

## PROPOSTA DE BLUP GENÔMICO COM EFEITOS ADITIVOS E DE DOMINÂNCIA EM AMBIENTE *R*

Vinicius Silva dos SANTOS<sup>1</sup>  
Sebastião MARTINS FILHO<sup>1</sup>  
Marcos Deon Vilela de RESENDE<sup>2</sup>  
Fabyano FONSECA E SILVA<sup>3</sup>

- RESUMO: Recentemente, efeitos de dominância, têm sido incluídos na seleção genômica de várias espécies, sendo o método GBLUP-D o mais utilizado. Esse método consiste em substituir, no procedimento REML/BLUP, as matrizes de parentesco baseadas no *pedigree* pelas matrizes com base nos marcadores moleculares. Este método pode ser realizado por meio do *software* GVCBLUP ou por meio do pacote BGLR do *software* R, o qual se baseia em regressão bayesiana via Kernel de Reprodução do Espaço de Hilbert. Objetivou-se nesse trabalho avaliar a possibilidade e efetividade de implementação do GBLUP-D via a função *lmekin* implementada no pacote *coxme* do *software* R por meio da inclusão das matrizes de parentesco genômicos aditivo e de dominância. Assim, comparou-se, via análises de dados simulados, os resultados obtidos pela função *lmekin* com os obtidos pelo *software* GVCBLUP e pacote BGLR. Os resultados mostraram que os métodos GBLUP e GBLUP-D ajustados via REML no *software* GVCBLUP e por meio da função *lmekin* são equivalentes. Assim, a função *lmekin* é uma alternativa eficiente para o ajuste, no *software* R, de modelos genômicos contemplando efeitos aditivos e de dominância.
- PALAVRAS-CHAVE: Amostrador de Gibbs; componentes de variância; modelo linear misto; SNPs.

### 1 Introdução

Tradicionalmente, no melhoramento animal e vegetal, o método mais utilizado de seleção é baseado nos valores genéticos esperados, que são estimados geralmente pelo procedimento REML/BLUP (estimação de componentes de variância por máxima verossimilhança residual - REML, e a predição de efeitos aleatórios pelo BLUP - melhor preditor linear não viesado), com base em informações fenotípicas e da matriz de parentesco (GODDARD e HAYES, 2007).

---

<sup>1</sup>Universidade Federal de Viçosa - UFV, Departamento de Estatística, Av. P. H. Rolfs, s/n – Campus Universitário, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: [vinicius.s.santos@ufv.br](mailto:vinicius.s.santos@ufv.br); [martinsfilho@ufv.br](mailto:martinsfilho@ufv.br).

<sup>2</sup>Embrapa Florestas/Universidade Federal de Viçosa - UFV, Caixa Postal 319, CEP: 83411-000, Colombo, PR, Brasil. E-mail: [marcos.deon@gmail.com](mailto:marcos.deon@gmail.com).

<sup>3</sup>Universidade Federal de Viçosa - UFV, Departamento de Zootecnia, Av. P. H. Rolfs, s/n – Campus Universitário, CEP: 36570-900, Viçosa, MG, Brasil. E-mail: [fabyanofonseca@ufv.br](mailto:fabyanofonseca@ufv.br).

No entanto, com o desenvolvimento e baixo custo dos marcadores moleculares, foi proposta, por Meuwissen *et al.* (2001), a Seleção Genômica Ampla (*Genomic Wide Selection - GWS*), a qual é definida como a seleção simultânea de milhares de marcadores cobrindo todo o genoma a fim de prever o valor genético dos indivíduos (MEUWISSEN *et al.*, 2001; RESENDE *et al.*, 2008).

Os primeiros trabalhos com GWS consideraram apenas o efeito genético aditivo. Recentemente, efeitos não-aditivos, como por exemplo, o efeito de dominância, foram incluídos na seleção genômica de várias espécies (SU *et al.*, 2012; ERTL *et al.*, 2014; MUÑOZ *et al.*, 2014). O método inicialmente proposto para a GWS com efeitos aditivos e de dominância foi o GBLUP (reportado na literatura como GBLUP-D), o qual é similar ao procedimento REML/BLUP, diferindo por substituir as matrizes de parentesco genético aditivo e de dominância, baseadas no *pedigree*, pelas matrizes baseadas nos marcadores. Esse procedimento é superior, pois captura a matriz de parentesco real e não uma matriz de parentesco médio associada ao *pedigree* (MEUWISSEN *et al.*, 2001; RESENDE *et al.*, 2008).

O método GBLUP-D está implementado no software denominado GVCBLUP (WANG *et al.*, 2014) e também no software R (R DEVELOPMENT CORE TEAM, 2016) por meio do pacote BGLR – Bayesian Generalized Linear Regression (PÉREZ e DE LOS CAMPOS, 2014) com base na Regressão Kernel de Reprodução do Espaço de Hilbert (*Reproducing Kernel Hilbert Spaces - RKHS*), usando a abordagem Bayesiana.

Santos *et al.* (2015) mostraram que o método GBLUP com efeitos aditivos pode ser realizado usando a função *lmekin* implementada no pacote *coxme* (THERNEAU, 2012) do software R. O objetivo deste trabalho foi estender esta metodologia incluindo o efeito de dominância (GBLUP-D) e comparar via dados simulados, os resultados obtidos usando a função *lmekin* com os obtidos no programa GVCBLUP e no pacote BGLR.

## 2 Material e métodos

Os dados analisados foram provenientes de um estudo de simulação realizado por Vitezica *et al.* (2013), em que uma característica quantitativa foi simulada com efeitos aditivos e de dominância, representando 20% e 10% da variância fenotípica, respectivamente. Os dados consistiram de 2100 indivíduos genotipados e fenotipados e 10.000 marcadores SNPs. A estimação dos componentes de variância e predição dos efeitos genéticos genômicos foi realizada por meio dos métodos GBLUP e GBLUP-D usando GVCBLUP, *lmekin* e BGLR.

### 2.1 Métodos GBLUP e GBLUP-D implementados no software GVCBLUP

O método GBLUP foi implementado no software GVCBLUP com base no seguinte modelo linear misto:

$$y = I\mu + Za + e \quad (1)$$

em que  $y$  é o vetor de fenótipos,  $\mu$  é a média geral,  $a$  é o vetor dos  $i$ -ésimos efeitos genéticos aditivos ( $i=1, \dots, n$ ),  $Z$  é a matriz de incidência para os efeitos aleatórios  $a$ , os quais apresentam distribuição normal com média 0 e matriz de covariância igual a  $G\sigma_a^2$ ; em que  $G$  é a matriz de parentesco genômico aditivo e  $\sigma_a^2$  é a variância genética aditiva. O

vetor de erros ( $e$ ) também possui distribuição normal com matriz de variâncias-covariâncias igual a  $I\sigma_e^2$ .

A matriz de parentesco genômico aditivo ( $G$ ) é dada por:

$$G = WW' / 2 \sum_{j=1}^m p_j (1 - p_j),$$

em que os valores  $w_{ij}$  da matriz  $W$  são iguais a  $0-2p$ ,  $1-2p$  e  $2-2p$  para os genótipos marcadores dos tipos  $mm$ ,  $Mm$  e  $MM$ , respectivamente, e  $p_j$  é a frequência alélica de  $M$  no loco  $j$  (VITEZICA *et al.*, 2013; RESENDE *et al.*, 2014).

Ao incluir o efeito de dominância, tem-se o método GBLUP-D com base no seguinte modelo linear misto:

$$y = I\mu + Za + Td + e \quad (2)$$

em que  $y$ ,  $\mu$ ,  $a$ ,  $Z$  e  $e$  são definidos da mesma forma que no modelo (1) e  $d$  é o vetor dos  $i$ -ésimos desvios de dominância ( $i=1, \dots, n$ ), sendo  $T$  a matriz de incidência dos respectivos efeitos aleatórios  $d$ , os quais apresentam distribuição normal com média 0 e matriz de covariância igual a  $D\sigma_d^2$ , onde  $\sigma_d^2$  é a variância genética de dominância e  $D$  é a matriz de parentesco genômico de dominância, dada pela seguinte expressão (VITEZICA *et al.*, 2013; RESENDE *et al.*, 2014):

$$D = SS' / \sum_{j=1}^m \{2p_j (1 - p_j)\}^2,$$

em que a matriz de incidência  $S$  contém os valores  $s_{ij}$  iguais a  $-2p^2$ ,  $2p(1-p)$  e  $-2(1-p)^2$  para os genótipos marcadores dos tipos  $mm$ ,  $Mm$  e  $MM$ , respectivamente; e  $p_j$  é o mesmo que na matriz  $G$ .

Ambas as matrizes foram obtidas no software GVCBLUP, o qual apresenta outras cinco diferentes formas de obtenção das matrizes de parentesco genômicos. Nesse trabalho, foi usada a definição 1 para as matrizes  $G$  e  $D$  (VANRADEN, 2008; WANG *et al.*, 2014).

A estimação dos componentes de variância em ambos os métodos foi realizada por meio do procedimento REML usando o método de maximização AI (Informação Média), o qual se baseia em derivadas parciais de primeira e segunda ordens (RESENDE *et al.*, 2014). Além do AI-REML, o método EM-REML (Esperança-Maximização) também está implementado no software GVCBLUP (WANG *et al.*, 2014). Os valores genéticos aditivos e desvios de dominância dos indivíduos foram preditos usando as equações de modelo misto, que para o caso do modelo completo com efeitos aditivos e de dominância são dadas por:

$$\begin{bmatrix} X'X & X'Z & X'Z \\ Z'X & Z'Z + G^{-1} \sigma_e^2 / \sigma_a^2 & Z'Z \\ Z'X & Z'Z & Z'Z + D^{-1} \sigma_e^2 / \sigma_d^2 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z'y \end{bmatrix}$$

em que  $\sigma_a^2$ ,  $\sigma_d^2$  e  $\sigma_e^2$  são as variâncias genéticas aditiva e de dominância e variância residual, respectivamente. Para o modelo com somente efeitos aditivos, as equações de modelo misto são dadas pela seguinte expressão:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \sigma_e^2 / \sigma_a^2 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

## 2.2 Métodos GBLUP e GBLUP-D usando o pacote BGLR

Os métodos GBLUP e GBLUP-D foram implementados no pacote BGLR como casos particulares do Modelo de Regressão Bayesiana Kernel de Reprodução do Espaço de Hilbert (*Reproducing Kernel Hilbert Spaces - RKHS*), uma vez que a matriz de *kernel*  $K$  pode ser computada a partir do *pedigree* e/ou dos marcadores (DE LOS CAMPOS *et al.*, 2009).

A forma de ajuste dos métodos é apresentada com base no modelo com efeitos aditivos e de dominância, sendo o método GBLUP um caso particular (quando não se estima o efeito de dominância). Para o método GBLUP-D via RKHS, foi definido o seguinte modelo (PÉREZ e DE LOS CAMPOS, 2014):

$$y = I\mu + Za + Td + e \quad (3)$$

em que  $y$ ,  $\mu$ ,  $Z$ ,  $T$  e  $e$  são definidos como no modelo (2),  $a$  é o vetor de valores genéticos aditivos individuais e  $d$  é o vetor de desvios de dominância. A distribuição condicional dos dados é normal multivariada e para  $\mu$ ,  $a$  e  $d$ , as seguintes distribuições *a priori* foram assumidas:

$$p(\mu) \propto \text{constante}; a|K_1, \sigma_a^2 \sim N(0, K_1 \sigma_a^2) \text{ e } d|K_2, \sigma_d^2 \sim N(0, K_2 \sigma_d^2),$$

em que  $K_1$  e  $K_2$  são as matrizes de *Kernel* com dimensão igual ao número de indivíduos, sendo que no método GBLUP-D,  $K_1$  e  $K_2$  foram substituídas pelas matrizes de parentesco genômico aditivo ( $G$ ) e de dominância ( $D$ ). No pacote BGLR, o método GBLUP-D via regressão RKHS é implementado ao fornecer as matrizes  $G$  e  $D$  diretamente ou por meio da decomposição de seus autovalores. Quando as matrizes de parentesco são fornecidas, a decomposição dos autovalores é feita internamente (PÉREZ e DE LOS CAMPOS, 2014). Para os componentes de variância  $\sigma_a^2$ ,  $\sigma_d^2$  e  $\sigma_e^2$ , foram assumidas as distribuições *a priori*:  $\sigma_i^2 \sim \chi^2(\sigma_i^2 | df_i, S_i)$ ; com graus de liberdade  $df_i$  e parâmetros de escala  $S_i > 0$  especificados de acordo com as configurações padrão do pacote (PÉREZ e DE LOS CAMPOS, 2014). Para o método GBLUP com somente efeitos aditivos, o processo é idêntico, considerando os efeitos de dominância ( $d$ ) iguais a 0 (DE LOS CAMPOS, 2009).

Os métodos GBLUP e GBLUP-D via RKHS foram implementados por meio da abordagem bayesiana usando o algoritmo *Gibbs sampler*, em que foram geradas 60.000 amostras, sendo as 10.000 iniciais descartadas visando à obtenção de estacionariedade do processo estocástico e, para assegurar a independência entre amostras foi considerado um espaçamento de 5 entre as amostras salvas. A convergência da cadeia foi verificada por meio do critério de Geweke (GEWEKE, 1992), implementado na biblioteca BOA (SMITH, 2007), do *software* R (R CORE TEAM, 2016).

### 2.3 Métodos GBLUP e GBLUP-D usando a função *lmekin* do R

O método GBLUP usando a função *lmekin* implementada no pacote *coxme* (THERNEAU, 2012) do software R foi definido da mesma forma que em (1). Ao incluir o efeito de dominância, a função não permite a predição de dois efeitos aleatórios simultaneamente. Uma alternativa para o método GBLUP-D foi considerar o seguinte modelo linear misto:

$$y = I\mu + Zg + e \quad (4)$$

em que  $y$ ,  $\mu$ , e  $e$  são como já definidos em (1) e (2), e  $g=a+d$  é o vetor dos  $i$ -ésimos valores genotípicos totais preditos, os quais são dados pela soma dos efeitos genéticos aditivos e desvios de dominância, assumidos com distribuição normal com média 0 e matriz de variâncias e covariâncias  $\Sigma = G\sigma_a^2 + D\sigma_d^2$ , com  $a$ ,  $d$ ,  $\sigma_a^2$ ,  $\sigma_d^2$ ,  $G$  e  $D$  definidos anteriormente. A equivalência entre os modelos 2 e 4 é mantida devido a pressuposição de que em ambos a covariância entre os efeitos genéticos aditivos e de dominância é nula (VITEZICA *et al.*, 2013). Conforme Mrode (2005, p.195), os valores genéticos aditivos individuais ( $a$ ) e dos desvios de dominância ( $d$ ) podem ser obtidos separadamente a partir dos valores genéticos totais ( $g$ ) pelas respectivas expressões:

$$\hat{a} = \sigma_a^2 G \Sigma^{-1} \hat{g} \text{ e } \hat{d} = \sigma_d^2 D \Sigma^{-1} \hat{g}.$$

Da mesma forma que no software GVCBLUP, os componentes de variância foram estimados em ambos os métodos, GBLUP e GBLUP-D, por meio do procedimento REML via a maximização da verossimilhança marginal dos resíduos a partir do ajuste de mínimos quadrados generalizados (VENABLES e RIPLEY, 2002). Para a maximização dessa função de verossimilhança, a função *lmekin* utiliza inicialmente o algoritmo EM e após um número adequado de iterações, passa-se a usar o algoritmo Newton-Raphson (PINHEIRO e BATES, 2000).

## 2.4 Estimativas de herdabilidade e comparação de modelos

As estimativas de herdabilidade aditiva ou no sentido restrito ( $h_a^2$ ), de dominância ( $h_d^2$ ) e no sentido amplo ( $H^2$ ), considerando as três abordagens descritas acima, foram obtidas pelas respectivas expressões:

$$h_a^2 = \left[ \sigma_a^2 / (\sigma_a^2 + \sigma_d^2 + \sigma_e^2) \right]; h_d^2 = \left[ \sigma_d^2 / (\sigma_a^2 + \sigma_d^2 + \sigma_e^2) \right]; H^2 = h_a^2 + h_d^2.$$

A inclusão do efeito de dominância no método GBLUP usando a função *lmekin* foi testada por meio do teste da razão de verossimilhança de Wilks (*Likelihood Ratio Test - LRT*), o qual é dado por:

$$\Lambda = 2\ln L(V) - 2\ln L(U),$$

em que  $L(V)$  é a verossimilhança do modelo com efeitos aditivos e de dominância e  $L(U)$  é a verossimilhança do modelo com somente efeitos aditivos (verossimilhança sob  $H_0$ ). A estatística do teste tem distribuição Qui-quadrado com graus de liberdade dados pela diferença entre o número de componentes de variância entre os dois modelos avaliados. O valor-p para o teste é dado por  $0,5[1 - P(\chi^2 \leq \lambda)]$ , onde  $\lambda$  é o valor observado de  $\Lambda$  (VISSCHER, 2006; ERTL *et al.*, 2014).

Para os modelos ajustados usando o pacote BGLR, a inclusão do efeito de dominância foi avaliada usando o critério DIC (*Deviance Information Criterion*), dado por:

$$DIC = \bar{D} + P_D$$

em que  $\bar{D}$  é a média *a posteriori* da deviance e  $P_D$  é o “número efetivo de parâmetros”, sendo o melhor modelo aquele que apresenta menor valor DIC (SPIEGELHALTER, 2002).

A capacidade preditiva dos métodos, em cada abordagem empregada, foi avaliada com base nas acurácias obtidas por meio da correlação entre os valores genéticos verdadeiros (VGV) e os valores genéticos preditos (VGP). Foram calculados também os coeficientes de regressão de VGV sobre VGP, os quais indicam se a predição foi viesada ou não (VITEZICA *et al.*, 2013; RESENDE *et al.*, 2014). Os resultados foram baseados na média das dez replicações do conjunto de dados disponíveis em <http://genoweb.toulouse.inra.fr/~zvitezic/simu> for [genetics.tar.gz](http://genetics.tar.gz) (VITEZICA *et al.*, 2013).

## 3 Resultados e discussão

A convergência da cadeia para os métodos bayesianos GBLUP (efeitos aditivos) e GBLUP-D (efeitos aditivos e de dominância) ajustados no pacote BGLR foi atendida, de acordo com o critério de Geweke, em que o valor-p obtido foi sempre maior que o nível de significância de 5%, não rejeitando a hipótese de igualdade entre as médias das 10% primeiras iterações após *burn-in* e as últimas 50% iterações da cadeia (NOGUEIRA *et al.*, 2004).

Pode ser notado tanto no método GBLUP (somente variância genética aditiva – Figura 1a) quanto no GBLUP-D (variâncias genéticas aditivas e de dominância - Figura 1b, 1c) que as variâncias se apresentaram de forma semelhante quando foram estimadas pela função *lmekin* e pelo software GVCBLUP, não ocorrendo o mesmo com o uso do pacote BGLR. Com o uso desse mesmo pacote para estimar as variâncias aditivas pelo método GBLUP, notou-se valores levemente maiores que às obtidas pela função *lmekin* e pelo software GVCBLUP (Figura 1a). Entretanto, quando foi utilizado o método GBLUP-D isto não ocorreu (Figura 1b).

Verificando ainda o uso do pacote BGLR com o método GBLUP-D (Figura 1c), observou-se que as estimativas das variâncias genéticas devido à dominância foram maiores que o valor simulado de 10 em todas as replicações do conjunto de dados simulados, com estimativas variando entre 11,34 e 18,12. Porém, com o uso do programa GVCBLUP e da função *lmekin*, essas estimativas variaram entre 5,81 e 14,81, e entre 4,81 e 16,07, respectivamente.

De acordo com a Tabela 1, as estimativas de herdabilidade obtidas por meio do software GVCBLUP e usando a função *lmekin* apresentaram valores idênticos, em torno de 0,202, 0,125 e 0,327 para as herdabilidades aditiva, de dominância e no sentido amplo, respectivamente, com desvios-padrão iguais a 0,03, concordando assim com os valores simulados e evidenciando a correta implementação dos métodos GBLUP e GBLUP-D usando a função *lmekin*.

Ao incluir o efeito de dominância, houve uma diminuição da herdabilidade aditiva somente com o uso do pacote BGLR, sendo esse decréscimo em torno de 8,0%. Segundo De Los Campos *et al.* (2009), os métodos GBLUP e GLUP-D ajustados pelo pacote BGLR são considerados casos particulares do método RKHS, em que as matrizes de parentesco genômicos *G* e *D* são usadas como *kernels* paramétricas.

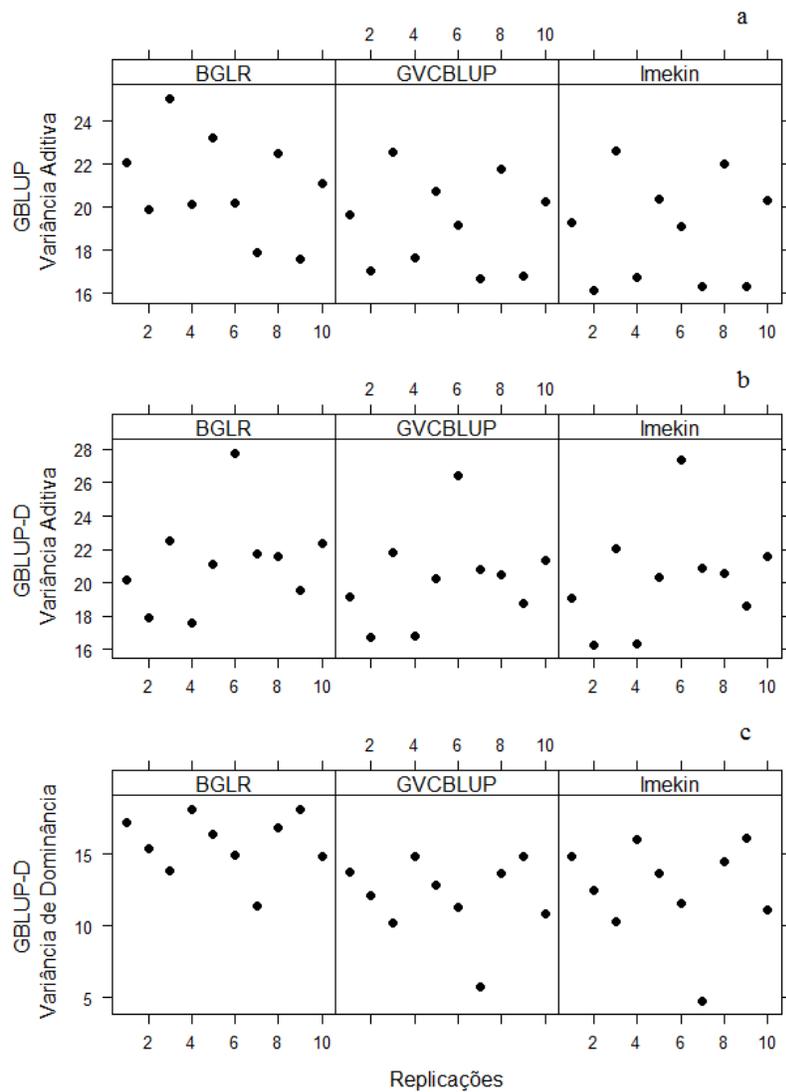


Figura 1 - Variâncias genéticas aditiva e de dominância estimadas com base nos métodos GBLUP (modelo somente com efeitos aditivos) e GBLUP-D (modelo com efeitos aditivos e de dominância) usando BGLR, GVCBLUP e *Imekin*.

As estimativas de herdabilidade obtidas pelo BGLR foram maiores que as estimativas obtidas pelo GVCBLUP e *Imekin*, em que a herdabilidade de dominância foi cerca de 23,2% maior, superestimando o verdadeiro valor simulado de 0,10. Isso ocorreu

devido a superestimação da variância de dominância, que foi igual a 15,68 com desvio-padrão de 2,09 com base nas dez replicações do conjunto de dados avaliado. Com isso, a herdabilidade no sentido amplo também foi superestimada, com valor de  $0,362 \pm 0,02$ . Resultado semelhante foi obtido por Morota et al. (2014), os quais atribuíram à falta de ortogonalidade entre as matrizes de parentesco genômicos  $G$  e  $D$  uma das possíveis causas dessa superestimação. Uma alternativa mencionada pelos autores seria a obtenção da matriz  $GD = X_{AD}X'_{AD}$ , em que  $X_{AD}$  é a matriz aditiva x dominância (XU et al., 2013).

Tabela 1 - Herdabilidades aditiva ( $h_a^2$ ), de dominância ( $h_d^2$ ), total ( $H^2$ ) e tempo de execução considerando o software GVCBLUP, função *lmekin* e o pacote BGLR em dados simulados

Modelos	$h_a^2$	$h_d^2$	$H^2$	Tempo*
Valor Verdadeiro	0,20	0,10	0,30	-
<b>GBLUP</b>				
BGLR	$0,226 \pm 0,03$	-	-	32.52
GVCBLUP	$0,202 \pm 0,03$	-	-	1.10
<i>lmekin</i>	$0,202 \pm 0,03$	-	-	21.04
<b>GBLUP-D</b>				
BGLR	$0,208 \pm 0,03$	$0,154 \pm 0,02$	$0,362 \pm 0,02$	49.25
GVCBLUP	$0,202 \pm 0,03$	$0,125 \pm 0,03$	$0,327 \pm 0,03$	2.41
<i>lmekin</i>	$0,202 \pm 0,03$	$0,125 \pm 0,03$	$0,327 \pm 0,03$	44.12

\* tempo em minutos usando um computador Intel Core i5 de 1.70 GHz e memória RAM de 4GB.

Vitezica et al. (2013) analisando esse mesmo conjunto de dados por meio de inferência bayesiana, encontraram estimativas de variância genética de dominância ( $13,2 \pm 3,5$ ) próximas as obtidas neste trabalho, usando a função *lmekin* e o software GVCBLUP.

Embora os resultados sejam idênticos, o software GVCBLUP possui a vantagem de predizer os valores genéticos mais rapidamente do que a função *lmekin*, conforme pode ser visto na Tabela 1, em que o software realizou o ajuste dos modelos em um tempo médio de 1 minuto e 10 segundos para o método GBLUP e 2 minutos e 41 segundos para o método GBLUP-D. O software GVCBLUP, assim como o R, é um software livre, disponível online em (<http://animalgene.umn.edu/>), o qual é programado em C++ e usa computação paralela com memória compartilhada na inversão de matrizes densas como as matrizes de correlação genética aditiva ( $G$ ) e de dominância ( $D$ ), otimizando assim o tempo computacional na realização dos métodos GBLUP e GBLUP-D (Wang et al., 2014).

Já na função *lmekin*, os mesmos modelos foram ajustados com aproximadamente 21 minutos (GBLUP) e 44 minutos (GBLUP-D). Dentre as três abordagens, a que levou mais tempo para ser realizada foi o ajuste dos modelos por meio do pacote BGLR, com um tempo médio de 33 e 49 minutos para os métodos GBLUP e GBLUP-D, respectivamente, em que foram usadas 60.000 iterações em cada método. A vantagem de usar a função *lmekin* é a possibilidade de estender a metodologia proposta considerando o modelo de fragilidade de Cox para dados censurados, conforme realizado por Santos et al. (2015) com efeitos aditivos. Outra vantagem em se usar a função *lmekin* é a disponibilidade dos

valores de log-verossimilhança, permitindo assim testar a significância do efeito de dominância por meio do LRT, conforme apresentado na Tabela 2.

Tabela 2 - Valores de  $-2\log$ -verossimilhança,  $\chi^2$  e seus correspondentes valores-p do teste da razão de verossimilhanças para os modelos ajustados usando a função *lmekin*, além dos valores DIC ajustados no pacote BGLR

Replicações	Modelos	$-2\log L$	valor- $\chi^2$	valor-p	DIC
1	GBLUP	15491,08			15411,40
	GBLUP-D	15472,41	18,671	$7,76*10^{-6}$	15282,81
2	GBLUP	15530,00			15465,98
	GBLUP-D	15515,4	14,599	$6,64*10^{-5}$	15368,96
3	GBLUP	15471,45			15372,69
	GBLUP-D	15459,89	11,554	0,00034	15283,43
4	GBLUP	15539,04			15473,90
	GBLUP-D	15519,36	19,677	$4,58*10^{-6}$	15346,65
5	GBLUP	15499,99			15413,64
	GBLUP-D	15485,44	14,545	$6,84*10^{-5}$	15295,45
6	GBLUP	15472,31			15347,86
	GBLUP-D	15458,72	13,591	0,00011	15227,57
7	GBLUP	15467,66			15378,73
	GBLUP-D	15465,57	2,093	0,07399	15325,05
8	GBLUP	15509,24			15427,04
	GBLUP-D	15490,48	18,761	$7,40*10^{-6}$	15293,45
9	GBLUP	15501,25			15427,79
	GBLUP-D	15481,31	19,938	$3,99*10^{-6}$	15285,03
10	GBLUP	15473,51			15379,00
	GBLUP-D	15462,24	11,271	0,00039	15279,20

Nessa mesma Tabela 2 também são apresentados os valores DIC para os métodos GBLUP e GBLUP-D ajustados pelo pacote BGLR, os quais indicaram como melhor modelo aquele que contém efeitos aditivos e de dominância.

Acurácias e coeficientes de regressão para cada modelo são apresentados na Tabela 3, na qual pode ser observado que ao incluir o efeito de dominância, as acurácias entre os valores genéticos preditos e verdadeiros permaneceram inalterados, em torno de  $0,72 \pm 0,02$ , apesar do efeito significativo de dominância (Tabela 2). Resultado semelhante foi reportado por Ertl *et al.* (2014), em que encontraram efeito significativo de dominância em cinco das nove características avaliadas, porém, ao incluir esse efeito, as acurácias permaneceram constantes.

Tabela 3 - Acurácias ( $r$ ) e coeficientes de regressão ( $b$ , medida de viés) entre valores genéticos preditos e verdadeiros para dez replicações de dados simulados, considerando o software GVCBLUP, função *lmekin* e o pacote BGLR

Modelos	$r$	$b$
GBLUP		
GVCBLUP	0,72 (0,02)	0,97 (0,06)
<i>lmekin</i>	0,72 (0,02)	0,97 (0,06)
BGLR	0,72 (0,02)	0,93 (0,05)
GBLUP-D		
GVCBLUP	0,72 (0,02)	0,97 (0,06)
<i>lmekin</i>	0,72 (0,02)	0,97 (0,06)
BGLR	0,72 (0,02)	0,96 (0,06)

Altos valores de acurácia possibilitam um maior ganho genético em curto prazo. Vários trabalhos no melhoramento animal e vegetal têm comprovado a superioridade da GWS em relação à seleção usual, em que foi observado, por exemplo, um aumento em torno de 33% nas acurácias seletivas em gado de corte (VANRADEN *et al.*, 2009) e de 25% em ovinos (DUCHEMIN *et al.*, 2012) ao considerar informações dos marcadores ao invés do pedigree somente.

Acurácias genômicas superiores a 90% foram obtidas em espécies vegetais como milho (FRITSCHÉ NETO *et al.*, 2012) e trigo (HESLOT *et al.*, 2012). Um dos fatores dessa superioridade da seleção genômica é devido ao uso da matriz de parentesco realizada ao invés da matriz de parentesco médio como ocorre na seleção usual (RESENDE, 2008).

De acordo com Resende *et al.*, 2014, a acurácia da GWS depende de cinco fatores: (i) do espaçamento entre marcadores, o qual depende do seu número e do tamanho do genoma; (ii) do número de indivíduos fenotipados e genotipados na população de treinamento; (iii) da herdabilidade do caráter; (iv) do número de locos controlando o caráter e da distribuição de seus efeitos; e (v) do tamanho efetivo da população.

Os coeficientes de regressão nas três abordagens empregadas foram próximos de 1, indicando assim que as predições foram não viesadas (RESENDE *et al.*, 2014). Da mesma forma que as acurácias, os coeficientes de regressão obtidos foram próximos aos reportados por Vitezica *et al.* (2013) ao avaliarem o mesmo conjunto de dados, porém, com desvios-padrão menores ( $0,96 \pm 0,08$ ).

No software R, até o momento, a única forma de implementação do método GBLUP-D via REML é por meio da função *lmekin*. Além dos programas aqui mostrados, outros softwares como GS3 (LEGARRA *et al.*, 2011; VITEZICA *et al.*, 2013; AZEVEDO *et al.*, 2015), ASReML (GILMOUR *et al.*, 2015; MUÑOZ *et al.*, 2014) e DMU (SU *et al.*, 2012; MADSEN, 2013) também permitem a predição genômica com base em modelos mistos considerando efeitos aditivos e não aditivos.

A estimação de efeitos de dominância usando a matriz de correlação genética, com base nos marcadores, tem sido de mais fácil implementação quando comparado ao ajuste dos modelos com base no *pedigree* (SU *et al.*, 2012; VITEZICA *et al.*, 2013). Além disso,

vários estudos considerando somente efeitos aditivos (VANRADEN, 2008; HESLOT *et al.*, 2012; RESENDE *et al.*, 2012) e mais recentemente, efeitos aditivos e não aditivos (VITEZICA *et al.*, 2013; ERTL *et al.*, 2014; MUÑOZ *et al.*, 2014; AZEVEDO *et al.*, 2015), têm comprovado a superioridade do método BLUP genômico (GBLUP-D) em relação ao BLUP convencional baseado no *pedigree*.

## Conclusões

Comprovou-se a possibilidade e efetividade de realizar a predição genômica de valores genéticos individuais considerando efeitos aditivos e de dominância por meio da função *lmekin* do software R. Os resultados obtidos foram idênticos aos do software GVCBLUP já disponível para esse tipo de análise. O método GBLUP-D implementado no pacote BGLR via Kernel de Reprodução do Espaço de Hilbert (RKHS) necessita ser mais bem investigado, uma vez que as estimativas de variância de dominância foram superestimadas.

## Agradecimentos

À CAPES pela bolsa concedida ao primeiro autor para doutorado no Brasil e doutorado sanduíche no exterior sob o processo de nº BEX 9415/14-9.

SANTOS, V. S., MARTINS FILHO, S., RESENDE, M. D. V., SILVA, F. F. Proposal for Genomic BLUP with additive and dominance effects in R environment. *Rev. Bras. Biom.*, Lavras, v.35, n.2, p. 361-375, 2017.

- **ABSTRACT:** *Recently, the dominance effects have been included in genomic selection of various species, being the GBLUP-D method the most widely used. This method consists in replacing, in the REML/BLUP procedure, the pedigree-based relationship matrices by marker-based relationship matrices. This method can be performed by means of GVCBLUP software or through R-package BGLR, which is based on Bayesian regression via Reproducing Kernel Hilbert Spaces – RKHS. In addition to these programs, we proposed the lmekin function implemented in the coxme package of R, which also allows the inclusion of additive and dominance genomic relationship matrices. In this context, we aimed to compare, via simulated data, the results from lmekin function with those from GVCBLUP software and BGLR. The results showed that the GBLUP and GBLUP-D methods fitted via REML in the GVCBLUP software and by means of the lmekin function are equivalent. The lmekin function is an efficient alternative for fitting genomic models with additive and dominance effects.*
- **KEYWORDS:** *Gibbs sampling; linear mixed model; SNPs; variance components.*

## Referências

AZEVEDO, C. F.; RESENDE, M. D. V.; SILVA, F. F.; VIANA, J. M. S.; VALENTE, M. S. F.; RESENDE, M. F. R.; MUÑOZ, P. Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genetics, London*, v.16, n.105, p.1-13, 2015.

DE LOS CAMPOS, G.; GIANOLA, D.; ROSA, G. J. M. Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. *Journal of Animal Science, Champaign*, v.87, n.6, p.1883-1887, 2009.

- DUCHEMIN, S. I.; COLOMBANI, C.; LEGARRA, A.; BALOCHE, G.; LARROQUE, H.; ASTRUC, J. M.; BARILLET, F.; ROBERT-GRANIÉ, C.; MANFREDI, E. Genomic selection in the French Lacaune dairy sheep breed. *Journal of Science Dairy*, Champaign, v.95, n.5, p.2723–2733, 2012.
- ERTL, J.; LEGARRA, A.; VITEZICA, Z.; VARONA, L.; EDEL, C.; EMMERLING, R.; GÖTZ, K. U. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genetics Selection Evolution*, London, v.46, n.1, p.40-49, 2014.
- FRITSCHÉ-NETO, R., RESENDE, M. D. V., MIRANDA, G. V., VALE, J. C. Seleção genômica ampla e novos métodos de melhoramento do milho. *Ceres*, Viçosa, v.59, n.6, p.794-802, 2015.
- GEWEKE, J. Evaluating the accuracy of sampling based approaches to the calculation of posterior moments. In: BERNARDO, J. O.; BERGER, J. M.; DAWID, A. P.; SMITH, A. F. M. (Ed.). *Bayesian statistics 4*. Oxford: Oxford University Press, 1992. p. 169–194.
- GILMOUR, A. R.; GOGEL, B. J.; CULLIS, B. R.; WELHAM, S. J.; THOMPSON, R. (2015). *ASReml User Guide Release 4.1 Structural Specification*, VSN International Ltd, Hemel Hempstead, HP11ES, UK <[www.vsni.co.uk](http://www.vsni.co.uk)>
- GODDARD, M. E.; HAYES, B. J. Genomic selection. *Journal of Animal Breeding and Genetics*, Austin, v.124, n.6, p.323-330, 2007.
- HESLOT, N.; YANG, H. P.; SORRELLS, M. E.; JANNINK, J. L. Genomic selection in plant breeding: a comparison of models. *Crop Science*, Madison, v.52, n.1, p.146-160, 2012.
- LEGARRA, A.; RICARD, A.; FILANGI, O. (2011). *GS3: Genomic Selection, Gibbs Sampling, Gauss-Seidel (and BayesCpi)*. <<http://genoweb.toulouse.inra.fr/~alegarra/>> Acesso em: 27 mar. 2015.
- MADSEN, P.; JENSEN, J. (2013). *A user's guide to DMU. A Package for Analysing Multivariate Mixed Models*. Version 6, release 5.2.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome wide dense marker maps. *Genetics*, Austin, v.157, n.4, p.1819-1829, 2001.
- MOROTA, G.; BODDHIREDDY, P.; VUKASINOVIC, N.; GIANOLA, D.; DENISE, S. Kernel-based variance component estimation and whole-genome prediction of pre-corrected phenotypes and progeny tests for dairy cow health traits. *Frontiers in Genetics*, Lausanne, v.5, n.56, doi: 10.3389/fgene.2014.00056, 2014.
- MRODE, R. A. *Linear models for the prediction of animal breeding values*. 2th Edition. Wallingford: CAB International, 2005. 358p.
- MUÑOZ, P. R.; RESENDE, M. F.; 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*, Austin, v.198, n.4, p.1759-1768, 2014.

- NOGUEIRA, D. A.; SAFÁDI, T.; FERREIRA, D. F. Avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov. *Revista Brasileira de Estatística*, Rio de Janeiro, v.65, n.224, p.59-88, 2004.
- PINHEIRO, J. C.; BATES, D.M. *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag, 2000. 537 p.
- PÉREZ, P.; DE LOS CAMPOS, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, Austin, v.198, n.2, p.483-495, 2014.
- R CORE TEAM. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, Version 3.2.4 (16-03-2016). Disponível em: <<http://www.R-project.org>>
- RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesquisa Florestal Brasileira*, Colombo, v.56, n.1, p.63-78, 2008.
- RESENDE JR., M. F. R.; MUÑOZ, P.; RESENDE, M. D. V.; GARRICK, D. J.; FERNANDO, R. L.; DAVIS, J. M.; JOKELA, E. J.; MARTIN, T. A.; PETER, G. F.; KIRST, M. Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine (*Pinus taeda* L.). *Genetics*, Austin, v.190, n.1, p.1503-1510, 2012.
- RESENDE, M. D. V.; SILVA, F. F.; AZEVEDO, C. F. *Estatística matemática, biométrica e computacional: modelos mistos, multivariados, categóricos e generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência*. Viçosa: Editora Suprema, 2014. 881 p.
- SANTOS, V. S.; MARTINS, F. S.; RESENDE, M. D.; AZEVEDO, C. F.; LOPES, P. S.; GUIMARÃES, S. E.; GLÓRIA, L.; SILVA, F. F. Genomic selection for slaughter age in pigs using the Cox frailty model. *Genetics and Molecular Research*, Ribeirão Preto, v.14, n.4, p.12616-12627, 2015.
- SMITH, B. J. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, Washington, v.21, n.11, p.1-37, 2007.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v.64, n.4, p.583-639, 2002.
- SU, G.; CHRISTENSEN, O.F.; HENRYON, M.; LUND, M.S. Estimating additive and non-additive genetic variances and prediction genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One*, San Francisco, v.7, n.9, p.e45293, 2012.
- THERNEAU, T. *coxme: Mixed effects Cox models (2012)*. R package version 2.2-3. <<http://CRAN.R-project.org/package=coxme>>. Acesso em: 04 mar. 2016.
- VANRADEN, P. M. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, Champaign, v.91, n.11, p.4414-4423, 2008.
- VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: reliability of

genomic predictions for North American Holstein bulls, *Journal of Dairy Science*, Champaign, v.92, n.1, p.16-24, 2009.

VENABLES, W. N.; RIPLEY, B. D. *Modern Applied Statistics with S-PLUS*. 4ed. New York: Springer Verlag, 2002. 504p.

VISSCHER, P. M. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin research and human genetics*, Cambridge, v.9, n.4, p.490-495, 2006.

VITEZICA, Z. G.; VARONA, L.; LEGARRA, A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, Austin, v.195, n.4, p.1223-1230, 2013.

WANG, C.; DA, Y. Quantitative genetics model as the unifying model for defining genomic relationship and inbreeding coefficient. *PloS One*, San Francisco, v.9, n.12, p.e114484, 2014.

XU, S. Mapping quantitative trait loci by controlling polygenic background effects. *Genetics*, Austin, v.195, n.4, p.1209-1222, 2013.

Recebido em 15.04.2016

Aprovado após revisão em 27.09.2016