  (Chen et al., 2011; Vitezica et al., 2011), which may be different by breed; however, the statistics for  $\mathbf{A}_{22}$  are computed across all the breeds included in the model. One way to possibly solve this issue is to scale  $\mathbf{G}$  differently based on  $\mathbf{A}_{22}$  constructed separately for each breed. Another way would be to use metafounders (Legarra et al., 2015), which scales  $\mathbf{A}_{22}$  to match  $\mathbf{G}$  and sets pseudoindividuals as proxies for the base animals.

Predictive abilities (i.e., validation reliability for cows) for validation cows in single-breed and multibreed analyses are shown in Table 4. The number of validation cows was large enough for most breeds, so the results for BS and GU were more stable for cows than bulls. Ayrshire still had a small number of validation cows ( $n = 181$ ), which resulted in unexpected predictivities for some scenarios and a small gain of 0.02 (milk and protein) and 0.03 (fat) when switching from BLUP to single-breed ssGBLUP. Overall, the gain in predictivity by using genomic single-breed models was 0.16 for BS, 0.09 for GU, 0.21 for HO, and 0.17 for JE. When moving from single to 5\_BREEDS scenarios, the average predictivity for HO and JE was similar, whereas the average predictivity for AY, BS, and GU cows decreased from 0.42 to 0.36. Again, this drop was mainly driven by the number and composition of the core animals in 5\_BREEDS, which did not have a good representation of AY, BS, and GU. Using 45k core animals in APY for the multibreed evaluation (5\_BREEDS\_45k) increased the average predictivity for AY, BS, and GU cows to 0.45. Predictivity for HO and JE did not change. When evaluations for AY, BS, and GU were done separately from HO and JE (i.e., AY\_BS\_GU, AY\_BS\_GU\_30k, and AY\_BS\_GU\_direct), predictivities were overall similar to the single scenario. Likewise, predictivities for HO and JE did not considerably change from single to HO\_JE scenarios. However, small fluctuations were observed across traits and breeds.

One of the concerns with multibreed evaluations is the increase in inflation of GEBV by combining breeds that have been selected differently. In this study, when breeds were combined in the same evaluation, we created breed-specific fixed effects and used UPG based on breed in addition to year of birth interval and sex. Thus, levels of effects were defined within breed (i.e., the effect of age by parity number 1 was assumed to be different in different breeds). Whereas UPG take care of

**Table 4.** Predictive ability for cows [i.e., correlation between adjusted phenotypes and EBV from BLUP or genomic EBV from single-step genomic BLUP (ssGBLUP)] in the 5 considered breeds<sup>1</sup>

Breed	Trait	N	BLUP	Single	ssGBLUP					HO_JE
					5_BREEDS	5_BREEDS_45k	AY_BS_GU	AY_BS_GU_30k	AY_BS_GU_direct	
AY	Milk	181	0.45	0.47	0.44	0.55	0.47	0.47	0.47	—
	Fat		0.51	0.54	0.39	0.52	0.53	0.53	—	
BS	Protein	2,423	0.52	0.54	0.45	0.55	0.54	0.54	0.54	—
	Milk		0.23	0.39	0.29	0.40	0.39	0.39	0.39	—
GU	Fat	750	0.27	0.42	0.33	0.41	0.42	0.42	0.42	—
	Protein		0.27	0.44	0.34	0.43	0.44	0.44	0.44	—
HO	Milk	577,340	0.24	0.35	0.37	0.44	0.33	0.35	0.35	—
	Fat		0.27	0.33	0.28	0.39	0.32	0.34	0.34	—
JE	Protein	90,666	0.31	0.34	0.31	0.39	0.32	0.34	0.34	—
	Milk		0.33	0.55	0.55	0.55	—	—	—	0.55
JE	Fat	90,666	0.34	0.55	0.54	0.55	—	—	—	0.55
	Protein		0.32	0.50	0.52	0.52	—	—	—	0.52
JE	Milk	90,666	0.32	0.50	0.50	0.51	—	—	—	0.52
	Fat		0.29	0.46	0.46	0.47	—	—	—	0.47
JE	Protein	90,666	0.35	0.51	0.50	0.52	—	—	—	0.47
	Fat		0.35	0.51	0.50	0.52	—	—	—	0.52

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

the genetic differences among breeds, the breed-specific fixed effects account for other nongenetic variation, such as different effects of parity number or different levels of average within breed heterosis. An initial investigation showed that not considering breed-specific fixed effects in the model resulted in lower validation reliability and predictivity and increased inflation of EBV and GEBV (results not shown).

The inflation and deflation of (G)EBV were assessed in this study by investigating the slope ( $b_1$ ) of the regression of DYD on (G)EBV for bulls (Table 5) and of adjusted phenotypes on (G)EBV for cows (Table 6). According to Tsuruta et al. (2011), estimates within 15% from 1 are acceptable, whereas estimates within 5% are considered good predictions. Across traits, EBV for bulls were inflated for most of the breeds. Average  $b_1$  for BS, GU, and HO was 0.63, 0.71, and 0.88. A deflation happened for milk and protein in JE. Adding genomic information in single-breed evaluations helped to reduce inflation and deflation. Combining the 5 breeds (5\_BREEDS) increased the average inflation of BS GEBV compared with the single scenario, and small fluctuations were observed across breeds and traits. When using 45k core animals (5\_BREEDS\_45k), inflation and deflation levels were, in general, similar to 5\_BREEDS. The only exception was AY that showed more inconsistencies across scenarios because of the small number of bulls. Overall, the AY\_BS\_GU scenarios resulted in  $b_1$  that were similar to the single scenario for those breeds. The  $b_1$  for HO did not vary much across 5\_BREEDS and HO\_JE scenarios compared with the single scenario, whereas small changes were observed for JE. Inflation and deflation patterns were very similar between bulls and cows, with  $b_1$  values from BS and GU cows in AY\_BS\_GU closer to the ones in the single scenario, stable  $b_1$  across scenarios for HO, and small variation for JE. The exception in this comparison between bulls and cows was AY that had less variation in the evaluations using 3 breeds compared with the single scenario.

Validation statistics for AY and GU in multibreed evaluations in the United States have not been reported before mainly because of the small number of validation bulls available and the late onset of genomic selection for those breeds. Whereas genomic selection was officially implemented for HO, JE, and BS in 2009, it only started in 2013 for AY and 2016 for GU. Olson et al. (2012) investigated the use of multibreed reference populations under multistep methods in the United States, but only genotypes for HO, JE, and BS were used. The authors concluded that predictivity across breeds was limited but breeds with fewer data slightly benefited from a multibreed estimation of SNP effects. Similar results were found by VanRaden et al. (2020),

**Table 5.** Regression coefficients ( $b_1$ ) of bull daughter yield deviations on EBV from BLUP and on genomic EBV from single-step genomic BLUP (ssGBLUP) for validation bulls<sup>1</sup>

Breed	Trait	N	BLUP	Single	ssGBLUP					HO_JE
					5_BREEDS	5_BREEDS_45k	AY_BS_GU	AY_BS_GU_30k	AY_BS_GU_direct	
AY	Milk	17	0.94	0.88	0.74	1.37	1.21	0.96	0.96	—
	Fat		0.79	0.86	1.20	1.25	1.15	0.92	0.92	—
	Protein		1.13	1.04	1.01	1.34	1.25	1.06	1.06	—
BS	Milk	107	0.64	0.94	0.56	0.81	0.94	0.93	0.94	—
	Fat		0.55	0.80	0.56	0.72	0.81	0.80	0.80	—
	Protein		0.70	0.93	0.68	0.82	0.92	0.92	0.92	—
GU	Milk	28	0.75	1.04	1.17	1.14	1.02	1.03	1.03	—
	Fat		0.67	0.84	0.76	0.89	0.84	0.85	0.85	—
	Protein		0.71	0.90	0.90	1.02	0.89	0.92	0.91	—
HO	Milk	3,278	0.86	1.03	1.05	1.03	—	—	—	1.04
	Fat		0.91	0.99	0.98	0.97	—	—	—	0.99
	Protein		0.87	1.03	1.02	1.03	—	—	—	1.02
JE	Milk	471	1.06	1.01	1.01	1.05	—	—	—	1.08
	Fat		0.89	0.86	0.93	0.92	—	—	—	0.97
	Protein		1.06	0.99	1.01	1.06	—	—	—	1.08

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

**Table 6.** Regression coefficients ( $b_1$ ) of cow-adjusted phenotypes on EBV from BLUP and on genomic EBV from single-step genomic BLUP (ssGBLUP) for validation cows<sup>1</sup>

Breed	Trait	N	BLUP	Single	ssGBLUP					HO_JE
					5_BREEDS	5_BREEDS_45k	AY_BS_GU	AY_BS_GU_30k	AY_BS_GU_direct	
AY	Milk	181	1.03	0.96	0.94	0.88	1.00	0.98	0.98	—
	Fat		0.95	0.89	0.81	0.76	0.92	0.90	0.90	—
	Protein		0.98	0.92	0.89	0.80	0.95	0.93	0.93	—
BS	Milk	2,423	0.72	0.96	0.79	0.90	0.97	0.97	0.97	—
	Fat		0.72	0.91	0.78	0.84	0.92	0.92	0.91	—
	Protein		0.73	0.96	0.83	0.89	0.96	0.96	0.96	—
GU	Milk	750	0.75	0.91	1.01	1.07	0.89	0.91	0.91	—
	Fat		0.71	0.74	0.79	0.93	0.75	0.77	0.77	—
	Protein		0.71	0.84	0.94	1.05	0.83	0.85	0.85	—
HO	Milk	577,340	1.01	1.10	1.13	1.11	—	—	—	1.10
	Fat		1.03	1.03	1.02	1.03	—	—	—	1.02
	Protein		1.02	1.08	1.07	1.08	—	—	—	1.07
JE	Milk	90,666	1.02	1.03	1.11	1.07	—	—	—	1.09
	Fat		0.89	0.88	0.99	0.93	—	—	—	0.97
	Protein		1.01	0.99	1.08	1.05	—	—	—	1.07

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

who included 5 breeds and many more animals. However, the estimation of SNP effects is still in a single-breed fashion for routine US dairy evaluations because of the small gains. This is true also for the evaluations in other countries (Su et al., 2010; Shabalina et al., 2020).

Although the estimation of SNP effects is single-breed for most of the dairy evaluations, the deregressed EBV used to estimate SNP effects in the multistep process are computed based on multibreed BLUP evaluations. VanRaden et al. (2007) developed an all-breed model for routine BLUP evaluations in the United States that considered AY, BS, GU, HO, JE, and Milking Shorthorn, in addition to crossbreds. The advantage of combining different breeds is to have updated breed differences, and a reason to include data for crossbred animals is to have extra, accurate information for both parents, which can increase their EBV reliability.

In our study, we used ssGBLUP, which is an extension of BLUP to include relationships based on genomic information. We showed that having multibreed genomic evaluations in a single run is possible because validation reliabilities, predictabilities, and regression coefficients were not considerably different from those in single-breed evaluations, especially for the breeds with enough validation animals. Multibreed evaluations should deliver GEBV with at least the same reliability and inflation-deflation level as the single-breed evaluations. Although we did not include data on crossbred animals in this study, several studies have shown the usefulness of ssGBLUP in multibreed and crossbred evaluations and that accuracy of predictions for crossbred animals greatly improves when their genotypes are in the evaluation (Lourenco et al., 2016; Xiang et al., 2016; Pocrnic et al., 2019). The current genomic evaluation for crossbred dairy in the United States is based on purebred direct genomic values (DGV) weighted by breed proportion (VanRaden et al., 2020), and the genomic predictions from this method are more accurate than parent average. Additional accuracy could possibly be added by having a multibreed and crossbred joint ssGBLUP evaluation, which is currently under investigation.

The use of ssGBLUP in a multibreed context, however, implies that SNP effects are the same for all the breeds in the data set. Similar to the other results, SNP estimates can also be more conditioned to the more numerically dominant breeds and different from those estimated in single-breed models. Because the number of genotypes is constantly increasing, DGV for the newly genotyped animals can be estimated by using indirect predictions instead of solving the whole MME (Lourenco et al., 2015). The DGV estimated using

SNP effects from multibreed models will be different from those of single-breed models, but they could be more accurate. In our laboratory, studies on indirect predictions and a new algorithm to approximate their theoretical reliabilities are under investigation.

Because of the large number of genotyped animals, APY was used to compute  $\mathbf{G}^{-1}$  for BS, HO, and JE as single breed, AY\_BS\_GU, AY\_BS\_GU\_30k, all 5\_BREEDS scenarios, and HO\_JE. Pocrnic et al. (2016) found the number of core animals in purebred cattle populations would vary from 10 to 19k; however, this may not apply to multibreed populations. In such a case, the number of core animals should represent the dimensionality within each breed, which implies using the number of core animals for each breed as the number of eigenvalues explaining 98% of the variance of  $\mathbf{G}$ . This was confirmed by the fact that using 15k core animals randomly sampled among all the genotyped animals caused some fluctuations in validation reliability, predictivity, and  $b_1$  for bulls and cows in the multi- compared with single-breed evaluations. Therefore, if the multibreed evaluation involves the 5 breeds, 45k core animals may be enough, whereas 30k is enough for AY\_BS\_GU and 30k for HO\_JE. Adding crossbred genotypes to the evaluation may not increase the number of core animals because of the sharing of independent chromosome segments with the purebred parents (Pocrnic et al., 2019). Independent chromosome segments can also be shared among different breeds, especially if there was recent introgression from the use of a few sires across breeds.

When combining multiple breeds in a joint ssGBLUP evaluation with APY, it is important to have a good representation of all breeds in the core (Mäntysaari et al., 2017; Vandenplas et al., 2018). Additionally, modeling the breed effects in a proper way guarantees more accurate predictions in any multibreed evaluations. In this study we added breed code to all the fixed effects to ensure animals from different breeds were not mistakenly in the same group. Breed effects were also added to the model as a fixed UPG. Therefore, GEBV resulting from the multibreed models were adjusted for breed-specific fixed effects but contained genetic breed effects combined with selection pathway by sex and year of birth (i.e., the original definition of UPG). This model ensured that predictions from multibreed and single breed were more similar. When the model does not correctly account for the breed and breed-specific effects, extremely high reliabilities from multi- compared with single-breed models may be observed due to the prediction of breed effects instead of GEBV (Bermann et al., 2021). However, this is not the case in this study because all validations are done within breed.

Table 7 shows number of rounds and computing time (i.e., only the preconditioning conjugate gradient with iteration on data) to reach convergence assuming a criterion of  $10^{-15}$  in all the scenarios. The computing cost for AY, BS, and GU was small in the single scenario and in the 3 AY\_BS\_GU scenarios. For HO, it took 2.7 h for BLUP and 7.4 h for ssGBLUP (3.4 million genotyped animals) in the single scenario. Larger computing time was recently reported by Cesarani et al. (2021b), who analyzed Holstein data up to 2018 with 861k genotypes. The larger computing costs reported in that previous work could be because of the use of selected genotypes only for cows with records and their parents, in addition to more UPG. When the 5 breeds were combined, the computing time increased 1.8 times for BLUP, 2.7 times for 5\_BREEDS, and 8.6 times for 5\_BREEDS\_45k compared with single HO. The largest number of iterations suggests a possible lack of compatibility between genomic and pedigree information after combining different breeds. Additionally, because scaling of  $\mathbf{G}$  is based on an average  $\mathbf{A}_{22}$  across breeds, the combination of the 5 breeds possibly resulted in an  $\mathbf{H}^{-1}$  that has a larger condition number. As far as the time needed for the creation of the relationship matrices, it took about 2.5 d to create the matrices for the 3.4 million HO, about 3.5 d for 5\_BREEDS, and about 5 d for 5\_BREEDS\_45k. About half of this time was used in the creation of  $\mathbf{A}_{22}$  for the scaling of  $\mathbf{G}$ . A new algorithm is under investigation to reduce the computing time required for the construction of  $\mathbf{A}_{22}$ . This algorithm considers the APY structure and sparsity, meaning that coefficients of  $\mathbf{A}_{22}$  are computed separately for core and noncore animals.

## CONCLUSIONS

Results of the present study demonstrated that large-scale ssGBLUP multibreed evaluations are computationally feasible, and that genomic EBV are as reliable as in single-breed evaluation when the number of genotyped animals in the APY core represent the dimensionality of the genomic information within the breeds. Accounting for genetic and nongenetic breed effects is important when combining several breeds in a joint evaluation; for instance, fixed effects need to be redefined within breed. Fine tuning is still needed to avoid a reduction in reliability for less represented breeds when numerically dominant breeds are in the evaluation. Having evaluations for Ayrshire, Brown Swiss, and Guernsey separate from Holstein and Jerseys may reduce inflation of GEV for the first 3 breeds.

## ACKNOWLEDGMENTS

This study was partially funded by Agriculture and Food Research Initiative Competitive Grant no. 2020-67015-31030 from the US Department of Agriculture's National Institute of Food and Agriculture. The authors thank the Council on Dairy Cattle Breeding (CDCB, Bowie, MD) for providing access to the data. The contribution of dairy producers who supplied data through their participation in the Dairy Herd Improvement program and the Dairy Records Processing Centers that edited and relayed information to the Council of Dairy Cattle Breeding are also acknowledged. Discussions with Duane Norman (CDCB, Bowie, MD) and Yutaka Masuda (Rakuno Gakuen University, Ebetsu,

**Table 7.** Computational costs for all considered scenarios in terms of number of rounds and seconds per round needed to reach a convergence criterion of  $10^{-15}$

Item <sup>1</sup>	Rounds		s/round		Time <sup>2,3</sup>	
	BLUP	ssGBLUP	BLUP	ssGBLUP	BLUP	ssGBLUP
AY	504	863	0.08	0.08	<1 min	~1 min
BS	364	867	0.18	0.45	1 min	~6 min
GU	345	757	0.07	0.07	<1 min	<1 min
HO	457	473	21.25	56.31	2.7 h	7.4 h
JE	586	432	2.00	5.58	~20 min	~40 min
AY_BS_GU	592	1,534	3.01	0.81	~3 min	~20 min
AY_BS_GU_30k		1,581		4.11		~1.4 h
AY_BS_GU_direct		1,529		2.22		~50 min
5_BREEDS	643	1,142	27.01	64.84	4.8 h	~20 h
5_BREEDS_45k		1,763		130.68		~64 h
HO_JE	479	1,403	40.57	81.71	5.4 h	~32 h

<sup>1</sup>AY = Ayrshire; BS = Brown Swiss; GU = Guernsey; HO = Holstein; JE = Jersey; 5\_BREEDS = all 5 breeds together.

<sup>2</sup>Time to reach the convergence (i.e., only time for the preconditioning conjugate gradient with iteration on data). ssGBLUP = single-step genomic BLUP.

<sup>3</sup>Computations were carried out on a Linux server (x86\_64) equipped with Intel Xeon E5-2683 v4 2.10 GHz processor with 32 cores. Parallel threads used between 12 and 20 cores.

Hokkaido, Japan) are greatly acknowledged. 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>.
- Aspilcueta-Borquis, R. R., F. R. Araujo Neto, D. J. A. Santos, N. A. Hurtado-Lugo, J. A. V. Silva, and H. Tonhati. 2015. Multiple-trait genomic evaluation for milk yield and milk quality traits using genomic and phenotypic data in buffalo in Brazil. *Genet. Mol. Res.* 14:18009–18017. <https://doi.org/10.4238/2015.December.22.27>.
- Bengtsson, C., H. Ståhlhammar, E. Strandberg, S. Eriksson, and W. F. Fikse. 2020. Association of genomically enhanced and parent average breeding values with cow performance in Nordic dairy cattle. *J. Dairy Sci.* 103:6383–6391. <https://doi.org/10.3168/jds.2019-17963>.
- Bermann, M., D. Lourenco, V. Breen, R. Hawken, F. Brito Lopes, and I. Misztal. 2021. Modeling genetic differences of combined broiler chicken populations in single-step GBLUP. *J. Anim. Sci.* 99:skab056. <https://doi.org/10.1093/jas/skab056>.
- Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *J. Anim. Breed. Genet.* 129:345–358. <https://doi.org/10.1111/j.1439-0388.2012.00991.x>.
- Cesarani, A., S. Biffani, A. Garcia, D. Lourenco, G. Bertolini, G. Neglia, I. Misztal, and N. P. P. Macciotta. 2021a. Genomic investigation of milk production in Italian buffalo. *Ital. J. Anim. Sci.* 20:539–547. <https://doi.org/10.1080/1828051X.2021.1902404>.
- Cesarani, A., G. Gaspa, F. Correddu, M. Cellesi, C. Dimauro, and N. P. P. Macciotta. 2019. Genomic selection of milk fatty acid composition in Sarda dairy sheep: Effect of different phenotypes and relationship matrices on heritability and breeding value accuracy. *J. Dairy Sci.* 102:3189–3203. <https://doi.org/10.3168/jds.2018-15333>.
- Cesarani, A., Y. Masuda, S. Tsuruta, E. L. Nicolazzi, P. M. VanRaden, D. Lourenco, and I. Misztal. 2021b. 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>.
- Colleau, J. J. 2002. An indirect approach to the extensive calculation of relationship coefficients. *Genet. Sel. Evol.* 34:409–421. <https://doi.org/10.1186/1297-9686-34-4-409>.
- 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>.
- Himmelbauer, J., H. Schwarzenbacher, and C. Fuerst. 2021. Implementation of single-step evaluations for fitness traits in the German and Austrian Fleckvieh and Brown Swiss populations. *Interbull Bull.* 56:82–89.
- Jibrila, I., J. Ten Napel, J. Vandenplas, R. F. Veerkamp, and M. P. Calus. 2020. Investigating the impact of preselection on subsequent single-step genomic BLUP evaluation of preselected animals. *Genet. Sel. Evol.* 52:42. <https://doi.org/10.1186/s12711-020-00562-6>.
- Khansefid, M., M. E. Goddard, M. Haile-Mariam, K. V. Konstantinov, C. Schrooten, G. de Jong, E. G. Jewell, E. O'Connor, E. J. Pryce, H. D. Daetwyler, and I. M. MacLeod. 2020. Improving genomic prediction of crossbred and purebred dairy cattle. *Front. Genet.* 11:598580. <https://doi.org/10.3389/fgene.2020.598580>.
- 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>.
- Liu, Z., and H. Alkholder. 2021. Application of a single-step SNP BLUP model to conformation traits of German Holsteins. *Interbull Bull.* 56:30–40.
- Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. *J. Dairy Sci.* 87:1896–1907. [https://doi.org/10.3168/jds.S0022-0302\(04\)73348-2](https://doi.org/10.3168/jds.S0022-0302(04)73348-2).
- Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, C. Y. Chen, W. O. Herring, and I. Misztal. 2016. Crossbred evaluations in single-step genomic best linear unbiased predictor using adjusted realized relationship matrices. *J. Anim. Sci.* 94:909–919. <https://doi.org/10.2527/jas.2015-9748>.
- 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>.
- 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>.
- Mäntysaari, E. A., R. D. 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>.
- Masuda, Y., S. Tsuruta, M. Bermann, 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., P. M. VanRaden, I. Misztal, and T. J. Lawlor. 2018. 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>.
- 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., S. Tsuruta, D. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for blupf90 family of programs. Accessed Jul. 2021. [http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90\\_all7.pdf](http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf).
- Mrode, R. A., and G. J. T. Swanson. 2004. Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. *Livest. Prod. Sci.* 86:253–260. <https://doi.org/10.1016/j.livprodsci.2003.09.001>.
- Olson, K. M., P. M. VanRaden, and M. E. Tooker. 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 95:5378–5383. <https://doi.org/10.3168/jds.2011-5006>.
- Patry, C., and V. Ducrocq. 2011. Accounting for genomic pre-selection in national BLUP evaluations in dairy cattle. *Genet. Sel. Evol.* 43:30. <https://doi.org/10.1186/1297-9686-43-30>.
- Pimentel, E. D. C. G., C. Edel, D. Krogmeier, R. Emmerling, and K. U. Götz. 2021. Effects of use of external information in Single-Step evaluations for linear type traits in Brown Swiss. *Interbull Bull.* 56:121–124.
- Pocrnic, I., D. Lourenco, C. Y. Chen, W. O. Herring, and I. Misztal. 2019. Crossbred evaluations using single-step genomic BLUP and algorithm for proven and young with different sources of data. *J. Anim. Sci.* 97:1513–1522. <https://doi.org/10.1093/jas/skz042>.
- Pocrnic, I., D. Lourenco, Y. Masuda, and I. Misztal. 2016. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. *Genet. Sel. Evol.* 48:82. <https://doi.org/10.1186/s12711-016-0261-6>.
- Quaas, R. L. 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).
- Shabalina, T., T. Yin, and S. König. 2020. Survival analyses in Holstein cows considering direct disease diagnoses and specific SNP

- marker effects. *J. Dairy Sci.* 103:8257–8273. <https://doi.org/10.3168/jds.2020-18174>.
- Su, G., B. Guldbbrandtsen, V. R. Gregersen, and M. S. Lund. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. *J. Dairy Sci.* 93:1175–1183. <https://doi.org/10.3168/jds.2009-2192>.
- Teissier, M., H. Larroque, and C. Robert-Granié. 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: A quantitative trait influenced by a major gene. *Genet. Sel. Evol.* 50:31. <https://doi.org/10.1186/s12711-018-0400-3>.
- 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., and I. Misztal. 2008. Computing options for genetic evaluation with a large number of genetic markers. *J. Anim. Sci.* 86:1514–1518. <https://doi.org/10.2527/jas.2007-0324>.
- 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, 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>.
- van den Berg, I., I. M. MacLeod, C. M. Reich, E. J. Breen, and J. E. Pryce. 2020. Optimizing genomic prediction for Australian Red dairy cattle. *J. Dairy Sci.* 103:6276–6298. <https://doi.org/10.3168/jds.2019-17914>.
- Vandenplas, J., M. P. L. Calus, and J. ten Napel. 2018. Sparse single-step genomic BLUP in crossbreeding schemes. *J. Anim. Sci.* 96:2060–2073. <https://doi.org/10.1093/jas/sky136>.
- 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](https://doi.org/10.3168/jds.S0022-0302(92)78077-1).
- VanRaden, P. M. 1997. Lactation yields and accuracies computed from test day yields and (co) variances by best prediction. *J. Dairy Sci.* 80:3015–3022. [https://doi.org/10.3168/jds.S0022-0302\(97\)76268-4](https://doi.org/10.3168/jds.S0022-0302(97)76268-4).
- 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. 2016. findhap.f90: Find haplotypes and impute genotypes using multiple chip sets and sequence data. Animal Improvement Program. Animal Genomics and Improvement Laboratory, ARS, USDA. <https://aipl.arsusda.gov/software/findhap>.
- VanRaden, P. M., M. E. Tooker, T. C. S. Chud, H. D. Norman, J. H. Megonigal Jr., I. W. Haagen, and G. R. Wiggans. 2020. Genomic predictions for crossbred dairy cattle. *J. Dairy Sci.* 103:1620–1631. <https://doi.org/10.3168/jds.2019-16634>.
- VanRaden, P. M., M. E. Tooker, J. B. Cole, G. R. Wiggans, and J. H. Megonigal Jr.. 2007. Genetic evaluations for mixed-breed populations. *J. Dairy Sci.* 90:2434–2441. <https://doi.org/10.3168/jds.2006-704>.
- VanRaden, P. M., M. E. Tooker, J. R. Wright, C. Sun, and J. L. Hutchison. 2014. Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. *J. Dairy Sci.* 97:7952–7962. <https://doi.org/10.3168/jds.2014-8489>.
- 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.* 93:357–366. <https://doi.org/10.1017/S001667231100022X>.
- Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustment for heterogeneous variance. *J. Dairy Sci.* 74:4350–4357. [https://doi.org/10.3168/jds.S0022-0302\(91\)78631-1](https://doi.org/10.3168/jds.S0022-0302(91)78631-1).
- Winkelman, A. M., D. L. Johnson, and B. L. Harris. 2015. Application of genomic evaluation to dairy cattle in New Zealand. *J. Dairy Sci.* 98:659–675. <https://doi.org/10.3168/jds.2014-8560>.
- Xiang, T., B. Nielsen, G. Su, A. Legarra, and O. F. Christensen. 2016. Application of single-step genomic evaluation for crossbred performance in pig. *J. Anim. Sci.* 94:936–948. <https://doi.org/10.2527/jas.2015-9930>.

## ORCID

- A. Cesarani  <https://orcid.org/0000-0003-4637-8669>  
D. Lourenco  <https://orcid.org/0000-0003-3140-1002>  
S. Tsuruta  <https://orcid.org/0000-0002-6897-6363>  
A. Legarra  <https://orcid.org/0000-0001-8893-7620>  
E. L. Nicolazzi  <https://orcid.org/0000-0002-5820-0799>  
P. M. VanRaden  <https://orcid.org/0000-0002-9123-7278>  
I. Misztal  <https://orcid.org/0000-0002-0382-1897>