



Factor analysis applied to genome prediction for high-dimensional phenotypes in pigs

F.R.F. Teixeira¹, M. Nascimento¹, A.C.C. Nascimento¹, F.F. e Silva²,
C.D. Cruz³, C.F. Azevedo¹, D.M. Paixão², L.M.A. Barroso¹, L.L. Verardo²,
M.D.V. de Resende⁴, S.E.F. Guimarães² and P.S. Lopes²

¹Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil

²Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil

³Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, Brasil

⁴Embrapa Florestas, Colombo, PR, Brasil

Corresponding author: F.R.F. Teixeira

E-mail: filipeformiga0@gmail.com

Genet. Mol. Res. 15 (2): [gmr.15028231](http://dx.doi.org/10.4238/gmr.15028231)

Received December 8, 2015

Accepted January 18, 2016

Published May 13, 2016

DOI <http://dx.doi.org/10.4238/gmr.15028231>

ABSTRACT. The aim of the present study was to propose and evaluate the use of factor analysis (FA) in obtaining latent variables (factors) that represent a set of pig traits simultaneously, for use in genome-wide selection (GWS) studies. We used crosses between outbred F2 populations of Brazilian Piau X commercial pigs. Data were obtained on 345 F2 pigs, genotyped for 237 SNPs, with 41 traits. FA allowed us to obtain four biologically interpretable factors: “weight”, “fat”, “loin”, and “performance”. These factors were used as dependent variables in multiple regression models of genomic selection (Bayes A, Bayes B, RR-BLUP, and Bayesian LASSO). The use of FA is presented as an interesting alternative to select individuals for multiple variables simultaneously in GWS studies; accuracy measurements of the factors

were similar to those obtained when the original traits were considered individually. The similarities between the top 10% of individuals selected by the factor, and those selected by the individual traits, were also satisfactory. Moreover, the estimated markers effects for the traits were similar to those found for the relevant factor.

Key words: Genome-enabled prediction; Multivariate analysis; SNP effects

INTRODUCTION

Meuwissen et al. (2001) idealized genome wide selection (GWS), aiming to directly incorporate molecular information on an individual's genetic merit for breeding purposes. This methodology can be used to improve the selection efficiency of traits for which phenotypic measurements are expensive or cannot be obtained on candidates under selection, e.g., carcass traits.

Recently, several studies on GWS were carried out in pigs. Badke et al. (2014) estimated the accuracy of genomic breeding values (GEBV) for three traits (backfat thickness, number of days to 250 lb, and loin muscle area) in a Yorkshire pig population; de Campos et al. (2015) predicted the GEBV for fat related traits (backfat thickness and loin depth) in a commercial pig population.

In general, GWS has been applied to individual traits, i.e., the results obtained are valid only for a single trait. However, in breeding programs, interest falls upon several traits simultaneously. Thus, an approach that simultaneously incorporates information from multiple traits in the selection process may be of interest.

A possible method for this approach is factor analysis (FA). This methodology allows the collection of latent variables (factors) that represent a set of original variables. In this context, further analysis can be more easily performed using the obtained factors. Specifically, in GWS, such methodology might allow the selection of individuals for a set of traits simultaneously, thus reducing the sometimes-extensive computational time required to obtain results.

de los Campos and Gianola (2007) employed FA successfully in assessments of the multitrait mixed model based on pedigree information. Silva et al. (2011) applied FA in order to detect quantitative trait loci (QTLs) in crosses between outbred populations, where the factors represented groups of carcass traits. However, to date, there have been no reports concerning the application of this technique in GWS studies.

The aim of this study was to propose and evaluate the use of FA in GWS studies by comparing individuals selected using FA or individual trait methods, using an F2 Piau X commercial pig population. The objective is also compare four Bayesian methods often used in GWS studies.

MATERIAL AND METHODS

Population assessment, phenotypic data, and genotypic information

Phenotypic data was obtained from the Pig Breeding Farm of the Department of Animal Science, Universidade Federal de Viçosa (UFV), MG, Brazil. A three-generation,

resource population was created and managed as described by Band et al. (2005). Briefly, two local breed Piau grand-sires were crossed with 18 grand dams from a commercial line composed of Large White, Landrace, and Pietrain breeds, to produce the F1 generation, from which 11 sires and 54 dams were selected. These F1 individuals were then crossed to produce the F2 population, of which 345 animals were phenotyped. The use of these animals was reviewed and approved by the Committee Bioethics of Veterinary Medicine Department (DVT - UFV) in accordance with the Guide for Care and Use of Experimental Animals of the Canadian Council of Animal Care.

A total of 237 markers, identified for F2 animals, are distributed as follows in the *Sus scrofa* chromosomes: SSC 1 (56), SSC4 (54), SSC7 (59), SSC8 (31), SSC17 (25), and SSCX (12). These markers were obtained only in regions where QTLs have been observed in previous studies in a similar population (Hidalgo et al., 2013). This is characteristic of a thin mapping, with only chromosomal regions of interest mapped, and explains the limited number of markers used (Azevedo et al., 2013b). Animals were genotyped using the Golden Gate[®] genotyping assay with Veracode[®] technology, which provides a robust and flexible platform, together with an Illumina BeadXpress reader. The variables included in the study, along with their abbreviations, are listed in Table 1.

Table 1. Description of the 41 phenotypic variables in the study.

Variable	Description	Variable	Description
CW	Hot carcass weight (kg)	TLW	Total (bone-in) loin weight (kg)
RHCH	Right half carcass weight (kg)	LW	Boneless loin weight (kg)
TWA	Total number of weaned alive	BCW	Bacon weight (kg)
TBA	Total number of bared alive	RW	Rib weight (kg)
SA	Slaughter age (days)	SLW	Sirloin weight (kg)
CY	Carcass yield (%)	AF	Abdominal fat (kg)
MBCC	Carcass length by the Brazilian carcass classification method (cm)	FI	Feed intake (kg)
MLC	Carcass length by the Brazilian carcass classification method (cm)	ADG	Average daily gain (kg)
SBT	Higher backfat thickness on the shoulder region (mm)	FG	Feed-gain ratio (kg/kg)
LR	Midline backfat thickness immediately after the last rib (mm)	TN	Total tea number
LL	Midline backfat thickness between last and next to last but one lumbar vertebrae (mm)	SW	Slaughter weight (kg)
L	Midline lower backfat thickness above the last lumbar vertebrae (mm)	BW	Birth weight (kg)
BFT	Backfat thickness (mm)	pH45	pH 45 min after slaughter
BCD	Bacon depth (mm)	pH24	pH 24 h after slaughter
LD	Loin depth (mm)	L	Brightness
LEA	Loin eye area (cm ²)	IMF	Intramuscular fat (%)
HEART	Heart weight (kg)	DL	Drip loss (%)
THW	Total ham weight (kg)	CL	Cooking loss (%)
HW	Skinless and fatless ham weight (kg)	SF	Shear force (kg/cm ²)
TBSW	Total Boston shoulder weight (kg)	C	Chroma
TPSW	Total picnic shoulder weight (kg)		

In order to study several traits simultaneously, FA was applied. This methodology enables the grouping of the original variables in subsets of variables mutually uncorrelated (called latent variables or factors), which can provide practical interpretations.

Factor analysis

The factorial model used for an observable phenotypic variable (trait) can be represented in equation (1), as follows:

$$X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \dots + l_{im}F_m + \varepsilon_i \quad (\text{Equation 1})$$

where $i = 1, 2, \dots, 41$, and , with 41 being the number of original observable phenotypic variables; the coefficient l_{ij} is the factor loading of the i -th phenotypic variable on the j -th factor; $j = 1, 2, \dots, m$; F_1, F_2, \dots, F_m are factors (unobservable random variables); ε_i is the random error vector that is only associated with the i -th phenotypic variable,. The phenotypic variables were corrected for fixed effects of sex, lot, and the presence or absence of the halothane gene.

The suitability of the proposed model was assessed using the Kaiser-Meyer-Olkin criterion (KMO) (Mingoti, 2007) and Bartlett's test (Ferreira, 2011). The number of factors was determined by considering a percentage that explains 70% of the total variability (Ferreira, 2011). After model validation, the allocation of the variables in each factor was made through the loadings (l_{ij}) following a varimax rotation.

Subsequently, the factor scores that were used for GWS analysis were obtained through the regression method given by equation 2, as follows:

$$\hat{F}_j = \hat{\Gamma}^T (\hat{\Gamma} \hat{\Gamma}^T + \hat{\psi})^{-1} (Y_j - \bar{Y}) \quad (\text{Equation 2})$$

where $\hat{\Gamma}$ is the loading matrix, $\hat{\psi}$ is the matrix of specific variances, Y_j is the original trait vector related to j -th sample unit ($j = 1, 2, \dots, 345$), and \bar{Y} is the mean vector regarding the 41 evaluated phenotypic variables. The $\hat{\Gamma}$ and $\hat{\psi}$ matrices were obtained using the main components method (Mingoti, 2007).

Genetic parameters and cross-validation

Four Bayesian methods (Bayes A, Bayes B, RR-Bayes, and Bayesian Lasso) were used to estimate the GEBVs, based on SNP effects estimates for the factors scores (considered as the dependent variable) obtained from the previous step. The accuracy for these GEBVs were calculated according to Resende et al. (2010), as follows in equation 3:

$$r_{q,\hat{q}} = \frac{r_{y,\hat{y}}}{\sqrt{h^2}} \quad (\text{Equation 3})$$

where $r_{q,\hat{q}}$ is the predictive ability of the model, and h^2 is the trait heritability given by equations 4 and 5, respectively:

$$r_{y,\hat{y}} = \frac{\text{Cov}(y,\hat{y})}{\sqrt{\text{Var}(y) \cdot \text{Var}(\hat{y})}} \quad (\text{Equation 4})$$

$$h^2 = \frac{VG}{VF} \quad (\text{Equation 5})$$

and VG and VF being the genetic and phenotypic variances, respectively. The calculations of predictive ability, heritability, and accuracy were based on the genetic merit of individuals (the GEBVs), given by equation 6, as follows:

$$\text{GEBV} = \mathbf{Z}\hat{\mathbf{b}} \quad (\text{Equation 6})$$

where \mathbf{Z} is the genotype matrix, comprising the numerical values assumed at each SNP (0, 1, and 2, respective to aa, aA, and AA); $\hat{\mathbf{b}}$ is the vector containing the estimates of the markers effects, according to the Bayesian methods.

To evaluate the goodness of fit, and in order that the marker effects are not overstated due to the estimation and validation of the same sample (Cruz et al., 2013), a cross-validation technique was used. The F2 pig population was divided into three distinct populations (folds), in which two of these folds were used to estimate the effects, and the third for validation, as described by Azevedo et al. (2013a).

After selecting the best individuals, corresponding to 10% of the validation population, the agreement between each variable and its respective factor was identified, i.e., the similarity between the individuals selected by factor and each original trait. The measure used to evaluate this agreement was Cohen's kappa coefficient (Cohen, 1960). This index can be measured by the equation 7:

$$\hat{k} = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (\text{Equation 7})$$

where $\text{Pr}(a) - \text{Pr}(e)$ represents the proportion of observations in which the correlation occurred beyond those randomly expected, and $1 - \text{Pr}(e)$ is the proportion of observations in which there was no agreement. This measure ranges from 0-1, and the higher the index, higher the agreement between the groups.

Computational tools

To observe the effects of each marker in the variables, and their chromosomes of interest, we used a Manhattan Plot. Statistical calculations were implemented using the R Project for Statistical Computing software (R Development Core Team, 2015). The codes and algorithms used are available at <http://www.det.ufv.br/~moyses/links.php>. For the FA we used the *factanal* function through the *psych* (Revelle, 2015) and *GPArotation* (Bernaards and Jennrich, 2005) packages. The Bayesian models were estimated using the *BGLR* function (with 100,000 iterations, 20,000 of burn-in and thin assuming the value 10), within the *BGLR* package (de los Campos and Rodriguez, 2014). Manhattan plots were obtained with the use of *mhtp* function, from the *gap* package (Zhao, 2007).

RESULTS

KMO index (0.85) and the Bartlett sphericity test, which was statistically significant ($p < 0.01$), showed that the data is suitable for application of FA. According to the criterion that the total variation should exceed 70% (Ferreira, 2011), ten factors, of which four have biological interpretations, were formed (Table 2).

The first factor grouped highly correlated variables related to the weight of the individual, and was named "weight" (Table 2). The second factor included variables related to backfat thickness: higher backfat thickness on the shoulder region (SBT), midline backfat

thickness immediately after the last rib (LR), midline backfat thickness between the last and the next but one lumbar vertebrae (LL), midline lower backfat thickness above the last lumbar vertebrae (L) and backfat thickness (BFT), bacon depth (BCD) and abdominal fat (AF); these variables are associated with “fat”, so this name was given to the second factor. The variables related to the sirloin trait: loin depth (LD), loin eye area (LEA) and boneless loin weight (LW) formed the third factor, denoted as “loin”. Finally, the fourth factor grouped the variables age at slaughter (SA), feed intake (FI), daily weight gain (ADG), and birth weight (BW); “performance” was the name given to this fourth factor. The variables FI, ADG and BW were positively correlated with the “performance” factor, but age of slaughter showed a negative correlation. This result was expected, as the higher the feed intake, daily weight gain, and birth weight of the animal, the greater the performance, and therefore the lower the age before slaughter.

Table 2. Interpretable factors and associated variables, with loadings in parentheses.

Factor	Variables associated
Weight	CW (0.87), RCW (0.87), MBCC (0.82), MLC (0.84), HEART (0.52), THW (0.77), HW (0.78), TBSW (0.71), TPSW (0.86), TLW (0.58), BCW (0.56), RW (0.59), SLW (0.50), and SW (0.85)
Fat	SBT (0.74), LR (0.82), LL (0.86), L (0.88), BFT (0.87), BCD (0.79), and AF (0.75)
Loin	LD (0.83), LEA (0.82), and LW (0.63)
Performance	SA (-0.78), FI (0.83), ADG (0.71), and BW (0.53)

CW = hot carcass weight; RCW = right carcass weight; MBCC = carcass length (Brazilian classification method); MLC = carcass length (American classification method); HEART = heart weight; THW = total ham weight; HW = skinless and fatless ham weight; TBSW = total Boston shoulder weight; TPSW = total picnic shoulder weight; TLW = total (bone-in) loin weight; BCW = bacon weight; RW = rib weight; SLW = sirloin weight; SW = slaughter weight; SBT = higher backfat thickness on the shoulder region; LR = midline backfat thickness immediately after the last rib; LL = midline backfat thickness between last and next to last but one lumbar vertebrae; L = midline lower backfat thickness above the last lumbar vertebrae; BFT = backfat; BCD = bacon depth; AF = abdominal fat; LD = loin depth; LEA = loin eye area; LW = boneless loin weight; SA = slaughter age; FI = feed intake; ADG = average daily gain; BW = birth weight.

Accuracy was measured, for each of the four factors, using the Bayes A, Bayes B, Bayesian ridge regression (Bayesian RR), and Bayesian LASSO (BLASSO) models. Figure 1 shows that the best methodology was BLASSO, and that the second factor, “fat”, showed the most accurate value (0.56) of all four factors.

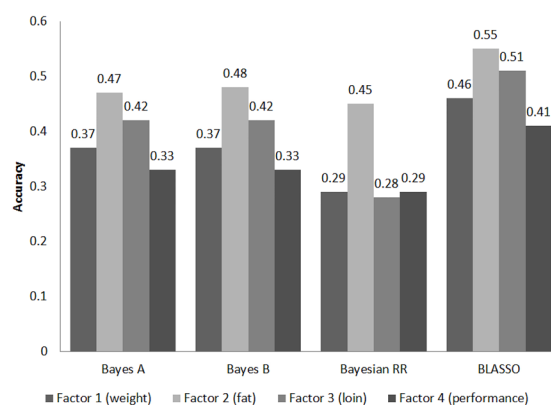


Figure 1. Accuracies for genomic estimated breeding values provided from Bayesian methods for each considered interpretable factor.

Given the higher accuracy, and its practical relevance for the pig industry (Silva et al., 2011; Paixão et al., 2012; Azevedo et al., 2013a, 2015), only the second factor (“fat”) will be discussed. The results for the other factors showed the same pattern, and are available at <http://www.det.ufv.br/~moyses/links.php>.

It is observed that the accuracy value for the factor “fat” (0.56) is similar to those obtained for the variables AF (0.49), BCD (0.63), BFT (0.58), L (0.59), and LR (0.55), that were analyzed individually using the BLASSO method (Figure 2). The accuracies of phenotypic variables LL (0.32) and SBT (0.28) showed values of less than 0.40.

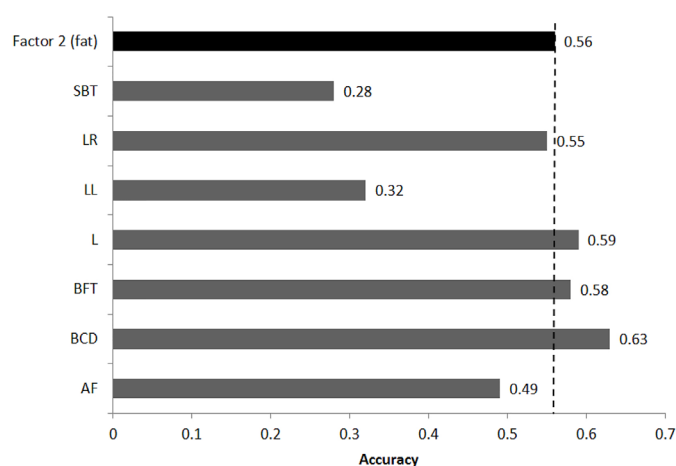


Figure 2. Accuracy observed on factor 2 (“fat”) and correlated variables (SBT - higher backfat thickness on the shoulder region; LR = midline backfat thickness immediately after the last rib; LL = midline backfat thickness between last and next to last but one lumbar vertebrae; L = midline lower backfat thickness above the last lumbar vertebrae; BFT = backfat; BCD - bacon depth; AF = abdominal fat).

Regarding the agreement between the top 10% of individuals selected through the factor “fat”, and separately through individual phenotypic variables (SBT, LR, LL, L, BFT, BCD and AF), it is possible to infer that satisfactory results were obtained, since the lowest coefficient (0.50) was for the thickness of the bacon and abdominal fat. The variables that agreed most were BFT (0.60), SBT (0.60) and LR (0.72). Table 3 shows all values obtained for the concordance index and their classification according to Landis and Koch (1977). These results again suggest that the selection of individuals by a common factor highly correlated with a variable group may be a valuable strategy to use when the objective is to select individuals for multiple variables simultaneously.

Finally, we investigated the markers effects over chromosomes positions, considering the original traits and the corresponding markers (Figure 3). In general, the effects of the markers for the original traits were similar to that found when considering the factor. Specifically, we can note that factor 2, “fat”, had higher effects on chromosomes 1, 4, 7, and 17, the same chromosomes highlighted for the trait SBT. Backfat thickness and the last rib traits were the most outstanding with regards the agreement in the selection, and showed higher values for chromosomes 1, 4, 7, and 17, exactly as observed for factor 2.

Table 3. Agreement coefficients related to each variable and the “fat” factor.

Variables	Cohen's Kappa	Classification
SBT	0.60	Good
LR	0.72	Very good
LL	0.60	Good
L	0.57	Good
BFT	0.60	Good
BCD	0.50	Good
AF	0.50	Good

SBT = higher backfat thickness on the shoulder region; LR = midline backfat thickness immediately after the last rib; LL = midline backfat thickness between last and next to last but one lumbar vertebrae; L = midline lower backfat thickness above the last lumbar vertebrae; BFT = backfat; BCD = bacon depth; AF = abdominal fat.

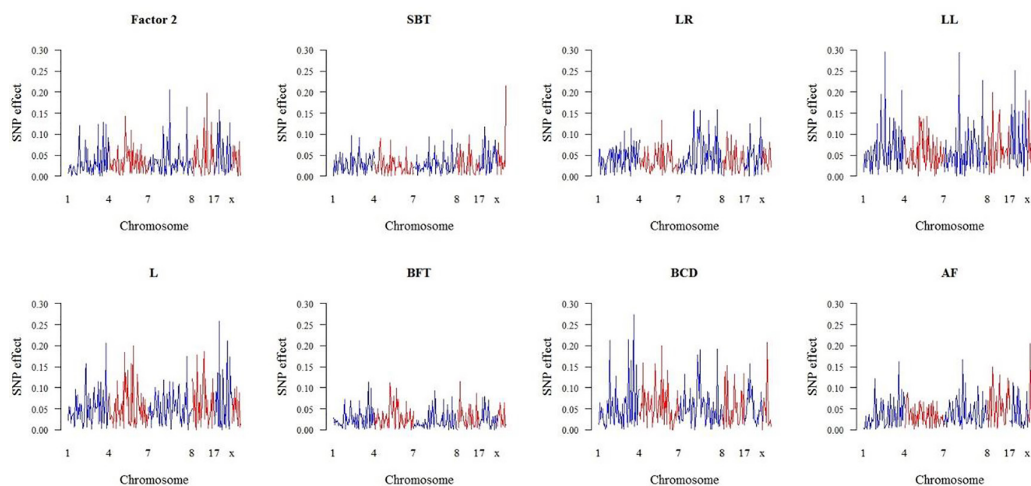


Figure 3. SNP effects for traits related to the second factor “fat” (SBT - higher backfat thickness on the shoulder region; LR - midline backfat thickness immediately after the last rib; LL = midline backfat thickness between last and next to last but one lumbar vertebrae; L = midline lower backfat thickness above the last lumbar vertebrae; BFT = backfat thickness; BCD = bacon depth; AF = abdominal fat).

DISCUSSION

Factor analysis was able to satisfactorily represent a set of highly correlated variables. Furthermore, the estimated accuracy values for the variables associated with the “fat” factor were, in general, very similar to the accuracy of the factor itself. This indicates that, when estimating the genetic value of the factor, we will also be reliably measuring the genetic value of the correlated variables.

The distribution of the variables of respective factors showed similar results, as found by Silva et al. (2011), because, for this study, the variables used for the MBCC (Método Brasileiro de Classificação de Carcaças) and MLC (Meat and Livestock Commission) classification systems were also part of the same factor. Two additional factors were found in the Silva et al. (2011) study: one grouped the LEA and LD variables, and other grouped features of backfat thickness; this is also consistent with the present work.

The superior performance of the BLASSO method was expected, since this result was observed in the study conducted by de los Campos et al. (2009).

In addition, the accuracy for the variables LR, LL, L, and BCD showed higher values than those obtained from GWS methodologies based on the dimensionality reduction presented by Azevedo et al. (2013b).

According to Landis and Koch (1977), values of the Kappa coefficient ≥ 0.5 may be considered good, since this measure ranges from 0-1. Therefore, the results for agreement were relevant; the lowest values were for the variables BCD and AF (0.5), and the coefficient value for LR can be considered very good (0.72). Additionally, a good agreement value was obtained for all other variables.

The positions of markers with higher effects for SBT are consistent with the study by Guo et al. (2008); they reported QTLs located in the same chromosomes (1 and 7) using data from two populations of Meishan x Large White pigs. Yin et al. (2012) also identified the presence of QTLs for this trait on chromosome 17.

Regarding the trait LR, results were in agreement with the studies of Fan et al. (2011), which identified a QTL on the same chromosome in a Large White population, and Hidalgo et al. (2013), which identified a QTL on chromosome 4 related to this same trait. Chromosome 1 also presented a higher importance for the ETUL, ETL, and ETO traits, similar to results found by Beeckmann et al. (2003), Rohrer and Keele (1998), and Azevedo et al. (2013b).

The use of latent variables from FA is a promising approach in GWS studies. The accuracy reported for the latent variable “fat” was similar to those obtained when the original traits were considered individually. The agreement between the top 10% of individuals selected by the factor (“Fat”), and those selected by individual traits, were also satisfactory. Moreover, the estimated markers effects for the traits were similar to those found for the considered factor.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the Foundation for Support of Universidade Federal de Viçosa (FUNARBE), Foundation for Research Support of Minas Gerais State (FAPEMIG - APQ #00825), the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), and the National Council for Scientific and Technological Development (CNPq), for financial support.

REFERENCES

- Azevedo CF, Resende MDV, Silva FF, Lopes PS, et al. (2013a). Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. *Pesquisa Agropecu. Bras.* 48: 619-626. <http://dx.doi.org/10.1590/S0100-204X2013000600007>
- Azevedo CF, Silva FF, Resende MDV, et al. (2013b). Quadrados mínimos parciais uni e multivariado aplicados na seleção genômica para características de carcaça em suínos. *Cienc. Rural* 43: 1642-1649. <http://dx.doi.org/10.1590/S0103-84782013000900017>
- Azevedo CF, Nascimento M, Silva FF, Resende MDV, et al. (2015). Comparison of dimensionality reduction methods to predict genomic breeding values for carcass traits in pigs. *Genet. Mol. Res.* 14: 12217-12227. <http://dx.doi.org/10.4238/2015.October.9.10>
- Badke YM, Bates RO, Ernst CW, Fix J, et al. (2014). Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. *C3.* 4: 623:631.

- Band GO, Guimarães SEF, Lopes PS, Peixoto JO, et al. (2005). Relationship between the Porcine Stress Syndrome gene and carcass and performance traits in F₂ pigs resulting from divergent crosses. *Genet. Mol. Biol.* 28: 92-96. <http://dx.doi.org/10.1590/S1415-47572005000100016>
- Beeckmann P, Schroffel J, Jr., Moser G, Bartenschlager H, et al. (2003). Linkage and QTL mapping for *Sus scrofa* chromosome 1. *J. Anim. Breed. Genet.* 120: 1-10. <http://dx.doi.org/10.1046/j.0931-2668.2003.00418.x>
- Bernaards CA and Jennrich RI (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educ. Psychol. Meas.* 65: 676-696. <http://dx.doi.org/10.1177/0013164404272507>
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20: 37-46. <http://dx.doi.org/10.1177/001316446002000104>
- Cruz CD, Salgado CC and Bhering LL (2013). *Genômica Aplicada*. Editora Suprema, Portugal.
- de Campos CF, Lopes MS, Silva FF, Veroneze R, et al. (2015). Genomic selection for boar taint compounds and carcass traits in a commercial pig population. *Livest. Sci.* 174: 10-17. <http://dx.doi.org/10.1016/j.livsci.2015.01.018>
- de los Campos G and Gianola D (2007). Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation. *Genet. Sel. Evol.* 39: 481-494. <http://dx.doi.org/10.1186/1297-9686-39-5-481>
- de los Campos G and Rodriguez PP (2014). BGLR: Bayesian Generalized Linear Regression. R package version 1.0.3. Available at [<http://CRAN.R-project.org/package=BGLR>].
- de los Campos G, Naya H, Gianola D, Crossa J, et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182: 375-385. <http://dx.doi.org/10.1534/genetics.109.101501>
- Fan B, Onteru SK, Du ZQ, Garrick DJ, et al. (2011). Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. *PLoS One* 6: e14726. <http://dx.doi.org/10.1371/journal.pone.0014726>
- Ferreira DF (2011). *Estatística Multivariada*. Editora UFLA, Lavras.
- Guo YM, Lee GJ, Archibald AL and Haley CS (2008). Quantitative trait loci for production traits in pigs: a combined analysis of two Meishan x Large White populations. *Anim. Genet.* 39: 486-495. <http://dx.doi.org/10.1111/j.1365-2052.2008.01756.x>
- Hidalgo AM, Lopes PS, Paixão DM, Silva FF, et al. (2013). Fine mapping and single nucleotide polymorphism effects estimation on pig chromosomes 1, 4, 7, 8, 17 and X. *Genet. Mol. Biol.* 36: 511-519. <http://dx.doi.org/10.1590/S1415-47572013000400009>
- Landis JR and Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics* 33: 159-174. <http://dx.doi.org/10.2307/2529310>
- Meuwissen THE, Hayes BJ and Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.
- Mingoti SA (2007). *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. Editora UFMG, Belo Horizonte.
- Paixão DM, Carneiro PLS, Paiva SR, Sousa KRS, et al. (2012). Mapeamento de QTL nos cromossomos 1, 2, 3, 12, 14, 15 e X em suínos: características de carcaça e qualidade de carne. *Arquivo Brasileiro Med. Veterinária Zootecnia* 64: 974-982. <http://dx.doi.org/10.1590/S0102-09352012000400026>
- R Development Core Team (2015). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at [<http://www.r-project.org>]. Accessed January 13, 2015.
- Resende MD, Silva FF, Viana JM, Peternelli LA, et al. (2010). *Computação da Seleção Genômica Ampla*. EMBRAPA Florestas, Brasília.
- Revelle W (2015). *psych: Procedures for Personality and Psychological Research*, Version 1.5.1. Northwestern University, Evanston, Illinois, USA. Available at [<http://CRAN.R-project.org/package=psych>]. Accessed December 5, 2015.
- Rohrer GA and Keele JW (1998). Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits. *J. Anim. Sci.* 76: 2247-2254.
- Silva FF, Rosa GJM, Guimarães SEF, Lopes PS, et al. (2011). Three-step Bayesian factor analysis applied to QTL detection in crosses between outbred pig populations. *Livest. Sci.* 142: 210-215. <http://dx.doi.org/10.1016/j.livsci.2011.07.012>
- Yin Q, Yang HW, Han XL, Fan B, et al. (2012). Isolation, mapping, SNP detection and association with backfat traits of the porcine CTNBL1 and DGAT2 genes. *Mol. Biol. Rep.* 39: 4485-4490. <http://dx.doi.org/10.1007/s11033-011-1238-8>
- Zhao JH (2007). gap: Genetic Analysis Package. *J. Stat. Softw.* 23: 1-18. <http://www.jstatsoft.org/v23/i08> Accessed December 05, 2015. <http://dx.doi.org/10.18637/jss.v023.i08>