

Seleção de Marcadores Para os Métodos RR-BLUP e BLASSO na Seleção Genômica Ampla

Italo Stefanine Correia Granato¹, Caillet Dornelles Marinho², Janeo Eustáquio de Almeida Filho², Marcos Deon Vilela de Resende³, Fabyano Fonseca e Silva⁴, Karina Carnielli Zamprogno Ferreira⁵, Leonardo Novaes Rosse⁵, Carolina Paola Sansaloni⁶, César Daniel Petroli⁷, Dario Grattapaglia⁸

Resumo:

Este trabalho foi realizado com o objetivo de avaliar a capacidade preditiva e o viés da seleção genômica ampla (GWS) para os métodos RR-BLUP e BLASSO, através da redução no número de marcadores, para definir o número ótimo de marcadores para os caracteres avaliados e o melhor método para estimação. Para tanto, foram utilizados dados de eucalipto, contendo 1000 clones fenotipados para diâmetro altura do peito (DAP), altura (ALT) e volume (VOL), e, 936 clones fenotipados para densidade básica pelo pilodyn® (PIL). Todos os clones foram genotipados com 2668 marcadores DArT. Os valores genéticos genômicos, a capacidade preditiva e o coeficiente de regressão do fenótipo no genótipo predito foram estimados via RR-BLUP e BLASSO. Os métodos RR-BLUP e BLASSO apresentaram aumento da capacidade preditiva com a redução do número de marcadores até um ponto ótimo, e a partir deste, houve redução na estimativa da capacidade preditiva. Os dois métodos, para os caracteres avaliados apresentaram estimativas de capacidades preditivas próximas, com uma leve superioridade do método RR-BLUP e com exceção do caráter volume analisado pelo BLASSO, em todas as análises realizada com redução do número de marcadores, não houve viés. Valores entre 70 e 90% de eliminação produzem melhores estimativas de capacidade preditiva e o método RR-BLUP apresentou superioridade.

Introdução

A seleção genômica ampla (GWS) proposta por Meuwissen et al. (2001) permite a predição simultânea dos efeitos dos marcadores que irão ser utilizados para seleção de indivíduos geneticamente superiores. Apresenta como uma solução viável à avaliação baseada somente em fenótipos e possui ampla utilização no melhoramento animal, mas sendo pouco explorada no melhoramento vegetal. No melhoramento de espécies florestais possui vantagem de realização de seleção precoce, acelerando o processo de melhoramento e proporciona maior ganho por unidade de tempo. Adicionalmente, a própria predição tende a ser mais acurada por considerar o real parentesco genético dos indivíduos em avaliação, em detrimento do parentesco médio esperado matematicamente (Resende 2007).

Na GWS são utilizadas três populações. A população de estimação, onde são avaliados os fenótipos e genótipos para estimação dos efeitos dos marcadores. A população de validação, realizada com indivíduos fenotipados e genotipados, que podem ser diferentes ou não da população de estimação, no entanto, utilizam-se os efeitos pré-estimados dos marcadores para prever os fenótipos. Na população de seleção, não há necessidade de fenotipar os indivíduos e as equações de predição derivadas na população de estimação são utilizadas para prever os fenótipos futuros.

A GWS, inicialmente proposta com alta densidade de marcadores, mas foi observado que o aumento do número de marcadores não aumenta linearmente a acurácia da GWS pelo método RR-BLUP (*Random*

¹ Mestrando do Programa de Pós-graduação em Genética e Melhoramento de Plantas – UFV/Viçosa. Bolsista do CNPq. E-mail: italo.granato@gmail.com

² Doutorando do Programa de Pós-graduação em Genética e Melhoramento de Plantas – UFV/Viçosa. E-mail: caillet.marinho@yahoo.com.br; janeo.eustaquio@ymail.com

³ Pesquisador Pós-doutor da Embrapa Floresta e Professor Credenciado do Departamento de Estatística – UFV/Viçosa. E-mail: marcos.deon@gmail.com

⁴ Professor Pós-doutor do Departamento de Estatística – UFV/Viçosa. E-mail: fabyanofonseca@ufv.br

⁵ Pesquisador (a) Doutor (a) da Empresa Veracel Celulose S.A. Salvador, BA. E-mail: karina.zamprogno@veracel.com.br; leonardo.rosse@veracel.com.br

⁶ Doutoranda do Programa de Pós-graduação em Biologia Molecular – UNB/Brasília. E-mail: carosansaloni@hotmail.com

⁷ Doutorando do Programa de Pós-graduação em Biologia Molecular – UNB/Brasília. E-mail: petrolic@hotmail.com

⁸ Pesquisador Pós-doutor da Embrapa Recursos Genéticos e Biotecnologia e Professor do curso de Pós-graduação em Ciências Genômicas e Biotecnologia – UCB/Brasília. E-mail: dario.grattapaglia@embrapa.br

Regression Best Linear Unbiased Predictor) (Fernando et al. 2007). Como observado por Cavalcanti e Resende (2011) na maioria dos casos, as capacidades preditivas diminuem conforme o número de marcadores excede o número ótimo. Assim, é necessário, para cada caráter, estabelecer o número ótimo de marcadores que maximizem a capacidade preditiva.

O objetivo deste trabalho foi avaliar e comparar os métodos RR-BLUP e BLASSO (*Bayesian Least Absolute Shrinkage and Selection Operator*) com seleção de marcadores quanto à capacidade preditiva na estimação de valores genéticos genômicos. Objetivou-se ainda definir o número ótimo de marcadores que afetam os caracteres avaliados e o melhor método para estimação.

Material e Métodos

Para o presente estudo foram utilizados dados de eucalipto, pertencentes ao projeto de seleção genômica desenvolvido pela Embrapa em conjunto com a Veracel Celulose, contendo 1000 clones fenotipados para diâmetro altura do peito (DAP), altura (ALT) e volume (VOL), e, 936 clones fenotipados para densidade básica pelo pilodyn® (PIL). Todos os clones foram genotipados com 2668 marcadores DArT (*Diversity Arrays Technology*) (Sansaloni et al. 2010).

As análises foram realizadas no software R, versão 2.14.1 (R Development Core Team 2012). A matriz de incidência genotípica foi parametrizada conforme recomendado por Resende et al. (2010), em que cada coluna de marcas foi centrada e padronizada.

Os valores genéticos genômicos $\hat{g}\hat{g}$ foram preditos via RR-BLUP com auxílio do pacote rrBLUP (Endelman 2011) e com o pacote BLR (Campos e Rodriguez 2012) para o BLASSO. A capacidade preditiva $r(\hat{g}f)r(\hat{g}f)$ foi calculada pela correlação entre o $\hat{g}\hat{g}$ e os fenótipos corrigidos e desregressados (f). Para verificar o viés da predição, a regressão de f em $\hat{g}\hat{g}$ foi obtida por $b_{f\hat{g}} = Cov(\hat{g}_v, f) / \sigma_{\hat{g}_v}^2$ $b_{f\hat{g}} = Cov(\hat{g}_v, f) / \sigma_{\hat{g}_v}^2$, em que, $\sigma_{\hat{g}_v}^2$ $\sigma_{\hat{g}_v}^2$ representa a variância dos valores genéticos genômicos preditos. Dessa forma, a melhor predição será aquela com $b_{f\hat{g}}b_{f\hat{g}}$ igual a 1.

O critério para eliminação foi o menor valor absoluto do efeito dos marcadores e a redução foi percentual, onde foram eliminados 95%, 90%, 85%, 80%, 70%, 60%, 40%, 20%, 10% e 5% dos marcadores de menor efeito absoluto.

Resultados e Discussão

Como observado na figura 1 os métodos RR-BLUP e BLASSO apresentaram aumento da capacidade preditiva com a redução do número de marcadores até um ponto ótimo, e a partir deste, houve redução na estimativa da capacidade preditiva. Como descrito por Cavalcanti e Resende (2011) a diminuição no número de alelos marcadores usados na GWS a partir do ponto de máxima acurácia faz com que ocorra uma redução na capacidade preditiva. Considerando o tempo para realização do mesmo, o método RR-BLUP apresenta superioridade, pois apesar de selecionar marcadores adicionará mais uma etapa no processo, o mesmo se tornará mais simples, além de necessitar de menos recursos computacionais. Resende Jr et al (2012) descreveram que selecionar marcas de maior efeito absoluto pode ser efetivo para características que não se encaixam no modelo infinitesimal e são regulados por genes de maior efeito.

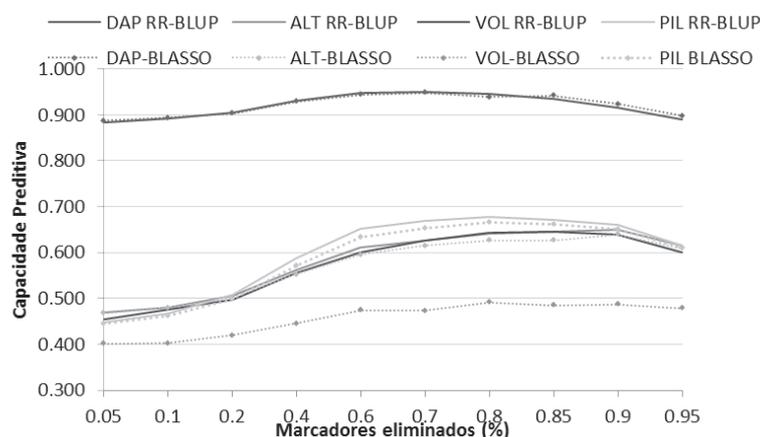


Figura 1 - Capacidade preditiva para os diferentes percentuais de eliminação de marcadores para as variáveis para diâmetro altura do peito (DAP), altura (ALT), volume (VOL) e densidade básica (PIL)

Os dois métodos, para os caracteres avaliados apresentaram estimativas de capacidades preditivas próximas, com uma leve superioridade do método RR-BLUP, e o caractere volume apresentou maiores diferenças de estimativa em relação ao BLASSO. Além da capacidade preditiva, a comparação entre os métodos foi baseada na regressão dos valores observados e preditos (figura 2). O valor esperado do coeficiente de regressão é um e indica que as avaliações são não viesadas e são efetivas em prever as reais magnitudes das diferenças entre os indivíduos em avaliação (Resende et al 2012), portanto, com exceção do caráter volume analisado pelo BLASSO, em todas as análises realizada com redução do numero de marcadores, não houve viés. O método BLASSO para o caráter volume apresentou coeficientes de regressão e acima de um e esses valores indicam que os valores genéticos estimados apresentam variabilidade aquém da esperada.

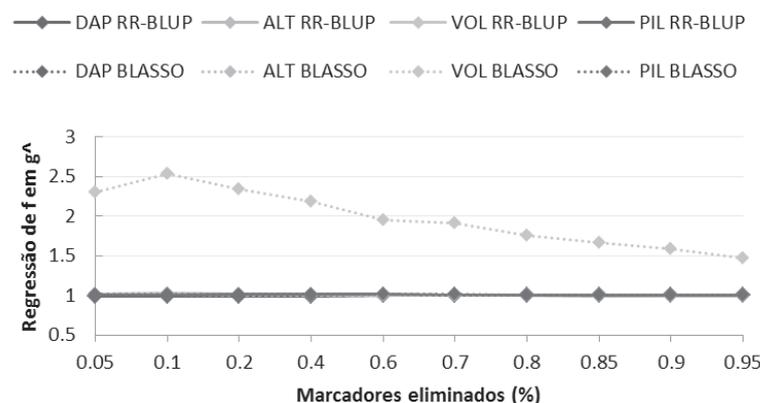


Figura 2. Coeficiente de regressão dos valores observados e preditos preditiva para os diferentes percentuais de eliminação de marcadores para as variáveis para diâmetro altura do peito (DAP), altura (ALT), volume (VOL) e densidade básica (PIL)

Agradecimentos

Os autores agradecem a Embrapa e a Veracel Celulose pela disponibilização dos dados. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pela concessão das bolsas de estudos.

Referências Bibliográficas

- Campos G and Rodriguez PP (2012). BLR: Bayesian Linear Regression. R package version 1.3. Disponível em: <<http://CRAN.R-project.org/package=BLR>>
- Cavalcanti JJV, Resende MDVd, Santos FHCd and Pinheiro CR (2012) Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. **Revista Brasileira de Fruticultura** 34:840-846.
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome** 4:250-255.
- Fernando RL, Habier D, Stricker C, Dekkers JCM and Tottir LR. 2007. Genomic selection. **Acta Agriculturae Scandinavica**, Section A – Animal Science 57: 192-195.
- Meuwissen T, Hayes B and Goddard M. (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics** 157:1819-1829.
- R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Resende Jr. MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF and Kirst M (2012) Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). **Genetics** 190:1503-1510.
- Resende MDV (2007) Matemática e estatística na análise de experimentos e no melhoramento genético. Colombo: EMBRAPA Florestas. 561p.