NRC Research Press

ARTICLE

A comprehensive comparison between single- and two-step GBLUP methods in a simulated beef cattle population

Mario L. Piccoli, Luiz F. Brito, José Braccini, Fernanda V. Brito, Fernando F. Cardoso, Jaime A. Cobuci, Mehdi Sargolzaei, and Flávio S. Schenkel

Abstract: The statistical methods used in the genetic evaluations are a key component of the process and can be best compared by using simulated data. The latter is especially true in grazing beef cattle production systems, where the number of proven bulls with highly reliable estimated breeding values is limited to allow for a trustworthy validation of genomic predictions. Therefore, we simulated data for 4980 beef cattle aiming to compare single-step genomic best linear unbiased prediction (ssGBLUP), which simultaneously incorporates pedigree, phenotypic, and genomic data into genomic evaluations, and two-step GBLUP (tsGBLUP) procedures and genomic estimated breeding values (GEBVs) blending methods. The greatest increases in GEBV accuracies compared with the parents' average estimated breeding values (EBV_{PA}) were 0.364 and 0.341 for ssGBLUP and tsGBLUP, respectively. Direct genomic value and GEBV accuracies when using ssGBLUP and tsGBLUP procedures were similar, except for the GEBV accuracies using Hayes' blending method in tsGBLUP. There was no significant or slight bias in genomic predictions from ssGBLUP or tsGBLUP (using VanRaden's blending method), indicating that these predictions are on the same scale compared with the true breeding values. Overall, genetic evaluations including genomic information resulted in gains in accuracy >100% compared with the EBV_{PA}. In addition, there were no significant differences between the selected animals (10% males and 50% females) by using ssGBLUP or tsGBLUP.

Key words: accuracy, genomic breeding value, beef cattle breeding, ssGBLUP, genomic selection, genomic data simulation.

Résumé : Les méthodes statistiques utilisées dans les évaluations génétiques sont des composantes clés du processus et peuvent être mieux comparées en utilisant des données simulées. Cette dernière affirmation est particulièrement vraie dans les systèmes de production de bovins à bœuf en pâturage, où le nombre de taureaux confirmés ayant des valeurs estimées de reproduction grandement fiables est limité pour permettre la validation fiable des prévisions génomiques. Donc, nous avons simulé les données pour 4980 bovins à bœuf dans le but de comparer les procédures de type meilleure projection linéaire génomique sans biais à étape unique (ssGBLUP — « single-step genomic best linear unbiased prediction »), qui incorpore simultanément les données de pedigree, de phénotype et de génotype dans les évaluations génomiques, et la méthode à deux étapes (tsGBLUP — « two-step genomic best linear unbiased prediction ») et les valeurs génomiques estimées de reproduction (GEBV — « genomic estimated breeding values ») des méthodes de mélange. Les plus grandes augmentations d'exactitude de GEBV par rapport aux valeurs estimées de reproduction de la moyenne des parents (EBV_{PA} — « parents' average

Received 1 November 2017. Accepted 22 January 2018.

M.L. Piccoli. Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91540-000, Brazil; GenSys Consultores Associados S/S, Porto Alegre, RS 90460-060, Brazil; Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON N1G 2W1, Canada.

L.F. Brito and F.S. Schenkel. Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON N1G 2W1, Canada.

J. Braccini. Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91540-000, Brazil; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasília, DF 71605-001, Brazil.

F.V. Brito. GenSys Consultores Associados S/S, Porto Alegre, RS 90460-060, Brazil.

F.F. Cardoso. Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasília, DF 71605-001, Brazil; Embrapa Pecuária Sul, Bagé, RS 96401-970, Brazil.

J.A. Cobuci. Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91540-000, Brazil.

M. Sargolzaei. Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON N1G 2W1, Canada; The Semex Alliance, Guelph, ON N1H 6J2, Canada.

Corresponding author: Flávio S. Schenkel (email: schenkel@uoguelph.ca).

estimated breeding values ») étaient de 0,364 et 0,341 pour ssGBLUP et tsGBLUP, respectivement. Les exactitudes de la valeur génomique directe et de GEBV lors d'utilisation des procédures ssGBLUP et tsGBLUP étaient similaires, sauf pour les exactitudes GEBV utilisant la méthode de mélange de Hayes dans tsGBLUP. Il n'y avait pas d'influence significative ni légère dans les prévisions génomiques de ssGBLUP ou tsGBLUP (utilisant la méthode de mélange de VanRaden), indiquant que ces prévisions se trouvent sur la même échelle lorsque comparées aux valeurs réelles de reproduction. De façon générale, les évaluations génétiques incluant l'information génétique se soldaient par des gains d'exactitude de plus de 100 % par apport aux EBV_{PA}. De plus, il n'y avait pas de différences significatives entre les animaux choisis (10 % mâles et 50 % femelles) en utilisant ssGBLUP ou tsGBLUP. [Traduit par la Rédaction]

Mots-clés : exactitude, valeur génomique de reproduction, reproduction de bovins à bœuf, ssGBLUP, sélection génomique, simulation de données génomiques.

Introduction

Genomic selection has shaped modern breeding programs and contributed substantially to the increase of genetic progress for a variety of economically important traits, especially in dairy cattle (Hayes et al. 2009; VanRaden et al. 2009; Harris and Johnson 2010; Su et al. 2012; Meuwissen et al. 2016). These gains are associated with shorter generation intervals, increased selection intensity, and greater selection accuracies (Meuwissen et al. 2001; Schaeffer 2006; Hayes et al. 2009; Aguilar et al. 2010).

The first studies to combine genomics and estimated breeding values (EBVs) data were based on two-step genomic best linear unbiased prediction (tsGBLUP) procedure, where direct genomic values (DGVs) generated based solely on genomic and phenotype information, and EBVs were combined using different indexes weighted by the accuracy of breeding values and heritability of the trait. The blending of DGVs and EBVs is a key step in genomic predictions because if the effect of the quantitative trait loci (QTL) is not captured by the genomic markers, it may be captured by the polygenic effects via EBVs (Hayes et al. 2009; VanRaden et al. 2009; Pryce et al. 2014). In the single-step GBLUP (ssGBLUP) approach (Aguilar et al. 2010; Christensen and Lund 2010), pedigree, phenotypes, and genotypes are used jointly to predict genomic estimated breeding values (GEBVs).

Previous studies presented by Aguilar et al. (2010), Chen et al. (2010), Garrick (2010), and Vitezica et al. (2010) showed both advantages and disadvantages when using ssGBLUP or tsGBLUP procedures. Other studies have reported an increase in accuracies of genomic breeding values when using ssGBLUP compared with tsGBLUP in various species such as cattle, dairy sheep, dairy goats, pigs, and chicken (Chen et al. 2011; Koivula et al. 2012; Harris et al. 2013; Přibyl et al. 2013; Baloche et al. 2014; Carillier et al. 2014; Legarra et al. 2014). However, in the majority of dairy cattle (and other livestock species) breeding programs, routine genomic evaluations have been performed through the tsGBLUP procedure (e.g., Hayes et al. 2009; VanRaden et al. 2009; Harris and Johnson 2010; Su et al. 2012; Brito et al. 2017; Piccoli et al. 2017).

In beef cattle, the application of ssGBLUP in genomic evaluations is still incipient (Onogi et al. 2014; Cardoso et al. 2015). Furthermore, beef cattle datasets are more complex compared with other livestock industries (e.g., dairy cattle) due to larger amount of missing pedigree information, smaller sib families, and greater influence of maternal effects (Legarra et al. 2014). Moreover, most beef cattle breeds have small number of genotyped animals compared with dairy cattle. The use of simulated data to compare statistical methods is of great value due to the possibility of making comparisons with the actual simulated true breeding values (TBVs). This is especially important for beef cattle in grazing production systems, where the number of proven bulls with highly reliable EBVs is limited to allow for a trustworthy validation of genomic predictions. Therefore, the objectives of this study were to compare ssGBLUP and tsGBLUP procedures and GEBV blending methods using beef cattle simulated data.

Materials and Methods

Animal Care and Use Committee approval was not obtained as all the data used for this study was computationally simulated (i.e., no biological samples collection or animals' involvement).

Data simulation

Studies with simulated data can be efficient when there is a need to compare different methodologies. Moreover, simulation analysis should use parameters based on the target populations to mimic real scenarios. The simulated data used in this study mimicked the extent of linkage disequilibrium (LD) in beef cattle and was previously described in Brito et al. (2011).

Defining the population structure

The beef cattle population was simulated based on forward-in-time process, using the QMSim software (Sargolzaei and Schenkel 2009) to generate 40K single nucleotide polymorphisms (SNPs) markers evenly distributed along the genome, and 750 QTL across the 29 *Bos taurus* autosomes. First, 1000 generations with a constant size of 1000 individuals were simulated. Subsequently, 1020 generations with a gradual decrease in population size from 1000 to 200 individuals were simulated to create initial LD and to establish mutation-drift equilibrium in historical generations. The number of males and females remained constant ($n = 200$) and the mating system was based on random

union of gametes, randomly sampled. In the second step, an expansion of the population was created by initially randomly selecting 100 founder males and 100 founder females from the last generation of the historical population. In the third step, to enlarge the population, eight generations were simulated with five offspring per dam. The mating was based on the random union of gametes and no presence of selection. In the fourth step, the two most recent generations were simulated from the last generation by selecting 640 males and 32 000 females (i.e., rate of 1 male to 50 females). The parameters used in the recent generations were chosen to mimic a practical production system: one progeny per dam per year, 50% of male/female progeny, only the animals with highest EBVs kept for breeding, and a replacement rate of 60% for sires and 20% for dams. Furthermore, sires and dams were randomly mated. The whole process was used to generate 10 independent populations to obtain 10 replicates for the study.

Genome simulation

The simulated genome consisted of 29 pairs of autosomes with length identical to the real bovine genome size (2333 cM) based on Btau_3.1 assembly (Snelling et al. 2007). The SNPs were evenly distributed along the genome to generate one density of segregating bi-allelic loci with minor allele frequency (MAF) >0.1. The markers were neutral regarding to their effect on the trait. A number of QTL were simulated to generate 750 segregating loci (randomly distributed along the genome) with two, three, or four alleles and MAF >0.1. Additive allelic effects were randomly sampled from gamma distribution with shape parameter equal to 0.4. The rate of missing marker genotypes was 0.01 and the rate of marker genotyping error was 0.005. A recurrent mutation rate of 10^{-5} for both markers and QTL was considered to establish mutation-drift equilibrium in historical generations. The same mutation rate was also applied in all subsequent generations.

Simulation of phenotypes and genetic values

A single trait with heritability of 0.25 (average heritability estimate for the majority of beef cattle traits) and phenotypic variance of 1.0 was simulated using the QMSim package (Sargolzaei and Schenkel 2009). The EBVs were predicted using the BLUPF90 software (Misztal et al. 2002) to fit an individual animal model, considering the true additive genetic variance. The rate of missing sire and dam information was 0.05. The TBV of an individual was equal to the sum of the additive effects of the QTL. The phenotypes were generated by adding random residuals to the TBVs.

Defining training and validation populations to be used in the tsGBLUP procedure

The training population (TP) was composed of 1920 sires under selection from generation three to eight of the current population and each sire had 50 or more progeny. The validation population included 3060 individuals randomly chosen from the 10th generation with parents born until the 8th generation (individuals from 9th generation were not included to simulate a genetic distancing between training and validation populations). Three sets of EBVs were generated using the BLUPF90 software (Misztal et al. 2002). The first set of EBVs was formed by all the animals born until 8th generation (TP). The other two sets were formed by the TP plus all animals born in the 10th generation. These EBVs were generated either including the 10th generation animals' phenotypes (VP2) or excluding their phenotypes (VP1) from the analyses. These last two sets of alternate EBVs were used for blending with DGVs to generate GEBVs. The VP2 scenario mimicked the situation where phenotypes are available prior to selection in beef cattle. The approach of VanRaden and Wiggans (1991) was used to calculate de-regressed EBVs (dEBVs) free of parent average effects using the EBVs and reliabilities of genotyped animals and their sires and dams. The EBVs and dEBVs of TP were used as pseudo-phenotypes to estimate SNP effects. The DGVs were estimated using the GEBV software (Sargolzaei et al. 2009) to fit a GBLUP model considering 5%, 10%, 15%, and 20% (package default) contribution of polygenic effects to the genomic relationship matrix (VanRaden 2008). The GBLUP model can be described as $\boldsymbol{y} = \mathbf{1_n}\,\mu + \mathbf{Z}\boldsymbol{g} + \boldsymbol{e}$, where $\boldsymbol{y}$ is the vector of EBVs or dEBVs for the trait, $\boldsymbol{\mu}$ is the overall mean, $\mathbf{1_n}$ is a vector of ones, $\mathbf{Z}$ is the design matrix that relates records to breeding values, $\boldsymbol{g}$ is the vector of DGVs to be predicted, and $\boldsymbol{e}$ is the vector of residual effects. We assumed that $\boldsymbol{g} \sim N(0, \mathbf{G} * \sigma_g^2)$, where σ_g^2 is the additive genetic variance and $\mathbf{G}$ is a genomic relationship matrix (VanRaden et al. 2009), and $\boldsymbol{e} \sim N(0, \mathbf{R} * \sigma_e^2)$, where σ_e^2 is the residual variance and $\mathbf{R}$ is a diagonal matrix whose elements account for the differences in reliabilities of the EBVs or dEBVs.

For tsGBLUP, we investigated two methods to combine DGVs with EBVs to generate GEBVs: (1) as described in Hayes et al. (2009), in which $\text{GEBV} = \frac{r^2_{\text{DGV}} \times \text{DGV} + r^2_{\text{EBV}} \times \text{EBV}}{r^2_{\text{DGV}} + r^2_{\text{EBV}}}$, and r^2_{DGV} and r^2_{EBV} are the reliabilities of DGV and EBV, respectively; and (2) as described by VanRaden et al. (2009), in which $\text{GEBV} = b_1 \times \text{DGV} + b_2\text{EBV}_1 + b_3 \times \text{EBV}$, and EBV_1 was predicted for the subset of genotyped animals using traditional relationship information and their respective dEBVs, excluding data from nongenotyped animals, and b_1, b_2, and b_3 are weights based on reliabilities of DGV, EBV_1, and EBV.

Defining population to be used in the ssGBLUP procedure

The pedigree information included up to the 10th generation. The phenotype information was used until the 8th generation in one of the analysis (VP1) and in the other one (VP2), phenotypes of genotyped animals of the validation population (10th generation) were also included. The genotyped individuals included in the

Table 1. Direct genomic value (DGV) accuracies (Acc) and standard deviations (SD) for genomic breeding values from alternative genomic prediction methods using different phenotypes.

	$_G95\text{-}_A5$[b]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
DGV[a]	Acc	SD		Acc	SD		Acc	SD		Acc	SD	
ssDGV	0.584	0.018	a,a	0.571	0.018	a,a	0.554	0.018	b,a	0.535	0.020	b,a
tsDGVdebv	0.577	0.018	a,a	0.573	0.018	a,a	0.569	0.019	a,a	0.564	0.019	a,a
tsDGVebv	0.576	0.019	a,a	0.573	0.019	a,a	0.569	0.019	a,a	0.563	0.019	a,a

Note: Different letters indicate significant differences ($P < 0.05$) by Scheffé's test. The letters before the comma indicate differences within rows, whereas the letters after the comma indicate differences within columns.

[a]ssDGV predicted by single-step procedure and tsDGVdebv and tsDGVebv predicted by two-step procedure using de-regressed estimated breeding values (dEBVs) or EBVs as pseudo-phenotypes.

[b]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

analysis were 1920 sires with 50 or more progeny (TP) plus 3060 (randomly selected) genotyped animals, as part of the validation population.

The DGVs (obtained by excluding pedigree information in the single-step analysis) and GEBVs were estimated using the BLUPF90 package (Misztal et al. 2002) by fitting a GBLUP model considering the same options for **H** matrix including different weights to create $\mathbf{G}^* = \mathbf{G} + \beta \times \mathbf{A}_{22}$, where β is the polygenic proportion added to the **G** matrix. The levels of beta investigated were 5% (package default), 10%, 15%, and 20%. In ssGBLUP procedure, GEBVs were obtained by directly combining phenotypic, genomic, and pedigree information (Misztal et al. 2009; Aguilar et al. 2010), where the traditional relationship matrix (**A**) was replaced by a matrix that also included the genomic information (**H**).

Comparison between the alternate prediction procedures

Three statistics were used to compare the alternate prediction procedures based on 10 simulation replicates, using analysis of variance procedure in SAS version 9.2 (SAS Institute Inc., Cary, NC, USA): (1) the accuracy measured by Pearson's correlation between either DGV or GEBV and TBV in the validation population; (2) the slope of the regression of TBV on either DGV or GEBV ($b1_{TBV,\ DGV\ or\ GEBV}$) in the validation population to evaluate the degree of inflation or deflation of genomic predictions; and (3) the percentage of ranking coincidence when selecting the best 10% and 50% of males and females for breeding, respectively.

Results

Accuracy of breeding values

There were no significant differences ($P > 0.05$) in DGV accuracies generated by ssGBLUP and tsGBLUP (using EBVs or dEBVs as pseudo-phenotypes) procedures at the same level of polygenic effect in the combined genomic relationship matrix. The different levels of polygenic effect added to **G** did not result in significant differences in the DGV accuracies, except for the DGV by ssGBLUP procedure in the levels of polygenic effect 15% and 20%. When ssGBLUP and tsGBLUP procedures considered the default of polygenic effects (5% in BLUPF90 and 20% in GEBV software), the DGV accuracies were 0.584 and 0.564, respectively (Table 1).

There were no significant differences ($P > 0.05$) in GEBV accuracies by ssGBLUP and tsGBLUP procedures within the same level of polygenic effect in the genomic relationship matrix using VP1 and VP2 validation populations. However, there were significant differences ($P < 0.05$) in GEBV accuracies between VanRaden and Hayes blending methods in the tsGBLUP procedure using VP1 and VP2 validation populations. The levels of polygenic effect added to **G** did not have significant ($P > 0.05$) effect on the GEBV accuracies, except for the GEBVs calculated using Hayes blending method in tsGBLUP procedure in the levels 15% and 20%, and using VP1 as validation population (Tables 2 and 3). Considering default contribution of polygenic effects to the genomic relationship matrix for single- (5%) and two-step (20%) GBLUP procedures, the GEBV accuracies for VP1 were 0.589 ssGBLUP and ranged from 0.542 to 0.604 for the alternate implementations of two-step procedure (Table 2), and for VP2 validation population they were 0.699 and ranged from 0.639 to 0.676, respectively (Table 3).

The accuracy of an EBV based on a candidate's parental average (EBV_{PA}) was 0.335 and improved to 0.534 when the phenotypic information of the selection candidates was available and included the genetic evaluation (EBV_{phe}), i.e., increasing the accuracy by 0.199 points (59.4%). When genotypes were added in the ssGBLUP procedure, the $ssGEBV_{PA}$ (GEBV generated by ssGBLUP including only pedigree and genotype information) and $ssGEBV_{phe}$ (GEBV generated by single-step procedure including pedigree, phenotype, and genotype information) accuracies (using default polygenic effect added to G) were 0.589 and 0.699, respectively. These results showed an increase of 0.254 (75.8%) and 0.055 (10.3%), and 0.364 (108.7%) and 0.165 (30.9%) compared with the EBV_{PA} and the EBV_{phe} accuracies, respectively (Table 4).

Table 2. Genomic estimated breeding value (GEBV) accuracies (Acc) and standard deviations (SD) using VP1 validation population.[a]

	$_G95\text{-}_A5$[c]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
GEBV[b]	Acc	SD		Acc	SD		Acc	SD		Acc	SD	
ssGEBV	0.589	0.019	a,ab	0.583	0.019	a,ab	0.577	0.019	a,ab	0.569	0.019	a,bc
tsGEBVv_debv	0.612	0.016	a,a	0.610	0.017	a,a	0.608	0.017	a,a	0.604	0.017	a,a
tsGEBVv_ebv	0.611	0.017	a,a	0.609	0.017	a,a	0.605	0.017	a,a	0.601	0.017	a,ab
tsGEBVh_debv	0.570	0.020	a,b	0.562	0.020	a,b	0.554	0.021	a,b	0.545	0.021	a,c
tsGEBVh_ebv	0.566	0.021	a,b	0.559	0.021	a,b	0.551	0.021	a,b	0.542	0.021	a,c

Note: Different letters indicate significant differences ($P < 0.05$) by Scheffé's test. The letters before the comma indicate differences within rows, whereas the letters after the comma indicate differences within columns.

[a]VP1 validation population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals.

[b]ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv, and tsGEBVh_ebv predicted by two-step procedure using de-regressed EBV (dEBV) or EBV as pseudo-phenotype, and VanRaden "v" or Hayes "h" blending method.

[c]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

Table 3. Genomic estimated breeding value (GEBV) accuracies (Acc) and standard deviations (SD) using VP2 validation population.[a]

	$_G95\text{-}_A5$[c]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
GEBV[b]	Acc	SD		Acc	SD		Acc	SD		Acc	SD	
ssGEBV	0.699	0.016	a,a	0.696	0.016	a,a	0.692	0.016	a,a	0.687	0.016	a,a
tsGEBVv_debv	0.685	0.016	a,a	0.683	0.016	a,a	0.680	0.016	a,a	0.676	0.016	a,a
tsGEBVv_ebv	0.684	0.016	a,a	0.681	0.016	a,a	0.678	0.016	a,a	0.674	0.016	a,a
tsGEBVh_debv	0.655	0.016	a,b	0.650	0.016	a,b	0.645	0.016	a,b	0.640	0.016	a,b
tsGEBVh_ebv	0.653	0.016	a,b	0.649	0.016	a,b	0.644	0.016	a,b	0.639	0.016	a,b

Note: Different letters indicate significant differences ($P < 0.05$) by Scheffé's test. The letters before the comma indicate differences within rows, whereas the letters after the comma indicate differences within columns.

[a]VP2 validation population formed with all animals born in the 10th generation and including the phenotypes of genotyped animals.

[b]ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv, and tsGEBVh_ebv predicted by two-step procedure using de-regressed EBV (dEBV) or EBV as pseudo-phenotype, and VanRaden "v" or Hayes "h" blending methods.

[c]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

Table 4. Accuracies (Acc) and standard deviations (SD) using single-step procedure compared with traditional estimated breeding value (EBV).[a]

	$_G95\text{-}_A5$[b]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
Genetic merit	Acc	SD		Acc	SD		Acc	SD		Acc	SD	
EBV_{PA}	0.335	0.019	d	0.335	0.019	d	0.335	0.019	d	0.335	0.019	d
EBV_{phe}	0.534	0.017	c	0.534	0.017	c	0.534	0.017	c	0.534	0.017	c
$ssDGV_{PA}$	0.584	0.018	b	0.571	0.018	b	0.554	0.018	bc	0.535	0.020	c
$ssGEBV_{PA}$	0.589	0.019	b	0.583	0.019	b	0.577	0.019	b	0.569	0.019	b
$ssDGV_{phe}$	0.695	0.016	a	0.686	0.015	a	0.674	0.014	a	0.661	0.014	a
$ssGEBV_{phe}$	0.699	0.016	a	0.696	0.016	a	0.692	0.016	a	0.687	0.016	a

Note: Different letters indicate significant differences ($P < 0.05$) by Scheffé's test. The letters indicate differences within columns. SD, standard deviations.

[a]EBV_{PA} are EBVs parent's average, EBV_{phe} are traditional EBVs, ssDGVpa and $ssDGV_{phe}$ are direct genomic values, $ssGEBV_{PA}$ and $ssGEBV_{phe}$ are genomic estimated breeding values. The "$_{PA}$" means that the analyses were performed with validation population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals (VP1) and "$_{phe}$" including the phenotypes of genotyped animals (VP2).

[b]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

Table 5. Accuracies (Acc) and standard deviations (SD) using two-step procedure compared with traditional estimated breeding value (EBV).[a]

	$_G95\text{-}_A5$[b]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
Genetic merit	Acc	SD		Acc	SD		Acc	SD		Acc	SD	
EBV_{PA}	0.335	0.019	f	0.335	0.019	f	0.335	0.019	f	0.335	0.019	e
EBV_{phe}	0.534	0.017	e	0.534	0.017	e	0.534	0.017	e	0.534	0.017	d
tsDGV	0.577	0.020	d	0.573	0.018	d	0.569	0.019	d	0.564	0.019	d
tsGEBVh_pa	0.570	0.018	d	0.562	0.020	de	0.554	0.021	de	0.545	0.021	d
tsGEBVv_pa	0.612	0.016	c	0.610	0.017	c	0.608	0.017	c	0.604	0.017	c
$tsGEBVh_{\_phe}$	0.655	0.016	b	0.650	0.016	b	0.645	0.016	b	0.640	0.016	b
$tsGEBVv_{\_phe}$	0.685	0.016	a	0.683	0.016	a	0.680	0.016	a	0.676	0.016	a

Note: Different letters indicate significant differences ($P < 0.05$) by Scheffé's test. The letters indicate differences within columns.

[a]EBV_{PA} are EBVs parent's average, EBV_{phe} are traditional EBVs, tsDGV is two-step direct genomic value, tsGEBVh_pa, tsGEBVv_pa, $tsGEBVh_{\_phe}$, and $tsGEBVv_{\_phe}$ are genomic estimated breeding value. The "pa" means that the analyses were performed with validation population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals (VP1) and "$_{phe}$" including the phenotypes of genotyped animals (VP2). The "h" means that the Hayes blending method and the "v" means that the VanRaden blending method.

[b]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

Table 6. Regression coefficients [b1; regression of true breeding value on direct genomic value (DGV)] and standard deviations (SD).[a]

	$_G95\text{-}_A5$[b]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
DGV	b1	SD		b1	SD		b1	SD		b1	SD	
ssDGV	0.944	0.033	ns	0.919	0.036	*	0.882	0.040	*	0.837	0.045	*
tsDGVdebv	0.982	0.032	ns	1.005	0.034	ns	1.026	0.036	ns	1.045	0.038	ns
tsDGVebv	1.038	0.036	ns	1.063	0.038	ns	1.085	0.041	*	1.106	0.043	*

Note: ns means statistically not different from 1.00 ($P > 0.05$) and * means statistically different from 1.00 ($P < 0.05$).

[a]ssDGV predicted by single-step procedure and tsDGVdebv and tsDGVebv predicted by two-step procedure using de-regressed estimated breeding value (dEBV) or EBV.

[b]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

By adding genotypes in tsGBLUP with VanRaden blending method and default polygenic effect in the genomic relationship matrix, the tsGEBVv_pa (GEBV generated by two-step procedure, including solely pedigree and genotype information and using the VanRaden blending method) and $tsGEBVv_{\_phe}$ (GEBV generated by two-step procedure, including pedigree, phenotype, and genotype information and using the VanRaden blending method) accuracies were 0.604 and 0.676, respectively, showing an increase of 0.269 (80.3%) and 0.07 (13.1%), and 0.341 (101.8%) and 0.142 (26.6%) in comparison to the EBV_{PA} and the EBV_{phe}, respectively. Using the Hayes blending method, there were decreases in accuracy of 0.06 and 0.04 compared with VanRaden blending method (Table 5).

Bias of DGV and GEBV predictions

Tables 6, 7, and 8 present the regression slopes, which ideally should be close to 1.00, indicating that DGV or GEBV predictions are not biased (inflated or deflated). Analysis with tsGBLUP procedure using dEBV as pseudo-phenotypes in the SNPs estimation yielded a slope of the regression on DGV statistically equal to 1.00 ($P > 0.05$), i.e., no evidence of inflation or deflation. However, for the ssGBLUP procedure the slope of the regression on DGVs was statistically <1.00 ($P < 0.05$), indicating slight inflation of the DGVs.

The slope of the regression on GEBV using tsGBLUP procedure did not show significant deviation from one using the VanRaden blending method. However, the slope of the regression showed deflated GEBVs when using Hayes blending method in both VP1 and VP2 validation populations. On the other hand, bias on GEBV was observed for ssGBLUP procedure in both VP1 and VP2 validation populations.

Selecting 10% of males and 50% of females

The ranking coincidence, when selecting 10% males and 50% females of highest genetic merit individuals

Table 7. Regression coefficients [b1; regression of true breeding value on genomic estimated breeding value (GEBV)] and standard deviations (SD) using VP1 validation population.[a]

	$_G95\text{-}_A5$[c]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
GEBV[b]	b1	SD		b1	SD		b1	SD		b1	SD	
ssGEBV	0.967	0.032	ns	0.993	0.034	ns	1.014	0.036	ns	1.032	0.037	ns
tsGEBVv_debv	1.035	0.026	ns	1.058	0.027	*	1.079	0.028	*	1.098	0.029	*
tsGEBVv_ebv	1.076	0.027	*	1.099	0.028	*	1.119	0.029	*	1.136	0.030	*
tsGEBVh_debv	1.256	0.053	*	1.271	0.056	*	1.281	0.060	*	1.288	0.063	*
tsGEBVh_ebv	1.298	0.057	*	1.312	0.061	*	1.322	0.064	*	1.327	0.068	*

Note: ns means statistically not different from 1.00 ($P > 0.05$) and * means statistically different from 1.00 ($P < 0.05$).

[a]VP1 validation population formed with all animals born in the 10th generation and not including the phenotypes of genotyped animals.

[b]ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv, and tsGEBVh_ebv predicted by two-step procedure using de-regressed EBV (dEBV) or EBV as pseudo-phenotype, and VanRaden "v" or Hayes "h" blending method.

[c]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

Table 8. Regression coefficients [b1; regression of true breeding value on genomic estimated breeding value (GEBV)] and standard deviations (SD) using VP2 validation population.[a]

	$_G95\text{-}_A5$[c]			$_G90\text{-}_A10$			$_G85\text{-}_A15$			$_G80\text{-}_A20$		
GEBV[b]	b1	SD		b1	SD		b1	SD		b1	SD	
ssGEBV	0.984	0.039	ns	0.995	0.040	ns	1.002	0.042	ns	1.007	0.043	ns
tsGEBVv_debv	1.006	0.025	ns	1.011	0.027	ns	1.013	0.029	ns	1.015	0.031	ns
tsGEBVv_ebv	1.039	0.027	ns	1.041	0.029	ns	1.041	0.031	ns	1.039	0.033	ns
tsGEBVh_debv	1.286	0.034	*	1.293	0.035	*	1.296	0.036	*	1.297	0.037	*
tsGEBVh_ebv	1.341	0.040	*	1.346	0.041	*	1.348	0.042	*	1.348	0.043	*

Note: ns means statistically not different from 1.00 ($P > 0.05$) and * means statistically different from 1.00 ($P < 0.05$).

[a]VP2 validation population formed with all animals born in the 10th generation and including the phenotypes of genotyped animals.

[b]ssGEBV predicted by single-step procedure and tsGEBVv_debv, tsGEBVh_debv, tsGEBVv_ebv, and tsGEBVh_ebv predicted by two-step procedure using de-regressed EBV (dEBV) or EBV as pseudo-phenotype, and VanRaden "v" or Hayes "h" blending method.

[c]$_G95\text{-}_A5$, $_G90\text{-}_A10$, $_G85\text{-}_A15$, and $_G80\text{-}_A20$ refers to 5%, 10%, 15%, and 20% of polygenic effect added to the genomic relationship matrix, respectively.

based on TBV compared with selection based on DGV or GEBV (VP1 and VP2 validation populations), was estimated as another comparison criterion for ssGBLUP and tsGBLUP procedures. There were no statistical differences ($P > 0.05$) in the percentage of coincidence between DGV or GEBV (VP1 and VP2 validation populations) when comparing ssGBLUP and tsGBLUP procedures (data not shown). For DGVs, the percentage of ranking coincidence was ~40% in males and ~70% in females, whereas for GEBV with VP1 validation population it was ~42% in males and ~70% in females and with VP2 validation population the percentage of coincidence was ~48% in males and ~73% in females. The overall TBV mean for the entire population (males and females) was 2.31 units, whereas the TBV average for the 10% elite males and 50% elite females was 3.07 and 2.66 units, respectively. The 10% elite males showed a TBV average 33% greater than the overall mean. When elite males' selection was based on DGV or GEBV the averages were 15% and 26% greater than the overall population mean, respectively. On the other hand, the 50% elite females showed a TBV average 15% greater than the overall population mean. When selecting the 50% elite females based on DGV or GEBV, their average breeding values were between 7% and 11% greater compared with the population mean.

Discussion

Using EBVs or dEBVs in the tsGBLUP procedure

As the TBVs of the individuals are unknown in practical situations, a variety of pseudo-phenotypes such as daughter yield deviations (DYD) (Guo et al. 2010; Ostersen et al. 2011; Baloche et al. 2014), dEBVs, and EBVs (Brito et al. 2011; Boddhireddy et al. 2014; Neves et al. 2014*a*) have been used to estimate SNP effects for

genomic prediction of breeding values. The average reliability for the dataset was 85% (sires in TP had between 50 and 250 progeny) and with this level of reliability, there is no need to de-regress the EBVs as there were no statistical differences in accuracies between the use of dEBV or EBV as pseudo-phenotypes for the estimation of the markers effects (DGVs). Guo et al. (2010) and Neves et al. (2014*a*), based on simulated data, did not observe differences in accuracies when using dEBVs, EBVs, or DYDs as pseudo-phenotypes in the genomic predictions of breeding values. A situation where the whole TP shows high reliability generally does not occur in commercial production systems of beef cattle. However, Boddhireddy et al. (2014) showed that the use of EBVs compared with dEBVs produced greater accuracy values in American Angus cattle.

Different levels of polygenic effect in the genomic relationship matrix

Genomic selection using medium- to high-density SNP chip panels does not cover the whole genome (de Roos et al. 2007) and therefore part of the genetic variation of the trait may not be accounted for by the markers, and could potentially be captured by polygenic effects. Considering that, we analyzed the accuracy of DGVs and GEBVs based on different levels of polygenic effect in the genomic relationship matrix (i.e., 0.05, 0.10, 0.15, and 0.20). With regards to tsGBLUP, the accuracies were always higher (but not statistically significant) when the weight for the markers effect was 0.95 (i.e., 0.05 polygenic effect). In the ssGBLUP procedure, the values were also higher when considering 95% of the markers effect in the prediction of breeding values. When a 20% polygenic effect was considered in predicting breeding values, there were lower accuracies. In contrast, Onogi et al. (2014) studying carcass traits in Black Japanese cattle using ssGBLUP with different weights to the polygenic effect observed greater GEBV accuracies for larger polygenic effect fractions. Moreover, Neves et al. (2014*b*), studying 15 traits in Brazilian Nellore, reported higher accuracies when tsGBLUP considered 20% of polygenic effects. Similar trend was reported by Gao et al. (2012) for 16 traits in Nordic Holstein population. Calus and Veerkamp (2007), working with simulated data and a variety of traits with different heritabilities, concluded that the inclusion of polygenic effect in the model increased the accuracy of DGVs. Liu et al. (2011) showed that, for German Holstein cattle, adjusting for the polygenic effect reduced GEBV bias and concluded that weighting for polygenic effect seems to differ between traits.

Accuracy of breeding values

The DGV accuracies presented in this study (0.584 for ssGBLUP and 0.564 for tsGBLUP), using default polygenic contributions to **G** showed that the use of genetic values obtained solely by the markers produced gains in accuracies of 74% and 68% compared with the EBVs parent's average for single- and two-step procedures, respectively. These accuracies were higher than the accuracies reported by Neves et al. (2014*b*) for traits with similar heritability in Brazilian Nellore and lower than those reported by Boddhireddy et al. (2014) in American Angus cattle. The DGV accuracies observed in this study for ssGBLUP and tsGBLUP procedures were statistically similar, which is in agreement with results reported by Vitezica et al. (2010), based on simulated data in the presence of artificial selection. However, gains in reliability for ssGBLUP compared with tsGBLUP have been reported in the literature (e.g., in Nordic Red cattle, Koivula et al. 2012; and in Lacaune sheep, Baloche et al. 2014). Přibyl et al. (2013) concluded that, if all available data are used in the traditional evaluations of Czech Holstein cattle and the genomic evaluation also exploits all the relevant traditional data and procedures, ssGBLUP should not cause a substantial increment in accuracy compared with tsGBLUP, which is consistent with our results.

The GEBV estimates from single- and two-step (VanRaden blending method) procedures were statistically equal and showed higher accuracies compared with the Hayes blending method in tsGBLUP. Cardoso et al. (2015) studied tick resistance in Hereford and Braford cattle in Brazil also reported greater GEBV accuracy by VanRaden blending method compared with the Hayes blending method. In addition, the authors reported greater GEBV accuracies when implementing ssGBLUP compared with tsGBLUP procedure. Similarly, Su et al. (2012) observed an increase in GEBV accuracies when implementing ssGBLUP compared with tsGBLUP (VanRaden blending method) for Nordic Red cattle.

A common practice in extensive production systems is to undertake the first culling of animals at weaning. Thus, it is possible to measure calves for a variety of traits (e.g., birth weight, weight gain between birth and weaning) to predict EBVs. In this scenario, information from parents and from the animals itself can be used in the EBVs prediction. If an animal was genotyped, this information would also be used for the prediction of breeding values. Therefore, we investigated this scenario in VP2. First, there were gains in accuracy of 59% using only the phenotypes combined with pedigree information with the accuracy increasing from 0.335 to 0.534. When genotypes were included there was a further improvement of 30.9% for ssGBLUP and 26.6% for tsGBLUP with the respective accuracies reaching 0.699 and 0.676, respectively. Based on the results of this study, it seems that the use of DGVs for selection of animals produce similar or greater accuracies in comparison with EBVs (i.e., generated solely based on pedigree and phenotype information).

Bias of DGV and GEBV predictions

The bias of genomic predictions is relevant to determine if DGV or GEBV of younger animals is on the same scale to be comparable with EBV or GEBV of older,

proven animals being, therefore, useful for predicting future differences in the EBVs or GEBVs once they are proven with progeny information. If the regression coefficient of DGV or GEBV on TBV <1.0, it indicates that genomic breeding values were overestimated, whereas regression coefficients >1.0 indicate underestimation of genomic predictions. Vitezica et al. (2011) have discussed the scale of breeding values under the effect of selection. If the parents of the next generation come from only genotyped selection candidates, they share a common mean for belonging to the same generation, then the bias would not be a concern. However, for different selection candidates there is a different amount of information (e.g., progeny tested males and newborn animals) and in the presence of bias, newborns could have an overestimated or underestimated genetic merit. Regarding to the DGV results, the regression coefficients indicated that tsGBLUP generated unbiased estimates when using dEBVs as pseudo-phenotypes in the estimation of SNP effects. In general, the regression coefficients for GEBV from ssGBLUP and tsGBLUP indicated small or no bias, except when using Hayes blending method in tsGBLUP. Similar regression coefficients were reported in the literature (Gao et al. 2012; Su et al. 2012).

Selecting 10% of males and 50% of females

Breeding programs use (traditional or genomic) breeding values to select individuals for breeding or culling. Therefore, we looked at the ranking coincidence with the ranking based on TBV for the 10% best replacement males and 50% best replacement females. Our findings indicated that the percentage of ranking coincidence in selected elite males (~40%) and females (~70%) was similar when using DGV or GEBV estimated by ssGBLUP or tsGBLUP to rank the best males and females (data not shown). These findings are in agreement with the accuracy results, which were similar for ssGBLUP and tsGBLUP procedures.

Our findings indicate that incorporation of genomic information into tradition genetic evaluations might result in gains of accuracy >100% compared with the EBV_{PA}. However, it is worth to notice that these gains could be even more substantial when a higher proportion of missing pedigree information exist than the simulated here (5%), as it would be the case of the use of multiple sire matings. Simulated data are of great value to compare methodologies under alternate scenarios, however, there might exist specificities in a population which will be identified only when using real data. Therefore, studies using real data for a variety of traits with different genetic architecture are warranted to corroborate the findings of this study.

The costs for implementing large-scale genomic selection are still high for beef cattle producers in developing countries. However, it has been already used for difficult or expensive-to-measure traits and traits measured late in life to select breeding candidates more accurately, particularly males. Therefore, the results of this study will help to decide the most appropriate genomic prediction methodology to use in the genomic evaluations with respect to accuracy and bias of genomic predictions.

Conclusions

Direct genomic values and GEBVs predicted by single- and two-step GBLUP procedures showed very similar accuracies, except for the GEBVs generated by Hayes blending method (in the two-step procedure), which were significantly lower. There was no significant or only slight bias on GEBV predictions from single- or two-step (using VanRaden blending) procedures, indicating that these predictions are on the same scale compared with the TBVs. The levels of contribution of polygenic effects to the genomic relationship matrix tested did not significantly affect the GEBV accuracies in single- and two-step procedures. Overall, genetic evaluations including genomic information resulted in gains of accuracy >100% compared with the EBV_{PA}. Moreover, there were also no significant differences between the coincidence of selected animals (10% males and 50% females) using single- and two-step procedures with actual best candidate list based on TBV. Therefore, using single-step or two-step (with VanRaden blending) GBLUP approaches for genomic prediction are equally recommended to increase accuracy of genetic evaluations, regardless the level of contribution of polygenes to the genomic relationship matrix (between 5% and 20%) adopted.

Acknowledgements

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) that provided the graduate fellowship for the first author.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., and Lawlor, T.J. 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. doi:10.3168/jds.2009-2730. PMID:20105546.

Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J.-M., Robert-Granié, C., and Barillet, F. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. J. Dairy Sci. **97**: 1107–1116. doi:10.3168/jds.2013-7135. PMID:24315320.

Boddhireddy, P., Kelly, M.J., Northcutt, S., Prayaga, K.C., Rumph, J., and DeNise, S. 2014. Genomic predictions in Angus cattle: comparisons of sample size, response variables, and clustering methods for cross-validation. J. Anim. Sci. **92**: 485–497. doi:10.2527/jas.2013-6757. PMID:24431338.

Brito, F.V., Neto, J.B., Sargolzaei, M., Cobuci, J.A., and Schenkel, F.S. 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. BMC Genet. **12**: 80. doi:10.1186/1471-2156-12-80. PMID:21933416.

Brito, L.F., Clarke, S.M., McEwan, J.C., Miller, S.P., Pickering, N.K., Bain, W.E., Dodds, K.G., Sargolzaei, M., and Schenkel, F.S. 2017. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. BMC Genet. **18**(1): 7. doi:10.1186/s12863-017-0476-8.

Calus, M., and Veerkamp, R. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J. Anim. Breed Genet. **124**: 362–368. doi:10.1111/j.1439-0388.2007.00691.x. PMID:18076473.

Cardoso, F.F., Gomes, C.C.G., Sollero, B.P., Oliveira, M.M., Roso, V.M., Piccoli, M.L., Higa, R.H., Yokoo, M.J., Caetano, A.R., and Aguilar, I. 2015. Genomic prediction for tick resistance in Braford and Hereford cattle. J. Anim. Sci. **93**: 2693–2705. doi:10.2527/jas.2014-8832. PMID:26115257.

Carillier, C., Larroque, H., and Robert-Granié, C. 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genet. Sel. Evol. **46**: 67. doi:10.1186/s12711-014-0067-3. PMID:25927866.

Chen, C.Y., Misztal, I., Aguilar, I., Tsuruta, S., and Meuwissen, T.H.E. 2010. Genome-wide marker assisted selection in chicken : making the most of all data, pedigree, phenotypic, and genomic in a simple one-step procedure. Proc. 9th World Congress of Genetics Applied to Livestock Production, Leipzig, Germany, 1–6 Aug. 2010.

Chen, C.Y., Misztal, I., Aguilar, I., Tsuruta, S., Meuwissen, T.H.E., Aggrey, S.E., Wing, T., and Muir, W.M. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. J. Anim. Sci. **89**: 23–28. doi:10.2527/jas.2010-3071. PMID:20889689.

Christensen, O.F., and Lund, M.S. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. **42**: 2. doi:10.1186/1297-9686-42-2. PMID:20105297.

de Roos, A.P.W., Schrooten, C., Mullaart, E., Calus, M.P.L., and Veerkamp, R.F. 2007. Breeding value estimation for fat percentage using dense markers on *Bos taurus* autosome 14. J. Dairy Sci. **90**: 4821–4829. doi:10.3168/jds.2007-0158. PMID:17881705.

Gao, H., Christensen, O.F., Madsen, P., Nielsen, U.S., Zhang, Y., Lund, M.S., and Su, G. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genet. Sel. Evol. **44**: 8. doi:10.1186/1297-9686-44-8. PMID:22455934.

Garrick, D.J. 2010. The nature, scope and impact of some whole-genome analyses in beef cattle. Proc. 9th World Congress of Genetics Applied to Livestock Production, Leipzig, Germany, 1–6 Aug. 2010.

Guo, G., Lund, M.S., Zhang, Y., and Su, G. 2010. Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. J. Anim. Breed Genet. **127**: 423–432. doi:10.1111/j.1439-0388.2010.00878.x. PMID:21077966.

Harris, B.L., and Johnson, D.L. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. **93**: 1243–1252. doi:10.3168/jds.2009-2619. PMID:20172244.

Harris, B.L., Winkelman, A.M., and Johnson, D.L. 2013. Impact of including a large number of female genotypes on genomic selection. Proc. Interbull Bulletin No. 47, Nantes, France, 23–25 Aug. 2013.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., and Goddard, M.E. 2009. Invited review: genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. **92**: 433–443. doi:10.3168/jds.2008-1646. PMID:19164653.

Koivula, M., Strandén, I., Su, G., and Mäntysaari, E.A. 2012. Different methods to calculate genomic predictions — comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). J. Dairy Sci. **95**: 4065–4073. doi:10.3168/jds.2011-4874. PMID:22720963.

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

Liu, Z., Seefried, F.R., Reinhardt, F., Rensing, S., Thaller, G., and Reents, R. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. **43**: 19. doi:10.1186/1297-9686-43-19. PMID:21586131.

Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, **157**: 1819–1829. PMID:11290733.

Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. 2016. Genomic selection: a paradigm shift in animal breeding. Anim. Front. **6**: 6–14. doi:10.2527/af.2016-0002.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D.H. 2002. BLUPF90 and related programs (BGF90). Proc. 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, 19–23 Aug. 2002.

Misztal, I., Legarra, A., and Aguilar, I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. **92**: 4648–4655. doi:10.3168/jds.2009-2064. PMID:19700728.

Neves, H.H.R., Carvalheiro, R., and Queiroz, S.A. 2014*a*. Adequacy of different pseudo-phenotypes for model training and validation of genomic predictions in beef cattle. Proc. 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada, 17–22 Aug. 2014.

Neves, H.H.R., Carvalheiro, R., O'Brien, A.M.P., Utsunomiya, Y.T., do Carmo, A.S., Schenkel, F.S., Sölkner, J., McEwan, J.C., Van Tassell, C.P., Cole, J.B., da Silva, M.V.G.B., Queiroz, S.A., Sonstegard, T.S., and Garcia, J.F. 2014*b*. Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. Genet. Sel. Evol. **46**: 17. doi:10.1186/1297-9686-46-17.

Onogi, A., Ogino, A., Komatsu, T., Shoji, N., Simizu, K., Kurogi, K., Yasumori, T., Togashi, K., and Iwata, H. 2014. Genomic prediction in Japanese Black cattle: application of a single-step approach to beef cattle. J. Anim. Sci. **92**: 1931–1938. doi:10.2527/jas.2014-7168. PMID:24782393.

Ostersen, T., Christensen, O.F., Henryon, M., Nielsen, B., Su, G., and Madsen, P. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. Genet. Sel. Evol. **43**: 38. doi:10.1186/1297-9686-43-38. PMID:22070746.

Piccoli, M.L., Brito, L.F., Braccini, J., Cardoso, F.F., Sargolzaei, M., and Schenkel, F.S. 2017. Genomic predictions for economically important traits in Brazilian Braford and Hereford beef cattle using true and imputed genotypes. BMC Genet. **18**(1): 2. doi:10.1186/s12863-017-0475-9.

Přibyl, J., Madsen, P., Bauer, J., Přibylová, J., Simečková, M., Vostrý, L., and Zavadilová, L. 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. J. Dairy Sci. **96**: 1865–1873. doi:10.3168/jds.2012-6157.

Pryce, J.E., Johnston, J., Hayes, B.J., Sahana, G., Weigel, K.A., McParland, S., Spurlock, D., Krattenmacher, N., Spelman, R.J., Wall, E., and Calus, M.P.L. 2014. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. J. Dairy Sci. **97**: 1799–1811. doi:10.3168/jds.2013-7368. PMID:24472132.

Sargolzaei, M., and Schenkel, F.S. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics, **25**: 680–681. doi:10.1093/bioinformatics/btp045. PMID:19176551.

Sargolzaei, M., Schenkel, F.S., and VanRaden, P.M. 2009. GEBV: genomic breeding value estimator for livestock.

Proc. Technical report to the Dairy Cattle Breeding and Genetics Committee, University of Guelph, Guelph, ON, Canada.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed Genet. **123**: 218–223. doi:10.1111/j.1439-0388.2006.00595.x. PMID:16882088.

Snelling, W.M., Chiu, R., Schein, J.E., Hobbs, M., Abbey, C.A., Adelson, D.L., Aerts, J., Bennett, G.L., Bosdet, I.E., Boussaha, M., Brauning, R., Caetano, A.R., Costa, M.M., Crawford, A.M., Dalrymple, B.P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C.A., Green, R.D., Holt, R., Jann, O., Jones, S.J., Kappes, S.M., Keele, J.W., de Jong, P.J., Larkin, D.M., Lewin, H.A., McEwan, J.C., McKay, S., Marra, M.A., Mathewson, C.A., Matukumalli, L.K., Moore, S.S., Murdoch, B., Nicholas, F.W., Osoegawa, K., Roy, A., Salih, H., Schibler, L., Schnabel, R.D., Silveri, L., Skow, L.C., Smith, T.P., Sonstegard, T.S., Taylor, J.F., Tellam, R., Van Tassell, C.P., Williams, J.L., Womack, J.E., Wye, N.H., Yang, G., and Zhao, S. 2007. A physical map of the bovine genome. Genome Biol. **8**: R165. doi:10.1186/gb-2007-8-8-r165. PMID:17697342.

Su, G., Madsen, P., Nielsen, U.S., Mäntysaari, E.A., Aamand, G.P., Christensen, O.F., and Lund, M.S. 2012. Genomic prediction for Nordic Red cattle using one-step and selection index blending. J. Dairy Sci. **95**: 909–917. doi:10.3168/jds.2011-4804. PMID:22281355.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. **91**: 4414–4423. doi:10.3168/jds.2007-0980. PMID:18946147.

VanRaden, P.M., and Wiggans, G.R. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. **74**: 2737–2746. doi:10.3168/jds.S0022-0302(91)78453-1. PMID:1918547.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., and Schenkel, F.S. 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. **92**: 16–24. doi:10.3168/jds.2008-1514. PMID:19109259.

Vitezica, Z.G., Aguilar, I., and Legarra, A. 2010. One-step vs. multi-step methods for genomic prediction in presence of selection. Proc. 9th World Congress of Genetics Applied to Livestock Production, Leipzig, Germany, 1–6 Aug. 2010.

Vitezica, Z.G., Aguilar, I., Misztal, I., and Legarra, A. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb). **93**: 357–366. doi:10.1017/S001667231100022X. PMID:21767459.