Iranian Society of Animal Science

Journal of Livestock Science and Technologies

Shahid Bahonar University of Kerman

ISSN: 2322-3553 (Print) ISSN: 2322-374X (Online)

Paper type: Original Research

The accuracy of breeding values for body size latent trait in pigs under different prediction models

Elahe Sanjari Banestani [1,], Ali Esmailizadeh [1], Mehdi Momen [2], Ahmad Ayatollahi Mehrgardi [1], Morteza Mokhtari [3]*

[1] *Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran*
[2] *Department of Surgical Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA*
[3] *Department of Animal Science, Faculty of Agriculture, University of Jiroft, P.O. Box 364, Jiroft, Iran*

[*]Corresponding author,
E-mail address:
elahesanjari1366@agr.uk.ac.ir

Received: 18 Jan 2023,
Accepted: 26 Apr 2023,
Published online: 04 May 2023,

ORCID
Elahe Sanjari Banestani
0000-0002-8205-3282
Ali Esmailizadeh
0000-0003-0986-6639
Mehdi Momen
0000-0002-2562-2741
Ahmad Ayatollahi Mehrgardi
0000-0001-5278-8459
Morteza Mokhtari
0000-0002-3971-628X

Abstract The present study was performed to quantify a latent variable for body size (BS) from the five linear body measurements, including body length (BL), body height (BH), chest width (CW), chest girth (CG), and tube girth (TG). The study population consisted of N= 5573 Yorkshire pigs, 592 individuals out of them were genotyped using a PorcineSNP80 BeadChip. The body size latent variable was determined using Confirmatory Factor Analysis (CFA). Then, the accuracy of breeding values was obtained using pedigree-based best linear unbiased prediction (PBLUP), genomic best linear unbiased prediction (GBLUP), and single-step genomic best linear unbiased prediction (ssGBLUP) models. The overall fit indices, including standardized root mean square residual (SRMR), root mean square error of approximation (RMSEA), Tucker-Lewis Index (TLI), and comparative fit index (CFI) were obtained for the BS as 0.03, 0.09, 0.93, and 0.96, respectively which imply the adequacy of the considered model for BS construct. The performance of models was measured in a 5-fold cross-validation with 10 repeats to get a more accurate measure of the model's performance. The accuracy of models was compared via the correlation between predicted breeding values (PBV) and estimated breeding values (EBV) metric which was 0.37, 0.30, and 0.28 for PBLUP, ssGBLUP, and GBLUP, respectively. Furthermore, the goodness of fit is measured by the mean square of error (MSE) and Pearson's correlations r(y, $\hat{y}$) between observed and predicted phenotypes. The lowest MSE and the highest Pearson's correlations were obtained under PBLUP while the highest MSE and the lowest Pearson's correlations were obtained under GBLUP. The obtained results showed the GBLUP method generally provided lower prediction accuracies than PBLUP and ssGBLUP methods, and also ssGBLUP generated lower prediction accuracy than traditional PBLUP. The performance of ssGBLUP and GBLUP was lower than expected mainly due to the small number of genotyped animals.

Keywords: body dimension, cross-validation, genetic evaluation, latent variable, pig

Introduction

Genetic improvement of the production traits in an animal population has mainly been achieved by selecting the best animals among the current generation to be served as parents of the next generation (Song et al., 2019). To further expedite genetic improvement, the genomic data can be integrated with the relevant statistical models for selection, called 'genomic selection' (Meuwissen et al., 2001). Generally, geno-

mic prediction employs information on all genotyped animals, however, in some breeding programs having genotypic information for all individuals may not be possible. The single-step BLUP (ssGBLUP) is a beneficial method that combines information from pedigree and genomic marker data which results in more accurate predictions of breeding values (Legarra et al., 2009; Christensen and Lund, 2010). Many important traits in domestic animals could be characterized by a set of traits, which often are quantitative (Leal-Gutierrez et al., 2018). In the animal breeding and genetics context, economically important traits are analyzed routinely by applying the multi-trait mixed models (MTM) which provide important insights regarding the genetic and phenotypic correlations among the traits (Silva et al., 2021).

A typical characteristic of the MTM is the increase in the number of parameters with the increase in the number of the traits involved in genetic analysis, which may compromise the feasibility of genetic analysis (Silva et al., 2021). On the other hand, certain relationships also exist among these traits given that they measure some common attributes of the system (Leal-Gutierrez et al., 2018). Dimension reduction models can be applied to phenotypic data to improve the prediction of complex traits and reduce computational complexity. For example, Silva et al. (2021) showed that a small group of latent variables can be used to reduce the data dimensionality and, consequently, the complexity attributed to the model over-parameterization. Latent variables are defined as variables that are not directly measurable but can be characterized by several observed phenotypes. Latent variable modeling provides the opportunity to investigate biologically complex phenomena by reducing at the same time data dimensionality because many phenotypes are combined to represent a few underlying concepts of interest (Leal-Gutierrez et al., 2018). The latent traits are assumed to be unobserved, but they are believed to explain the covariation among the observed variables. Confirmatory Factor Analysis (CFA) is a method that can be used to reduce the dimensionality of phenotypic data in genomic prediction. CFA is usually used to test the hypothesis that a set of observed variables are related to a smaller number of latent factors, also known as latent variables or latent traits (Momen et al., 2021).

The body size (BS) trait is an important trait that can reflect the overall appearance of animals (Liu et al., 2021). Body size is a typical quantitative (or complex) trait; understanding the genetic mechanism of body size differences among individuals can effectively help control the growth and production of animals (Niu et al., 2013). Compared with the description of physical appearance, body size traits can objectively reflect the response of pigs to their environment and other aspects (Ohnishi and Satoh, 2018). In pig breeding, the body shape character index is often used as the most direct production index of a pig (Liu et al., 2021). Body measurement and morphological traits such as body length, body height, chest width, rump width, and heart girth have been considered as measures of body size in sheep (Kominakis et al., 2017, goat (Rahmatalla et al., 2018) and pig (Song et al., 2019; Liu et al., 2021) species.

The purpose of this study is to identify and quantify a latent variable for body size (BS) from five linear body measurements of Yorkshire pigs and compare the performance of the GBLUP, ssGBLUP, and PBLUP methods. Identifying which of the models used provides the most accurate breeding values, which can be used to make informed decisions in the breeding program of the pig industry to improve the body size characteristics.

Materials and methods

Data source and phenotypes

The data used in the present study originated from an elite Chinese pig breeding farm that is a descendant of American Yorkshire populations and was downloaded from http://figshare.com/articles/single-step-strategies/7434203. The phenotypic records comprised linear body measurements including body length (BL), body height (BH), chest width (CW), chest girth (CG), and tube girth (TG) of 5573 pigs. In total, 7,020 animals were traced back to construct a pedigree relationship matrix. CFC program was used for checking errors in the pedigree, preparing it for the subsequent analyses, and computing inbreeding coefficients of animals (Sargolzaei et al., 2006). The pedigree structure of the data set used is shown in Table 1. Descriptive statistics for the considered body measurement traits used for constructing BS latent variable are shown in Table 2. More detailed information about the animals and phenotypes was presented by Song et al. (2019) and Liu et al. (2021).

Genotype data, quality control, and imputation

In the present study, genotype data on 592 out of 5573 Yorkshire pigs (Song et al., 2019) were used. Animals were genotyped using the PorcineSNP80 Bead Chip (Illumina, San Diego, CA, United States), which includes 68,528 SNPs across the whole pig genome. PLINK software version 2.0 (Purcell et al., 2007) was used for quality control of genotype data and SNPs with a maximum missing rate of 0.10, a maximum individual missing rate of 0.10, a minor allele frequency ≤ 0.01, and Hardy-Weinberg equilibrium with a P-value < 10^{-7} were excluded. Missing genotypes were imputed using Beagle software version 5.4 (Browning and Browning, 2009). After quality control, all genotyped individuals remained and 52,710 SNPs were finally used for further analysis.

Latent variable modeling by confirmatory Factor Analysis (CFA)

The confirmatory factor analysis (CFA) technique is used to extract the underlying latent factors that contribute to variations in body size measurements. To conduct a CFA for extracting body size latent factor, one needs to specify a measurement model, which defines the relationships between the observed variables and the latent factor(s). The following CFA model was considered in this study:

$$\mathbf{X} = \mathbf{\Lambda\xi} + \mathbf{\delta}$$

where **X** is the matrix of measured variables (BL, BH, CW, CG, and TG), ξ is the vector of latent factors, **Λ** matrix contains the factor loadings that associate these factors to the measured variables, and δ the residuals vector. The model was fitted by applying the maximum likelihood estimation. The lavaan package (R Core Team, 2021; Rosseel, 2012) was employed to fit the above CFA model.

The overall fit of the models was evaluated by using fit indices, including standardized root mean square residual (SRMR) (Bentler., 1990), root mean square error of approximation (RMSEA) (Steiger., 1990), Tucker-Lewis Index (TLI) (Bentler., 1990), and comparative fit index (CFI) (Bentler., 1990).

Statistical prediction models

Pedigree-based BLUP

The values obtained for BS latent variable were standardized for the subsequent analyses. The breeding values for BS were predicted using the traditional animal model with a pedigree-based relationship matrix as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}$$

where, **y** is the vector of records for BS, **b**, **g**, and **e,** are the vectors of fixed effects (herd-year-season-sex), additive genetic effects, and residual effects, respectively. The incidence matrices of **X** and **Z** associate **b** and **g** with **y**, respectively. Furthermore, the body weight of animals was considered a covariate. It is assumed that additive genetic effects followed a normal distribution of $\mathbf{g} \sim N(\mathbf{0}, \mathbf{A}\sigma_g^2)$, in which **A** is the matrix of additive pedigree-based relationships and σ_g^2 is the variance of additive genetic effects. **e** is the vector of random residuals with the distribution of $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, in which **I** is the identity and σ_e^2 is the residual variance. The BW of pigs was considered a linear covariate for BS.

Genomic BLUP (GBLUP)

The GBLUP model uses a genomic relationship matrix (**G**) derived from the SNP markers instead of the pedigree-based numerator relationship matrix. To predict GEBV of all genotyped individuals the following GBLUP (VanRaden, 2008) model was used:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

where **y**, **X**, **b**, and **e** are as defined previously. **a** is a vector of genomic breeding values and **Z** is a design matrix associating genomic breeding values to records.

It is assumed that genomic breeding values followed a normal distribution of $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G}\sigma_a^2)$, in which **G** is the genomic relationship matrix (VanRaden, 2008) and σ_a^2 is the variance of genetic effects for this model.

Single step GBLUP (ssGBLUP)

The ssGBLUP model exploits the information of both genotyped and non-genotyped animals by joint using the marker and pedigree information for genetic evaluations. The single-trait ssGBLUP has the same model as BLUP, except vector **g** is assumed to follow a normal distribution $\mathbf{g} \sim N(\mathbf{0}, \mathbf{H}\sigma_g^2)$. Following Aguilar et al. (2010), the **H** was defined as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A_{11}} + \mathbf{A_{12}}\mathbf{A_{22}^{-1}}(\mathbf{G_w} - \mathbf{A_{22}})\mathbf{A_{22}^{-1}}\mathbf{A_{12}'} & \mathbf{A_{12}}\mathbf{A_{22}^{-1}}\mathbf{G_w} \\ \mathbf{G_w}\mathbf{A_{22}^{-1}}\mathbf{A_{12}'} & \mathbf{G_w} \end{bmatrix}$$

in which $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, and $\mathbf{A}_{22}$ were the sub-matrices of **A**, and subscripts 1 and 2 denote non-genotyped and genotyped animals, respectively. The inverse of H was:

$$\mathbf{H^{-1}} = \begin{bmatrix} \mathbf{G_w^{-1}} - \mathbf{A_{22}^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \mathbf{A^{-1}}$$

For avoiding problems due to singularity, $\mathbf{Gw} = 0.95\mathbf{G_a} + 0.05\mathbf{A_{22}}$ (Aguilar et al., 2010). $\mathbf{G_a}$ is an adjusted **G**, to avoid the differences in scale and location between the coefficients of **G** and the pedigree relationship matrix ($\mathbf{A}_{22}$), the **G** matrix was adjusted ($\mathbf{G}_{adj}$) according to Christensen et al. (2012):

$$\mathbf{G}_{adj} = \beta\mathbf{G} + \alpha$$

where α and β are adjustment factors derived from the following equations:

Avg. diag $(\mathbf{G})\beta + \alpha$ = Avg. diag($\mathbf{A}_{22}$)

Avg. offdiag $(\mathbf{G})\beta + \alpha$ = Avg. offdiag($\mathbf{A}_{22}$)

where Avg. diag is the average of the diagonal elements, and Avg. offdiag is the average of the off-diagonal elements.

Model's comparison

The goodness of fit for the considered models was evaluated using two statistical measures: mean square of error (MSE) and Pearson's correlation coefficient r(y, $\hat{y}$), between observed and predicted records. The MSE is a measure of the deviation between the observed and predicted values, while Pearson's correlation coefficient measures the strength of the linear relationship between the observed and predicted values. A five-fold cross-validation scheme was conducted to evaluate the models' performance. This method of evaluation involves dividing the data into five equal parts, where one part is held out as a validation set and the model is trained on the remaining four parts. This process is repeated five times, with each of the five parts held out once. This allows for a more robust evaluation of the model's performance and helps to avoid overfitting.

The cross-validation process was then replicated ten times, resulting in ten averaged accuracies of genomic prediction metrics. For PBLUP, GBLUP, and ssGBLUP,-

the validation set was the same in each replicate of the five-fold cross-validation.

The accuracy of genomic prediction was evaluated via two methods, the first as $r_{(PBV, EBV)}$, the correlation between predicted breeding values (PBV) in the validation population, and estimated breeding value (EBV). The second method was by using the following formula:

$$\sqrt{1 - \frac{PEV}{\sigma_a^2}}$$

where, PEV is prediction error variance and σ_a^2 is additive genetic variance. Model fitting and genetic analysis were carried out using the WOMBAT program (Meyer, 2013).

Results and discussion

Table 1 presents the pedigree structure of the population used in this study. Registered animals originated from 194 sires and 1444 dams. Among them, 23.33 % had progeny while the remaining had no progeny. Animals with both parents known, both parents unknown, and one parent known comprised 96.60%, 3.18%, and 0.22 %, respectively, implying high quality of the pedigree for genetic analysis. Average inbreeding coefficients in all and inbred animals were 1.38% and 2.17%, respectively. The average inbreeding coefficient in pigs can vary depending on the population and breeding program. In commercial pig populations, the average inbreeding coefficient is generally between 0.01 and 0.05, with some higher values being observed in specialized or closed populations. The range of the inbreeding coefficient in pigs can also vary, but it usually falls between 0 and 0.2 (Dekkers et al., 2011; Lopes et al., 2019).

Table 1. Pedigree structure of the population

Item	Numbers
Individuals in total	7019
Inbreds in total	4484
Sires in total	194
Dams in total	1444
Individuals with progeny	1638
Individuals with no progeny	5381
Founders	223
Individuals with both parents known	6781
Individuals with both parents unknown	223
Individuals with one parent unknown	15
Average inbreeding coefficients (%)	1.38
Average inbreeding coefficients in the inbreds (%)	2.17
Maximum of inbreeding coefficients (%)	31.25
Minimum of inbreeding coefficients	0.04

Figure 1 provides a comprehensive representation of the underlying relationships within the system under investigation and incorporates BS as a latent variable in the center of the five measured variables. The goodness of fit metrics including CFI (0.96), TLI (0.93), RMSEA (0.09), and SRMR (0.03) indicated the adequacy of the confirmatory factor model proposed for the latent variable of BS. Leal-Gutierrez et al. (2018) applied a similar methodology to that used in the present study for defining the latent variable of carcass quality by applying observed variables of quality grade, fat over ribeye, and marbling in beef cattle.

Factors lodings or structural coefficients for BL, CG, TG, BH, and CW were 0.62, 0.84, 0.57, 0.54, and, 0.63, respectively. Positive loading factors indicate that an increase in the observed variable is associated with an increase in the latent variable, while negative loading factors indicate the opposite relationship. In this study, the CG variable had the strongest relationship with the BS latent variable based on the estimated loading factors. The direction of the loading factors with BS is positive and agrees with the theoretical model where the considered observed variables are positively related to BS. The positive direction was expected given that the BS construct reflects mainly body measurement traits (Kominakis et al., 2017).

Table 3 presents the goodness of fit metrics for the evaluated models, which were determined using both mean square error (MSE) and Pearson's correlation coefficient between observed and predicted body size (BS). The lowest MSE and the highest Pearson's correlation values were obtained under the PBLUP model while the highest MSE and the lowest correlation were obtained under GBLUP. In other words, the considered models can be ranked as PBLUP, ssGBLUP, and GBLUP in terms of goodness of fit.

The prediction accuracies for BS using different models are shown in Table 4. For the cross-validation assessment scheme, a subset of approximately 100 genotyped individuals was randomly chosen and considered as the validation set. The remaining genotyped individuals were considered as the reference population, and all non-genotyped animals were included in the analysis for both PBLUP and ssGBLUP models. Due to the limited size of the genotyped population, the correlation coefficient (r) between the estimated breeding values (EBVs) and the predicted breeding values (PBVs) as a measure of the accuracy of predictions under the GBLUP method was found to be lower (0.28) than the c-

orresponding values under both PBLUP (0.37) and ssGBLUP (0.30) models. This suggests that the PBLUP method performed better than GBLUP in terms of the accuracy of predictions. A similar trend was also observed applying the second method ($\sqrt{(1- (PEV/\sigma_a^2)}$) used for computing the accuracies of breeding values under the considered models. As presented in Table 4, the PBLUP method yielded the highest accuracy with a value of 0.74, followed by the ssGBLUP method with 0.58 and the GBLUP method with 0.52. Forni et al. (2011) reported that there was no large difference in the accuracy of breeding values obtained from the genomic relationship (0.799) and traditional pedigree-based (0.791) matrices in a pig population.

In the present study, for taking advantage of all available information, all the genotyped and non-genotyped individuals were considered to construct the H matrix in ssGBLUP. As both the genomic and pedigree information was utilized by ssGBLUP, previous studies have shown that the ssGBLUP method was superior to both the GBLUP method (Christensen et al., 2012; Song et al., 2019) and traditional pedigree-based BLUP (Gao et al., 2012; Koivula et al., 2015), in which only genomic or pedigree information was used.

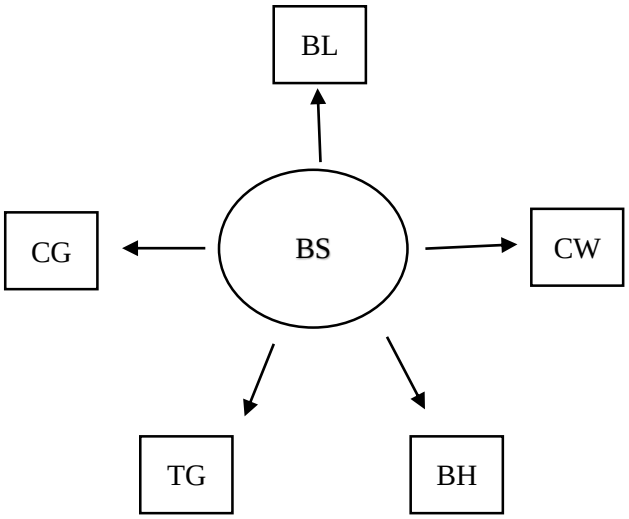

Fig. 1. Relationship between observed variables and the latent variable of body size (BS) in the final structural equation model. Body length (BL), body height (BH), chest width (CW), chest girth (CG), and tube girth (TG)

Table 2. Descriptive statistics for the body measurement traits

Trait [a]	No. of records	Mean	S.D.	C.V. (%)	Min.	Max.
BL (cm)	5573	108.89	6.18	5.67	88	134
BH (cm)	5573	62.87	2.92	4.64	51	75
CW (cm)	5573	29.75	2.31	7.76	19	38
CG (cm)	5573	104.58	5.75	5.50	85	126
TG (cm)	5573	17.98	1.03	5.73	13	23

[a] body length (BL), body height (BH), chest width (CW), chest girth (CG), and tube girth (TG)

Table 3. Comparison of the goodness of fit metrics for PBLUP, ssGBLUP, and GBLUP genomic prediction models for body Size latent trait (BS)

Model [a]	Measure	
	r(y, $\hat{y}$) [b]	MSE [b]
PBLUP	0.859 ***	0.356
GBLUP	0.315 ***	0.943
ssGBLUP	0.660 ***	0.566

[a] PBLUP: Pedigree-based BLUP (based on pedigree relationship matrix), GBLUP: Genomic BLUP (based on genomic relationship matrix), ssGBLUP: single-step GBLUP.

[b] r(y, $\hat{y}$) : Pearson's correlation coefficient between the observed and predicted values of body size, MSE: mean square of the error

Table 4. Accuracy of the breeding values of animals under the considered models

Model [a]	r (EBV and PBV) [b]	$\sqrt{1 - \frac{PEV}{\sigma_a^2}}$ [c]
P-BLUP	0.37	0.74
G-BLUP	0.28	0.52
ss-GBLUP	0.30	0.58

[a] PBLUP: Traditional BLUP (based on pedigree relationship matrix), GBLUP: Genomic BLUP (based on genomic relationship matrix), ssGBLUP: single-step GBLUP.

[b] Based on a five-fold cross-validation evaluation, EBV: estimated breeding value (considering all animals).

[c] PBV: predicted breeding value (in the validation set) σ_a^2: additive genetic variance

The lower performance of ssGBLUP might be explained by several reasons. Firstly, the genotyped individuals were not large enough to improve the accuracy of genomic predictions. In other words, approximately 400 genotyped reference individuals could not probably provide more extra information in comparison with the pedigree information available from about 7,000 individuals. Similar to us, the lower accuracy of GBLUP than PBLUP was reported by Song et al. (2019) for body measurement traits in pigs. In a simula-

tion study, they also demonstrated that GBLUP yielded lower prediction accuracy than PBLUP, even when the genotyped reference population size reached 3,000, but it was still very small compared to non-genotyped individuals of 26,000 utilized by the PBLUP. Lourenco et al. (2014) also reported that GBLUP performed worse than PBLUP and ssGBLUP in terms of the accuracy of breeding values in a relatively small genotyped dairy population. In contrast, a study by Choi et al. (2017) that compared the accuracy of breeding values for the intramuscular fat in Hanwoo (Korean beef cattle) using different genomic and pedigree-based relationship matrices, found that the accuracy of the genomic-based model was 1.5 times higher than that of the pedigree-based model.

Secondly, in our study, the heritability estimate for the latent variable of BS was a high value of 0.45, which can obtain sufficient accuracy for the traditional BLUP method, and improvement from genomic prediction was not large as expected. Goddard and Hayes (2009) pointed out that for the traits with medium to high heritably small progress will be obtained when using a small reference population. Therefore, it is important to consider the population size and heritability when choosing a genomic prediction method and interpreting the results.

Conclusion

Latent variable modeling can be used to identify latent variables, which are variables that are not directly measurable but can be characterized by several observed phenotypes. This can help researchers in genetics to understand the underlying concepts that drive the covariation among the observed variables and make more accurate predictions about complex traits. Here, we showed that the CFA can be used to reduce the dimensionality of phenotypic data in genomic prediction and also to test the hypothesis that a set of observed variables are related to a smaller number of latent factors, also known as latent variables or latent traits like body size in pigs. Because the body size in pigs can objectively reflect the response of animals to their environment and other aspects. In pig breeding, the body shape character index is often used as the most direct production index of a pig. Overall, we showed that dimension reduction models and their application to phenotypic data can provide you with the tools and knowledge to analyze large sets of data and extract meaningful insights in the field of genetics and animal breeding. This can help you make more accurate predictions, understand the underlying concepts that drive complex traits, and make more informed decisions in research and breeding.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., Lawlor, T.J., 2010. Hot topic: A unified -- approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93, 743-752.

Bentler, P. M., 1990. Comparative fit indexes in structural models. *Psychological Bulletin* 107, 238-246.

Browning, B. L., Browning, S. R., 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics* 84, 210-223.

Choi, T., Lim, D., Park, B, Sharma, A., Kim, J.J., Kim, S., Lee, S.H., 2017. Accuracy of genomic breeding value prediction for intramuscular fat using different genomic relationship matrices in Hanwoo (Korean cattle). *Asian-Australasian Journal of Animal Science*s 30, 907-911.

Christensen, O.F., Lund, M.S., 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42, 2.

Christensen, O.F., Madsen, P., Nielsen, B., Ostersen, T., Su, G., 2012. Single-step methods for genomic evaluation in pigs. *Animal* 6, 1565-1571.

Dekkers, J.C.M., Mathur, P.K., Knol, E.F., 2011. Genetic improvement of the pig. In: Rothschild, M.F., Ruvinsky, A. (Eds.), The Genetics of the Pig, CAB International, UK, pp. 390-425.

Forni, S., Aguilar, I., Misztal, I., 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43:1.

Gao, H.D., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., Guosheng, S., 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics Selection Evolution* 44:8.

Goddard, M.E., Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Review Genetics* 10, 381-391.

Koivula, M., Stranden, I., Poso, J., Aamand, G.P., Mantysaari, E.A., 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. *Journal of Dairy Science* 98, 2775-2784.

Kominakis, A., Hager-Theodorides, A.L., Zoidis, E., Saridaki, A., Antonakos, G., Tsiamis, G., 2017. Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep. *Genetics Selection Evolution* 49, 41.

Leal-Gutierrez, J.D., Rezende, F.M., Elzo, M.A., Johnson, D., Penagaricano, F., Mateescu, R.G., 2018. Structural equation modeling and whole-genome scan-

s uncover chromosome regions and enriched pathways for carcass and meat quality in beef. *Frontiers in Genetics* 9, 532.

Liu, H., Song, H., Jiang, Y., Jiang, Y., Zhang, F., Liu, Y., Shi, Y., Ding, X., Wang, C., 2021. A single-step genome wide association study on body size traits using imputation-based whole-genome sequence data in Yorkshire pigs. *Frontiers in Genetics* 12:629049.

Legarra, A., Aguilar, I., Misztal, I., 2009. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92, 4656-4663.

Lopes, J.S., Rorato, P.R.N., Mello, F.C.B., Freitas, M.S.D., Prestes, A.M., Garcia, D.A., Oliveira, M.M.D., 2019. Strategies to control inbreeding in a pig breeding program: a simulation study. *Ciencia Rural*, 49.

Lourenco, D.A.L., Misztal, I., Tsuruta, S., Aguilar, I., Ezra, E., Ron, M., Shirak, A., Weller, J.I., 2014. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. *Journal of Dairy Science* 97, 1742–1752.

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

Meyer, K., 2013. WOMBAT- A Programme for Mixed Model Analyses by Restricted Maximum Likelihood. User Notes, Animal Genetics and Breeding Unit, Armidale, Australia.

Momen, M., Bhatta, M., Hussain, W., Yu, H., Morota, G., 2021. Modeling multiple phenotypes in wheat using data-driven genomic exploratory factor analysis and Bayesian network learning. *Plant Direct* 5, p.e00304.

Niu, P., Kim, S., Choi, B., Kim, T., Kim, J., Kim, K., 2013. Porcine insulinlike growth factor 1 (IGF1) gene polymorphisms are associated with body size variation. *Genes Genomics* 35, 523-528.

Ohnishi, C., Satoh, M., 2018. Estimation of genetic parameters for performance and body measurement traits in Duroc pigs selected for average daily gain, loin muscle area, and backfat thickness. *Livestock Science* 214, 161-166.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., Sham, P.C., 2007. Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81, 559-575.

R Development Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rahmatalla, S.A., Arends, D., Reissmann, M., Wimmers, K., Reyer, H., Brockmann, G.A., 2018. Genome-wide association study of body morphological traits in Sudanese goats. *Animal Genetics* 49, 478-482.

Rosseel, Y., 2012. lavaan: an R package for structural equation modeling. *Journal of Statistical Software 48*, 1-36.

Sargolzaei, M., Iwaisaki, H., Colleau, J.J., 2006. CFC: A tool for monitoring genetic diversity, Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil.

Silva, H.T., Paiva, J.T., Botelho, M.E., Carrara, E.R., Lopes, P.S., Silva, F.F., Veroneze, R., Ferraz, J.B.S., Eler, J.P., Mattos, E.C., Gaya, L.G., 2021. Searching for causal relationships among latent variables concerning performance, carcass, and meat quality traits in broilers. *Journal of Animal Breeding and Genetics* 139, 181-192.

Song, H., Zhang, J., Zhang, Q., Ding, X., 2019. Using different single-step strategies to improve the efficiency of genomic prediction on body measurement traits in pig. *Frontiers in Genetics* 9, 730.

Steiger, J. H., 1990. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 25, 173-180.

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414-4423.