PAB
Pesquisa Agropecuária Brasileira

Genetics/ Original Article

Genomic prediction with the additive-dominant model by dimensionality reduction methods

Abstract – The objective of this work was to evaluate the application of different dimensionality reduction methods in the additive-dominant model and to compare them with the genomic best linear unbiased prediction (G-BLUP) method. The dimensionality reduction methods evaluated were: principal components regression (PCR), partial least squares (PLS), and independent components regression (ICR). A simulated data set composed of 1,000 individuals and 2,000 single-nucleotide polymorphisms was used, being analyzed in four scenarios: two heritability levels × two genetic architectures. To help choose the number of components, the results were evaluated as to additive, dominant, and total genomic information. In general, PCR showed higher accuracy values than the other methods. However, none of the methodologies are able to recover true genomic heritabilities and all of them present biased estimates, under- or overestimating the genomic genetic values. For the simultaneous estimation of the additive and dominance marker effects, the best alternative is to choose the number of components that leads the dominance genomic value to a higher accuracy.

Index terms: dominance effect, G-BLUP, independent components regression, partial least squares, principal components regression.

Predição genômica com o modelo aditivo-dominante por métodos de redução de dimensionalidade

Resumo – O objetivo deste trabalho foi avaliar a aplicação de diferentes métodos de redução de dimensionalidade no modelo aditivo-dominante e compará-los ao método genômico da melhor predição linear não viesada (G-BLUP). Os métodos de redução avaliados foram: regressão via componentes principais (PCR), quadrados mínimos parciais (PLS) e regressão via componentes independentes (ICR). Utilizou-se um conjunto de dados simulados composto por 1.000 indivíduos e 2.000 polimorfismos de nucleotídeo único, analisados em quatro cenários: dois níveis de herdabilidade × duas heranças genéticas. Para auxiliar na escolha do número de componentes, os resultados foram avaliados quanto às informações genômicas aditiva, dominante e total. De modo geral, a PCR apresentou maiores valores de acurácia em comparação aos demais métodos. No entanto, nenhuma das metodologias consegue capturar as herdabilidades genômicas reais e todas apresentam estimativas viesadas, tendo subestimado ou superestimado os valores genéticos genômicos. Para a estimação simultânea dos efeitos de marcadores aditivos e devidos à dominância, a melhor alternativa é a escolha do número de componentes que conduz o valor genômico devido à dominância à maior acurácia.

Termos para indexação: efeito de dominância, G-BLUP, regressão via componentes independentes, quadrados mínimos parciais, regressão via componentes principais.

Jaquicele Aparecida da Costa[1] (iD),
Camila Ferreira Azevedo[1] (✉) (iD),
Moysés Nascimento[1] (iD),
Fabyano Fonseca e Silva[2] (iD),
Marcos Deon Vilela de Resende[3] (iD) and
Ana Carolina Campana Nascimento[1] (iD)

[1] Universidade Federal de Viçosa, Departamento de Estatística, Avenida Peter Henry Rolfs, s/nº, Campus Universitário, CEP 36570-977 Viçosa, MG, Brazil. E-mail: jaquicele.costa@ufv.br, camila.azevedo@ufv.br, moysesnascim@ufv.br, ana.campana@ufv.br

[2] Universidade Federal de Viçosa, Departamento de Zootecnia, Avenida Peter Henry Rolfs, s/nº, Campus Universitário, CEP 36570-977 Viçosa, MG, Brazil. E-mail: fabyanofonseca@ufv.br

[3] Embrapa Florestas, Estrada da Ribeira, Km 111, Guaraituba, Caixa Postal 319, CEP 83411-000 Colombo, PR, Brazil. E-mail: marcos.deon@gmail.com

✉ Corresponding author

Introduction

The genome-wide selection (GWS), conceived by Meuwissen (2001), assumes the existence of a linkage disequilibrium between markers and quantitative trait loci (QTLs), which makes it possible to estimate the genomic values of individuals from the estimation of marker effects on their phenotype, capturing the genotypic information that may influence their phenotypic variability (Goddard & Hayes, 2007). However, the implementation of GWS to assess individual genomic estimated breeding values (GEBVs) faces some statistical challenges, such as multicollinearity and highly correlated markers, which decreases the probability of a single nucleotide occurring independently of another in the same position of the genome (Gianola et al., 2003). Moreover, according to the same authors, due to the high cost of individual genotyping techniques, the number of individual observations is generally much lower than the number of markers.

Several methods – including Bayesian methods as the Bayesian least absolute shrinkage and selection operator (BLASSO) method, the mixed-model method, the genomic best linear unbiased predictor (G-BLUP) method, and dimensionality reduction methods such as principal components regression (PCR), partial least squares (PLS), and independent components regression (ICR) – have been applied to GWS and are recommended for genomic prediction (de los Campos et al., 2013; Azevedo et al., 2014, 2015b). These methodologies guarantee the absence of multicollinearity between their components and consider the marker effects as fixed, being a solution for the statistical problems related to the high dimensionality of GWS (Resende et al., 2012).

The dimensionality reduction methods also stand out for presenting a great applicability and relatively simple theory when compared with the other methods applied to GWS (Long et al., 2011; Azevedo et al., 2013). However, those methodologies have only considered additive genomic effects (Long et al., 2011; Azevedo et al., 2013, 2014, 2015a; Du et al., 2018).

The inclusion of dominance effects in statistical-genomic models is essential for the selection of crosses and clones, because it is an effective way of increasing genetic gain by capitalizing on heterosis (Toro & Varona, 2010; Wellmann & Bennewitz, 2012; Denis & Bouvet, 2013). Several studies have been conducted on the relevance of additive-dominant models in genomic prediction (Bennewitz & Meuwissen, 2010; Wellmann & Bennewitz, 2012; Vitezica et al., 2017; Varona et al., 2018), using G-BLUP (Su et al., 2012; Muñoz et al., 2014; Wang & Da, 2014), regression ridge (RR-BLUP), and Bayesian methods (Toro & Varona, 2010; Zeng et al., 2013; Azevedo et al., 2015b; Almeida Filho et al., 2016). However, the methodologies based on the dimensionality reduction methods have not yet been analyzed under the additive-dominant model.

The objective of this work was to evaluate the application of different dimensionality reduction methods in the additive-dominant model and to compare them with the G-BLUP method.

Materials and Methods

The used data set was simulated by Azevedo et al. (2015b) and described by Costa (2018). A population of 5,000 individuals from 100 families, generated from the random mating of two linkage equilibrium populations, was subjected to five generations of random mating without selection, mutation, or migration. The final population is an advanced generation composite in Hardy-Weinberg equilibrium and linkage disequilibrium. For marker density, Azevedo et al. (2015b) simulated a total of 2,000 equidistant single-nucleotide polymorphisms (SNPs) as biallelic markers, separated by 0.10 cM, on ten chromosomes. Marker alleles had a minor allele frequency greater than 5%. Of the 2,000 markers, 100 were randomly chosen to be QTLs. The linkage disequilibrium between the markers and QTLs was calculated according to Goddard et al. (2011) and was equal to 0.95. A total of 1,000 individuals from 20 full-sib families, with 50 individuals each one, were phenotyped for four traits and then genotyped. This simulation was made to mimic an elite breeding population of plant species with an effective size of approximately 40.

Traits were simulated with two genetic architectures: one following an infinitesimal model (polygenic inheritance) and the other with five major effect genes, responsible for 50% of the genetic variability (mixed inheritance). In the first genetic architecture, low-magnitude effects on phenotype were assumed for each of the 100 QTLs, whereas, in the second, large effects were assigned to 5 QTLs representing 50% of

the genetic variability and small effects were assigned to the remaining 95 QTLs.

The additive and dominant effects (SNPs and QTLs) were normally distributed with zero mean and genetic variance according to the desired heritability levels. The additive variances were 35, 35, 49, and 49, while the dominance variances were approximately 18, 23, 25, and 33 (Table 1). The simulations assumed independence between additive and dominance effects. For each trait, the average degree of dominance level was approximately 1 (complete dominance) in a population with intermediate allele frequencies. The obtained genotypic values were within the limits of Gmax = 100(m + a) and Gmin = 100(m - a), which are the maximum and minimum values, respectively, where m is the mean of the genotypic values and a is the homozygote genotypic value.

In order to obtain the phenotypic value, an environmental effect was added to the genotypic value. This effect was obtained from the normal distribution $N(0, \sigma^2_e)$, where the σ^2_e variance was defined according to two levels of narrow-sense heritability (additive heritabilities of 0.20 and 0.30, respectively) and two levels of broad-sense heritability (additive heritability plus dominance heritability of 0.30 and 0.50, respectively) (Table 1). Heritability levels were chosen to represent traits with low and moderate heritabilities, in which case GWS is expected to be superior to phenotypic selection (Azevedo et al., 2015b). Therefore, for the populations of full-sib families, four scenarios were studied: two broad-sense heritability levels of approximately 0.30 and 0.50 × two genetic architectures (Table 1). Each scenario was simulated ten times.

The general model for genomic selection was given by: $y = 1\mu + Wm_a + Sm_d + e$, where y is the vector of phenotypic observations with dimension I × 1, with I

being the number of individuals; μ is the general mean of the trait and 1 is a vector of the same dimension of y where all elements are equal to one; W is the incidence matrix of additive effects with dimension I × J, with J being the number of markers; m_a are the additive effects of the markers; S is the incidence matrix of the dominance effect with dimension I × J; m_d are the dominance marker effects; and e is the vector of random errors with a variance structure given by $e \sim N(0, I_e\sigma^2_e)$, with I_e being the identity matrix and σ^2_e, the residual variance. The W and S matrices were coded according to Vitezica et al. (2013) and their juxtaposition was defined by the X matrix as X = [W|S] (dimension I × 2J) and the marker effects, as $m = [m_a|m_d]'$ (2J × 1). For the estimation of these effects, the dimensionality reduction methods and G-BLUP were used as detailed below.

The PCR method defined the principal components (PC) as Z = XP, where Z is the matrix with the first n_{PCR} PC, which are orthogonal, and P is the matrix with the first n_{PCR} eigenvectors of the variance matrix of (Ferreira, 2018).

The PLS method decomposes the X matrix and the y vector simultaneously. For this, the Y and Xj's variables are centered on the mean, defining variables U_1 and V_{1j}, where: $u_1 = y - \bar{y}$ and $V_{1(j)} = x_j - \bar{x}_j$, for j = 1,...,2J, respectively. Then, the S_1 variable is written as $s_1 = V_1'u_1$ (dimension 2J × 1), where $V_1 = [v_{11} \quad v_{12} \quad ... \quad v_{12J}]$, and applies to the singular value decomposition in the s_1 vector as follows: $s_1 = L_1kq'_1$, where L_1 is a unit matrix (2J × 2J) with the first column vector equal

$$\frac{s_1}{\|s_1\|}$$

to (normalized s_1 vector), k_1 is a vector (2J × 1) with the first value equal to $\|s_1\|$ (norm of vector s_1), and q_1 is a scalar equal to 1. The first component, T_1, is defined by $t_1 = V_1L_1$ (Garthwaite, 1994). The information about

Table 1. Description of the scenarios and genetic architectures used for the analysis of the simulated data set, mimicking an elite breeding population of plant species[1].

Scenario	Genetic architecture[2]	σ^2_a	h^2_a	σ^2_d	h^2_d	σ^2_e
Scenario 1	Polygenic inheritance	35	0.20	18	0.10	124
Scenario 2	Polygenic inheritance	35	0.30	23	0.20	59
Scenario 3	Mixed inheritance	49	0.20	25	0.10	171
Scenario 4	Mixed inheritance	49	0.30	33	0.20	81

[1], simulated additive variances; , narrow-sense heritability; , simulated dominance variances; , proportion of dominance variance to phenotypic variance; and , simulated environmental variance. [2]Polygenic inheritance, traits controlled by genes of small effect; and Mixed inheritance, traits controlled by genes of small and greater effect.

variables X_j and Y that are not covered by component T_1 can be estimated by the residuals of the regression between the X_j and T_1 variables or, equivalently, by the regression between the V_{1j} and T_1 latent (unobservable) variables (Garthwaite, 1994).

Therefore, to define the second component, T_2, the $V_{2(j)}$ and U_2 variables were determined, respectively, by: $\hat{v}_{2(j)} = v_{1(j)} - t_1\hat{r}_1'$ and $\hat{u}_2 = u_1 - t_1\hat{p}_1'$, where $\hat{v}_{2(j)}$ and $\hat{u}_2$ are the residuals, and r_1 and p_1 are the coefficients obtained from the regression between u_1 and t_1 and $v_{1(j)}$ and t_1, in that order. The S_2 variable is defined as: $s_2 = V_2'u_2$, and the procedure applied in s_1 is repeated to construct the T_2 component. The $t_3,...,t_{nPLS}$ ($1 \leq n_{PLS} \leq \min(I,2J) - 1$) components were determined successively and analogously to the above, all being considered orthogonal (Garthwaite, 1994).

For ICR, proposed under the context of GWS in additive models by Azevedo et al. (2013), but also valid for additive-dominant models, the X data matrix, whose values are centered on the mean, is decomposed into $X = SA'$, where $S(I \times \min(2J,I) - 1)$ is the matrix of independent components and $A(2J \times \min(2J,I) - 1)$ is the mixture matrix. The A matrix is a function of two matrices, K and R; the first is obtained by the whitening process, making the covariance matrix of X equal to the identity matrix, so that matrix A is orthogonal (Yao et al., 2012), and the second is obtained by an iterative algorithm based on the principle of maximum entropy (Hyvärinen, 1998). After the convergence of this algorithm, the R matrix that guarantees the independence of the columns of was obtained. Therefore, the independent components are defined as $S = XKR$.

Subsequently, a multiple linear regression was performed between the Y variable and the Z, T, and S components obtained by PCR, PLS, and ICR, respectively. The following predictions were assumed:
$$\hat{y} = Z\hat{\alpha}, \quad \hat{y} = T\hat{\beta}, \quad \text{and} \quad \hat{y} = S\hat{\gamma}, \quad \text{where}$$
$$\hat{\alpha}_m = (m = 1,...,n_{PCR}), \quad \hat{\beta}_m (m = 1,...,n_{PLS}), \text{ and}$$
$\hat{\gamma}_m (m = 1,...,n_{ICR}$ being $1 \leq n_{PCR}, n_{PLS}, n_{ICR} \leq \min(2J,I)-1)$ are the parameter estimates associated with the components and calculated by the ordinary least squares method. These estimates — $\hat{\alpha}_m$, $\hat{\beta}_m$, and $\hat{\gamma}_m$ — obtained previously are not associated with the original variables (molecular markers), that is, they do not have a biological interpretation. The original estimates of the marker effects (fixed effects in the reduction methods) are given by: $\hat{m}_{pcr} = P\hat{\alpha}$, $\hat{m}_{PLS} = L(R'L)^{-1}\hat{\beta}$, and

$\hat{m}_{ICR} = KR\hat{\gamma}$ through PCR, PLS, and ICR, respectively (Azevedo et al., 2013). It should be noted that the expressions for estimating the marker effects depend on the choice of the number of components.

In this work, to determine the number of components to be used for PCR, PLS, and ICR, the exhaustive criterion adopted by Azevedo et al. (2013, 2014, 2015a) in the context of additive models was used. In additive-dominant models, this criterion consists of choosing the number of components that leads to a higher accuracy in the prediction of additive (a), dominance (d), or total (g = a + d) genomic values. However, ICR, even in additive models, requires a high computational demand to perform the analyzes by this criterion (Costa, 2018). Therefore, an alternative criterion for ICR, called optimized criterion, was adopted in the present study. It was proposed by Costa (2018) in the context of additive models and consists in obtaining a number of independent components equal to the number of principal components.

G-BLUP was based on an individual-level model given by: $y = 1\mu + Za + Zd + e$, where a is the vector of the additive genomic effects of the individuals (I × 1), with an incidence matrix Z (I × 1), for which the variance structure is given by $a \sim N(0, G_a\sigma_a^2)$, with σ_a^2 being the additive variance and $G_a(I \times 1)$, the additive genomic relationship matrix; d is the vector of the dominance genomic effects of the individuals, with an incidence matrix Z (I × I), for which the variance structure is given by $d \sim N(0, G_d\sigma_d^2)$, with σ_d^2 being the variance due to dominance and G_d (I × I), the dominance genomic relationship matrix; and e is the vector of random residual effects, with $e \sim N(0, I_e\sigma_e^2)$, where σ_e^2 is the residual variance. Additive and dominance genomic values can be estimated via mixed-model equations, and the variance components $(\sigma_e^2, \sigma_a^2 \text{ and } \sigma_d^2)$ are estimated by restricted maximum likelihood. The individual-level model is equivalent to the marker-level model, as: $a = Wm_a$ and $d = Sm_d$, where m_a and m_d are the additive and dominance marker effects, respectively.

The dimensionality reduction methods and G-BLUP were compared through a cross-validation process in which the first nine replicates were assumed as estimation populations (different populations) used to estimate the marker effects (additive or dominance) and the tenth replicate was assumed as a population of validation. All efficiency measures (heritability,

accuracy, regression coefficients between phenotypes and simulated genetic values, and relative efficiency in relation to G-BLUP) were obtained for each replicate in each scenario, considering all three genetic information (additive, dominance, or total), and the general results were reported as average values. The expressions for additive heritability (narrow-sense heritability), dominance heritability (proportion of dominance variance to phenotypic variance), total heritability (broad-sense heritability), accuracy (additive and dominance), regression coefficient between phenotypes and simulated genetic values (additive and dominance), and relative efficiency are shown in Table 2 and were also used by Azevedo et al. (2015b).

Table 2. Expressions used for calculating the following efficiency measures: additive and dominance accuracies $\left(r_{\hat{a}a} \text{ and } r_{\hat{d}d}\right)$, regression coefficients $\left(b_{y\hat{a}} \text{ and } b_{y\hat{d}}\right)$, heritabilities $\left(h^2_{aM} \text{ and } h^2_{dM}\right)$, and relative efficiency (ER_a and ER_d).

Efficiency measure	Genetic information	Expression[1]
Accuracy	Additive	$r_{\hat{a}a} = \mathrm{Cor}\left(a, \hat{a}\right)$
	Dominance	$r_{\hat{d}d} = \mathrm{Cor}\left(d, \hat{d}\right)$
Regression coefficients	Additive	$b_{y\hat{a}} = \dfrac{\mathrm{Cov}\left(y, \hat{a}\right)}{\mathrm{Var}\left(\hat{a}\right)}$
	Dominance	$b_{y\hat{d}} = \dfrac{\mathrm{Cov}\left(y, \hat{d}\right)}{\mathrm{Var}\left(\hat{d}\right)}$
Heritability	Additive	$h^2_{aM} = \dfrac{\sigma^2_{a_M}}{\sigma^2_{a_M} + \sigma^2_{d_M} + \sigma^2_e}$
	Dominance	$h^2_{dM} = \dfrac{\sigma^2_{d_M}}{\sigma^2_{a_M} + \sigma^2_{d_M} + \sigma^2_e}$
	Total	$h^2_{gM} = h^2_{am} + h^2_{dM}$
Relative efficiency	Additive	$ER_a = \dfrac{r_{\hat{a}a\ RM}}{r_{\hat{a}a\ G\text{-}BLUP}}$
	Dominance	$ER_d = \dfrac{r_{\hat{d}d\ RM}}{r_{\hat{d}d\ G\text{-}BLUP}}$

[1]a, vector of the true additive genetic effect of an individual; â, vector of the estimate of the additive genetic effect of an individual; d, vector of the true dominance genetic effect of an individual; $\bar{d}$, vector of the estimate of the dominance genetic effect of an individual; y , vector of phenotypes; $\sigma^2_{a_M} = \Sigma^j_{j=1} 2 p_j q_j m^2_{aj}$, additive genomic variance, and $\sigma^2_{d_M} = \sum^j_{j=1}\left(2 p_j q_j\right)^2 m^2_{dj}$, dominance genomic variance, where J is the number of markers, and p_j and q_j are the allele frequencies of the jth marker; σ^2_e, error variance; $r_{\hat{a}aRM}$ and $r_{\hat{a}aG-BLUP}$ respectively, additive accuracies of the dimensionality reduction methods and of G-BLUP; and $r_{\hat{d}dRM}$ and $r_{\hat{d}dG-BLUP}$, respectively, dominance accuracies of the dimensionality reduction methods and of G-BLUP.

All computational routines were implemented with the R software (R Core Team, 2019), using the packages: sommer, for G-BLUP; pls, for PLS and PCR; and caret, for ICR. The computer configuration was Intel (R) Core (TM) i7-6500, 2.50 GHz, 16 Gb RAM.

Results and Discussion

When the dimensionality reduction methods were evaluated only considering the number of components based on dominance genomic values, ICR, under both the optimized and exhaustive criteria, presented, on average, a lower accuracy, followed by PLS and lastly by PCR (Tables 3, 4, 5, and 6). Considering all scenarios, PLS was, on average, 1.25% more efficient than G-BLUP for additive values and 6.5% less efficient for dominance values. PCR was, on average, 4.25 and 14.75% more efficient than G-BLUP for additive and dominance values, respectively. Specifically, considering additive effects, PLS was more efficient than G-BLUP in scenarios 1 and 2 of polygenic inheritance, whereas PCR was more efficient than G-BLUP in scenarios 1, 2, and 4. When considering dominance effects, PLS was more efficient than G-BLUP in scenarios 1 and 3, both with a heritability of 0.30, and PCR was more efficient than G-BLUP in all scenarios. Furthermore, PCR presented better results than ICR when explanatory variables (molecular markers) only showed a linear dependence (Azevedo et al., 2014). Therefore, the obtained results suggest that there is a greater proportion of linear dependence in the linkage disequilibrium structure between loci (Smith, 2020).

In addition to accuracy, other measures that can be evaluated are heritability and prediction bias (de los Campos et al., 2013; Daetwyler et al., 2013; Gianola, 2013). The PLS and PCR methods underestimated the additive and dominance heritabilities, while G-BLUP underestimated additive heritability and overestimated dominance heritability in most scenarios. Therefore, these methods were not able to capture the additive and dominance heritabilities that were simulated. Regarding bias, it was defined as one minus the regression coefficient between the estimated genetic value (additive and dominance) and the phenotype, that is, coefficients equal to 1 indicated nonbiased genomic values. For regression coefficients below one (< 1), it was understood that the

Table 3. Additive, dominance, and total simulated heritabilities (h_s^2), number of components (Nc), additive and dominance heritabilities (h_{aM}^2 and h_{dM}^2), additive and dominance accuracies ($r_{a\hat{a}}$ and $r_{d\hat{d}}$), regression coefficients ($\hat{b}_{y\hat{a}}$ and $\hat{b}_{y\hat{d}}$), and relative additive and dominance efficiencies (EF$_a$ and EF$_d$) obtained for the dimensionality reduction methods and the genomic best linear unbiased prediction (G-BLUP) method in scenario 1 of polygenic inheritance (with traits controlled by small gene effects), considering the additive, dominance, and total genomic values as targets.

Genomic value	h_s^2	Method[1]	Criterion[2]	Nc	h_{aM}^2	$r_{a\hat{a}}$	$\hat{b}_{y\hat{a}}$	h_{dM}^2	$r_{d\hat{d}}$	$\hat{b}_{y\hat{d}}$	EF$_a$	EF$_d$
Additive	0.20	ICR	Exhaustive	20±0.00	0.05±0.01	0.58±0.04	1.30± 0.12	0.10±0.01	0.22±0.05	1.80±0.17	1.07	0.71
		ICR	Optimized	56±22	0.07±0.02	0.57±0.04	1.20± 0.09	0.03±0.01	0.19±0.08	1.70±0.13	1.06	0.61
		PCR	Exhaustive	56±22	0.12±0.02	0.60±0.03	1.00± 0.10	0.02±0.01	0.32±0.11	2.50±0.70	1.11	1.03
		PLS	Exhaustive	2.80±1.50	0.22±0.07	0.59±0.05	0.80± 0.09	0.06±0.06	0.31±0.08	2.10±1.10	1.09	1.00
Dominance	0.10	ICR	Exhaustive	211±0.00	0.10±0.01	0.53±0.04	0.89± 0.07	0.04±0.02	0.24±0.05	1.30±0.09	0.98	0.77
		ICR	Optimized	120±120	0.29± 0.04	0.54±0.03	1.10± 0.23	0.04± 0.01	0.23±0.04	1.50±0.33	1.00	0.74
		PCR	Exhaustive	120±120	0.16±0.08	0.58±0.03	0.91± 0.20	0.04±0.04	0.37±0.04	1.90±0.76	1.07	1.19
		PLS	Exhaustive	1.10±0.33	0.13±0.02	0.58±0.04	0.88± 0.06	0.01±0.01	0.33±0.07	4.60±0.83	1.07	1.06
Total	0.30	ICR	Exhaustive	60±0.00	0.07±0.01	0.56±0.04	1.20± 0.09	0.03±0.02	0.19±0.05	1.60±0.13	1.04	0.61
		ICR	Optimized	78±32	0.29± 0.02	0.54±0.03	1.10± 0.23	0.03± 0.01	0.23±0.04	1.50±0.33	1.00	0.74
		PCR	Exhaustive	78±32	0.14±0.03	0.60±0.04	0.97± 0.04	0.03±0.01	0.37±0.06	2.10±0.40	1.11	1.19
		PLS	Exhaustive	1.60±0.53	0.12±0.02	0.59±0.04	0.87± 0.07	0.02±0.01	0.26±0.07	3.50±1.20	1.10	0.84
		G-BLUP	-	-	0.15±0.05	0.54±0.06	0.54± 0.20	0.13±0.06	0.31±0.08	2.90±0.50	1.00	1.00

[1]ICR, independent components regression; PCR, principal components regression; and PLS, partial least squares. [2]The exhaustive criterion consists in obtaining a number of components equal to the number of components that leads ICR, PCR, and PLS to a higher accuracy. The optimized criterion consists in obtaining a number of independent components equal to the number of main components that leads PCR to a higher accuracy.

Table 4. Additive, dominance, and total simulated heritabilities (h_s^2), number of components (Nc), additive and dominance heritabilities (h_{aM}^2 and h_{dM}^2), additive and dominance accuracies ($r_{a\hat{a}}$ and $r_{d\hat{d}}$), regression coefficients ($\hat{b}_{y\hat{a}}$ and $\hat{b}_{y\hat{d}}$), and relative additive and dominance efficiencies (EF$_a$ and Ef$_d$) obtained for the dimensionality reduction methods and the genomic best linear unbiased prediction (G-BLUP) method in scenario 2 of polygenic inheritance (with traits controlled by small gene effects), considering the additive, dominance, and total genomic values as targets.

Genomic value	h_s^2	Method[1]	Criterion[2]	Nc	h_{aM}^2	$r_{a\hat{a}}$	$\hat{b}_{y\hat{a}}$	h_{dM}^2	$r_{d\hat{d}}$	$\hat{b}_{y\hat{d}}$	EF$_a$	EF$_d$
Additive	0.30	ICR	Exhaustive	158±0.00	0.37±0.01	0.62±0.04	1.20±0.06	0.06±0.05	0.25±0.04	1.70±0.09	1.07	0.68
		ICR	Optimized	270±110	0.38± 0.02	0.62±0.03	1.10± 0.12	0.07±0.01	0.27±0.05	1.50±0.17	1.07	0.73
		PCR	Exhaustive	270±110	0.32± 0.02	0.68±0.03	0.96± 0.09	0.11±0.04	0.40±0.06	1.40±0.38	1.17	1.08
		PLS	Exhaustive	4.80±1.60	0.38±0.04	0.66±0.03	0.93± 0.04	0.14±0.04	0.23±0.05	1.30±0.22	1.14	0.62
Dominance	0.20	ICR	Exhaustive	444±0.00	0.35±0.01	0.59±0.03	0.88±0.04	0.09±0.04	0.31±0.03	1.30±0.05	1.02	0.84
		ICR	Optimized	210±120	0.36± 0.02	0.60±0.03	1.10± 0.12	0.07±0.01	0.26±0.03	1.60±0.17	1.03	0.70
		PCR	Exhaustive	210±120	0.29±0.07	0.66±0.04	0.98± 0.09	0.09±0.05	0.42±0.04	1.60±0.50	1.14	1.14
		PLS	Exhaustive	1.00±0.00	0.19±0.02	0.60±0.03	1.00±0.04	0.01±0.01	0.33±0.05	5.50±0.34	1.03	0.89
Total	0.50	ICR	Exhaustive	315±0.00	0.37±0.01	0.61±0.03	1.00±0.06	0.08±0.01	0.29±0.03	1.50±0.08	1.05	0.78
		ICR	Optimized	280±90	0.36±0.02	0.60±0.03	1.10±0.12	0.07±0.01	0.26±0.03	1.60±0.17	1.03	0.70
		PCR	Exhaustive	280±90	0.33±0.05	0.68±0.03	0.96± 0.08	0.12±0.04	0.41±0.04	1.30±0.28	1.17	1.11
		PLS	Exhaustive	2.70±1.10	0.29±0.09	0.65±0.03	0.97± 0.05	0.06±0.05	0.33±0.08	2.70±1.40	1.12	0.89
		G-BLUP	-	-	0.27±0.03	0.58±0.03	0.63±0.12	0.20±0.03	0.37±0.06	3.20±0.54	1.00	1.00

[1]ICR, independent components regression; PCR, principal components regression; and PLS, partial least squares. [2]The exhaustive criterion consists in obtaining a number of components equal to the number of components that leads ICR, PCR, and PLS to a higher accuracy. The optimized criterion consists in obtaining a number of independent components equal to the number of main components that leads PCR to a higher accuracy.

Table 5. Additive, dominance, and total simulated heritabilities (h_s^2), number of components (Nc), additive and dominance heritabilities (h_{aM}^2 and h_{dM}^2), additive and dominance accuracies ($r_{a\hat{a}}$ and $r_{d\hat{d}}$), regression coefficients ($\hat{b}_{y\hat{a}}$ and $\hat{b}_{y\hat{d}}$), and relative additive and dominance efficiencies (EF$_a$ and EF$_d$) obtained for the dimensionality reduction methods and the genomic best linear unbiased prediction (G-BLUP) method in scenario 3 of mixed inheritance (with traits controlled by major and small gene effects), considering the additive, dominance, and total genomic values as targets.

Genomic value	h_s^2	Method[1]	Criterion[2]	Nc	h_{aM}^2	$r_{a\hat{a}}$	$\hat{b}_{y\hat{a}}$	h_{dM}^2	$r_{d\hat{d}}$	$\hat{b}_{y\hat{d}}$	EF$_a$	EF$_d$
Additive	0.20	ICR	Exhaustive	17±0.00	0.26±0.01	0.55±0.05	1.10±0.11	0.02±0.01	0.10±0.09	1.60±0.15	0.92	0.29
			Optimized	85±67	0.28±0.02	0.53±0.04	1.00±0.17	0.02±0.02	0.16±0.04	1.50±0.25	0.88	0.46
		PCR	Exhaustive	85±67	0.12±0.04	0.59±0.04	0.94±0.15	0.03±0.03	0.31±0.10	2.00±1.50	0.98	0.89
		PLS	Exhaustive	3.30±1.40	0.26±0.08	0.59±0.06	0.70±0.14	0.09±0.05	0.36±0.08	1.30±1.30	0.98	1.03
Dominance	0.10	ICR	Exhaustive	487±0.00	0.16±0.01	0.40±0.04	0.50±0.05	0.06±0.01	0.22±0.04	0.71±0.07	0.67	0.63
			Optimized	170±110	0.23±0.02	0.48±0.04	0.89±0.18	0.04±0.01	0.21±0.04	1.30±0.26	0.80	0.60
		PCR	Exhaustive	170±110	0.18±0.07	0.56±0.05	0.81±0.16	0.06±0.05	0.40±0.06	1.30±0.63	0.93	1.14
		PLS	Exhaustive	1.00±0.00	0.12±0.01	0.59±0.04	0.87±0.06	0.01±0.01	0.36±0.04	3.40±0.86	0.98	1.03
Total	0.30	ICR	Exhaustive	18±0.00	0.26±0.01	0.51±0.05	1.10±0.11	0.02±0.01	0.10±0.09	1.60±0.15	0.85	0.29
			Optimized	120±95	0.23±0.02	0.48±0.04	0.89±0.18	0.04±0.01	0.21±0.04	1.60±0.26	0.80	0.60
		PCR	Exhaustive	120±95	0.14±0.06	0.58±0.05	0.90±0.17	0.04±0.04	0.36±0.11	1.60±1.10	0.97	1.03
		PLS	Exhaustive	1.60±1.30	0.15±0.07	0.58±0.05	0.84±0.11	0.03±0.05	0.29±0.08	2.90±1.30	0.97	0.83
		G-BLUP	-	-	0.14±0.03	0.60±0.04	0.71±0.15	0.13±0.04	0.35±0.07	2.70±0.87	1.00	1.00

[1]ICR, independent components regression; PCR, principal components regression; and PLS, partial least squares. [2]The exhaustive criterion consists in obtaining a number of components equal to the number of components that leads ICR, PCR, and PLS to a higher accuracy. The optimized criterion consists in obtaining a number of independent components equal to the number of main components that leads PCR to a higher accuracy.

Table 6. Additive, dominance, and total simulated heritabilities (h_s^2), number of components (Nc), additive and dominance heritabilities (h_{aM}^2 and h_{dM}^2), additive and dominance accuracies ($r_{a\hat{a}}$ and $r_{d\hat{d}}$), regression coefficients ($\hat{b}_{y\hat{a}}$ and $\hat{b}_{y\hat{d}}$), and relative additive and dominance efficiencies (EF$_a$ and EF$_d$) obtained for the dimensionality reduction methods and the genomic best linear unbiased prediction (G-BLUP) method in scenario 4 of mixed inheritance (with traits controlled by major and small gene effects), considering the additive, dominance, and total genomic values as targets.

Genomic value	h_s^2 ,	Method[1]	Criterion[2]	Nc	h_{aM}^2	$r_{a\hat{a}}$	$\hat{b}_{y\hat{a}}$	h_{dM}^2	r_{dd}	$\hat{b}_{y\hat{d}}$	EF$_a$	EF$_d$
Additive	0,30	ICR	Exhaustive	203±0.00	0.30±0.01	0.55±0.05	0.96±0.04	0.06±0.01	0.25±0.04	1.40±0.05	0.90	0.61
			Optimized	290±140	0.29±0.02	0.54±0.05	0.89±0.11	0.06±0.01	0.25±0.04	1.30±0.15	0.89	0.61
		PCR	Exhaustive	290±140	0.30±0.08	0.66±0.03	0.82±0.12	0.13±0.07	0.43±0.05	1.10±0.35	1.08	1.05
		PLS	Exhaustive	4.70±1.00	0.36±0.04	0.64±0.04	0.78±0.06	0.16±0.03	0.38±0.04	1.00±0.20	1.05	0.93
Dominance	0.20	ICR	Exhaustive	350±0.00	0.28±0.01	0.53±0.04	0.83±0.05	0.08±0.01	0.28±0.04	1.20±0.07	0.87	0.68
			Optimized	270±120	0.28±0.02	0.53±0.04	0.89±0.09	0.07±0.09	0.27±0.04	1.30±0.13	0.87	0.66
		PCR	Exhaustive	270±120	0.29±0.06	0.63±0.03	0.81±0.10	0.12±0.07	0.46±0.03	1.10±0.32	1.03	1.12
		PLS	Exhaustive	1.10±0.33	0.18±0.11	0.59±0.03	0.88±0.05	0.01±0.01	0.41±0.06	3.90±0.56	0.97	0.76
Total	0.50	ICR	Exhaustive	288±0.00	0.29±0.01	0.54±0.04	0.89±0.04	0.08±0.03	0.26±0.04	1.30±0.06	0.89	0.63
			Optimized	280±93	0.23±0.02	0.53±0.05	0.89±0.09	0.07±0.01	0.27±0.04	1.30±0.13	0.87	0.66
		PCR	Exhaustive	280±93	0.30±0.06	0.65±0.03	0.82±0.10	0.13±0.05	0.45±0.03	1.10±0.27	1.07	1.10
		PLS	Exhaustive	2.90±1.90	0.28±0.11	0.62±0.04	0.85±0.07	0.09±0.08	0.39±0.10	2.40±1.50	1.02	0.95
		G-BLUP	-	-	0.25±0.06	0.61±0.02	0.74±0.20	0.18±0.04	0.41±0.04	3.20±0.68	1.00	1.00

[1]ICR, independent components regression; PCR, principal components regression; and PLS, partial least squares. [2]The exhaustive criterion consists in obtaining a number of components equal to the number of components that leads ICR, PCR, and PLS to a higher accuracy. The optimized criterion consists in obtaining a number of independent components equal to the number of main components that leads PCR to a higher accuracy.

predicted values had been overestimated, whereas, for those above one (> 1), it was concluded that the predicted values had been underestimated (Azevedo et al., 2015b). The PLS, PCR, and G-BLUP methods overestimated the additive values and underestimated those due to dominance. In additive models, Azevedo et al. (2014) and Azevedo et al. (2015a) found that PCR and PLS also led to overestimated additive genomic values in the genomic prediction of carcass traits in pigs. Although the bias is essential to determine the genetic merit of individuals, it does not influence their ranking and, subsequently, the selection process (Resende et al., 2012).

The main difference between the PCR and PLS methods is that PCR uses only the explanatory variables (molecular markers) for the construction of the components and PLS uses the explanatory variables and the response of the variables (phenotypes) (Garthwaite, 1994). Since the exhaustive criterion aims to choose the number of components associated with a higher accuracy, it is expected that PLS require fewer components than PCR (Du et al., 2018). However, even with a higher number of components, PCR provided a reduction of 86.50% in the original data in scenario 2, when considering the dominance genomic value as a target. It should be noted that cross-validation was carried out to protect overfitting and overparameterization (James et al., 2013).

On average, higher additive and dominance accuracies were observed for ICR (with the exhaustive and optimized criteria), PCR, and PLS (Table 7), when considering the additive and total genomic values as targets. According to Huang & Mackay (2016), the additive variance explains a greater proportion of genetic variance, even under dominant gene action, as shown by the parameterization of the marker incidence matrix for additive and dominance effects used in the present study. This is because the additive variance is maximized initially and the dominance variance is the residue of the total genetic variation. In this way, Huang & Mackay (2016) showed that prioritizing nonadditive gene actions can capture the majority of genetic variation. Moreover, Falconer & Mackay (1996) reported that an accurate estimation of dominance effects can improve genetic gain in breeding programs. Therefore, if the interest is specifically one of the additive or dominance effects, the analysis must be based on the target genomic

information; however, the obtained results suggest that, based on the dominance genomic values, it is possible to better estimate both information simultaneously. This shows that the effective selection of parents, crosses, and clones can occur based only on the efficient estimation of the additive and dominance effects (Azevedo et al., 2015b).

In additive models, ICR presented better results than the other dimensionality reduction methods (Azevedo et al., 2013, 2014, 2015a). Although, in the present study, the additive-dominant models had the worst performance, further researches with ICR are still necessary, especially in cases of nonlinear dependence between markers. However, ICR requires a high computational effort, which is not feasible in the genomic prediction of breeding programs, requiring a reduction in computational time without a relevant loss in the efficiency of the method. In the present work, for ICR, the computational time for each replicate in each scenario lasted 221 hours using the exhaustive criterion, but was drastically reduced to about 0.18 hour with the optimized criterion. It is worth mentioning that, in most analyzes, this reduction in time did not result in a relevant loss of accuracy, as also observed for additive models by Costa (2018). The G-BLUP method presented a shorter computational analysis time of about 0.09 hour, followed by PCR, with 0.10 hour, and PLS, with 0.12 hour.

The average results of the ICR method showed that, considering the dominance genomic value as a target, the optimized criterion led to an additive accuracy equal to or higher than that of the exhaustive criterion in scenario 3 of mixed inheritance and heritability of 0.30, with additive accuracies of 0.48 and 0.40, respectively. The dominance accuracies obtained by the optimized and exhaustive criteria were equal, except in the scenario with polygenic inheritance and heritability of 0.50; in this case, the additive accuracies were 0.26 and 0.31, respectively.

Regarding biases, in general, both the exhaustive and the optimized criteria under the additive-dominant model led to biased additive and dominance genomic values, that is, showing a regression coefficient far from 1 (Resende et al., 2012). In additive models, Azevedo et al. (2014, 2015a) concluded that the ICR with the exhaustive criterion also resulted in biased additive genomic values. Methods or criteria that lead to the same accuracy provide the same classification of individuals, even if one results in nonbiased genomic

Table 7. Mean results of the ratio between the additive and dominance accuracies of the dimensionality reduction methods in relation to the target genomic value (GV) in the evaluated scenario, according to the used criteria.

Scenario[1]	Method[2]	Criterion[3]	GV	Ratio	
				Additive	Dominance
Scenario 1	ICR	Exhaustive	Additive	1.00	0.92
			Dominance	0.91	1.00
			Total	0.97	0.79
	ICR	Optimized	Additive	0.98	0.79
			Dominance	0.93	0.96
			Total	0.93	0.96
	PCR	Exhaustive	Additive	1.00	0.86
			Dominance	0.97	1.00
			Total	1.00	1.00
	PLS	Exhaustive	Additive	1.00	0.94
			Dominance	0.98	1.00
			Total	1.00	0.79
Scenario 2	ICR	Exhaustive	Additive	1.00	0.81
			Dominance	0.95	1.00
			Total	0.98	0.94
	ICR	Optimized	Additive	1.00	0.87
			Dominance	0.97	0.84
			Total	0.98	0.84
	PCR	Exhaustive	Additive	1.00	0.95
			Dominance	0.97	1.00
			Total	1.00	0.98
	PLS	Exhaustive	Additive	1.00	0.70
			Dominance	0.91	1.00
			Total	0.98	1.00
Scenario 3	ICR	Exhaustive	Additive	1.00	0.45
			Dominance	0.73	1.00
			Total	0.93	0.45
	ICR	Optimized	Additive	0.96	0.73
			Dominance	0.87	0.95
			Total	0.87	0.95
	PCR	Exhaustive	Additive	1.00	0.78
			Dominance	0.95	1.00
			Total	0.98	0.90
	PLS	Exhaustive	Additive	1.00	1.00
			Dominance	1.00	1.00
			Total	0.98	0.81
Scenario 4	ICR	Exhaustive	Additive	1.00	0.89
			Dominance	0.96	1.00
			Total	0.98	0.93
	ICR	Optimized	Additive	0.98	0.89
			Dominance	0.96	0.96
			Total	0.96	0.96
	PCR	Exhaustive	Additive	1.00	0.93
			Dominance	0.95	1.00
			Total	0.98	0.98
	PLS	Exhaustive	Additive	1.00	0.93
			Dominance	0.92	1.00
			Total	0.97	0.95

[1]Scenario 1, traits controlled by 375 small gene effects, with polygenic inheritance; scenario 2, traits controlled by small gene effects, with polygenic inheritance; scenario 3, traits controlled by 398 major and small gene effects, with mixed inheritance; and scenario 4, traits controlled by 413 major and small gene effects, with mixed inheritance. [2]ICR, independent components regression; PCR, principal components regression; and PLS, partial least squares. [3]An exhaustive criterion consists in obtaining a number of components equal to the number of components that leads ICR, PCR, and PLS to a higher accuracy. The optimized criterion consists in obtaining a number of independent components equal to the number of main components that leads PCR to a higher accuracy.

values and the other in biased genomic values (Resende et al., 2012). However, in the presence of bias, the individual's genetic merit and genetic gain are over- or underestimated (Vitezica et al., 2011).

Considering the dominance genomic value as a target, the additive heritabilities estimated using the optimized criterion were either similar in scenarios 2 and 4, with a heritability of 0.50, or greater in scenarios 1 and 3, with a heritability of 0.30, when compared with the exhaustive criterion. Using both criteria, the dominance heritability estimates were similar in all scenarios. However, ICR was not able to capture the additive and dominance heritabilities that were simulated. Since the reduction methods assume that the markers in the model are fixed effects, an alternative for calculating heritability is precisely by the estimates of marker effects. However, if the genomic values are biased, it is inferred that the effects of the markers will also be biased. It is likely that this bias exists in the additive-dominance models because the estimates of dominance variance are less accurate and require much more information to be estimated (Toro & Varona, 2010). These results suggest that the ICR method was inefficient for estimating additive and dominance heritabilities in the evaluated scenarios. Regarding the two genetic architectures and the two levels of heritability simulated, it was not possible to find any pattern in the obtained results.

Conclusions

1. Under the additive-dominant model, the relative efficiency of the principal components regression is higher in terms of accuracy, compared with the genomic best linear unbiased prediction (G-BLUP) and the other dimensionality reduction methods evaluated.

2. None of the assessed methods (G-BLUP, principal components regression, independent components regression, and partial least squares) capture the simulated heritabilities and all of them show biased additive and dominance genomic values.

Acknowledgments

References

ALMEIDA FILHO, J.E. de; GUIMARÃES, J.F.R.; SILVA, F.F. e; RESENDE, M.D.V. de; MUÑOZ, P.; KIRST, M.; RESENDE JR, M.F.R. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. **Heredity**, v.117, p.33-41, 2016. DOI: https://doi.org/10.1038/hdy.2016.23.

AZEVEDO, C.F.; NASCIMENTO, M.; SILVA, F.F.; RESENDE, M.D.V.; LOPES, P.S.; GUIMARÃES, S.E.F.; GLÓRIA, L.S. Comparison of dimensionality reduction methods to predict genomic breeding values for carcass traits in pigs. **Genetics and Molecular Research**, v.14, p.12217-12227, 2015a. DOI: https://doi.org/10.4238/2015.October.9.10.

AZEVEDO, C.F.; RESENDE, M.D.V. de; SILVA, F.F. e; LOPES, P.S.; GUIMARÃES, S.E.F. Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. **Pesquisa Agropecuária Brasileira**, v.48, p.619-626, 2013. DOI: https://doi.org/10.1590/S0100-204X2013000600007.

AZEVEDO, C.F.; RESENDE, M.D.V. de; SILVA, F.F. e; VIANA, J.M.S.; VALENTE, M.S.F.; RESENDE JR, M.F.R.; MUÑOZ, P. Ridge, Lasso and Bayesian additive-dominance genomic models. **BMC Genetics**, v.16, art.105, 2015b. DOI: https://doi.org/10.1186/s12863-015-0264-2.

AZEVEDO, C.F.; SILVA, F.F.; RESENDE, M.D.V. de; LOPES, M.S.; DUIJVESTEIJN, N.; GUIMARÃES, S.E.F.; LOPES, P.S.; KELLY, M.J.; VIANA, J.M.S.; KNOL, E.F. Supervised independent component analysis as an alternative method for genomic selection in pigs. **Journal of Animal Breeding and Genetics**, v.131, p.452-461, 2014. DOI: https://doi.org/10.1111/jbg.12104.

BENNEWITZ, J.; MEUWISSEN, T.H.E. The distribution of QTL additive and dominance effects in porcine F2 crosses. **Journal of Animal Breeding and Genetics**, v.127, p.171-179, 2010. DOI: https://doi.org/10.1111/j.1439-0388.2009.00847.x.

COSTA, J.A. da. **Predição genômica via redução de dimensionalidade em modelos aditivo-dominante**. 2018. 107p. Dissertação (Magister Scientiae) - Universidade Federal de Viçosa, Viçosa.

DAETWYLER, H.D.; CALUS, M.P.L.; PONG-WONG, R.; DE LOS CAMPOS, G.; HICKEY, J.M. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v.193, p.347-365, 2013. DOI: https://doi.org/10.1534/genetics.112.147983.

DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALLUS, M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v.193, p.327-345, 2013. DOI: https://doi.org/10.1534/genetics.112.143313.

DENIS, M.; BOUVET, J.-M. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. **Tree Genetics & Genomes**, v.9, p.37-51, 2013. DOI: https://doi.org/10.1007/s11295-012-0528-1.

DU, C.; WEI, J.; WANG, S.; JIA, Z. Genomic selection using principal component regression. **Heredity**, v.121, p.12-23, 2018. DOI: https://doi.org/10.1038/s41437-018-0078-x.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th ed. Edinburgh: Pearson, 1996. 464p.

FERREIRA, D.F. **Estatística multivariada**. 3.ed. Lavras: UFLA, 2018. 624p.

GARTHWAITE, P.H. An interpretation of partial least squares. **Journal of the American Statistical Association**, v.89, p.122-127, 1994. DOI: https://doi.org/10.1080/01621459.1994.10476452.

GIANOLA, D. Priors in whole-genome regression: the Bayesian alphabet returns. **Genetics**, v.194, p.573-596, 2013. DOI: https://doi.org/10.1534/genetics.113.151753.

GIANOLA, D.; PEREZ-ENCISO, M.; TORO, M.A. On marker-assisted prediction of genetic value: beyond the ridge. **Genetics**, v.163, p.347-365, 2003.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, v.124, p.323-330, 2007. DOI: https://doi.org/10.1111/j.1439-0388.2007.00702.x.

GODDARD, M.E.; HAYES, B.J.; MEUWISSEN, T.H.E. Using the genomic relationship matrix to predict the accuracy of genomic selection. **Journal of Animal Breeding and Genetics**, v.128, p.409-421, 2011. DOI: https://doi.org/10.1111/j.1439-0388.2011.00964.x.

HUANG, W.; MACKAY, T.F.C. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. **PLoS Genetics**, v.12, e1006421, 2016. DOI: https://doi.org/10.1371/journal.pgen.1006421.

HYVÄRINEN, A. New approximations of differential entropy for independent component analysis and projection pursuit. **Advances in Neural Information Processing Systems**, v.10, p.273-279, 1998.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**: with applications in R. New York: Springer, 2013. 426p. DOI: https://doi.org/10.1007/978-1-4614-7138-7.

LONG, N.; GIANOLA, D.; ROSA, G.J.M; WEIGEL, K.A. Dimension reduction and variable selection for genomic selection: application to predicting milk yield in Holsteins. **Journal of Animal Breeding and Genetics**, v.128, p.247-257, 2011. DOI: https://doi.org/10.1111/j.1439-0388.2011.00917.x.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, p.1819-1829, 2001.

MUÑOZ, P.R.; RESENDE JR., M.F.R.; GEZAN, S.A.; RESENDE, M.D.V.; DE LOS CAMPOS, G.; KIRST, M.; HUBER, D.; PETER, G.F. Unraveling additive from nonadditive effects using genomic relationship matrices. **Genetics**, v.198, p.1759-1768, 2014. DOI: https://doi.org/10.1534/genetics.114.171322.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2019. Available at: <https://www.R-project.org/>. Accessed on: July 20 2020.

RESENDE, M.D.V. de; SILVA, F.F. e; LOPES, P.S.; AZEVEDO, C.F. **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial**. Viçosa: Universidade Federal de Viçosa, 2012. 291p.

SMITH, R.D. The nonlinear structure of linkage disequilibrium. **Theoretical Population Biology**, v.134, p.160-170, 2020. DOI: https://doi.org/10.1016/j.tpb.2020.02.005.

SU, G.; CHRISTENSEN, O.F.; OSTERSEN, T.; HENRYON, M.; LUND, M.S. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. **PLoS One**, v.7, e45293, 2012. DOI: https://doi.org/10.1371/journal.pone.0045293.

TORO, M.A.; VARONA, L. A note on mate allocation for dominance handling in genomic selection. **Genetics Selection Evolution**, v.42, art.33, 2010. DOI: https://doi.org/10.1186/1297-9686-42-33.

VARONA, L.; LEGARRA, A.; TORO, M.A.; VITEZICA, Z.G. Non-additive effects in genomic selection. **Frontiers in Genetics**, v.9, art.78, 2018. DOI: https://doi.org/10.3389/fgene.2018.00078.

VITEZICA, Z.G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. **Genetics Research**, v.93, p.357-366, 2011. DOI: https://doi.org/10.1017/S001667231100022X.

VITEZICA, Z.G.; LEGARRA, A.; TORO, M.A.; VARONA, L. Orthogonal estimates of variances for additive, dominance and epistatic effects in populations. **Genetics**, v.206, p.1297-1307, 2017. DOI: https://doi.org/10.1534/genetics.116.199406.

VITEZICA, Z.G.; VARONA, L.; LEGARRA, A. On the additive and dominance variance and covariance of individuals within the genomic selection scope. **Genetics**, v.195, p.1223-1230, 2013. DOI: https://doi.org/10.1534/genetics.113.155176.

WANG, C.; DA, Y. Quantitative genetics model as the unifying model for defining genomic relationship and inbreeding coefficient. **PLoS One**, v.9, e114484, 2014. DOI: https://doi.org/10.1371/journal.pone.0114484.

WELLMANN, R.; BENNEWITZ, J. Bayesian models with dominance effects for genomic evaluation of quantitative traits. **Genetics Research**, v.94, p.21-37, 2012. DOI: https://doi.org/10.1017/S0016672312000018.

YAO, F.; COQUERY, J.; LÊ CAO, K.-A. Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets. **BMC Bioinformatics**, v.13, art.24, 2012. DOI: https://doi.org/10.1186/1471-2105-13-24.

ZENG, J.; TOOSI, A.; FERNANDO, R.L.; DEKKERS, J.C.M.; GARRICK, D.J. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. **Genetics Selection Evolution**, v.45, art.11, 2013. DOI: https://doi.org/10.1186/1297-9686-45-11.