ADRIANO GARCIA

IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA PARA IDADE AO PRIMEIRO PARTO EM FÊMEAS NELORE

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação, Mestrado Profissional em Zootecnia, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
G216i
2019

Garcia, Adriano, 1970-
Implementação da seleção genômica para idade ao primeiro parto em fêmeas Nelore / Adriano Garcia. – Viçosa, MG, 2019.
vi, 19 f. : il. ; 29 cm.

Orientador: Henrique Torres Ventura.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 18-19.

1. Genômica. 2. Nelore (Bovino) - Seleção. I. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-Graduação em Zootecnia. II. Título.

CDD 22. ed. 636.20821

ADRIANO GARCIA

IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA PARA IDADE AO PRIMEIRO PARTO EM FÊMEAS NELORE.

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação, Mestrado Profissional em Zootecnia, para obtenção do título de *Magister Scientiae*.

APROVADA: 12 de julho de 2019.

Fabyano Fonseca e Silva

Daniele B. D. Marques

Daniele Botelho Diniz Marques

Henrique Torres Ventura
(Orientador)

AGRADECIMENTOS

A **Deus**, por me permitir desfrutar e viver um pouco mais desta vida.

Aos meus **Pais**, pelo amor e a educação que me ofereceram com seus exemplos de retidão e humildade. Apoiando-me sempre com muito carinho, respeito e amor.

À minha esposa, **Angeliana Martini Garcia** e meu filho, **Luis Eduardo Martini Garcia**, pela paciência com minhas ausências, mas, principalmente, pelo amor e incentivo.

Ao **Dr. Henrique Torres Ventura**, meu orientador.

Ao **Prof. Dr. Fabyano Fonseca e Silva** pela amizade e apoio incondicional e imprescindível.

Aos amigos, companheiros de trabalho e aulas, **Fábio Eduardo Ferreira**, **Roberto Winkler**, **Luis Antônio Josahkian**, **Eric Costa Marques**. Valeu os desafios, descontração, alegrias, trocas de experiências e estudos.

À **Ednira Gleida Marques**, por acreditar em mim e me conceder a oportunidade de fazer este mestrado.

Em especial, à **Jackeline Gomes Viegas**, pela amizade, companheirismo, perseverança, compromisso, dedicação e principalmente por ser uma pessoa iluminada, meu **Anjo da Guarda**.

Aos **professores da UFV**, que ao longo desse curso, me ensinaram e ajudaram em meu crescimento intelectual e profissional.

À **ABCZ** pelo incentivo e pela oportunidade de investir em meu crescimento profissional.

LISTA DE TABELAS

RESUMO

GARCIA, Adriano, M.Sc., Universidade Federal de Viçosa, julho de 2019. **Implementação da seleção genômica para idade ao primeiro parto em fêmeas Nelore.** Orientador: Henrique Torres Ventura.

Objetivou-se utilizar a seleção genômica (via duas metodologias: *single-step* e o índice de seleção combinado) e a análise tradicional baseada em pedigree para uma avaliação genética intra-rebanho considerando a característica idade ao primeiro parto (IPP) em fêmeas Nelore. Os resultados foram comparados tendo em vista a acurácia e o viés dos valores genéticos preditos. Também foram obtidas estimativas dos componentes de variância e herdabilidade para IPP. Para a análise considerando índice de seleção combinado, vários métodos estatísticos que consideram a seleção de covariáveis (problema de multicolinearidade) e a regularização do processo de estimação (problema de dimensionalidade) foram utilizados (RR-BLUP, Bayes A, Bayes B e Bayes LASSO). Foram usados dados fenotípicos (IPP) e genotípicos de 714 animais, e o pedigree contemplou um total de 4.133 animais. O método *single-step* superou todos os outros, e dentre os métodos *two-step*, o Bayes B se sobressaiu em relação aos demais. A melhor performance do método *single-step* pode ser explicada pelo fato do mesmo absorver informações de parentesco de indivíduos não genotipados na predição dos valores genéticos. Ambos os métodos, single-step e pedigree, reportaram o mesmo valor de herdabilidade, sendo este 0,13, porém o erro-padrão foi de 0,03 e 0,06, respectivamente para os dois métodos descritos. A avaliação genética intra-rebanho realizado neste estudo incluindo informações genômicas simultaneamente com avaliações tradicionais baseadas em pedigree (metodologia denominada *single-step*) apresentou desempenho superior às demais metodologias testadas em relação ao aumento médio da acurácia e redução do viés de predição. Dentre os métodos Bayesianos de seleção genômica, o Bayes B superou os demais, porém o mesmo não é recomendado devido à necessidade de análises two-step, o que torna a avaliação genética onerosa em termos de tempo de execução das análises estatísticas, e também por ter gerado resultados de predição inferiores ao método *single-step*.

ABSTRACT

GARCIA, Adriano, M.Sc., Universidade Federal de Viçosa, July, 2019. **Genomic selection implementation for age at first calving on Nelore females.** Adviser: Henrique Torres Ventura.

The objective was to use genomic selection (via two methodologies: single-step and combined selection index), and traditional pedigree-based analysis for an intra-herd genetic evaluation considering the trait age at first calving (AFC) in Nelore females. The results were compared considering both prediction accuracy and bias of the estimated breeding values. Variance components and heritabilities for AFC were also calculated. Regarding the combined selection index analysis, we used several statistical methods that reflects on covariates selection (multicollinearity problems) and regularization of estimation process (dimensionality problem). The used methods used were: RR-BLUP, Bayes A, Bayes B and Bayes LASSO. Phenotypic (AFC) and genotypic data from 714 animals were used, and the pedigree included a total of 4,133 animals. All the methods were outperformed by the single-step methodology; and among the two-step methods, the best results were showed by Bayes B. The better performance of single-step method can be explained by the fact that it absorbs relationship information from non-genotyped individuals in prediction of breeding values. Both single-step and pedigree methods reported the same heritability value, which was 0.13, but the standard error was 0.03 and 0.06, respectively for both described methods. The intra-herd genetic evaluation performed in this study, which included genomic information simultaneously with traditional pedigree-based evaluations (methodology called single-step), performed better than the other methodologies tested in relation to the average increase in accuracy and reduction of prediction bias. Among the Bayesian methods of genomic selection, Bayes B outperformed the others, but, it is not recommended due to the need for two-step analysis, which makes genetic evaluation costly in terms of statistical analysis execution time, and also by have provided lower prediction results than single-step method.

SUMÁRIO

1- INTRODUÇÃO

O Brasil se destaca no mercado mundial como um dos maiores produtores e exportadores de carne bovina, e este sucesso se deve principalmente ao alto desempenho da raça Nelore. Embora nos últimos anos a utilização de inovações tecnológicas levou a raça a atingir importantes marcas de produtividade em relação à quantidade de carne produzida, os aspectos reprodutivos ainda deixam a desejar, pois as baixas taxas de concepção e prenhez ainda se caracterizam como gargalos para o pleno sucesso da bovinocultura de corte. Melhorias têm sido propostas principalmente na área de manejo reprodutivo, a qual requer alto investimento com mão de obra e infraestrutura. Assim, uma alternativa viável é o investimento em programas de melhoramento genético visando o aumento da performance reprodutiva de machos e fêmeas por meio de cruzamentos específicos entre os indivíduos geneticamente superiores.

Devido ao alto investimento com alimentação e manejo de uma novilha até o seu primeiro parto, é recomendável adicioná-la ao rebanho de cria tão logo esteja fisiologicamente disponível (MARQUES, 2018). Normalmente, os criadores verificam a precocidade sexual nas fêmeas indiretamente pela idade ao primeiro parto (IPP), uma vez que esta é uma característica de baixo custo e de fácil mensuração (MARTIN et al.,1992; GRESSLER, 2004; VIEIRA et al., 2010; PIRES et al., 2015). Em termos gerais, a redução da IPP aumenta a lucratividade da atividade devido à redução de custos de manutenção de novilhas improdutivas, antecipa a idade produtiva das vacas, possibilita maior intensidade de seleção nas fêmeas e reduz o intervalo entre gerações, possibilitando maiores ganhos genéticos (MARQUES, 2018).

Embora a IPP demonstre considerável relevância econômica, suas estimativas de herdabilidade variam de baixas a moderadas, 0,05 a 0,26 (GRESSLER, 2004), de forma que as baixas herdabilidades geralmente são consequência do manejo e dos critérios adotados para escolha das novilhas aptas à reprodução, principalmente o peso corporal.

Portanto, investir em programas de melhoramento visando o aumento do desempenho reprodutivo por meio de identificação de indivíduos geneticamente superiores e respectivos acasalamentos entre os mesmos, é uma alternativa viável para melhorar geneticamente a IPP. Porém, o sucesso de tais programas depende de predições acuradas dos valores genéticos para que se possa realmente ter alta confiabilidade na identificação dos melhores indivíduos sob o ponto de vista genético. Neste sentido, visto que as características reprodutivas em animais jovens apresentam baixas acurácias devido à pouca precisão na obtenção dos fenótipos (ainda não têm filhos devido à pouca idade), e por esses indivíduos não possuírem descendentes diretos avaliados, a seleção genômica (SG) se justifica por permitir a identificação precoce destes animais com potencial para serem futuros reprodutores da raça Nelore.

De forma geral, a utilização da SG apresenta como principal vantagem o aumento da acurácia na avaliação genética de animais (MEUWISSEN et al., 2001) para características cuja expressão fenotípica é muito influenciada por fontes de variação não controladas, como a IPP, e de indivíduos não fenotipados e/ou que não apresentam (ou apresentam poucos) descendentes avaliados, como aquelas restritas ao sexo (tal como IPP).

Mesmo que vários projetos tenham sido propostos no Brasil com o objetivo de utilizar a SG para aumentar as acurácias das avaliações genéticas em gado de corte, a maioria deles investem em comparações de metodologias de predição de valores genéticos genômicos (GEBVs) e poucos abordam realmente como incorporar estes resultados em avaliações genéticas combinando resultados da SG com aqueles das avaliações genéticas tradicionais a fim de se constatar na prática o real aumento da acurácia de predição dos valores genéticos. Dentre as possíveis formas de combinar estes resultados baseados na avaliação genética genômica destacam-se o single-step (LEGARRA et al., 2009) e o índice de seleção combinado (VanRaden et al., 2009).

Diante do exposto, o presente trabalho vai ao encontro dos interesses da bovinocultura de corte do país no que diz respeito à seleção precoce de indivíduos geneticamente superiores para características reprodutivas, neste caso a IPP. Esta seleção irá contribuir para o sucesso da raça Nelore, e

consequentemente da pecuária nacional, uma vez que pelo menos 80% do rebanho brasileiro possui algum grau de composição genética zebuína e que a eficiência reprodutiva da fêmea tem um impacto econômico considerável nos sistemas de produção. Além disso, observa-se ainda uma demanda cada vez maior de profissionais aptos a analisarem dados na área de SG, uma vez que no Brasil grandes investimentos científicos já foram realizados para a obtenção de tais dados em rebanhos experimentais e comerciais. Assim, o treinamento de profissionais da área de melhoramento genético de bovinos de corte em modernas técnicas de análise de dados desta natureza é imprescindível para o sucesso da SG em nosso país.

Tendo em vista as informações apresentadas, objetivou-se utilizar a SG (via duas metodologias: single-step e o índice de seleção combinado) e a análise tradicional baseada em pedigree para uma análise intra-rebanho considerando a característica idade ao primeiro parto em fêmeas Nelore. Os resultados foram comparados tendo em vista as acurácias dos valores genéticos preditos bem como as estimativas dos componentes de variância e herdabilidade. Em resumo, a justificativa geral para o presente projeto reside na importância de se expandir o nível tecnológico das avaliações genéticas intra-rebanho via inclusão de informações genômicas, principalmente para características reprodutivas. Destaca-se ainda a importância da formação de recursos humanos por parte do grupo de técnicos da ABCZ, visto que a área de Genética e Melhoramento Animal da UFV vem realizando pesquisas de alto nível em SG aplicada às características reprodutivas de bovinos de corte.

2- MATERIAL E MÉTODOS

2.1 Dados fenotípicos, genotípicos e de pedigree

Foram utilizados dados fenotípicos (IPP) e genotípicos de 714 animais, oriundos do projeto de pesquisa aprovado pela CAPES (5/2013/CBE/CGBP/DRI/CAPES) intitulado "*Genomic selection for reproductive traits in Nelore cattle"* e disponibilizados pelo grupo de pesquisa da UFV com o intuito de avaliar a inclusão dessas informações nas avaliações genéticas de características reprodutivas do Programa de Melhoramento Genético de Zebuínos da Associação Brasileira dos Criadores de Zebu - ABCZ. A coleta de material biológico para extração do DNA foi realizada conjuntamente por alunos da UFV e técnicos da Fazenda Derribadinha (Carlos Chagas-MG), os quais foram treinados para executar coletas de pelo e armazenamento em kits apropriados.

Uma vez coletado o material biológico, estes foram enviados para empresas terceirizadas para realizar a genotipagem via SNPchip customizado de 70K. Após a análise de controle de qualidade de marcadores ("call rate">0,95 e MAF>0,05), restaram 57.053 SNPs, os quais foram utilizados nos métodos para predição dos GEBVs.

Informações de pedigree cedidas pela ABCZ foram inicialmente compostas por 6.341 indivíduos. Após análises de qualidade de pedigree ("pruning"), na qual são removidos indivíduos não aparentados e sem informações fenotípicas, restaram um total de 4.133 animais.

2.2 Análise tradicional

Foram realizadas análises tradicionais baseadas em BLUP a fim de predizer os valores genéticos dos indivíduos considerando informações fenotípicas e genealógicas apresentadas no item 2.1.

O modelo utilizado na análise tradicional foi dado por:

IPP = GC + IDADE + IDADE^2 + CLASSEVACA + EGD + ER,

em que: GC = Grupo de contemporâneo (FIXO), IDADE = Idade do animal (COVARIÁVEL), IDADE^2 = Idade do animal ao quadrado (COVARIÁVEL), CLASSEVACA/SEXO = Classe de idade da mãe ao parto e sexo do filho (FIXO), CLASSEVACA = Classe de idade da mãe ao parto (FIXO), EGD= Efeito genético direto, e ER = Efeito residual. Para o ajuste do modelo foi utilizado a família de programas BLUPF90 (http://nce.ads.uga.edu/~ignacy/newprograms.html)

Neste modelo foi assumido que EGD ~ N(0, $\mathbf{A}\sigma^2_a$) e ER ~ N(0,$\mathbf{I}\sigma^2_e$), sendo **A** à matriz de parentesco tradicional entre os indivíduos e **I** uma matriz identidade.

2.3 Análise genômica

A utilização da grande quantidade de informações genômicas (57.053 SNPs) ainda é um desafio, pois na maioria das vezes não é possível estimar livremente o efeito de cada SNP sobre o fenótipo devido a problemas de multicolinearidade (diferentes marcadores com o mesmo perfil genotípico) e de dimensionalidade (número de marcadores muito maior que o número de animais genotipados, ou seja, o número de parâmetros a serem estimados é muito maior que o número de observações). De acordo com Gianola et al. (2006), tal situação demanda a utilização de métodos estatísticos que considerem a seleção de covariáveis (problema de multicolinearidade) e a regularização do processo de estimação (problema de dimensionalidade). Nestes métodos, os fenótipos considerados foram os corrigidos para todos os efeitos fixos a fim de facilitar a análise de validação-cruzada para comparação das metodologias. O modelo geral para seleção genômica segundo Meuwissen et al. (2001) foi dado por:

$$\mathbf{y} = \mathbf{1}\mu + \sum_i \mathbf{x}_i g_i + \mathbf{e}, \qquad [1]$$

em que: **y** é o vetor de fenótipos corrigidos, **1** é o vetor de mesma dimensão de **y** com todas as entradas iguais a 1, μ é a média da característica estudada, g_i é o efeito do marcador SNP (i=1,2,...,p), $\mathbf{x}_i$ é matriz de incidência

de cada marcador i, e **e** é o vetor de resíduos do modelo.

Os seguintes métodos foram testados, como descrito nos tópicos abaixo.

2.3.1 Método RR-BLUP/GWS

Nesta abordagem, o modelo [1], o qual é representado matricialmente em [2], admite a pressuposição de que os efeitos de marcadores são considerados aleatórios, com distribuição normal e variância homogênea. Esta variância, bem como, a variância residual, são consideradas desconhecidas e podem ser estimadas juntamente com os efeitos dos marcadores mediante resolução das equações de modelos mistos via método da Máxima Verossimilhança Restrita (REML). O seguinte modelo linear misto geral é ajustado para estimar os efeitos dos marcadores

$$\mathbf{y} = \mathbf{Wb} + \mathbf{Xg} + \mathbf{e}, \qquad [2]$$

em que: $\mathbf{y}$ é um vetor de observações fenotípicas corrigidas, $\mathbf{b}$ é um vetor de efeitos fixos (ao usar fenótipos fcd corrigidos para efeitos fixos se reduz a um vetor unitário), $\mathbf{g}$ é o vetor de efeitos dos marcadores assumidos como aleatórios e $\mathbf{e}$ se refere ao vetor de erros aleatórios. $\mathbf{W}$ e $\mathbf{X}$ são as matrizes de incidência para $\mathbf{b}$ e $\mathbf{g}$, respectivamente. A matriz de incidência $\mathbf{X}$ contém os valores 0, 1 e 2, respectivamente para aa, aA e AA. .Assim, as equações de modelo misto para a predição de $\mathbf{g}$ por meio do método RR-BLUP/GWS equivalem a:

$$\begin{bmatrix} \mathbf{W'W} & \mathbf{W'X} \\ \mathbf{X'W} & \mathbf{X'X} + \mathbf{I}\dfrac{\sigma_e^2}{(\sigma_g^2 / n)} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{W'y} \\ \mathbf{X'y} \end{bmatrix},$$

em que: σ_g^2 se refere a variância genética da característica, σ_e^2 a variância residual e *n* é função do número total de marcadores ponderados por suas frequências alélicas, sendo dado por $n=2\sum_{i=1}^{p} q_i(1-q_i)$, em que, q_i é a frequência do alelo *i*.

Neste método é considerado que cada loco explica $(1/n)\sigma_g^2$, ou seja, partes iguais da variância genética são atribuídas a todos os locos. Além disso, o valor direto genômico (DGV – *direct genomic breeding values*) do indivíduo j foi dado por: $DGV_j=\hat{y}_j=\hat{\mu}+\sum_i x_{ij}\hat{g}_i = \hat{\mu} + x_{1j}\hat{g}_1 + x_{2j}\hat{g}_2 + \cdots + x_{pj}\hat{g}_p$.

2.3.2 Método Bayes A

Meuwissen et al. (2001) apresentam também uma metodologia para estimar por abordagem Bayesiana os parâmetros do modelo [1] extendido, no qual diferentes componentes de variância σ_{gi}^2 são atribuídos para cada marcador considerado na análise. Considerou-se, em um segundo nível hierárquico, um modelo para σ_{gi}^2 com o intuito de realizar inferências a respeito destes parâmetros. Os autores assumiram, como densidades *a priori,* distribuição normal para os efeitos de marcadores, uma constante para a média geral e distribuição qui-quadrado invertida com parâmetro de escala para a variância residual. Uma distribuição da mesma família desta última foi considerada *a priori* para a variância do efeito de cada segmento, com o intuito de que a distribuição *a posteriori* de cada variância seja uma combinação daquela com as informações contidas nas observações.

As distribuições utilizadas na construção da densidade *a posteriori* conjunta resultam em condicionais completas a posteriori com forma conhecida, o que possibilita a utilização de amostrador de Gibbs para gerar amostras da densidade conjunta *a posteriori* (e por consequência, das marginais *a posteriori* de interesse). Ao final do processo MCMC, obteve-se as estimativas dos efeitos de cada marcador, e consequentemente as estimativas dos valores diretos genômicos de cada animal: $DGV_j=\hat{y}_j=\hat{\mu}+\sum_i x_{ij}\hat{g}_i = \hat{\mu} + x_{1j}\hat{g}_1 + x_{2j}\hat{g}_2 + \cdots + x_{pj}\hat{g}_p$.

2.3.3 Método Bayes B

Adicionalmente aos modelos já apresentados, Meuwissen et al (2001) desenvolveram uma abordagem Bayesiana alternativa. Os autores

reconheceram como um problema no método Bayes A o fato de que a distribuição das variâncias dos efeitos de marcadores, não apresentavam uma massa de densidade no valor 0. Esta característica seria interessante para esta distribuição, uma vez que a maior parte dos segmentos não apresenta variância genética (não apresentam segregação). O método Bayes B utiliza densidade *a priori* com massa de densidade em σ^2_{gi} =0. Considera-se que σ^2_{gi} =0 com probabilidade π, enquanto σ^2_{gi} ~ *inv* χ^2 (u, *S*) com probabilidade $1-\pi$.

Considerando **y** como vetor de observações livre de efeitos da média e efeitos genéticos com exceção do marcador *i*, a solução para a amostragem de g_i e σ^2_{gi} pode ser feita por meio de $p(\sigma^2_{gi}, g_i \mid \mathbf{y}) \sim p(\sigma^2_{gi} \mid \mathbf{y})\, p(g_i \mid \sigma^2_{gi}, \mathbf{y})$, de modo que a amostragem de σ^2_{gi} não seja função de g_i. Entretanto, os autores não obtiveram $p(\sigma^2_{gi} \mid \mathbf{y})$ de forma fechada, ou seja, como sendo uma distribuição de probabilidade conhecida. Assim, foi necessária a utilização do algoritmo Metropolis-Hastings para obter amostras de $p(\sigma^2_{gi} \mid \mathbf{y})$. As estimativas dos valores diretos genômicos de cada animal também foi obtida mediante estimativas de efeitos de marcadores: $DGV_j = \hat{y}_j = \hat{\mu} + \sum_i x_{ij}\hat{g}_i = \hat{\mu} + x_{1j}\hat{g}_1 + x_{2j}\hat{g}_2 + \cdots + x_{pj}\hat{g}_p$.

2.3.4 Método LASSO Bayesiano

Conforme já apresentado por Meuwissen et al. (2001), a regressão Bayesiana pode ser utilizada nas situações em que se tem mais marcadores (covariáveis) do que observações, uma vez que determinadas distribuições *a priori* impõem regularização no ajuste do modelo, sob forma de encurtamento dos coeficientes de regressão (*shrinkage*).

Uma forma interessante de executar este encurtamento é por meio da regressão LASSO (*Least Absolute Shrinkage and Selection Operator*), a qual combina seleção de variáveis e regularização via encurtamento dos coeficientes de regressão. A versão Bayesiana da regressão LASSO para seleção genômica foi idealizada por de los Campos et al. (2009). De forma geral, esta consiste na obtenção de estimadores de coeficientes de regressão do modelo [1] que resolvam o seguinte problema de otimização:

$\min\{[\mathbf{y} - (\mathbf{1}\mu + \sum_i \mathbf{x}_i g_i)]'[\mathbf{y} - (\mathbf{1}\mu + \sum_i \mathbf{x}_i g_i)] + \lambda \sum_i^p |g_i| \}$, em que $\sum_i^p |g_i|$ é a soma dos valores absolutos dos coeficientes de regressão e λ é o parâmetro que controla a força da regularização, de forma que quando λ = 0 não há regularização.

Na implementação Bayesiana do LASSO (de Los Campos et al., 2009) impõe-se como distribuição marginal *a priori* dos p coeficientes de regressão um produto de densidades exponenciais duplas: $p(\mathbf{g}|\lambda) = \prod_{i=1}^{p} \frac{\lambda}{2} \exp(-\lambda|g_i|)$. Por sua vez, os métodos Bayes A e B utilizam distribuição normal: $p\left(\mathbf{g}\middle|\sigma_{gi}^2\right) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\sigma_{gi}^2}} \exp\left(-\frac{g_i}{2\sigma_{gi}^2}\right)$. As estimativas dos valores diretos genômicos também foram obtidas por meio da expressão: $\mathrm{DGV}_j = \hat{y}_j = \hat{\mu} + \sum_i x_{ij}\hat{g}_i = \hat{\mu} + x_{1j}\hat{g}_1 + x_{2j}\hat{g}_2 + \cdots + x_{pj}\hat{g}_p$.

2.4 Recursos computacionais utilizados nas análises SG

Os métodos dos itens 2.3.1 a 2.3.4 foram implementados no software R (R Development Core Team, 2018). Na Tabela 1 são mostrados os pacotes e as funções usadas na implementação de cada um dos métodos propostos.

Tabela 1. Métodos propostos e respectivas ferramentas de implementação no software R.

Métodos	Pacote	Função
RR-BLUP/GWS	rrBLUP	*mixed.solve (Ridge regression-BLUP)*
Bayes A, Bayes B e LASSO Bayesiano	*BGLR*	*BGLR (Bayesian Generalized Linear Regression)*

2.5 Comparações das metodologias SG (Seleção Genômica)

Os métodos apresentados nos itens 2.3.1 a 2.3.4 foram comparados por meio da análise de acurácia via validação cruzada com três populações oriundas da partição aleatória da população original de 714 animais. Assim, cada sub-população foi composta por 238 animais. Em cada repetição, uma sub-população foi removida do conjunto de dados para compor a população de

validação e as outras duas utilizados na estimação dos valores genômicos na população de treinamento. Assim, uma vez estimados todos os efeitos de marcadores ($\hat{g}_1, \hat{g}_2, ..., \hat{g}_p$) na população de treinamento, estes foram aplicados na população de validação para predizer o valor direto genômico ($DGV_j=\hat{y}_j$) para cada indivíduo da população de validação. Dessa forma, ao final da análise, foram obtidos três valores de correlação (entre observados e preditos), cuja média representou a acurácia de cada método testado.

2.6 Combinando a análise tradicional com SG

Foram comparadas também duas estratégias de se incorporar as informações genômicas nos resultados da avaliação genética tradicional (item 2.2), uma vez que esta combinação de informações gerou o GEBV e sua respectiva acurácia. Tal abordagem possibilitou uma comparação real com os resultados provenientes apenas da análise tradicional (baseada em pedigree). Tal comparação foi efetuada por meio da verificação de qual delas apresentou maior valor de acurácia (tal como descrito no item 2.5) em relação à análise tradicional.

2.6.1 Índice de seleção combinado (ou *two-step*)

VanRaden et al. (2009) propuseram o cálculo do GEBV como sendo um índice de seleção combinando resultados das análises tradicionais e genômicas, o qual é dado por:

$$GEBV=b_1*DGV+b_2*PI_s+b_3*PI_t, \quad [3]$$

em que:

GEBV é o valor genético genômico;

DGV é o valor direto genômico obtido da análise GWS considerando apenas os indivíduos genotipados (será utilizado apenas DGV proveniente do melhor método entre aqueles apresentados nos itens 3.4.1 a 3.4.5)

PI_t é o valor genético predito pela análise tradicional ao se considerar todos os indivíduos da população (neste caso o número total de animais especificados no item 2.2);

PI_s é o valor genético predito pela análise tradicional considerando apenas informação de parentesco entre os animais genotipados (714 animais),

b_1, b_2 e b_3 são os pesos atribuídos a cada fator que serão estimados via sistema de equação apresentado a seguir.

O seguinte sistema linear foi utilizado para a obtenção dos pesos:

$b_1*V_{11} + b_2*V_{12} + b_3*V_{13} = V_{11}$,

$b_1*V_{12} + b_2*V_{22} + b_3*V_{23} = V_{22}$,

$b_1*V_{13} + b_2*V_{23} + b_3*V_{33} = V_{33}$,

em que:

V_{11}, V_{22}, e V_{33} são, respectivamente, a confiabilidade (acurácia ao quadrado, r^2) de DGV, PI_t e PI_s.

De acordo com VanRaden et al. (2009), define-se $V_{12} = V_{22}$, $V_{23} = V_{22}$ e $V_{13} = V_{22} + (V_{11} - V_{22})\ (V_{33} - V_{22})/(1 - V_{22})$.

Os valores de confiabilidade para o DGV (V_{11}) foram obtidos de acordo com Kachman (2008) considerando cada um dos métodos apresentados nos itens 2.3.1 a 2.3.4. Uma vez resolvido o sistema linear em questão por meio do pacote *deSolve* do software R, os valores obtidos para b_1, b_2 e b_3 foram utilizados em [3] para se obter o GEBV.

2.6.2 Método *Single Step*

Legarra et al. (2009) propuseram uma abordagem baseada no modelo misto tradicional na qual a matriz de co-variância dos efeitos aleatórios (denominada de matriz **H**) contempla simultaneamente informações de parentesco baseadas em pedigree (matriz **A** tradicional) e baseadas em marcadores SNPs (matriz de parentesco genômico, **G**). A matriz **G** é dada por: $\mathbf{G}=\mathbf{MM}'/2\sum_i q_i(1-q_i)$ VanRaden et al. (2009), sendo **M** a, matriz de genótipos

(N linhas e p colunas, em que N é o número de animais genotipados e p é o número de marcadores) e q_i a menor frequência alélica de cada marcador i.

Dessa forma, o modelo denominado de *single-step* é dado por:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zu} + \mathbf{e},$$

em que:

y é o vetor de fenótipos corrigidos (de todos os indivíduos sob análise) ;

μ representa a média geral;

u representa diretamente o vetor GEBV, assumindo que $\mathbf{u} \sim N(0, \mathbf{H}\sigma_u^2)$;

e representa o vetor de resíduos, assumindo que $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$;

A matriz **H** é definida por:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12} & \mathbf{G} \end{bmatrix} = \mathbf{A} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G} - \mathbf{A}_{22} \end{bmatrix},$$

em que: $\mathbf{A}_{11}$ é a matriz de parentesco tradicional entre os indivíduos não genotipados, $\mathbf{A}_{12}$ é a matriz de parentesco tradicional entre os indivíduos genotipados e não genotipados, $\mathbf{A}_{22}$ é a matriz de parentesco tradicional entre os indivíduos genotipados e **G** é a matriz de parentesco genômico entre indivíduos genotipados. Esta metodologia foi utilizada por meio do software AIREMLF90 com a função PREGSF90.

A grande vantagem deste método é que os valores de GEBV e suas acurácias são obtidos diretamente do sistema de equações do modelo misto de Henderson, não havendo necessidade de separar em dois passos distintos (análises GWS e tradicional). Porém, neste enfoque é assumido que todos os SNPs explicam a mesma fração da variância genética aditiva, o que apenas faz sentido para características de caráter extremamente poligênico. Por outro lado, o método de índice de seleção (item 2.6.1) quando considera os métodos Bayes A e B e LASSO Bayesiano, assume que cada SNP explica uma fração particular da variância genética aditiva, o que pode apresentar vantagens para características nas quais alguns *loci* apresentam maior relevância que os demais.

3- RESULTADOS E DISCUSSÃO

3.1 Resultados das análises SG (Seleção Genômica)

Na Tabela 2 são apresentados os resultados de acurácia e viés para a característica IPP. Tais resultados foram obtidos conforme descrição metodológica do item 2.5, no qual é informado que a correlação entre os valores fenótipos corrigidos (y) e os valores diretos genômicos (DGV) é utilizada como medida de acurácia de predição. A medida de viés que também compõe a tabela em questão foi obtida por meio do coeficiente de regressão linear de DGV em função de y.

Também é relevante comentar que os desvios-padrão (Tabela 2) foram obtidos porque considerou-se a validação cruzada com três sub-populações, de forma que as acurácias e viés foram calculados usando três repetições. Estes desvios foram similares para todos os valores calculados, indicando que as três sub-populações obtidas aleatoriamente não foram efetivamente diferentes em termos genéticos.

Tabela 2. Acurácia de predição e viés (com os respectivos desvios-padrão entre parênteses) provenientes das análises de seleção genômica (realizadas via diferentes métodos) para a característica idade ao primeiro parto (IPP) em fêmeas Nelore.

	Métodos Bayesianos							
Estatística	RR-BLUP		Bayes A		Bayes B		Bayes LASSO	
Acurácia	0,40	(0,15)	0,40	(0,13)	**0,43**	**(0,12)**	0,41	(0,14)
Viés	0,15	(0,10)	0,15	(0,11)	**0,13**	**(0,11)**	0,16	(0,12)

Em relação aos métodos comparados, nota-se que o Bayes B (BB) superou de forma discreta os demais métodos em termos de acurácia e viés. Este tem sido o método Bayesiano mais utilizado em seleção genômica, talvez por ter sido indicado como o mais apropriado no trabalho seminal de Meuwissen et al. (2001). No trabalho em questão, os autores justificaram a superioridade do método Bayes B em termos de sua maior flexibilidade para acomodar diferentes arquiteturas genéticas de diferentes características de interesse econômico para o melhoramento animal.

Embora sejam escassos estudos contemplando a avaliação de predições genômicas para características reprodutivas em gado Nelore, algumas pesquisas têm reportado valores animadores de acurácias, principalmente quando comparadas com a metodologia tradicional. Boddhireddy et al. (2014) obtiveram, via método Bayes C, estimativa de acurácia para IPP igual a 0,64 considerando uma população de 2.241 animais Nelore genotipados via tecnologia BovineSNP50. Estes mesmos autores reportaram acurácias para Stayability variando entre 0,58 e 0,59. Por outro lado, Zhang et al. (2014) reportaram acurácia de 0,33 para IPP via GBLUP em uma população Brahman (4.782 animais) genotipada via tecnologia Illumina 7K. Trabalhando com dados de fêmeas Nelore, Mota et al. (2018) reportaram acurácia de predição de 0,38 para IPP usando o método RR-BLUP.

3.2 Resultados das análises combinando SG com o método tradicional

Conforme descrito no tópico Material e Métodos (item 2.6), também foram comparadas duas estratégias (índice de seleção, ou *two-step*, e *single-step*) de se incorporar as informações genômicas nos resultados da avaliação genética tradicional baseada em pedigree. Isto possibilitou uma comparação real (em termos de acurácia e viés) com os resultados provenientes apenas da análise tradicional (baseada em pedigree). Na Tabela 3 são apresentados os resultados de acurácia e viés para a característica IPP obtidos pelos métodos descritos.

Tabela 3. Acurácia de predição e viés (com os respectivos desvios-padrão entre parênteses) provenientes das análises *two-step* e *single-step* para a característica idade ao primeiro parto (IPP).

Métodos*	Acurácia	Viés	Métodos*	Acurácia	Viés
Pedigree	0,41 (0,11)	0,89 (0,16)	*Two-step* BA	0,47 (0,16)	0,95 (0,19)
Single-step	0,51 (0,13)	0,03 (0,15)	*Two-step* BB	0,48 (0,19)	0,04 (0,19)
Two-step RR	0,46 (0,17)	0,95 (0,18)	*Two-step* BL	0,47 (0,19)	0,95 (0,19)

Bayes A (BA), Bayes B (BB), Bays LASSO (BL)

Nota-se na Tabela 3 que a inclusão das informações genômicas aumentou a acurácia e reduziu o viés. O método *single-step* superou todos os outros, e dentre os métodos *two-step*, o BB se sobressaiu em relação aos demais tal como observado nas análises SG (Tabela 2) considerando apenas informações de indivíduos genotipados.

A melhor performance do método single-step pode ser explicada pelo fato do mesmo absorver informações de parentesco de indivíduos não genotipados na predição dos valores genéticos. Em relação à análise tradicional, o ganho de acurácia ao se utilizar informações genômicas já foi discutido no parágrafo anterior. Porém, em relação aos métodos two-step, esta vantagem é reduzida, uma vez que estes métodos contemplam a quantidade PIt (valor genético predito pela análise tradicional ao se considerar todos os indivíduos da população) no índice. De certa forma, tal informação também permite absorver informações de todos os indivíduos da população, porém o método single-step é mais efetivo para explorar esta propriedade (Legarra et al., 2009).

Ainda em relação aos métodos two-step, ressalta-se que a vantagem dos mesmos em relação ao método single-step são as pressuposições realizadas a respeito dos efeitos dos marcadores, uma vez que nos métodos Bayes A, Bayes B, Bayes LASSO e RKHS tais marcadores apresentam variâncias individuais, ao passo que no single-step e no RR-BLUP somente uma variância é assumida para todos os marcadores (ou seja, assume-se que a característica seja poligênica, de forma que cada marcador explica a mesma fração da variância genética aditiva total). De forma geral, dado a melhor performance do método single-step, podemos inferir indiretamente que a IPP pode ser descrita como poligênica.

3.3 Estimativas de herdabilidade

Visto que o método *single-step* superou os demais nas comparações de acurácias, o mesmo foi utilizado para a estimação de herdabilidade (via REML) para IPP. Tal estimação também foi realizada via método tradicional (Pedigree) a fim de verificar se a inclusão de informações genômicas reflete em aumento da herdabilidade.

Ambos os métodos, single-step e pedigree, reportaram o mesmo valor de herdabilidade, sendo este 0,13, porém o erro-padrão foi de 0,03 e 0,06, respectivamente para os dois métodos descritos. Em termos gerais, a inclusão de informações genômicas geralmente não afeta as estimativas de parâmetros genéticos, sendo que o aumento do ganho genético via seleção genômica (Garrick, 2011) está relacionado ao aumento da acurácia (no numerador) e a redução do intervalo de geração (no denominador). Porém, é importante mencionar que o erro-padrão obtido via *single-step* foi quase a metade daquele obtido via método tradicional para todas as características estudadas.

Em relação às estimativas obtidas, ressalta-se que todas elas estão inseridas dentro do intervalo de confiança calculado para a raça (Oliveira et al., 2017). Tais intervalos foram obtidos via meta-análise utilizando centenas de estudos de estimação de parâmetros genéticos na raça Nelore, e desta forma, pode ser utilizado como referência para validação de estimativas de herdabilidade.

4- CONCLUSÕES

A avaliação genética intra-rebanho realizado neste estudo incluindo informações genômicas simultaneamente com avaliações tradicionais baseadas em pedigree (metodologia denominada *single-step*) apresentou desempenho superior às demais metodologias testadas em relação ao aumento médio da acurácia e redução do viés de predição. Dentre os métodos Bayesianos de seleção genômica, o Bayes B superou os demais, porém o mesmo não é recomendado devido a necessidade de análises two-step, o que torna a avaliação genética onerosa em termos de tempo de execução das análises estatísticas, e também por ter gerado resultados de predição inferiores ao método *single-step*.

5- REFERÊNCIAS BIBLIOGRÁFICAS

BODDHIREDDY, P. K PRAYAGA, P BARROS, R LÔBO, S DENISE. Genomic Predictions of Economically Important Traits in Nelore Cattle of Brazil. **10th World Congress on Genetics Applied to Livestock Production**. 2014. Vancouver, Canada.

de los CAMPOS, G., NAYA, H., GIANOLA, D. et al. Predicting quantitative traits with regression models for dense molecular makers. **Genet.** 182: 375-385, 2009.

GARRICK, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. **Genet. Sel. Evol.** 15, 43:17.

GARRICK, D. J., TAYLOR, J. F., FERNANDO, R. L. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genet Sel Evol**. v.41(1).

GIANOLA, D., FERNANDO, R. L., STELLA, A. Genomic assisted prediction of genetic value with semi-parametric procedures. **Genet.**, v. 173, p. 1761-1776, 2006.

GRESSLER, S. L. **Fatores ambientais e genéticos do perímetro escrotal e da idade ao primeiro parto em novilhas Nelore desafiadas tradicional ou precocemente**. Belo Horizonte: Universidade Federal de Minas Gerais, 2004. 139 p. Tese (Doutorado em Ciência Animal). Escola de Veterinária, 2004.

KACHMAN, S. Incorporation of marker scores into national genetic evaluations. In: **GENETIC PREDICTION WORKSHOP OF THE BEEF IMPROVEMENT FEDERATION, PREDICTION OF GENETIC MERIT OF ANIMALS FOR SELECTION**, 9., 2008, Kansas City. Proceedings... Kansas City, 2008.

LEGARRA, A., AGUILAR, I., MISZTAL, I. A relationship matrix including full pedigree and genomic information**. J. Dairy Sci**. v. 92, p. 4656–4663, 2009.

MARQUES, E. G. **Evolução Fenotípica dos Animais com Registro Genealógico na Associação Brasileira dos Criadores de Zebu**. Viçosa: Universidade Federal de Viçosa, 2018, 68p. Dissertação (Mestrado Profissional em Zootecnia). Universidade Federal de Viçosa, 2018.

MARTIN, L.C.; BRINKS, J.S.; BOURDON, R.M. et al. Genetic effects on beef heifer puberty and subsequent reproduction. **J. Anim. Sci.**, v. 70, p. 4006-4017,1992.

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

MOTA, R. R., SILVA, F. F., GUIMARÃES, S. E. F., et al. Benchmarking Bayesian genome enabled-prediction models for age at first calving in Nellore cows. **Liv. Sci.**, 211, 75-79, 2018.

OLIVEIRA, H. R., VENTURA, H. T., COSTA, E. V. et al. 2017. Meta-analysis of genetic-parameter estimates for reproduction, growth and carcass traits in Nellore cattle by using a random-effects model. **Anim. Prod. Sci.** http://dx.doi.org/10.1071/AN16712.

PIRES, A. V.; OLIVEIRA, D. C. F.; OLIVEIRA, L. T. et al. Precocidade Reprodutiva em Bovinos de Corte. **Cad. Ciênc. Agrá.**, v. 7, n. 1, p. 1-14, 2015.

R CORE TEAM (2018). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SARGOLZAEI, M., J. P. CHESNAIS AND F. S. SCHENKEL. 2014. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, 15:478 (DOI: 10.1186/1471-2164-15-478).

VANRADEN, P. Invited review: Reliability of genomic predictions for North American Holstein bulls. **J. Dairy Sci.**, v. 92, n. 1, p. 16–24. 2009.

VIEIRA, D.H., MEDEIROS, L. F. D., BARBOSA, C. G. et al. Efeitos não genéticos sobre as características reprodutivas de fêmeas da raça Nelore. II – idade à primeira parição e intervalo de parto. **Rev. Bras. Med. Vet.**, v.32, n.2, p.79-88, 2010.

ZHANG Y. D., JOHNSTON D. J., BOLORMAA S. et al. Genomic selection for female reproduction in Australian tropically adapted beef cattle. **Anim. Prod. Sci.**, v. 54, p. 16–24, 2014.