Genomic prediction using different estimation methodology, blending and cross-validation techniques for growth traits and visual scores in Hereford and Braford cattle¹

Gabriel Soares Campos,^{*,2} Fernando Antônio Reimann,^{*} Leandro Lunardini Cardoso,[†] Carlos Eduardo Ranquetat Ferreira,[‡] Vinicius Silva Junqueira,[§],
Patricia Iana Schmidt,^{*} José Braccini Neto,^{*,1} Marcos Jun Iti Yokoo,[†] Bruna Pena Sollero,[†] Arione Augusti Boligon,^{*,1} and Fernando Flores Cardoso^{*,†,1}

^{*}Departamento de Zootecnia, Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul 96010-900, Brazil;

[†]Embrapa Pecuária Sul, Bagé, Rio Grande do Sul 96401-970, Brazil; [‡]Faculdade de Medicina Veterinária, Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul 96010-900, Brazil; [§]Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil; and; [#]Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul 91540-000, Brazil

ABSTRACT: The objective of the present study was to evaluate the accuracy and bias of direct and blended genomic predictions using different methods and cross-validation techniques for growth traits (weight and weight gains) and visual scores (conformation, precocity, muscling, and size) obtained at weaning and at yearling in Hereford and Braford breeds. Phenotypic data contained 126,290 animals belonging to the Delta G Connection genetic improvement program, and a set of 3,545 animals genotyped with the 50K chip and 131 sires with the 777K. After quality control, 41,045 markers remained for all animals. An animal model was used to estimate (co)variance components and to predict breeding values, which were later used to calculate the deregressed estimated breeding values (DEBV). Animals with genotype and phenotype for the traits studied were divided into 4 or 5 groups by random and *k*-means clustering cross-validation strategies. The values of accuracy of the direct genomic values (DGV) were moderate to high magnitude for at weaning and at yearling traits, ranging from 0.19 to 0.45 for the *k*-means and

0.23 to 0.78 for random clustering among all traits. The greatest gain in relation to the pedigree BLUP (PBLUP) was 9.5% with the BayesB method with both the *k*-means and the random clustering. Blended genomic value accuracies ranged from 0.19 to 0.56 for *k*-means and from 0.21 to 0.82 for random clustering. The analyses using the historical pedigree and phenotypes contributed additional information to calculate the GEBV, and in general, the largest gains were for the single-step (ssGBLUP) method in bivariate analyses with a mean increase of 43.00% among all traits measured at weaning and of 46.27% for those evaluated at yearling. The accuracy values for the marker effects estimation methods were lower for *k*-means clustering, indicating that the training set relationship to the selection candidates is a major factor affecting accuracy of genomic predictions. The gains in accuracy obtained with genomic blending methods, mainly ssGBLUP in bivariate analyses, indicate that genomic predictions should be used as a tool to improve genetic gains in relation to the traditional PBLUP selection.

Key words: accuracy prediction, beef cattle, genomic selection, visual evaluation, weight gain

¹The research was supported by Coordination for the Improvement of Higher Education Personnel—CAPES/Embrapa grant 15/2014—Proposal 77 and by Embrapa-Brazilian Agricultural Research Corporation Grants 02.13.10.002 and 02.13.14.014. A.A.B., J.B.N., and F.F.C. are scholars of the National Council for

Scientific and Technological Development (CNPq). The authors acknowledge the Delta G Connection for providing data for this research.

²Corresponding author: gabrielsoarescampos@hotmail.com

Received January 30, 2018.

Accepted May 5, 2018.

© The Author(s) 2018. Published by Oxford University Press on behalf of the American Society of Animal Science. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

J. Anim. Sci. 2018.96:2579–2595

doi: 10.1093/jas/sky175

INTRODUCTION

With the advancement in DNA sequencing and genotyping technologies, it has become possible to evaluate animals using millions of genetic variants and to estimate their effect on phenotype of interest (Aguilar et al., 2010). The marker effects can be added up to predict direct genomic values (DGV) of any genotyped animal, and this can also be combined with the traditional breeding value to obtain the genomic enhanced breeding value (GEBV; VanRaden et al., 2009). Multistep (VanRaden et al., 2009; MacNeil et al., 2010) and single-step (Misztal et al., 2009; Aguilar et al., 2010) procedures have been proposed and are being tested to identify the most appropriate strategy for the implementation of genomic selection in genetic evaluation and breeding programs.

Growth traits and visual scores are widely used as selection criteria in beef cattle evaluation, and, therefore, it is expected that the use of genomic predictions for these traits may improve the efficiency of breeding programs. This benefit could be obtained as an additional product of ongoing efforts to use genomics in traits that are difficult to measure, such as the tick resistance (Cardoso et al., 2015), because the animals that have been genotyped also have growth and visual score measures.

The hypothesis of the study was that methods for DGV blending with the traditional genetic evaluation enables to predict more accurate breeding values for growth and visual score measures

compared with the methods that use only markers or pedigree. Therefore, aim of this study was to evaluate the prediction accuracy by different methods of estimating the markers effects and blending DGV with historical information using 2 cross-validation strategies for growth traits and visual scores in Hereford and Braford cattle breeds.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because no animals were used. The data were extracted from existing databases.

Phenotype, Genotype, and Pedigree Data

Phenotypic records of 126,290 animals of the Hereford and Braford cattle born between 1991 and 2012, belonging to Conexão Delta G breeding program, located in the Brazilian state of Rio Grande do Sul, were used. The studied traits were as follows: birth weight (BW0), weaning weight adjusted to 205 d of age (WW205), weaning conformation (WC), weaning precocity (WP), weaning muscling (WM), weaning size (WS), yearling weight adjusted to 550 d of age (YW550), postweaning weight gain adjusted for 345 d of age (WWG345), yearling conformation (YC), yearling precocity (YP), yearling muscling (YM), and yearling size (YS; Table 1).

The animals were weighted at birth, at weaning (about 205 d of age) and at yearling (about 550 d

Table 1. Description of the data set used in the analyses for growth and visual scores traits evaluated in Hereford and Braford breeds

Traits	Number of observations	Mean $\pm$ SD	Number of CG ¹	Number of sires	Number of dams
Birth weight (kg)	105,596	33.00 $\pm$ 5.32	1,045	2,162	41,279
Weaning weight (kg)	113,022	179.53 $\pm$ 32.87	2,884	2,226	43,886
Weaning conformation (1 to 5)	112,696	3.11 $\pm$ 1.04	2,864	2,219	43,767
Weaning precocity (1 to 5)	112,723	3.18 $\pm$ 1.07	2,868	2,219	43,783
Weaning muscling (1 to 5)	112,735	3.08 $\pm$ 1.08	2,870	2,219	43,798
Weaning size (1 to 5)	100,975	3.04 $\pm$ 0.98	2,535	2,006	39,956
Yearling weight (kg)	60,422	315.87 $\pm$ 77.37	3,285	1,971	30,546
Postweaning weight gain (kg)	59,617	133.80 $\pm$ 65.08	3,261	1,965	30,295
Yearling conformation (1 to 5)	60,641	3.20 $\pm$ 0.96	3,278	1,964	30,701
Yearling precocity (1 to 5)	60,879	3.20 $\pm$ 1.01	3,304	1,964	30,784
Yearling muscling (1 to 5)	60,848	3.14 $\pm$ 1.02	3,300	1,964	30,773
Yearling size (1 to 5)	53,849	3.17 $\pm$ 0.95	2,902	1,789	27,559

¹CG = contemporary groups.

of age), in addition to being visually evaluated at weaning and at yearling, receiving individual scores from 1 to 5 for conformation, precocity, muscling, and size, where 5 refers to the maximum expression of each trait.

A total of 3,545 genotyped animals with the Illumina BovineSNP50 Bead Chip (50K; Illumina, San Diego, CA) and 131 genotyped sires with the Illumina High-Density Beef Chip Array (HD; Illumina) were used. Genotype quality control was implemented using R/SNPStats package (Clayton, 2014). Samples with genotyping rate [call rate (CR)] < 0.90, heterozygosity with 3 SD above or below the observed mean, mismatching sex, and duplicate records were removed. Only SNPs mapped to autosomes with CR > 0.98, minor allele frequencies (MAF) > 0.03, and not in highly significant deviation Hardy–Weinberg equilibrium ($P > 10^{-7}$) were considered in the analyses. In addition, only the SNP with highest MAF was retained when SNPs were observed in the same position or the genotypes were highly correlated ($r > 0.98$). Data from HD chip were filtered to select only SNPs that were also present on the 50K panel. After editing and merging the 2 bases of markers, HD and 50K, 41,045 SNPs and 3,592 samples remained to carry out the analysis, including 2,934 Braford and 658 Hereford animals. Missing genotypes (0.86% of all genotypes) were imputed using the FImpute software (Sargolzaei et al., 2011).

Statistical Models and Analyses

The (co)variance components and genetic parameters were estimated using Bayesian inference by Gibbs sampling, with the Gibbs1f90 program (Misztal et al., 2002) in single-trait analyses. These estimated (co)variance components were used to predict breeding values using the BLUPF90 software (Misztal et al., 2002). The following linear mixed model was applied:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Spe} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{u}$ is the vector of random direct additive genetic effects, $\mathbf{m}$ is the vector of random maternal additive genetic effects, $\mathbf{pe}$ is the vector of maternal permanent environmental effects, $\mathbf{e}$ is the vector of random residual effects, and $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$, and $\mathbf{S}$ are incidence matrices relating the observations to fixed, direct additive genetic, maternal additive genetic, and maternal permanent environmental effects, respectively. According to preliminary analyses, the maternal genetics and permanent

environmental effects were not significant for yearling traits studied (YW550, WWG345, YC, YP, YM, and YS); therefore, these effects were not included in the model for these traits.

The fixed effects of contemporary groups (CG—same farm, sex, year and season of birth, management group, and date of weaning or yearling), age of dam at calving (15 age classes, in which class 1 corresponded to cows with 2 yr of age and 15 to cows older than 16 yr), linear and quadratic effects of animal age at measurement (in days), and linear covariates for breed and heterozygosity effects were considered. In addition, linear maternal breed and heterozygosity effects were fit only for weaning traits (BW0, WW205, WC, WP, WM, and WS).

For continuous traits (BW0, WW205, WWG345, and YW550), CG with less than 5 animals and data exceeding 3.5 SD above or below the mean of the CG were excluded. For visual scores, CG with less than 5 animals and without variability were eliminated (Table 1).

Genomic Prediction

From the EBV obtained by model 1, the approach proposed by Garrick et al. (2009) was used to calculate deregressed breeding values (DEBV) and corresponding weights ($\mathbf{w}$) = $\{w_j\}$ for all animals, free of parent average breeding values and adjusted for the respective reliabilities. The weighted DEBV were used as response variables to estimate marker effects through different methods. In all cases, the same linear model was fitted:

$$\text{DEBV}_j = \mu + \sum_{k=1}^k M_{jk} g_k + \varepsilon_j \quad (2)$$

where DEBV_j is the DEBV of animal j , μ is the overall mean, $k = 41,045$ is the total number of SNP markers in the panel, M_{jk} is a covariate representing the genotype of animal j at marker k by the number of alleles in the Illumina A/B coding (Illumina, 2006), g_k is the random effect of marker k , and ε_j is a residual term with normal distribution and heterogeneous variance, $\varepsilon_j \sim N(0, w_j^{-1} \sigma_e^2)$, in which σ_e^2 is the residual variance and w_j is the weight as a function of the reliability of DEBV_j (Garrick et al., 2009).

Because the number of markers with effect on the traits exceeds the number of available phenotypic observations, restrictions must be applied in the model to make the marker effects estimable (Gianola et al., 2009). The differences in the methods used in genomic selection result from different

assumptions regarding the distributions of marker effects: 1) SNP_BLUP (Meuwissen et al., 2001), the marker effects are considered to belong to the same normal distribution and assuming that all markers have the common variance σ_g^2 , that is, $g_k \sim N(0, \sigma_g^2)$ for all k ; 2) BayesB (Meuwissen et al., 2001) assumes a mixture model, in which a fraction of the markers has null effect ($\pi = 0.95$), whereas the remaining fraction of markers ($1 - \pi$) has effects drawn from a normal distributions, $g_k \sim N(0, \sigma_{g_k}^2)$, with site specific variances $\sigma_{g_k}^2$, derived from a priori inverse scaled chi-square distribution, that is, $\sigma_{g_k}^2 \sim \chi_{(\%S^2)}^{-2}$; and 3) BayesC (Habier et al., 2011) is also a mixture model similar to BayesB but assumes common variance for the distribution of the nonzero markers effects ($1 - \pi$), that is, $g_k \sim N(0, \sigma_g^2)$.

The DGV were obtained as $DGV_j = \sum_{k=1}^k M_{jk} \hat{g}_k$,

in which the estimated SNP effects were represented by their posterior means in the Bayesian methods.

The SNP_BLUP was implemented by equivalent genomic BLUP (GBLUP; Strandén and Garrick, 2009) using the BLUPf90 software (Miszta et al., 2002). The Bayesian methods (BayesB and BayesC) were implemented using Gensel software (Fernando and Garrick, 2009). For Bayesian methods using Gensel program, was used a chain size of 42,000 samples after 2,000 burn-in cycles, and hyperparameters for variance component distributions were specified based on estimates for model 1. The convergence was verified with the criterion proposed by Geweke (1992), through “boa” package, in R program (Smith, 1997).

Blending of Traditional Information With Genomics

Three alternative approaches were used to combine the DGV with the pedigree and performance data to obtain GEBV.

The first approach was performed based on the selection index (SI) proposed by VanRaden et al. (2009). The first set of indexes is DGVs; the second set is calculated for a subset of the genotyped animals using traditional relationships and their DEBV, excluding the data from nongenotyped animals (SPA); and the third set is the traditional EBV obtained from model 1 using historical data (PA). This approach allows calculating the GEBV of each animal as the difference between the genomic and traditional predictions added to the full data PA by the following index (VanRaden et al., 2009):

$$GEBV = b_1 DGV + b_2 SPA + b_3 PA \quad (3)$$

where $\mathbf{b} = [b_1 \ b_2 \ b_3] = \mathbf{c}'\mathbf{V}^{-1}$ and $\mathbf{c}'$ is a vector with the reliabilities corresponding to the 3 terms of the index and $\mathbf{V}$ is a 3×3 symmetric matrix with elements of $\mathbf{c}$ on the diagonals and the following functions of those 3 reliabilities on the off-diagonals:

$$V_{12} = V_{23} = r_{SPA}^2, \text{ and}$$

$$V_{13} = r_{SPA}^2 + (r_{DGV}^2 - r_{SPA}^2)(r_{PA}^2 - r_{SPA}^2) / (1 - r_{SPA}^2).$$

For all animals, $V_{11} (r_{DGV}^2)$ and $V_{33} (r_{PA}^2)$ were constrained to be greater than $V_{22} (r_{SPA}^2)$ to ensure positive-definite matrices (VanRaden et al., 2009).

In a second approach, GEBV was performed based on a bitrait analysis between the phenotype and the DGV using the traditional pedigree (MacNeil et al., 2010). Different linear models were assumed for each trait. The same model 1 was used for the phenotypic data, whereas for the DGV, the following model was assumed: $DGV = \mu + \alpha_j + \varepsilon_j$, where μ is an overall mean, α_j is the genetic additive effect of j th animal, and ε_j is the residual effect. In the bivariate model, we assumed for the additive genetic effects of the phenotypes ($\mathbf{a}$) and for $\alpha = \{\alpha_j\}$ the following distribution:

$$\begin{bmatrix} \mathbf{a} \\ \alpha \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{A}\sigma_a^2 & \mathbf{A}\sigma_{a\alpha} \\ \mathbf{A}\sigma_{a\alpha} & \mathbf{A}\sigma_\alpha^2 \end{bmatrix} \right),$$

where σ_a^2 is the additive genetic variance for phenotype, σ_α^2 is the additive genetic variance for DGV, and $\sigma_{a\alpha}$ is the genetic covariance between phenotype and DGV and was assumed for random residual effects for the phenotype ($\mathbf{e}$) and for DGV $\varepsilon = \{\varepsilon_j\}$ the following distribution:

$$\begin{bmatrix} \mathbf{e} \\ \varepsilon \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}\sigma_e^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_\varepsilon^2 \end{bmatrix} \right),$$

where in the bitrait model, there is a borrowing information between the traditional breeding value and the DGV through covariance ($\sigma_{a\alpha}$), and both estimates of a_j and α_j are candidate values to represent the GEBV of animal j .

The third and last approach used was to combine directly the phenotypic, pedigree, and genomic information using the single-step GBLUP (ssGBLUP; Miszta et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). In ssGBLUP, the same linear model 1 was used, but the traditional relationship matrix ($\mathbf{A}$) was replaced by $\mathbf{H}$ matrix, which includes the genomic information. Although $\mathbf{H}$ is complex (Legarra et al., 2009), its inverse has a simple form (Aguilar et al., 2010):

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

where $\mathbf{G}$ is the genomic relationship matrix constructed as shown by VanRaden (2008) using current allelic frequencies and $\mathbf{A}_{22}$ is a numerator relationship matrix only for the genotyped animals.

The analyses were performed using the BLUPf90 software, and ssGBLUP program was implemented using preGSf90 software with default options for adjusting the matrices $\mathbf{G}$ and $\mathbf{A}_{22}$ and making $\mathbf{G}$ positive-definite (Aguilar and Misztal, 2014). An alternative approach was tested using ssGBLUP program and bivariate animal model (MTssGBLUP), to evaluate the effect of the preselection in the GEBV accuracy of yearling traits. The following bivariate analyses were performed: WW205 with YW550 and with PWG345, and each visual score measured at weaning with the same score at yearling (WC-YC, WP-YP, WM-YM, and WS-YS).

Cross-validation and Prediction Accuracy

The genotyped animals with phenotypes for the studied traits were divided into 5 groups for the traits at weaning and 4 groups for the traits at yearling by 2 cross-validation strategies using the R program (R Core Team, 2013): grouping according to genomic relationship by k -means clustering or at random according to Saatchi et al. (2011) for the purpose of forming mutually exclusive training and validation groups. These strategies were used to represent 2 scenarios of genomic selection where animals are more distantly (k -means) or more closely (random) related to the training set.

The k -means method clustered the individuals according to genomic relationship distance, so that the genetic relationship within each group was greater than between groups. The other strategy partitioned the animals by selecting individuals at random (Saatchi et al., 2011).

Cross-validation was performed in the group not used in the training, that is, in each analysis, data were excluded from a one group (validation set) and data were used to estimate the DGVs and GEBVs from the remaining groups (training set). The same cross-validation strategy was applied using traditional pedigree BLUP (PBLUP) as a way of comparing genomic predictions and to generate GEBVs in genomic enhanced methods.

The prediction accuracies of the genotyped animals from the validation sets were derived as the genetic correlation between the observed phenotype

(y) and the DGV or GEBV, estimated in a bivariate animal model, as demonstrated below:

$$\begin{bmatrix} \mathbf{y} \\ \text{DGV / GEBV} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of fixed effects (same effects used in Eq. 1 for $\boldsymbol{\beta}_1$ and class effect of the groups used in the validation for $\boldsymbol{\beta}_2$); $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are vectors of random direct additive genetic effects for the 2 traits, $\text{Var}(\boldsymbol{\alpha}_1) = \mathbf{A}^* \boldsymbol{\sigma}_{\alpha_1}^2$, $\text{Var}(\boldsymbol{\alpha}_2) = \mathbf{A}^* \boldsymbol{\sigma}_{\alpha_2}^2$, and $\text{Cov}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) = \mathbf{A}^* \boldsymbol{\sigma}_{\alpha_1, \alpha_2}$, where $\mathbf{A}^*$ is a pedigree numerator relationship matrix, in which covariance between individuals in different groups was set to zero (Saatchi et al., 2013):

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{44} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{55} \end{bmatrix}$$

The (co)variance components and genetic correlations were estimated by Bayesian inference based on the Markov chain Monte Carlo methods using 100,000 cycles and a conservative burn-in period of 50,000 samples, using Gibbs2f90 software (Misztal et al., 2002).

Within a given cluster (c), prediction accuracies were estimated as the correlation ($\hat{r}_{a\hat{a}g}$) between predicted ($\hat{\mathbf{a}}$) and true ($\mathbf{a}$) breeding values, as proposed by Legarra et al. (2008):

$$\hat{r}_{a\hat{a}g} = r(\mathbf{y}_g^*, \hat{\mathbf{y}}_g^*) / h \quad (4)$$

where $\mathbf{y}_g^*$ is the vector of phenotype of animals belonging from group c adjusted for the fixed effects as in Eq. 1 and $\hat{\mathbf{y}}_g^*$ is the predicted values from cross-validation, the DGV for direct prediction and GEBV for the genomic enhanced methods, and h is the square root of the heritability of the trait. For each method, the slope coefficient for the regression of $\mathbf{y}_g^*$ on $\hat{\mathbf{y}}_g^*$ was calculated to evaluate the degree of inflation/deflation of the genomic predictions (DGV/GEBV). Nonbiased models are expected to have a regression coefficient of 1.

RESULTS AND DISCUSSION

The (co)variance components and heritability estimated by traditional method for growth traits and visual scores are in concordance with the

values reported in the literature for Hereford and Braford cattle (De Mattos et al., 2000; Cardoso and Tempelman, 2004; Biegelmeyer et al., 2017; Teixeira et al., 2018). The additive direct heritability ranged from 0.13 to 0.28 and maternal heritability ranged from 0.07 to 0.10 for growth traits and visual scores (Table 2).

Clustering for Cross-validation

Clustering by the k -means method produced groups with unbalanced number of animals for the weaning traits (Table 3), showing that the average of genomic relationship (G_{ij}) within a group was greater than the average between groups. The same trend was observed for yearling traits (results not shown). Differently from Saatchi et al. (2011), who used the traditional relationship matrix for group formation, in this study was used the genomic relationship between the animals for the clustering according to Boddhireddy et al. (2014), due to the presence of incomplete pedigree information for part of the animals considered. The results show that the k -means clustering partitioned the groups primarily by the breed composition (one group containing practically Hereford breed animals) and then, for the Braford animals, the groups were formed according to the degree of inbreeding and genomic relationship.

The results for the clustering at random were very different in relation to the k -means method. In this clustering methodology, the groups had a balanced number of animals with a similar average inbreeding of 0.01 and breed composition

of 30% zebu proportion. There was also no difference in G_{ij} values within and between groups, being close to zero (results not shown), due to the centralization of the genomic relationship matrix (VanRaden, 2008). Therefore, the objective of forming distinct groups to represent scenarios with animals of more distant relationship (k -means) and another, with closer relationship (random), to the training and validation populations was accomplished.

Accuracy and Bias of Direct Genomic Prediction

In general, the accuracies between phenotypes and DGV obtained from the cross-validation, using different methodologies to estimate the markers effects, were of moderate to high magnitude for weaning and yearling traits, ranging from 0.19 and 0.78 for the k -means and random clustering methods (Tables 4 and 5). Greatest accuracies were estimated with the BayesB method, of 0.74 and 0.78 with the random grouping for WC and BW0, respectively. As expected, due to larger genomic relationship between training and validation sets in the randomized groups, for all traits, greater accuracies were observed in the cross-validation in relation to the k -means, with an overall mean accuracy of 0.29 and 0.56 for k -means and random, respectively. According to Habier et al. (2007, 2010), the prediction accuracy is affected by the additive genetic relationship between training sets, and the reduction in accuracy occurs as genetic additive relationships between animals decrease between the reference and the validation sets.

Table 2. Posteriori means and SD ($\pm$ SD) of direct additive genetic variances (σ_a^2), maternal genetic variances (σ_m^2), genetic additive-maternal (co)variances (σ_{am}), maternal permanent environment variances (σ_{pe}^2) and residual variances (σ_e^2), and direct heritability (h_a^2) and maternal heritability (h_m^2)

Traits ¹	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_e^2	$h_a^2(\pm\text{SD})$	$h_m^2(\pm\text{SD})$
BW0	4.98 ($\pm$ 0.25)	1.23 ($\pm$ 0.13)	-0.93 ($\pm$ 0.14)	0.61 ($\pm$ 0.08)	12.21 ($\pm$ 0.18)	0.28 ($\pm$ 0.01)	0.07 ($\pm$ 0.01)
WW205	123.39 ($\pm$ 9.06)	50.29 ($\pm$ 5.17)	-27.48 ($\pm$ 5.22)	133.34 ($\pm$ 3.85)	284.03 ($\pm$ 6.20)	0.22 ($\pm$ 0.01)	0.09 ($\pm$ 0.01)
WC	0.129 ($\pm$ 0.01)	0.077 ($\pm$ 0.01)	-0.047 ($\pm$ 0.01)	0.136 ($\pm$ 0.01)	0.569 ($\pm$ 0.01)	0.15 ($\pm$ 0.01)	0.09 ($\pm$ 0.01)
WP	0.134 ($\pm$ 0.01)	0.061 ($\pm$ 0.01)	-0.046 ($\pm$ 0.01)	0.135 ($\pm$ 0.01)	0.659 ($\pm$ 0.01)	0.15 ($\pm$ 0.01)	0.06 ($\pm$ 0.01)
WM	0.159 ($\pm$ 0.01)	0.066 ($\pm$ 0.01)	-0.045 ($\pm$ 0.01)	0.142 ($\pm$ 0.01)	0.655 ($\pm$ 0.01)	0.16 ($\pm$ 0.01)	0.07 ($\pm$ 0.01)
WS	0.109 ($\pm$ 0.01)	0.054 ($\pm$ 0.01)	-0.016 ($\pm$ 0.01)	0.105 ($\pm$ 0.01)	0.497 ($\pm$ 0.01)	0.15 ($\pm$ 0.01)	0.07 ($\pm$ 0.01)
YW550	273.14 ($\pm$ 11.69)	—	—	—	761.30 ($\pm$ 10.55)	0.26 ($\pm$ 0.01)	—
WWG345	96.35 ($\pm$ 7.17)	—	—	—	649.44 ($\pm$ 7.30)	0.13 ($\pm$ 0.01)	—
YC	0.148 ($\pm$ 0.01)	—	—	—	0.622 ($\pm$ 0.01)	0.19 ($\pm$ 0.01)	—
YP	0.164 ($\pm$ 0.01)	—	—	—	0.701 ($\pm$ 0.01)	0.19 ($\pm$ 0.01)	—
YM	0.168 ($\pm$ 0.01)	—	—	—	0.704 ($\pm$ 0.01)	0.19 ($\pm$ 0.01)	—
YS	0.195 ($\pm$ 0.01)	—	—	—	0.531 ($\pm$ 0.01)	0.27 ($\pm$ 0.01)	—

¹BW0 = birth weight; WW205 = weaning weight adjusted to 205 d of age; WC = weaning conformation; WP = weaning precocity; WM = weaning muscling; WS = weaning size; YW550 = yearling weight adjusted to 550 d of age; WWG345 = postweaning weight gain adjusted for 345 d of age; YC = yearling conformation; YP = yearling precocity; YM = yearling muscling; YS = yearling size.

Table 3. Number of animals each group (N) and averages ($\pm$ SD) of zebu proportion, inbreeding coefficient¹, and within- and between-group genomic relationship (G_{ij}) for k -means clustering in traits measured at weaning

Traits ²	Group	N	Zebu proportion	Inbreeding coefficient	G_{ij} within group	G_{ij} between group
BW0	1	597	0.01 $\pm$ 0.06	0.08 $\pm$ 0.03	0.13 $\pm$ 0.03	-0.02 $\pm$ 0.03
	2	270	0.37 $\pm$ 0.04	-0.01 $\pm$ 0.03	0.10 $\pm$ 0.03	-0.001 $\pm$ 0.03
	3	656	0.37 $\pm$ 0.06	0.004 $\pm$ 0.02	0.05 $\pm$ 0.04	-0.01 $\pm$ 0.05
	4	524	0.33 $\pm$ 0.09	-0.003 $\pm$ 0.03	0.01 $\pm$ 0.04	-0.002 $\pm$ 0.05
	5	1,145	0.36 $\pm$ 0.06	-0.007 $\pm$ 0.03	0.01 $\pm$ 0.04	-0.004 $\pm$ 0.03
WW205	1	603	0.01 $\pm$ 0.07	0.08 $\pm$ 0.02	0.13 $\pm$ 0.04	-0.02 $\pm$ 0.06
	2	271	0.37 $\pm$ 0.05	-0.01 $\pm$ 0.02	0.10 $\pm$ 0.07	-0.001 $\pm$ 0.04
	3	669	0.37 $\pm$ 0.06	0.004 $\pm$ 0.03	0.05 $\pm$ 0.05	-0.01 $\pm$ 0.05
	4	545	0.33 $\pm$ 0.09	-0.002 $\pm$ 0.03	0.01 $\pm$ 0.03	-0.002 $\pm$ 0.03
	5	1,167	0.36 $\pm$ 0.06	-0.007 $\pm$ 0.03	0.01 $\pm$ 0.03	-0.004 $\pm$ 0.03
WC, WP, and WM	1	603	0.01 $\pm$ 0.07	0.08 $\pm$ 0.02	0.13 $\pm$ 0.04	-0.02 $\pm$ 0.05
	2	271	0.36 $\pm$ 0.05	-0.01 $\pm$ 0.02	0.10 $\pm$ 0.06	-0.009 $\pm$ 0.03
	3	668	0.36 $\pm$ 0.06	0.004 $\pm$ 0.03	0.05 $\pm$ 0.03	-0.01 $\pm$ 0.03
	4	545	0.37 $\pm$ 0.09	0.002 $\pm$ 0.03	0.01 $\pm$ 0.04	-0.001 $\pm$ 0.05
	5	1,168	0.34 $\pm$ 0.06	-0.007 $\pm$ 0.03	0.01 $\pm$ 0.03	-0.003 $\pm$ 0.03
WS	1	568	0.01 $\pm$ 0.07	0.08 $\pm$ 0.02	0.13 $\pm$ 0.04	-0.02 $\pm$ 0.05
	2	266	0.37 $\pm$ 0.05	-0.01 $\pm$ 0.02	0.11 $\pm$ 0.07	-0.009 $\pm$ 0.03
	3	657	0.38 $\pm$ 0.06	0.003 $\pm$ 0.03	0.04 $\pm$ 0.03	-0.01 $\pm$ 0.03
	4	520	0.33 $\pm$ 0.09	-0.002 $\pm$ 0.03	0.01 $\pm$ 0.04	-0.001 $\pm$ 0.04
	5	1,126	0.36 $\pm$ 0.06	-0.007 $\pm$ 0.03	0.01 $\pm$ 0.03	-0.003 $\pm$ 0.03

¹Diagonal elements of genomic relationship matrix minus 1 ($G_{ii} - 1$).

²BW0 = birth weight; WW205 = weaning weight adjusted to 205 d of age; WC = weaning conformation; WP = weaning precocity; WM = weaning muscling; WS = weaning size.

For growth traits, using the k -means clustering, the BayesB and GBLUP methodologies had the greatest accuracies, with value equal to 0.42 for BW0. For the random grouping, the greater accuracies were obtained using the BayesB method, of 0.72 and 0.78 for YW550 and BW0, respectively. The lowest estimated values were obtained for PWG345 and WW205, which may be due to lower heritability of these traits rather than BW and YW550.

For BW0 and WW205, accuracies estimated in this study were similar to those obtained by Saatchi et al. (2013), who reported means ranging from 0.25 to 0.42, using BayesB and BayesC, for Hereford cattle data and using k -means clustering. However, for YW550, the accuracy estimated in this study was greater than that of 0.27 reported by Saatchi et al. (2013). On the other hand, Saatchi et al. (2011, 2012) estimated greater accuracy values compared with our study, working with Angus, Simmental, and Limousin cattle, respectively, for growth traits using similar grouping formation criteria and estimation of the markers effects with Bayesian methodology. Differences in accuracies estimated between studies may occur due to different heritabilities, amount of information, and genetic architecture of each population and trait.

For visual scores measured at weaning and yearling, the greatest values of accuracies were estimated using Bayesian methodologies, with values of 0.38 for WS and 0.45 for YS, using the k -means clustering. When the groups were formed at random, the greatest values were also for these methods, of 0.74 and 0.73 for WC and YS, respectively. Analyzing data from Nelore cattle, Neves et al. (2014) estimated mean accuracy values of 0.21, 0.46, 0.46, 0.29, 0.72, and 0.69 for WC, WP, WM, YC, YP, and YM, respectively, using different methods. The difference in the values found by Neves et al. (2014) in comparison to the present study may be due to their smaller set of genotyped animals compared with our research, validation set using the forward prediction scheme and greater heritabilities found by Neves et al. (2014) for visual scores. Another reason for different values of accuracy could be the different pattern of linkage disequilibrium between *indicus* and *taurine* beef cattle populations (O'Brien et al., 2014).

In general, for some traits studied, Bayesian methods were superior to the GBLUP, a fact that may be due to the different assumption for the markers effects in each methodology. The GBLUP method assumes an infinitesimal model with many locus of small effects and that the QTL effects have

Table 4. Accuracies of the phenotype adjusted for fixed effects with the direct genomic values (DGV) of cross-validation predictions (r_{aa}) and regression coefficients (β)¹ using different methods for traits measured at weaning

Traits ²	Methods ³	<i>k</i> -means clustering		Random clustering	
		r_{aa}	β	r_{aa}	β
BW0	PBLUP	0.40 ± 0.15	0.76	0.63 ± 0.20	0.91
	GBLUP	0.42 ± 0.11	0.72	0.76 ± 0.17	0.86
	BayesC	0.41 ± 0.09	0.79	0.77 ± 0.16	0.92
	BayesB	0.42 ± 0.08	0.71	0.78 ± 0.17	0.85
WW205	PBLUP	0.29 ± 0.05	1.03	0.45 ± 0.12	1.18
	GBLUP	0.23 ± 0.03	0.64	0.45 ± 0.09	0.75
	BayesC	0.27 ± 0.05	1.03	0.45 ± 0.12	1.03
	BayesB	0.28 ± 0.05	0.84	0.46 ± 0.13	0.83
WC	PBLUP	0.33 ± 0.12	1.15	0.70 ± 0.18	1.23
	GBLUP	0.35 ± 0.08	1.13	0.73 ± 0.19	1.07
	BayesC	0.35 ± 0.09	1.19	0.74 ± 0.17	1.01
	BayesB	0.34 ± 0.09	1.31	0.73 ± 0.18	1.10
WP	PBLUP	0.20 ± 0.06	1.08	0.24 ± 0.08	1.12
	GBLUP	0.19 ± 0.06	0.77	0.23 ± 0.07	0.80
	BayesC	0.20 ± 0.06	1.26	0.24 ± 0.08	0.96
	BayesB	0.23 ± 0.06	0.95	0.28 ± 0.07	0.84
WM	PBLUP	0.27 ± 0.20	1.15	0.70 ± 0.19	1.24
	GBLUP	0.22 ± 0.06	0.91	0.73 ± 0.21	1.02
	BayesC	0.29 ± 0.07	0.85	0.71 ± 0.20	1.04
	BayesB	0.31 ± 0.19	1.07	0.72 ± 0.22	1.09
WS	PBLUP	0.31 ± 0.07	1.07	0.39 ± 0.09	1.14
	GBLUP	0.34 ± 0.09	1.07	0.38 ± 0.10	1.01
	BayesC	0.38 ± 0.11	1.23	0.42 ± 0.12	1.09
	BayesB	0.38 ± 0.11	1.22	0.42 ± 0.11	1.10

¹Estimated regression slope of phenotype adjusted for the fixed effects on its predicted values based on cross-validation DGV.

²BW0 = birth weight; WW205 = weaning weight adjusted to 205 d of age; WC = weaning conformation; WP = weaning precocity; WM = weaning muscling; WS = weaning size.

³PBLUP = pedigree BLUP; GBLUP = genomic BLUP; BayesB = Bayesian mixture of *t*-distribution and point mass on zero with probability ($\pi = 0.95$); BayesC = Bayesian mixture of normal distribution and point mass on zero with probability ($\pi = 0.95$).

normal distribution and constant variance for all chromosomal segments. In Bayesian methods, the observed data are combined with a priori information on the unknown parameters, thus generating a posteriori probabilities of these parameters. These methods allow for assumptions of different variances for the effects of each set of markers in a given chromosomal segment and consider that only a fraction of these markers have nonzero effect, as in BayesB and BayesC (Meuwissen et al., 2001; Garrick, 2010). Therefore, BayesB and BayesC may better represent reality in relation to the SNPs effects, considering that genomic analyses usually involve thousands of markers, many are expected to have insignificant effect.

The average of the regression coefficients of y^* on DGV, calculated as indicators of the bias prediction, ranged from 0.44 to 1.43 for the *k*-means clustering method and from 0.77 to 1.48 for the random methodology (Tables 5 and 6),

suggesting lesser underestimation of DGV prediction under random groups. In general, the coefficients for traits measured at weaning were greater than 1 for Bayesian methods; therefore, genomic predictions were underestimated. For the GBLUP methodology, the coefficients were less than 1 such that predictions were overestimated for both *k*-means and random clustering methods. For yearling traits, predictions were overestimated for both frequentist and Bayesian methods by clustering *k*-means and underestimated for BayesC, BayesB, and GBLUP methods with groups selected randomly. Prediction bias may be a problem, particularly if predictions of genotyped (DGV) and nongenotyped (EBV) animals need to be compared for animal classification and selection decisions. In the case of overestimation, the DGV predictions would contain extreme values, favoring individuals who have genomic information (Cardoso et al., 2015).

Table 5. Accuracies of the phenotype adjusted for fixed effects with the direct genomic values (DGV) of cross-validation predictions (r_{act}) and regression coefficients (β)¹ using different methods for traits measured at yearling

Traits ²	Methods ²	<i>k</i> -means clustering		Radom clustering	
		r_{act}	β	r_{act}	β
YW550	PBLUP	0.30 ± 0.14	0.85	0.68 ± 0.18	0.86
	GBLUP	0.31 ± 0.09	0.76	0.70 ± 0.18	0.89
	BayesC	0.33 ± 0.10	1.04	0.71 ± 0.17	1.14
	BayesB	0.33 ± 0.10	0.78	0.72 ± 0.17	0.89
WWG345	PBLUP	0.21 ± 0.08	1.06	0.44 ± 0.22	0.98
	GBLUP	0.21 ± 0.09	0.86	0.45 ± 0.22	1.18
	BayesC	0.23 ± 0.10	0.52	0.41 ± 0.25	1.40
	BayesB	0.24 ± 0.11	0.71	0.45 ± 0.22	0.88
YC	PBLUP	0.28 ± 0.13	0.87	0.67 ± 0.19	0.95
	GBLUP	0.26 ± 0.09	0.78	0.68 ± 0.19	0.91
	BayesC	0.26 ± 0.08	1.08	0.68 ± 0.19	1.28
	BayesB	0.27 ± 0.08	0.82	0.70 ± 0.12	1.03
YP	PBLUP	0.33 ± 0.24	0.78	0.47 ± 0.22	0.88
	GBLUP	0.34 ± 0.22	0.73	0.60 ± 0.20	1.03
	BayesC	0.35 ± 0.21	0.74	0.61 ± 0.20	1.11
	BayesB	0.36 ± 0.21	0.70	0.61 ± 0.21	1.06
YM	PBLUP	0.22 ± 0.05	0.88	0.52 ± 0.18	1.03
	GBLUP	0.24 ± 0.04	0.90	0.55 ± 0.16	1.06
	BayesC	0.24 ± 0.04	0.97	0.56 ± 0.17	1.18
	BayesB	0.26 ± 0.05	0.91	0.56 ± 0.16	1.09
YS	PBLUP	0.35 ± 0.11	0.76	0.63 ± 0.20	0.86
	GBLUP	0.42 ± 0.08	0.96	0.72 ± 0.15	1.03
	BayesC	0.42 ± 0.07	1.02	0.72 ± 0.14	1.09
	BayesB	0.45 ± 0.07	1.04	0.73 ± 0.15	1.09

¹Estimated regression slope of phenotype adjusted for the fixed effects on its predicted values based on cross-validation DGV.

²YW550 = yearling weight adjusted to 550 d of age; WWG345 = postweaning weight gain adjusted for 345 d of age; YC = yearling conformation; YP = yearling precocity; YM = yearling muscling; YS = yearling size.

³PBLUP = pedigree BLUP; GBLUP = genomic BLUP; BayesB = Bayesian mixture of *t*-distribution and point mass on zero with probability ($\pi = 0.95$); BayesC = Bayesian mixture of normal distribution and point mass on zero with probability ($\pi = 0.95$).

The accuracy within the groups, calculated on the basis of Eq. 4, for all traits was lower for the group composed mostly of Herefords by the *k*-means clustering between all methodologies (Fig. 1). The same pattern was observed for all methods; for this reason, only one method has been shown as example. When using random clustering, where breeds are balanced distributed across groups, we do not observe these accented differences, and accuracies are similar for all groups (results not shown). These results indicating that the training set composed of the majority of Braford animals will not estimate accurate predictions for the Hereford animals in the validation set. Such outcome was already expected because Cardoso et al. (2015), analyzing tick resistance trait in the same data set and using the same methodology for clustering, also found lower accuracy for the group with only Hereford breed animals. Saatchi et al. (2013), using a training population

with North American cattle data from Hereford, also did not produce accurate predictions for herds of the same breed in Uruguay and Argentina. In the same way, Habier et al. (2007) showed that the accuracy of genomic predictions in the selection candidates decreases as their average relationship with the training set decreases.

In general, greater gains in accuracy using the methodologies to estimate the effects of the markers in relation to the PBLUP method (Table 6) were obtained with the Bayesian methods (BayesB and BayesC) for all traits. For weaning traits, greater gain was achieved with the BayesB method for BW, being of 23.81% with random grouping. At yearling, greater gain was obtained with the BayesB and BayesC methodologies, of 29.81%, for YP, again for random clustering. These results indicate that Bayesian methods tend to be more suitable genomic evaluation, yielding greater accuracy values among the tested methods.

Table 6. Accuracy gain (%) of the methods of estimation of markers effects in relation to the traditional pedigree BLUP in the traits measured at weaning and at yearling

Traits ¹	<i>k</i> -means clustering			Random clustering		
	GBLUP ²	BayesC	BayesB	GBLUP	BayesC	BayesB
BW0	5.00	2.50	5.00	20.63	22.23	23.81
WW205	-20.69	-6.90	-3.45	0.00	0.00	2.22
WC	6.06	6.06	3.03	4.29	5.71	4.29
WP	-5.00	0.00	15.00	-4.17	0.00	16.67
WM	-18.52	7.41	14.81	4.29	1.43	2.86
WS	9.68	22.58	22.58	-2.56	7.69	7.69
YW550	3.33	10.00	10.00	2.94	4.41	5.88
WWG345	0.00	9.52	14.29	2.27	-6.82	2.27
YC	-7.14	-7.14	-3.57	1.49	1.49	4.48
YP	3.03	6.06	9.09	27.66	29.79	29.79
YM	9.09	9.09	18.18	5.77	7.69	7.69
YS	20.00	20.00	28.57	14.29	14.29	15.87
Average	0.40	6.60	11.13	6.41	7.33	10.29

¹BW0 = birth weight; WW205 = weaning weight adjusted to 205 d of age; WC = weaning conformation; WP = weaning precocity; WM = weaning muscling; WS = weaning size; YW550 = yearling weight adjusted to 550 d of age; WWG345 = postweaning weight gain adjusted for 345 d of age; YC = yearling conformation; YP = yearling precocity; YM = yearling muscling; YS = yearling size.

²GBLUP = genomic BLUP; BayesB = Bayesian mixture of *t*-distribution and point mass on zero with probability ($\pi = 0.95$); BayesC = Bayesian mixture of normal distribution and point mass on zero with probability ($\pi = 0.95$).

Figure 1. Accuracies of direct genomic predictions using BayesB method for each *k*-means clustering cross-validation group for birth weight (BW0), weaning weight adjusted for 205 d of age (WW205), weaning conformation (WC), weaning precocity (WP), weaning muscling (WM), weaning size (WS), yearling weight adjusted for 550 d of age (YW550), postweaning weight gain adjusted for 345 d of age (WWG345), yearling conformation (YC), yearling precocity (YP), yearling muscling (YM), and yearling size (YS).

For most traits, the traditional pedigree method (PBLUP) performed similarly to the DGV, particularly using the *k*-means cross-validation. Saatchi et al. (2011) also found similar results comparing the accuracy between an adjusted parent average and DGV. According to these authors, the relative magnitude of PA (parent average) accuracies compared to DGV accuracies depends on the amount information available to calculated PA, which for the sires in their study was large involving patrilineages with grand-progeny records. Another fact that may have occurred in relation to the similar gains with the PBLUP and DGV is the greater molecular genetic distance of the *k*-means clustering, influencing the smaller estimated accuracy for the DGV

that is dependent on the strength of the genomic relationships between the training and validation sets (Saatchi et al., 2011).

Blending of Traditional Information and Genomic Predictions

Incorporation of historical pedigree and phenotype information with DGV was performed using the GBLUP and BayesB methods for GEBV prediction. The BayesB method was chosen to combine traditional information among Bayesian methods because it was the method with the greatest gain among Bayesians and because it has more distinct assumptions for the markers effects in

relation to GBLUP (Cardoso et al., 2015). Three alternative methods, SI (VanRaden et al., 2009), blending by bivariate analysis (MacNeil et al., 2010), and ssGBLUP (Misztal et al., 2009; Aguilar et al., 2010), were used, and for most of the traits, the genetic correlations between the phenotype and the genomic prediction were improved (Tables 7 and 8). These results indicate that the historical information contributed considerably as additional information for the calculation of GEBVs. Nevertheless, Saatchi et al. (2011) did not obtain a consistent increase in accuracy by combining PA and DGV to estimate the GEBV, suggesting that the parent average is not totally independent of the Mendelian sampling. This gain in accuracy depends very much on the veracity of the available information. For example, if the pedigree is incorrect, the correlation between the DGV and the EBV is misleading and it is impossible to have accurate inference about GEBVs.

The first approach used was the SI proposed by VanRaden et al. (2009), which is a function of the mean reliabilities of the animals used in the prediction. In this index, subtracting the subgroup of the genotyped animals (SPA), only the additional information contained in the genomic predictions is added to the complete set (PA), thus avoiding the double counting of the predictions of the traditional pedigree in the subset of data used for the training markers effects (VanRaden et al., 2009). For all traits, genetic correlation was improved when compared with the traditional method using the parent average (PBLUP), by GBLUP and BayesB methods, with an increase in accuracy. The greatest gains in relation to PBLUP, which ranged from 0.18 to 0.19, were for YW550 and WWG345, respectively, with the *k*-means clustering method and with randomly formed groups, varying from 0.18 to 0.19 for YP and WW205 traits, respectively.

In the bivariate analysis (second approach), DGV was considered as a correlated trait as proposed by MacNeil et al. (2010) and produced 2 GEBVs for each animal, one for phenotypic data (a) and the other for DGV (α). In this methodology, there is an information exchange through $\sigma_{a\alpha}$, which is directly dependent on the prediction accuracy (Tables 7 and 8). In the present study, GEBV (a) was a better predictor, with an increase in genetic correlation of up to 0.14 and 0.18 in relation to PBLUP, using the *k*-means and random clustering, respectively, for traits measured at weaning. At yearling, there was an increase of 0.21 for *k*-means and 0.18 for random groups. In this bivariate approach, for some traits, the combination of

DGV with traditional information was similar to PBLUP, mainly for GEBV (α). These results may have occurred because the accuracy of the DGV was for some traits similar to the genetic value estimated by the PBLUP, not contributing much to the gain in accuracy of the GEBV (α).

For most of the traits, the greatest accuracy values among the methodologies tested were for the ssGBLUP method using single or bivariate trait model (Tables 7 and 8). The greatest increase was obtained using MTssGBLUP for YW550 with *k*-means clustering (0.26) and for WW205 when clusters were randomly formed (0.25). The ssGBLUP involves a submatrix corresponding to the genotyped animals and the traditional pedigree A^{-1} , which is modified by adding the difference between the inverse of the genomic relationship matrix and the inverse of the relationship matrix for the genotyped animals (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). Thus, this method considers that all loci have the same weight and variance (Saatchi et al., 2013). Despite this, ssGBLUP methodology used in this study to combine all information sources, phenotype, pedigree, and genomics, in one step showed the greater gain in accuracy mainly using bivariate analysis. Multitrait analysis can allow for more accurate predictions due to correlation structure between the traits and minimize effect of selection bias (Bijma, 2012; Hayashi and Iwata, 2013). In the present study, greater increases in prediction accuracy were observed for the traits measured at yearling rather than at weaning. In the same way, using multitrait ssGBLUP in Angus cattle, Lourenço et al. (2015) estimated a larger increase in predictive ability for postweaning gain trait compared with weaning traits in relation to traditional pedigree BLUP. According to Lourenço et al. (2015), the selection bias is not easy to be accounted for in beef cattle, as the selection process is sequential and affected by genetic correlations between traits and specific indexes used for selection.

The patterns of the accuracies within groups for GEBVs using Eq. 4 were similar to those obtained with DGV using *k*-means clusters. Again, the group 1 having the greatest number of Hereford animals had the smaller accuracies compared with the other groups composed of mostly Braford animals (Fig. 2). The same pattern was observed for all blending methods. This happened despite the accuracy values of GEBV being larger than those of DGV.

The regression coefficients for the adjusted phenotype on the GEBV for both weaning and yearling traits in any of the methods used were, on average,

Table 7. Accuracies of the phenotype adjusted for fixed effects with the genomic enhanced breeding values (GEBV) of cross-validation predictions (r_{au}) and regression coefficients (β)¹ using different methods for traits measured at weaning

Traits ²	Methods ³	<i>k</i> -means clustering		Random clustering	
		r_{au}	β	r_{au}	β
BW0	GBLUP_Biv(a)	0.43 ± 0.09	0.62	0.63 ± 0.18	0.51
	GBLUP_Biv(α)	0.43 ± 0.11	0.73	0.75 ± 0.17	0.87
	SI_GBLUP	0.53 ± 0.17	0.77	0.77 ± 0.15	0.88
	BayesB_Biv(a)	0.43 ± 0.09	0.62	0.63 ± 0.18	0.53
	BayesB_Biv(α)	0.40 ± 0.08	0.72	0.76 ± 0.16	0.87
	SI_BayesB	0.52 ± 0.13	0.78	0.77 ± 0.13	0.92
	ssGBLUP	0.53 ± 0.12	0.76	0.78 ± 0.16	1.12
	MTssGBLUP	0.53 ± 0.09	0.71	0.77 ± 0.17	0.70
WW205	GBLUP_Biv(a)	0.40 ± 0.03	1.35	0.63 ± 0.03	1.23
	GBLUP_Biv(α)	0.24 ± 0.04	0.76	0.65 ± 0.16	0.86
	SI_GBLUP	0.34 ± 0.05	0.73	0.65 ± 0.11	0.87
	BayesB_Biv(a)	0.40 ± 0.03	1.34	0.65 ± 0.15	1.23
	BayesB_Biv(α)	0.25 ± 0.04	0.75	0.63 ± 0.04	0.86
	SI_BayesB	0.36 ± 0.06	0.79	0.64 ± 0.12	0.82
	ssGBLUP	0.40 ± 0.06	0.86	0.65 ± 0.12	0.94
	MTssGBLUP	0.44 ± 0.06	0.96	0.70 ± 0.14	1.03
WC	GBLUP_Biv(a)	0.38 ± 0.09	1.14	0.82 ± 0.17	0.72
	GBLUP_Biv(α)	0.33 ± 0.07	1.19	0.82 ± 0.18	1.15
	SI_GBLUP	0.42 ± 0.11	1.02	0.79 ± 0.13	1.06
	BayesB_Biv(a)	0.38 ± 0.09	0.09	0.82 ± 0.17	0.72
	BayesB_Biv(α)	0.33 ± 0.07	0.07	0.82 ± 0.17	1.19
	SI_BayesB	0.36 ± 0.14	1.08	0.79 ± 0.12	1.06
	ssGBLUP	0.43 ± 0.11	1.08	0.80 ± 0.13	1.14
	MTssGBLUP	0.44 ± 0.11	1.05	0.79 ± 0.13	1.12
WP	GBLUP_Biv(a)	0.31 ± 0.07	1.31	0.36 ± 0.07	1.19
	GBLUP_Biv(α)	0.20 ± 0.06	1.02	0.22 ± 0.06	0.91
	SI_GBLUP	0.30 ± 0.07	0.83	0.37 ± 0.10	0.90
	BayesB(a)	0.35 ± 0.06	1.21	0.38 ± 0.08	1.19
	BayesB(α)	0.19 ± 0.05	0.84	0.21 ± 0.05	0.86
	SI_BayesB	0.32 ± 0.08	0.89	0.38 ± 0.10	0.94
	ssGBLUP	0.33 ± 0.08	0.89	0.39 ± 0.09	0.96
	MTssGBLUP	0.33 ± 0.09	0.76	0.40 ± 0.10	0.83
WM	GBLUP_Biv(a)	0.39 ± 0.18	1.17	0.73 ± 0.20	0.79
	GBLUP_Biv(α)	0.22 ± 0.07	0.94	0.72 ± 0.21	1.06
	SI_GBLUP	0.28 ± 0.09	0.86	0.80 ± 0.17	1.03
	BayesB_Biv(a)	0.41 ± 0.20	0.96	0.67 ± 0.23	0.75
	BayesB_Biv(α)	0.32 ± 0.19	1.08	0.71 ± 0.22	1.09
	SI_BayesB	0.35 ± 0.18	0.91	0.81 ± 0.16	1.05
	ssGBLUP	0.46 ± 0.17	1.01	0.82 ± 0.16	1.10
	MTssGBLUP	0.48 ± 0.12	0.98	0.82 ± 0.15	1.09
WS	GBLUP_Biv(a)	0.43 ± 0.05	1.29	0.48 ± 0.04	1.24
	GBLUP_Biv(α)	0.37 ± 0.09	1.15	0.39 ± 0.11	1.11
	SI_GBLUP	0.37 ± 0.06	0.98	0.43 ± 0.08	0.99
	BayesB_Biv(a)	0.37 ± 0.06	0.91	0.39 ± 0.06	0.92
	BayesB_Biv(α)	0.38 ± 0.11	1.25	0.42 ± 0.12	1.12
	SI_BayesB	0.35 ± 0.05	0.92	0.44 ± 0.08	0.94
	ssGBLUP	0.47 ± 0.10	1.08	0.54 ± 0.12	1.09
	MTssGBLUP	0.46 ± 0.08	0.81	0.53 ± 0.15	1.04

¹Estimated regression slope of phenotype adjusted for the fixed effects on its predicted values based on cross-validation GEBV.

²BW0 = birth weight; WW205 = weaning weight adjusted to 205 d of age; WC = weaning conformation; WP = weaning precocity; WM = weaning muscling; WS = weaning size.

³GBLUP = genomic BLUP; BayesB = Bayesian mixture of *t*-distribution and point mass on zero with probability ($\pi = 0.95$); Biv = bivariate analysis blending with 2 GEBV, one for phenotype (a) and the other for the direct genomic value data (α); SI = selection index weighting DGV, PA is pedigree parent average calculated using data from all animals and PA derived just using data from genotyped animals (SPA), by functions of their reliabilities (VanRaden et al., 2009); ssGBLUP = single-step GBLUP; MTssGBLUP = bivariate single-step GBLUP.

Table 8. Accuracies of the phenotype adjusted for fixed effects with the genomic enhanced breeding values (GEBV) of cross-validation predictions (r_{aa}) and regression coefficients (β)¹ using different methods for traits measured at yearling

Traits ²	Method ³	<i>k</i> -means clustering		Random clustering	
		(r_{aa})	β	(r_{aa})	β
YW550	GBLUP_Biv(a)	0.39 ± 0.12	0.73	0.68 ± 0.19	0.49
	GBLUP_Biv(α)	0.32 ± 0.10	0.80	0.70 ± 0.18	0.93
	SI_GBLUP	0.48 ± 0.14	0.85	0.75 ± 0.15	0.88
	BayesB_Biv(a)	0.40 ± 0.12	0.74	0.68 ± 0.18	0.50
	BayesB_Biv(α)	0.33 ± 0.10	0.81	0.72 ± 0.18	0.94
	SI_BayesB	0.49 ± 0.15	0.89	0.76 ± 0.14	0.91
	ssGBLUP	0.52 ± 0.17	0.79	0.78 ± 0.14	0.80
	MTssGBLUP	0.56 ± 0.13	0.91	0.85 ± 0.12	0.96
WWG345	GBLUP_Biv(a)	0.24 ± 0.09	0.86	0.44 ± 0.22	0.65
	GBLUP_Biv(α)	0.21 ± 0.09	0.90	0.48 ± 0.22	1.20
	SI_GBLUP	0.35 ± 0.12	1.06	0.52 ± 0.22	0.95
	BayesB_Biv(a)	0.26 ± 0.10	0.85	0.43 ± 0.22	0.70
	BayesB_Biv(α)	0.24 ± 0.11	0.75	0.45 ± 0.22	0.92
	SI_BayesB	0.39 ± 0.14	0.98	0.53 ± 0.22	0.84
	ssGBLUP	0.25 ± 0.09	0.73	0.57 ± 0.22	0.66
	MTssGBLUP	0.26 ± 0.06	0.67	0.60 ± 0.17	0.71
YC	GBLUP_Biv(a)	0.37 ± 0.12	0.78	0.66 ± 0.19	0.58
	GBLUP_Biv(α)	0.29 ± 0.09	0.81	0.68 ± 0.19	0.94
	SI_GBLUP	0.39 ± 0.13	0.79	0.74 ± 0.16	0.90
	BayesB_Biv(a)	0.37 ± 0.12	0.79	0.66 ± 0.19	0.58
	BayesB_Biv(α)	0.27 ± 0.08	0.84	0.70 ± 0.18	1.05
	SI_BayesB	0.40 ± 0.12	0.82	0.75 ± 0.16	0.97
	ssGBLUP	0.42 ± 0.11	0.78	0.78 ± 0.15	0.92
	MTssGBLUP	0.49 ± 0.13	0.98	0.80 ± 0.14	1.09
YP	GBLUP_Biv(a)	0.35 ± 0.23	0.69	0.48 ± 0.19	0.66
	GBLUP_Biv(α)	0.36 ± 0.22	0.75	0.60 ± 0.20	1.05
	SI_GBLUP	0.46 ± 0.22	0.87	0.65 ± 0.19	1.02
	BayesB(a)	0.35 ± 0.22	0.70	0.48 ± 0.20	0.65
	BayesB(α)	0.34 ± 0.21	0.72	0.61 ± 0.20	1.08
	SI_BayesB	0.45 ± 0.23	0.86	0.65 ± 0.19	1.04
	ssGBLUP	0.47 ± 0.22	0.79	0.65 ± 0.19	0.94
	MTssGBLUP	0.51 ± 0.23	0.84	0.69 ± 0.19	0.97
YM	GBLUP_Biv(a)	0.28 ± 0.05	0.83	0.51 ± 0.15	0.76
	GBLUP_Biv(α)	0.24 ± 0.05	0.93	0.56 ± 0.16	1.09
	SI_GBLUP	0.32 ± 0.06	0.92	0.63 ± 0.16	1.02
	BayesB_Biv(a)	0.28 ± 0.05	0.83	0.51 ± 0.14	0.75
	BayesB_Biv(α)	0.24 ± 0.04	0.95	0.57 ± 0.17	1.12
	SI_BayesB	0.32 ± 0.06	0.93	0.63 ± 0.17	1.04
	ssGBLUP	0.33 ± 0.05	0.81	0.66 ± 0.16	0.99
	MTssGBLUP	0.35 ± 0.05	0.91	0.69 ± 0.16	1.05
YS	GBLUP_Biv(a)	0.43 ± 0.08	0.76	0.68 ± 0.15	0.66
	GBLUP_Biv(α)	0.42 ± 0.08	1.03	0.73 ± 0.15	1.07
	SI_GBLUP	0.55 ± 0.17	0.85	0.75 ± 0.14	1.14
	BayesB_Biv(a)	0.45 ± 0.09	0.78	0.75 ± 0.14	1.55
	BayesB_Biv(α)	0.43 ± 0.08	1.11	0.77 ± 0.15	1.14
	SI_BayesB	0.55 ± 0.17	0.89	0.75 ± 0.15	1.01
	ssGBLUP	0.56 ± 0.10	0.91	0.78 ± 0.13	1.02
	MTssGBLUP	0.58 ± 0.06	0.85	0.79 ± 0.15	0.92

¹Estimated regression slope of phenotype adjusted for the fixed effects on its predicted values based on cross-validation GEBV.²YW550 = yearling weight adjusted to 550 d of age; WWG345 = postweaning weight gain adjusted for 345 d of age; YC = yearling conformation; YP = yearling precocity; YM = yearling muscling; YS = yearling size.³GBLUP = genomic BLUP; BayesB = Bayesian mixture of *t*-distribution and point mass on zero with probability ($\pi = 0.95$); Biv = bivariate analysis blending with 2 GEBV, one for phenotype (a) and the other for the direct genomic value data (α); SI = selection index weighting DGV, PA calculated using data from all animals (PA) and PA derived just using data from genotyped animals (SPA), by functions of their reliabilities (VanRaden et al., 2009); ssGBLUP = single-step GBLUP; MTssGBLUP = bivariate single-step GBLUP.

Figure 2. Accuracies of genomic enhanced breeding value predictions using single-step GBLUP (ssGBLUP) method for each *k*-means clustering cross-validation group for birth weight (BW0), weaning weight adjusted for 205 d of age (WW205), weaning conformation (WC), weaning precocity (WP), weaning muscling (WM), weaning size (WS), yearling weight adjusted for 550 d of age (YW550), postweaning weight gain adjusted for 345 d of age (WWG345), yearling conformation (YC), yearling precocity (YP), yearling muscling (YM), and yearling size (YS).

Table 9. Accuracy gain (%) of the methods of blending genomic with historical information in relation to the traditional pedigree BLUP in the traits measured at weaning and at yearling

Traits ¹	GBLUP_Biv ²	SI_GBLUP	BayesB_Biv	SI_BayesB	ssGBLUP	MTssGBLUP
<i>k</i> -means clustering						
BW0	7.50	32.50	7.50	30.00	32.50	32.50
WW205	37.93	17.24	37.93	24.14	37.93	51.72
WC	15.15	27.27	15.15	9.09	30.30	33.33
WP	55.00	50.00	75.00	60.00	65.00	65.00
WM	44.44	3.70	51.85	29.63	70.37	77.78
WS	38.71	19.35	22.58	12.90	51.61	48.39
YW550	30.00	60.00	33.33	63.33	73.33	86.67
WWG345	14.29	66.67	23.81	85.71	19.05	28.57
YC	32.14	39.29	32.14	42.86	50.00	75.00
YP	9.09	39.39	6.06	36.36	42.42	54.55
YM	27.27	45.45	27.27	45.45	50.00	59.09
YS	22.86	57.14	28.57	57.14	62.92	65.71
Average	22.86	38.16	30.09	41.38	48.78	56.52
Random clustering						
BW	19.05	22.22	20.63	22.22	23.81	22.22
WW205	40.00	44.44	44.44	42.22	44.44	55.56
WC	17.14	12.86	17.14	12.86	14.29	12.86
WP	50.00	54.17	58.33	58.33	62.50	66.67
WM	4.29	14.29	1.43	15.71	17.14	17.14
WS	23.08	10.26	7.69	12.82	38.46	32.90
YW550	2.94	10.29	5.88	11.76	14.71	25.00
WWG345	9.09	18.18	2.27	20.45	29.55	36.36
YC	1.49	10.45	4.48	11.94	16.42	19.48
YP	27.66	38.30	29.79	38.30	38.30	46.81
YM	7.69	21.15	9.62	21.15	26.92	32.69
YS	15.87	19.05	22.22	19.05	23.81	25.40
Average	18.19	22.97	18.66	23.90	29.19	32.75
Overall mean	23.03	30.56	24.38	32.64	39.00	44.64

¹BW0 = birth weight; WW205 = weaning weight adjusted to 205 d of age; WC = weaning conformation; WP = weaning precocity; WM = weaning muscling; WS = weaning size; YW550 = yearling weight adjusted to 550 d of age; WWG345 = postweaning weight gain adjusted for 345 d of age; YC = yearling conformation; YP = yearling precocity; YM = yearling muscling; YS = yearling size.

²GBLUP = genomic BLUP; BayesB = Bayesian mixture of *t*-distribution and point mass on zero with probability ($\pi = 0.95$); Biv = bivariate analysis; SI = selection index; ssGBLUP = single-step GBLUP; MTssGBLUP = bivariate single-step GBLUP.

less than 1, thus producing overestimated predictions. To reduce the bias of genomic predictions, future research should be done using, for example, the single-step Bayesian regression method (ssBR; Fernando et al., 2014, 2016), which considers different assumptions for the effects of markers, unlike ssGBLUP. Lee et al. (2017) comparing different single-step methodologies concluded that in general the results were similar between ssGBLUP and ssBR; however, the ssBR was superior when the trait is associated with QTL with large effects.

For all traits, gains in accuracy were obtained using blending methods in relation to the traditional PBLUP method (Table 9). The blending method with the best performance was the MTssGBLUP, with a mean accuracy gain of 48.78% and 29.19% among all traits with *k*-means and random clustering, respectively. According to Aguilar et al. (2010), the single-step procedure provides a unified structure, eliminates several assumptions and parameters, and reduces the bias of the predictions by using all information sources allowing to calculate genomic evaluations more accurate than the procedures of multiple steps, as SI and bivariate analysis.

Among the methods tested, despite the ssGBLUP have had the best accuracy results for most traits, according to Cardoso et al. (2015), the choice of the best method will depend on the business model used for genomic testing and the structure of genetic evaluation programs. For example, if private companies deliver the DGV without the individual genotypes, as happened since 2009 in the American Angus Association (AAA), multistep analysis such as the bivariate model proposed by MacNeil et al. (2010) are indicated, as well as the SI, used in dairy cattle genetic evaluations (VanRaden et al., 2009). To improve prediction and simplify procedures, AAA recently changed to ssGBLUP implementation (Lourenço et al., 2016). According to the authors, initial tests with ssGBLUP showed an increase in accuracy of 25% for growth traits compared with traditional evaluations. In addition, ssGBLUP approach has been used in genetic evaluations of milk and other beef cattle breeds (Saatchi et al., 2013).

CONCLUSIONS

In some cases, higher accuracy values were estimated using the Bayesian methods rather than the other methodologies evaluated for estimating the markers effects. In general, the BayesB was the best method of the DGV prediction accuracy. The accuracies for the groups that contained a greater number of Hereford animals were generally lower than

the values for the groups that presented Braford animals, suggesting the inclusion of more Hereford animals with phenotypes and genotypes in the training population would improve the estimation of the markers effects specifically for this breed.

Smaller accuracies were estimated using the *k*-means clustering, indicating that the training set should have individuals closely related to the candidate selection. Random clustering provided higher accuracies, indicating that DGV/GBEV accuracy varies according to additive genetic relationships between the training and prediction sets.

The gains obtained by blending of traditional and genomic information in relation to the PBLUP method indicate that the genomic predictions should be used as a tool to improve the genetic gains for Hereford and Braford growth and linear scores. In this sense, ssGBLUP method especially in multi-trait analyses seems to be a suitable alternative.

LITERATURE CITED

- Aguilar, I., and I. Misztal. 2014. PreGSF90. <http://nce.ads.uga.edu/wiki/doku.php?id=readme.seekparentf90> (Accessed 25 November 2015.)
- 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. doi:10.3168/jds.2009–2730
- Biegelmeyer, P., C. C. Gulias-Gomes, V. M. Roso, N. J. L. Dionello, and F. F. Cardoso. 2017. Tick resistance genetic parameters and its correlations with production traits in Hereford and Braford cattle. *Livest. Sci.* 202:96–100. doi:10.1016/j.livsci.2017.05.019
- 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. doi:10.1111/j.1439-0388.2012.00991.x
- Bodhiredy, P., M. J. Kelly, S. Northcutt, K. C. Prayaga, J. Rumph, and S. DeNise. 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
- Cardoso, F. F., C. C. Gomes, B. P. Sollero, M. M. Oliveira, V. M. Roso, M. L. Piccoli, R. H. Higa, M. J. Yokoo, A. R. Caetano, and I. Aguilar. 2015. Genomic prediction for tick resistance in Braford and Hereford cattle. *J. Anim. Sci.* 93:2693–2705. doi:10.2527/jas.2014-8832
- Cardoso, F. F., and R. J. Tempelman. 2004. Hierarchical Bayes multiple-breed inference with an application to genetic evaluation of a Nelore–Hereford population. *J. Anim. Sci.* 82:1589–1601. doi:10.2527/2004.8261589x
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. doi:10.1186/1297-9686-42-2
- Clayton, D. 2014. snpStats: snpMatrix and XSnpmatrix classes and methods. R package version 1.18.0. <http://>

- www.bioconductor.org/packages/release/bioc/html/snp-Stats.html (Accessed 20 November 2015.)
- De Mattos, D., I. Misztal, and J. K. Bertrand. 2000. Variance and covariance components for weaning weight for Herefords in three countries. *J. Anim. Sci.* 78:33–37. doi:10.2527/2000.78133x
- Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet. Sel. Evol.* 48:96. doi:10.1186/s12711-016-0273-2
- Fernando, R. L., J. C. Dekkers, and D. J. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet. Sel. Evol.* 46:50. doi:10.1186/1297-9686-46-50
- Fernando, R. L., and D. J. Garrick. 2009. GenSel: User manual for a portfolio of genomic selection related analyses. <http://www.biomedcentral.com/content/supplementary/1471-2105-12-186-S1.pdf> (Accessed 1 August 2015.)
- Garrick, D. J. 2010. The nature, scope and impact of some whole-genome analyses in beef cattle. In: *Proc. 9th World Congress Genet. Appl. Livest. Prod.*, 1–6 August, Leipzig, German.
- Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55. doi:10.1186/1297-9686-41-55
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian statistic*. Oxford University, New York, NY. p. 625–631.
- Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347–363. doi:10.1534/genetics.109.103952
- Habier, D., R. L. Fernando, and J. C. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397. doi:10.1534/genetics.107.081190
- Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi:10.1186/1471-2105-12-186
- Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:5. doi:10.1186/1297-9686-42-5
- Hayashi, T., and H. Iwata. 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics* 14:34. doi:10.1186/1471-2105-14-34
- Illumina. 2006. “TOP/BOT” strand and “A/B” allele: A guide to Illumina’s method for determining strand and allele for the GoldenGate and Infinium assays. Technical note. http://res.illumina.com/documents/products/technotes/technote_topbot.pdf (Accessed 15 October 2011.)
- Lee, J., H. Cheng, D. Garrick, B. Golden, J. Dekkers, K. Park, D. Lee, and R. Fernando. 2017. Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle. *Genet. Sel. Evol.* 49:2. doi:10.1186/s12711-016-0279-9
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663. doi:10.3168/jds.2009-2061
- Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. *Genetics* 180:611–618. doi:10.1534/genetics.108.088575
- Lourenço, 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, et al. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *J. Anim. Sci.* 93:2653–2662. doi:10.2527/jas.2014-8836
- Lourenço, D. A. L., S. Tsuruta, B. D. Fragomeni, Y. Masuda, I. Pocnic, I. Aguilar, J. K. Bertrand, D. W. Moser, and I. Misztal. 2016. Issues in commercial application of single-step genomic BLUP for genetic evaluation in American Angus. *J. Anim. Sci.* 94(E. Suppl 5):144–145. doi:10.2527/jam2016-0303
- MacNeil, M. D., J. D. Nkrumah, B. W. Woodward, and S. L. Northcutt. 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. *J. Anim. Sci.* 88:517–522. doi:10.2527/jas.2009-2022
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648–4655. doi:10.3168/jds.2009-2064
- Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). In: *Proc. 7th World Genet. Appl. Livest. Prod.*, Montpellier, France. CD-ROM Communication Number 28-07.
- Neves, H. H., R. Carvalheiro, A. M. O’Brien, Y. T. Utsunomiya, A. S. do Carmo, F. S. Schenkel, J. Sölkner, J. C. McEwan, C. P. Van Tassell, J. B. Cole, et al. 2014. Accuracy of genomic predictions in *Bos indicus* (Nelore) cattle. *Genet. Sel. Evol.* 46:17. doi:10.1186/1297-9686-46-17
- O’Brien, A. M. P., Y. T. Utsunomiya, G. Mészáros, D. M. Bickhart, G. E. Liu, C. P. Van Tassell, T. S. Sonstegard, M. V. Da Silva, J. F. Garcia, and J. Sölkner. 2014. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet. Sel. Evol.* 46:19. doi:10.1186/1297-9686-46-19
- R Core Team. 2013. The R project for statistical computing. <http://www.R-project.org/> (Accessed 14 May 2013.)
- Saatchi, M., M. C. McClure, S. D. McKay, M. M. Rolf, J. Kim, J. E. Decker, T. M. Taxis, R. H. Chapple, H. R. Ramey, S. L. Northcutt, et al. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Sel. Evol.* 43:40. doi:10.1186/1297-9686-43-40
- Saatchi, M., R. D. Schnabel, M. M. Rolf, J. F. Taylor, and D. J. Garrick. 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genet. Sel. Evol.* 44:38. doi:10.1186/1297-9686-44-38
- Saatchi, M., J. Ward, and D. J. Garrick. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *J. Anim. Sci.* 91:1538–1551. doi:10.2527/jas.2012-5593
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute: An efficient imputation algorithm for dairy cattle populations. *J. Dairy Sci.* 94:421.

- Smith, B. J. 1997. BOA: An R package for MCMC output convergence assessment and posterior inference. *J. Stat. Softw.* 21:1–37. doi:10.18637/jss.v021.i11
- Strandén, I., and D. J. Garrick. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* 92:2971–2975. doi:10.3168/jds.2008-1929
- Teixeira, B. B. M., M. D. MacNeil, R. F. da Costa, N. J. L. Dionello, M. J. Yokoo, and F. F. Cardoso. 2018. Genetic parameters and trends for traits of the Hereford and Braford breeds in Brazil. *Livest. Sci.* 208:60–66. doi:10.1016/j.livsci.2017.12.008
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. doi:10.3168/jds.2007-0980
- 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. doi:10.3168/jds.2008-1514