

XI Simpósio Brasileiro de Melhoramento Animal Santa Maria, RS – 07 e 08 de setembro de 2015

Acurácia das predições genômicas em Hereford e Braford a partir de genótipos imputados

Mario Luiz Piccoli¹, José Braccini², Fernando Flores Cardoso³, Medhi Sargolzaei⁴, Flávio Scharamm Schenkel⁴

¹GenSys Consultores Associados S/S, Porto Alegre/RS. e-mail: mariopiccoli@gensys.com.br

²Departamento de Zootecnia – UFRGS, Porto Alegre/RS. e-mail: jose.braccini@ufrgs.br

³Embrapa Pecuária Sul – Bagé/RS. e-mail: fernando.cardoso@embrapa.br

⁴Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada e-mail: schenkel@uoguelph.ca

Resumo: Diferentes percentagens de animais com genótipos imputados na população de treinamento foram utilizadas para investigar o efeito do uso destes genótipos imputados na acurácia dos valores genômicos diretos (DGV) em caracteres de importância econômica em animais Hereford e Braford. A população de treinamento foi formada pelos animais nascidos entre os anos de 2008 a 2011 e por 130 touros. A população de validação foi formada pelos animais nascidos em 2011. Valores genéticos desregredidos foram utilizados como pseudo fenótipos em um modelo GBLUP usando dois diferentes painéis de baixa densidade derivados a partir do painel de 50K (painel comercial de 8K - Beef LD GeneSeek Genomic Profiler e o painel não comercial de 15K - desenvolvido a partir do Beef LD GeneSeek Genomic Profiler). Estes dois painéis foram posteriormente imputados para o painel de 50K. A acurácia média de predição dos DGV na população de validação variaram entre 0,38 a 0,40. As perdas médias em acurácia dos valores genéticos genômicos (GEBV) em relação ao verdadeiro painel de 50K variaram entre -0,0002 a -0,0012. A percentagem de animais com genótipos imputados na população de treinamento não influenciou a acurácia dos DGV na população de validação enquanto que as perdas de acurácia dos GEBV foram menores quando foi utilizado o painel de 15K imputado para 50K em relação ao painel de 8K imputado para 50K.

Palavras-chave: Seleção genômica, Bovinos de corte, Imputação, Acurácia.

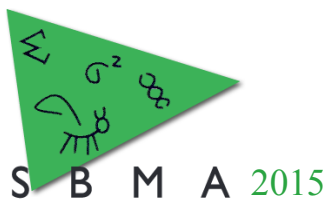
Accuracy of genomic prediction in Braford and Hereford beef cattle from imputed genotypes

Abstract: Alternate scenarios composed by different percentages of animals with imputed genotypes of the training population were generated to investigate the effect of the use of imputed genotypes in the accuracy of genomic selection in economic important traits in Brazilian Braford and Hereford beef cattle breeds. The training population was formed by animals born between 2008 to 2011 plus 130 sires and the validation population by animals born in 2011. Deregressed EBVs were used as pseudo phenotypes in a GBLUP model using two different mimicked panels derived from the 50K panel, the 8K panel (Beef LD GeneSeek Genomic Profiler) and the 15K panel (no commercial panel), which were subsequently imputed to the 50K panel. DGV accuracy in the prediction population averaged over traits ranged from 0.38 to 0.40 in the different scenarios. The average losses in GEBV accuracy relative to the actual 50K genotypes ranged from -0.0002 to -0.0012. The percentage of animals with genotypes imputed in the training population did not influence the DGV accuracy. The losses in GEBV accuracy were lower when using the 50K panel imputed from the 15K panel than from the 8K panel.

Keywords: Genomic selection, Beef cattle, Imputation, Accuracy.

Introdução

O melhoramento genético animal para características de importância econômica tem sido praticado ao longo dos anos com base em fenótipos e nas relações de parentesco entre os indivíduos. A incorporação dos marcadores moleculares em programas de melhoramento genético deverá resultar em ganhos substancialmente mais elevados em curtos período de tempo (Meuwissen et al., 2001). Estes ganhos estariam associados com a diminuição do intervalo entre gerações, com o aumento da acurácia da seleção e também com a possibilidade de incorporar novas características de importância econômica (VanRaden, 2008, Goddard et al., 2010). Particularmente, em bovinos de leite, esta nova tecnologia tem proporcionado economias financeiras substanciais em testes de progênie (Schaeffer, 2006). A acurácia do valor genômico direto (DGV), calculada a partir dos efeitos estimados de marcadores moleculares, depende de muitos fatores, tais como: a) do nível de desequilíbrio de ligação entre marcadores e QTL (locos de característica quantitativa); b) do número de animais na população de treinamento; c) da



XI Simpósio Brasileiro de Melhoramento Animal Santa Maria, RS – 07 e 08 de setembro de 2015

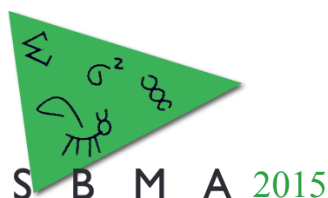
herdabilidade da característica; d) da distribuição dos efeitos do QTL. Os custos de genotipagem por animal é a principal barreira para a completa incorporação dos marcadores moleculares, em larga escala, nos programas de melhoramento genético animal. Os produtores, provavelmente, irão optar por usar em seus programas de melhoramento os painéis de menor densidade por possuírem custos mais acessíveis. Estes painéis de baixa densidade, seriam então, imputados para um painel de densidade média ou alta (Piccoli et al., 2014) e seriam usados para prever os DGV e GEBV. O objetivo deste estudo foi investigar a acurácia das predições genômicas quando da utilização de diferentes porcentagens de animais com genótipos imputados para 50K a partir dos painéis de 8K e 15K em um rebanho formado por animais Braford e Hereford.

Material e Métodos

Fenótipos e genótipos provenientes do Programa de Melhoramento da Conexão Delta G (Conexão Delta G, Dom Pedrito/RS, Brasil) foram utilizados neste estudo. O conjunto de fenótipos continha aproximadamente 520.000 animais provenientes de 97 fazendas localizadas nas regiões Sul, Sudeste, Centro-Oeste e Norte do Brasil. O conjunto de genótipos foi formado por 683 Hereford e 2.997 Braford nascidos entre 2008 e 2011 e por 130 touros. Os SNP do painel de 50K que apresentaram problema de genotipagem no laboratório ($=0.46\%$) foram previamente imputados. O controle de qualidade visando a imputação dos SNP incluiu: GenCall score (≥ 0.15); Call Rate (≥ 0.90); Equilíbrio Hardy-Weinberg ($P \geq 10^{-6}$) e utilização somente dos SNP dos cromossomos autossômicos. O controle de qualidade para as amostras incluiu: GenCall Score (≥ 0.15); Call Rate (≥ 0.90); desvio da heterozigose (± 3 SD); amostras repetidas; erros de paternidade. Os painéis de 8K e 15K foram imputados para o painel de 50K usando o software FImpute v.2.2 (Sargolzaei et al., 2014). Para estimação dos efeitos dos marcadores moleculares, além dos critérios acima mencionados, foi utilizado o critério de Minor Allele Frequency (≥ 0.05). A população de treinamento foi formada pelos animais nascidos entre 2008 e 2011 mais os 130 touros enquanto que a população de validação foi formada pelos animais nascidos em 2011. Cenários com diferentes porcentagens de animais com genótipos imputados (10%, 20%, 30%, 40%, 50%, 60%) foram analisados. Dois conjuntos de valores genéticos (EBV) foram preditos (modelo animal) utilizando um pacote escrito em linguagem Fortran, desenvolvido pela GenSys (GenSys Consultores Associados, Porto Alegre, Brasil). O primeiro, foi estimado utilizando todas as informações disponíveis, enquanto que no segundo foi estimado excluindo os dados dos animais nascidos em 2011. Estes dois conjuntos de EBV foram então usados para a validação e treinamento, respectivamente. Os EBV da população treinamento foram desregredidos e usado como pseudo fenótipos para estimação dos efeitos dos marcadores (VanRaden & Wiggans, 1991). Os DGV foram estimados usando o método GBLUP através do software GEBV (Sargolzaei, et al., 2009).

Resultados e Discussão

As acurácias médias de predição (correlação entre DGV e EBV) das características foram de 0,29 a 0,31 (Tabela 1). Não houveram diferenças significativas de acurácia na população de validação para as diferentes porcentagem de animais imputados ($P > 0,05$) em todas as características. As acurácias obtidas neste estudo apresentaram valores inferiores aos reportados por Saatchi et al. (2011) e Boddhireddy et al. (2014) ambos trabalhando com Angus nos Estados Unidos. Comparando com os resultados de Neves et al. (2014), que trabalharam com Nelore no Brasil e com o mesmo conjunto de características também encontraram maiores valores de acurácia do que os reportados neste estudo, exceto para ganho de peso e conformação pré desmama. Estes valores mais baixos de acurácia neste estudo também estão relacionados com os menores valores de acurácia dos EBV na população treinamento ($r=0,64$). As acurácias foram maiores para as características de maior herdabilidade (características de pós desmama). Estes resultados foram semelhantes aos encontrados por Brito et al. (2011), trabalhando com dados simulados em bovinos de corte, e por Khatkar et al. (2012), trabalhando com bovinos de leite na Austrália. A circunferência escrotal apresentou maior herdabilidade mas mostrou a menor acurácia, muito provavelmente devido ao menor número de animais na população de treinamento ($n=708$). As perdas em acurácia dos GEBV quando do uso do painel de 8K foram maiores na comparação ao painel de 15K. Estes resultados estão associados com a maior taxa de concordância de imputação no painel 50K imputado a partir do painel de 15K (Piccoli et al., 2014). Esse mesmo



XI Simpósio Brasileiro de Melhoramento Animal Santa Maria, RS – 07 e 08 de setembro de 2015

comportamento foi encontrado por Segelke et al. (2012) quando eles analisaram as perdas na confiabilidade de painéis imputados de duas densidades diferentes de SNP em bovinos de leite na Alemanha. Quando a porcentagem de animais com genótipos imputados na população de treinamento aumentava as perdas também aumentavam.

Tabela 1. Acurácia dos valores genômicos diretos (DGV) na população de validação¹²

Caracteres ³	8K ⁴						15K ⁴						
	10	20	30	40	50	60	10	20	30	40	50	60	50K ⁴
GD	0.33	0.33	0.32	0.33	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
GS	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
CD	0.29	0.29	0.29	0.29	0.29	0.29	0.28	0.28	0.28	0.28	0.29	0.29	0.29
CS	0.27	0.26	0.26	0.26	0.25	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
PD	0.31	0.31	0.31	0.31	0.31	0.32	0.31	0.30	0.30	0.31	0.31	0.31	0.31
PS	0.34	0.34	0.34	0.33	0.33	0.34	0.34	0.34	0.34	0.34	0.33	0.34	0.34
MD	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
MS	0.35	0.35	0.35	0.34	0.34	0.35	0.35	0.35	0.34	0.34	0.34	0.35	0.35
PE	0.18	0.19	0.18	0.19	0.19	0.18	0.18	0.19	0.19	0.19	0.19	0.19	0.18
Média	0.30	0.30	0.30	0.30	0.29	0.30	0.29	0.29	0.29	0.29	0.29	0.30	0.30

¹ A acurácia do DGV foi medida pela correlação de Pearson entre o DGV e o EBV na população de validação;

² Cenários com 10%, 20%, 30%, 40%, 50% e 60% dos animais com genótipos imputados na população de treinamento;

³ GD: ganho de peso pré desmama (kg); GS: ganho de peso pós desmama (kg); CD: conformação na desmama (escores 1-5); CS: conformação pós desmama (escores 1-5); PD: precocidade na desmama (escores 1-5); PS: precocidade pós desmama (escores 1-5); MD: musculatura na desmama (escores 1-5); MS: musculatura pós desmama (escores 1-5); PE: perímetro escrotal pós desmama (cm);

⁴ 8K: painel de 8K imputado para 50K; 15K: painel de 15K imputado para 50K; 50K: painel original de 50K.

Conclusões

A porcentagem de animais com genótipos imputados na população de treinamento não influenciou significativamente as acurácias dos DGV. As perdas em acurácia dos GEBV devido ao uso de genótipos imputados foram menores quando se utilizou o painel de 15K imputado para 50K em relação ao painel de 8K imputado para o painel de 50K.

Agradecimentos

Os autores agradecem a Conexão Delta G, a CAPES, ao CNPq e a Embrapa.

Literatura citada

- Brito, F. V.; Neto, J. B.; Sargolzaei, M.; Cobuci, J. A. and Schenkel, F. S. 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genetics* 12:80.
- Meuwissen, T. H.; Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Piccoli, M. L.; Braccini, J.; Cardoso, F. F.; Sargolzaei, M. and Schenkel, F. S. 2014. Accuracy of genome-wide imputation in Braford and Hereford beef cattle. *BMC Genetics* 15:157.
- Sargolzaei, M.; Chesnais, J. P. and Schenkel, F. S. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478.
- Sargolzaei, M.; Schenkel, F. S. and VanRaden, P. M. 2009. *gebv*: Genomic breeding value estimator for livestock. In: Technical report to the Dairy Cattle Breeding and Genetics Committee University of Guelph. Guelph.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91:4414–4423.