

Estudo de diferentes estratégias de genotipagem seletiva na predição de valores genômicos usando dados simulados

Arione Augusti Boligon¹, Nanye Long², Lucia Galvão de Albuquerque^{3,4}, Kent Weigel², Daniel Gianola²,
Guilherme Jordão Magalhães Rosa²

¹Pós-doutoranda do Programa de Pós-Graduação em Genética e Melhoramento Animal da FCAV/UNESP. Jaboticabal/SP. Bolsista FAPESP. e-mail: arioneboligon@yahoo.com.br

²Department of Dairy Science, University of Wisconsin, Madison, United States. E-mail: nanye.long@gmail.com; kweigel@wisc.edu; gianola@ansci.wisc.edu; grosa@wisc.edu

³Docente da Faculdade de Ciências Agrárias e Veterinárias - FCAV/UNESP. Jaboticabal/SP. e-mail: lgalb@fcav.unesp.br

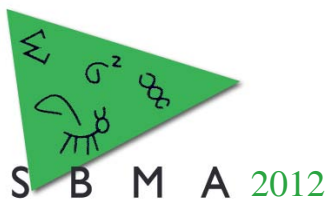
⁴Pesquisadora do CNPq e INCT-CA

Resumo: Foram utilizados dados simulados para avaliar a habilidade de predição genômica considerando diferentes estratégias de genotipagem seletiva. A população de referência (G_0) foi gerada por acasalamento ao acaso, considerando 50 machos e 50 fêmeas. Na geração G_0 , os animais foram selecionados baseando-se nos valores fenotípicos (maiores valores fenotípicos) ou escolhidos aleatoriamente, considerando diferentes intensidades de seleção para produzir a geração G_1 . Cinco estratégias de genotipagem foram consideradas na escolha de 500 animais na geração G_0 (considerado como *training set*): amostra aleatória (Aleatória); maiores valores fenotípicos (Melhores); menores valores fenotípicos (Piores); valores fenotípicos extremos (menores e maiores) (Extremos) e uma amostra de animais menos aparentados (Menos Aparentados). O número de animais nas gerações G_0 e G_1 foram fixados em 2.500 e, as características simuladas apresentavam herdabilidades de 0,10, 0,25 e 0,50. As estratégias de genotipagem seletiva foram comparadas em termos de habilidade de predição dos valores genéticos dos animais pertencentes a geração G_1 , usando um modelo Lasso Bayesiano, considerando 2.020 marcadores simultaneamente. Para todas as características simuladas, a menor correlação entre os valores genéticos preditos e verdadeiros (VGP e VGV, respectivamente) foi obtida na utilização da estratégia de genotipagem Piores. Para as estratégias Extremos, Aleatórios e Menos Aparentados, a correlação entre o VGP e VGV foi levemente superior com a redução da intensidade de seleção, sendo maior quando não foi aplicada seleção. Essas estratégias foram superiores que a Melhores. A estratégia de genotipagem Extremos apresentou a melhor habilidade de predição de valores genéticos, indicando que animais com valores fenotípicos extremos na população de referência são os mais informativos.

Palavras-chave: habilidade de predição, Lasso bayesiano, marcadores moleculares, seleção genômica

Selective genotyping strategies for prediction of breeding values using simulated data

Abstract: A simulation study was used to evaluate the predictive ability of genomic selection under different strategies of selective genotyping. A reference population (G_0) was generated by a full factorial mating between the 50 males and 50 females. In G_0 animals were selected based on their phenotypes (highest phenotypic values) or at random, at different selection intensities, to produce generation G_1 . Five genotyping strategies were considered to choose 500 animals in G_0 as a training set: random sampling (R); highest phenotypic values (H); lowest phenotypic values (L); extreme (low and high) phenotypic values (E); and subset of less related animals (I). The number of individuals in G_0 and G_1 was fixed at 2,500, and heritability was set to 0.10, 0.25 and 0.50. The selective genotyping strategies were compared in terms of their ability of predicting the genetic values of the animals in G_1 , using a Bayesian Lasso model with all 2020 markers simultaneously. For all simulated traits, the lowest correlation between predicted and true breeding values (GEBV and TBV, respectively) was obtained when using the L genotyping strategy. For E, R, and I strategies, the correlation between GEBV and TBV became slightly higher as selection intensity decreased, and was largest when no selection occurred. These three strategies were better than H. Overall, genotyping strategy E led to the best predictive ability of breeding values, indicating that animals with extreme phenotypic values in a reference population are the most informative when training genomic selection models.



Keywords: Bayesian Lasso, genomic selection, molecular markers, predictive ability

Introdução

A utilização de informações de marcadores moleculares distribuídos ao longo do genoma pode aumentar a acurácia associada aos valores genéticos preditos de animais jovens, candidatos à seleção. Em rebanhos comerciais, devido ao custo da genotipagem, apenas um subconjunto de animais selecionados são genotipados, sendo essa estratégia conhecida como genotipagem seletiva. Apesar do custo da genotipagem limitar o seu uso geralmente aos melhores animais, pouco se sabe se esta estratégia é a melhor em termos de capacidade de predição do mérito genético dos animais na geração seguinte.

O presente estudo foi desenvolvido com o objetivo de se avaliar a habilidade de predição de valores genômicos de animais candidatos à seleção, com base em diferentes estratégias de genotipagem seletiva, utilizando dados simulados.

Material e Métodos

O genoma estudado foi composto por cromossomos considerando a seguinte constituição: dois marcadores, seguido por um suposto QTL, mais dois marcadores, um suposto QTL, e assim por diante (M1-M2-Q1-M3-M4-Q2-M5-M6...). Na população inicial, todos os loci foram fixados no alelo “0”, considerando posteriormente que mutações ocorreram aleatoriamente, sendo representadas como “1”.

A população simulada foi inicialmente constituída de 100 animais (50 machos e 50 fêmeas). Para considerar os possíveis efeitos de mutação e deriva genética, os animais foram acasalados ao acaso por aproximadamente 5 mil gerações ($t = 5000$). Depois desses sucessivos acasalamentos, o número de progênies nascidas na geração $t = 5001$ foi aumentada para 2.500 animais por meio de cruzamentos entre cada um dos 50 touros com cada vaca (50 vacas), considerando uma progênie por acasalamento. Os animais pertencentes a essa última geração foram considerados como a população de referência (*training set*). Posteriormente, diferentes estratégias de seleção foram utilizadas para decidir quais animais pertencentes a população de referência seriam acasalados, produzindo uma nova geração, sendo constituída de sua progênie (*testing set*). As estratégias utilizadas foram: seleção dos melhores animais (aqueles com maiores valores fenotípicos) ou acasalamentos ao acaso (sem seleção). No primeiro caso, foram consideradas diferentes intensidades de seleção. Os animais escolhidos seguindo essas estratégias foram acasalados aleatoriamente, sendo sua progênie considerada como *testing set*.

Cinco estratégias de genotipagem seletiva foram estudadas em indivíduos pertencentes a população de referência e, somente parte dos animais foram genotipados para cada estratégia. As diferentes estratégias foram: 1) Aleatória: animais escolhidos aleatoriamente; 2) Melhores: selecionados os melhores animais (com melhores fenótipos); 3) Piores: selecionados os piores animais (com piores fenótipos); 4) Extremos: selecionados os animais extremos (com os melhores e piores fenótipos) e 5) Menos aparentados: escolhidos os animais menos aparentados. Os animais candidatos à seleção (*testing set*) não apresentavam registros de fenótipo, mas possuíam informações dos marcadores. As cinco estratégias de genotipagem seletiva foram avaliadas considerando diferentes níveis de herdabilidade (0,10; 0,25 e 0,50).

Os procedimentos necessários para a simulação da população foram desenvolvidos com o programa R. Os valores genômicos foram preditos utilizando o modelo Lasso bayesiano, proposto por Park e Casella (2008). Dessa forma, as diferentes estratégias de genotipagem seletiva na população de referência foram comparadas quanto a sua habilidade de predição de valores genômicos de animais candidatos à seleção (pertencentes a geração seguinte). A habilidade de predição foi avaliada pela correlação entre valores genômicos preditos e verdadeiros (acurácia). A simulação e análises foram repetidas 50 vezes para cada combinação de intensidade de seleção e estratégia de genotipagem em G_0 .

Resultados e Discussão

Na Figura 1 são apresentadas as correlações entre os valores genéticos preditos (VGP) e os valores genéticos verdadeiros (VGV) de animais pertencentes a geração G_1 , para os diferentes níveis de herdabilidade. Considerando as cinco diferentes estratégias de genotipagem seletiva, as correlações variaram de 0,14 a 0,46 (herdabilidade = 0,10); 0,15 a 0,57 (herdabilidade = 0,25) e 0,17 e 0,66 (herdabilidade = 0,50).

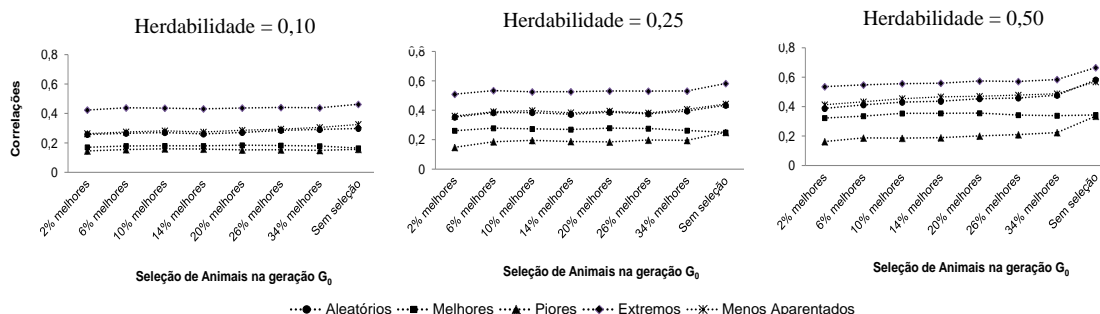


Figura 1. Correlações entre os valores genéticos preditos e verdadeiros na geração (G_1) com base em diferentes intensidades de seleção e estratégias de genotipagem seletiva aplicados à população de referência (G_0). Diferentes níveis de herdabilidade foram considerados (0,10; 0,25 e 0,50). Cada ponto representa a média de 50 repetições.

Para todas as características simuladas, menores correlações entre VGP e VGV foram obtidas com a estratégia Piores, independentemente da intensidade de seleção considerada (exceto quando não foi aplicada seleção, em que o desempenho das estratégias Piores e Melhores forem iguais). Uma possível razão para isso é que diferentes intensidades de seleção (2, 6, 10, 14, 20, 26 e 34% dos animais com melhores fenótipos) foram consideradas para produzir a geração G_1 . Por outro lado, no caso de acasalamentos ao acaso (sem seleção), os animais foram escolhidos aleatoriamente na geração G_0 para serem utilizados como pais da próxima geração (G_1), o que explica as correlações semelhantes quando as estratégias de genotipagem Piores e Melhores são aplicadas.

Para todas as estratégias de genotipagem em G_0 , com exceção de Piores, as correlações entre os VGP e os VGV aumentaram levemente com a diminuição da intensidade de seleção, atingindo o maior valor quando não foi aplicada seleção (acasalamentos aleatórios), sendo mais evidente na característica com maior herdabilidade.

Para todas as intensidades de seleção e níveis de herdabilidade estudados, as estratégias de genotipagem Aleatórios, Extremos e Menos Aparentados foram melhores em relação à estratégia Melhores. Entretanto, essas diferenças foram menores quando a seleção foi aplicada na geração G_0 , em relação a não considerar efeitos de seleção. Em um estudo de simulação, Ehsani et al. (2010) mostraram que a genotipagem seletiva dos melhores animais forneceu piores previsões de valores genéticos de animais pertencentes a uma geração seguinte, em comparação com os animais genotipados serem escolhidos ao acaso. Além disso, esses autores não encontraram diferenças importantes entre a genotipagem de animais com maiores valores fenotípicos e menores valores fenotípicos na população de referência.

Os resultados deste estudo mostram que a capacidade preditiva de valores genéticos utilizando marcadores moleculares dependerá, entre outros fatores, da escolha dos animais a serem genotipados na população de referência. Considerando a habilidade de predição, a melhor estratégia foi a genotipagem de animais fenotípicamente extremos na população de referência.

Conclusões

A genotipagem de animais com fenótipos extremos apresentou a melhor habilidade de predição de valores genéticos, indicando que animais com valores fenotípicos extremos na população de referência são os mais informativos e devem ser preferidos na escolha dos animais a serem genotipados.

Literatura citada

- EHSANI, A., L. JANSSEN, AND O. F. CHRISTENSEN. 2010. Effects of Selective Genotyping on Genomic Prediction. Proc. 9th World Congr. Genet. Appl. Livest. Prod. Leipzig (Germany).
- PARK, T., CASELLA, G. The Bayesian Lasso. *Journal of the American Statistical Association*, v.103, p.681-686, 2008.