

Parametrizações em marcadores dominantes para seleção genômica ampla em eucalipto

Janeu Eustáquio de Almeida Filho¹, Camila Ferreira Azevedo², Caillet Dornelles Marinho¹, 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

O objetivo desse trabalho foi analisar quatro diferentes parametrizações na matriz de incidência de marcadores dominantes, sendo elas 0 e 1 (I), 0 e 4/3 (II), 0 e 16/9 (III) para ausência e presença da marca respectivamente e valores centralizados e padronizados (IV) sobre a parametrização (I). No trabalho foram utilizados dados de 1000 clones fenotipados para as características: Altura de plantas (ALT), diâmetro na altura do peito (DAP) e Volume (VOL) e 936 clones fenotipados para densidade básica pelo pilodyn® (PIL), todos os clones foram genotipados com 2668 marcadores DArTs. A parametrização IV foi a mais indicada para seleção clonal em presença de dominância, uma vez que esta se destacou para capacidade preditiva e herdabilidade capturada. As parametrizações II e III foram semelhantes entre si, essas foram deduzidas de forma a capturar o valor aditivo esperado, com isso é indicada sua avaliação com diferentes gerações, para a seleção de genitores. A distribuição do efeito dos marcadores foi mais leptocúrtica com a parametrização IV, seguido da III, II e I. Em geral o método RR-BLUP foi superior ao BLASSO com *prioris* não informativas.

Introdução

Com o advento das técnicas de genotipagem densa em larga escala, a seleção genômica proposta por Meuwissen et al. (2001) se tornou uma alternativa bastante promissora para o melhoramento, tanto animal quanto vegetal pela possibilidade de selecionar indivíduos com base em marcadores moleculares, que estão em desequilíbrio de ligação com os QTLs das características de interesse.

Para a seleção é estimado o valor genômico de cada indivíduo, que consiste no valor predito de um modelo de regressão cujo fenótipo é regredido nos marcadores, com isso o efeito estimado de marcadores codominante é o seu efeito médio de substituição alélica, o que é bastante interessante uma vez que se torna possível estimar o valor genético aditivo ou GEBV (valor de melhoramento genômico esperado).

Contudo para marcadores dominantes, a parametrização convencional considera 0 para ausência e 1 para presença da marca, portanto, na estimação do GEBV existem efeitos não aditivos confundidos. Com isso Resende et al. (2010) denomina em modelos gaméticos e genotípicos, os modelos que consideram marcadores codominantes e dominantes respectivamente.

Com intuito de explorar melhor a fração aditiva do GEBV em marcadores dominantes, novas parametrizações considerando propriedades estatísticas e genéticas da população podem ser modeladas (Resende 2008; Resende et al., 2012), bem como padronizações com intuito de melhor capturar o efeito de alelos raros associados ao caráter (Meuwissen et al., 2011).

Nesse sentido, esse trabalho consiste na avaliação de diferentes padronizações na matriz de marcadores dominantes, para seleção genômica em eucalipto.

¹ Doutorando do Programa de Pós-graduação em Genética e Melhoramento – UFV/Viçosa. E-mail: janeu.eustaquio@ymail.com; caillet.marinho@yahoo.com.br

² Doutoranda do Programa de Pós-graduação em Estatística e Biometria – UFV/Viçosa. E-mail: camila.azevedo1504@gmail.com

³ Pesquisador Pós-doutor da Embrapa Florestas 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

Material e métodos

Foram utilizados dados de eucalipto pertencentes ao projeto de seleção genômica desenvolvido pela Embrapa em conjunto com a Veracel Celulose. Os dados utilizados nesse trabalho provieram da média clonal de 1000 genótipos fenotipados para as características altura (ALT), diâmetro na altura do peito (DAP) e volume (VOL) e 936 clones fenotipados para densidade básica pelo pilodyn® (PIL), todos os clones foram genotipados com 2668 marcadores DArTs (*Diversity Arrays Technology*).

Todos os fenótipos foram corrigidos para efeitos fixos e desregressados para efeito de genitores, em seguida foram predito os valores genômicos (GEBV) pelo método RR-BLUP utilizando o pacote do *software* R rrBLUP (Endelman 2011) e pelo método BLASSO com o pacote BLR (de los Campos e Peres 2012), utilizando 100000 iterações, *burn-in* igual a 25000, *thin* igual a 2 e priores não informativas. Tanto para o RR-BLUP quanto para o BLASSO foi considerado quatro parametrizações para a matriz de marcadores.

Na primeira parametrização (I) foi utilizando 0 e 1 para ausência e presença da marca, respectivamente. A segunda parametrização (II) considerou 0 e 4/3 para ausência e presença da marca, respectivamente, conforme preconizado por Resende (2008). Segundo o autor como em marcadores dominantes não é possível diferenciar entre homozigotos dominantes e heterozigotos, então seja Z uma variável aleatória que representa o número de alelos "A", com isso considerando frequências iguais entre os alelos $E(Z|Presença da Marca) = 2 \cdot 1/3 + 1 \cdot 2/3 = 4/3$.

A terceira parametrização (III) considerou 0 e 16/9 para ausência e presença da marca, respectivamente, como proposto por Resende et al. (2012). Essa parametrização foi deduzida de forma que o valor predito para o efeito de marcas seja o efeito médio de substituição alélica esperado.

A quarta parametrização (IV) da matriz de marcadores consiste em centrar na média e padronizar os valores da parametrização I, com isso o método permite detectar efeito de QTLs em alelos raros na população (Meuwissen et al., 2011; Resende et al., 2012). De acordo com Resende et al. (2012) se o RR-BLUP for estendido para seu modelo equivalente G-BLUP, a utilização dessa parametrização reflete positivamente na matriz de parentesco, uma vez que as relações de parentesco são ponderadas pelo inverso da variância do erro de predição.

A comparação das parametrizações foi baseada na capacidade preditiva (r_{fg}) que é a correlação entre os GEBVs e os valores fenotípicos, essa é uma medida do quanto a predição será eficiente para seleção. Também foi utilizado o coeficiente de regressão do fenótipo regredido no genótipo predito (b_{fg}), que é uma medida de viés, segundo Resende et al. (2012) a predição é não viesada quando esse coeficiente é igual a 1. Por fim foi considerado a herdabilidade capturada pelo método (h^2).

Os métodos foram validados com a validação cruzada, para as características ALT, DAP e VOL, foram retirados grupos de 10 e 50 indivíduos para RR-BLUP e BLASSO respectivamente, enquanto que para PIL foram retirados 12 e 52 indivíduos para RR-BLUP e BLASSO respectivamente. Em cada ciclo foi estimado os parâmetros do modelo com os indivíduos remanescentes e predizendo o GEBV dos indivíduos retirados, o processo de validação termina quando todos os indivíduos foram retirados da estimação, sendo que após o grupo de indivíduos ser retirado, ele retorna no ciclo seguinte e cada grupo é retirado apenas em um ciclo. No final do processo todos os indivíduos tem seu GEBV predito e com isso é possível estimar r_{fg} , b_{fg} e h^2 , utilizando os valores médios dos componentes de variância ao longo dos ciclos (Resende et al., 2012).

Resultados e Discussão

Pelos valores estimados para b_{fg} (Tabela 1), foi observado que as predições do efeito dos marcadores pelo método RR-BLUP foi não viesada em todos os casos. Com o BLASSO a predição não foi viesada para ALT em todas as parametrizações e para DAP e PIL com a parametrização IV. Nas outras parametrizações utilizadas para DAP e PIL (I, II e III) os resultados do BLASSO indicam superestimação do efeito de marcadores ($b_{fg} < 1$), enquanto que, para VOL é indicado subestimação ($b_{fg} > 1$) do efeito das marcas em todas as parametrizações (Resende et al., 2012).

A superioridade do método RR-BLUP, também foi assegurada por apresentar estimativas para r_{fg} e h^2 superiores na maioria dos casos, sendo equivalente ao BLASSO em apenas alguns casos. Com isso, supõe-se que um modelo infinitesimal é mais adequado para explicar as características avaliadas nesse estudo, ou também que, o uso de *priors* não informativas para o BLASSO que é um método que preconiza a existência de genes com efeito maior, restringiu que este fosse superior ao RR-BLUP.

Tabela 1. Resultados da predição do parâmetros capacidade preditiva (r_{fg}), regressão do GEBV nos valores fenotípicos (b_{fg}) e herdabilidade capturada (h^2) pelo método de seleção genômica utilizando 2668 marcadores DArTs, para as variáveis Altura (ALT), Diâmetro na altura do peito (DAP), densidade básica pelo pilodyn® (PIL) e Volume (VOL).

Característica	Parametrização	BLASSO			RR-BLUP		
		r_{fg}	b_{fg}	h^2	r_{fg}	b_{fg}	h^2
ALT	I*	0,45	0,98	0,25	0,45	1,00	0,25
	II	0,44	0,99	0,16	0,45	1,00	0,16
	III	0,43	0,95	0,10	0,45	1,00	0,09
	IV	0,46	1,03	0,26	0,46	1,01	0,26
DAP	I	0,84	0,90	0,87	0,87	0,98	0,88
	II	0,85	0,93	0,79	0,87	0,98	0,80
	III	0,84	0,90	0,68	0,87	0,98	0,69
	IV	0,88	0,97	0,87	0,88	0,98	0,87
PIL	I	0,38	0,86	0,36	0,41	1,00	0,37
	II	0,38	0,90	0,25	0,41	1,00	0,25
	III	0,37	0,86	0,16	0,41	1,00	0,16
	IV	0,42	1,01	0,38	0,43	1,01	0,38
VOL	I	0,36	3,20	0,01	0,43	1,00	0,31
	II	0,37	2,87	0,01	0,43	1,00	0,20
	III	0,38	2,62	0,01	0,43	1,00	0,12
	IV	0,40	2,43	0,03	0,44	1,01	0,31

* Parametrizações I: 0 e 1; II: 0 e 4/3; III: 0 e 16/9 para ausência e presença da marca, respectivamente; IV: valores para incidência de marcadores centradas na média e padronizados.

A parametrização IV apresentou valores ligeiramente superiores para a capacidade preditiva em todos os casos. Para h^2 as parametrizações I e IV apresentaram as maiores estimativas considerando todas as características. Sendo assim, presume-se a possibilidade da existência de alelos pouco frequentes que são importantes para as características em questão, conseqüentemente, a parametrização IV se mostrou a melhor alternativa para seleção de clonal em presença de dominância.

As parametrizações II e III foram deduzidas de modo a capturar a fração aditiva na predição dos GEBVs com marcadores dominantes, com isso era previsto que a herdabilidade capturada fosse menor que nas outras parametrizações.

Para seleção de genitores, espera-se que as parametrizações II e III sejam superiores que as parametrizações I e IV, contudo uma forma de averiguar essa questão seria com predição dos efeitos dos marcadores em uma geração, e validação nas progênies dos indivíduos que foram usados na população de estimação.

Em relação a distribuição do valor predito para efeito dos marcadores, na parametrização IV a distribuição foi a mais leptocúrtica, seguido pelas parametrizações III, II e I (Figura 1). Como na parametrização IV foi realizada a padronização, os marcadores com menor polimorfismo tendem a contribuir da mesma forma que os marcadores mais polimórficos e conseqüentemente, os valores preditos entre os marcadores vão apresentar menor divergência. No caso das parametrizações II e III, espera-se que na predição do efeito de marcadores, seja capturado apenas o efeito médio de substituição alélica, com isso é razoável prever que a dispersão do efeito predito considerando as parametrizações II e III, será menor que na parametrização I, que tende a capturar o efeito genotípico (fração aditiva e não aditiva).

A distribuição do efeito predito dos marcadores foi menos disperso com o método BLASSO do que com o RR-BLUP, sendo essa diferença pronunciada apenas para VOL. Esses resultados são coerentes com a filosofia dos métodos, o método BLASSO pressupõe que o efeito dos marcadores segue distribuição exponencial dupla com variâncias heterogêneas, enquanto que o método RR-BLUP pressupõe que o efeito de marcadores segue distribuição normal com variâncias homogêneas, dessa forma, o BLASSO considera que existam alguns genes com maior efeito para característica e conseqüentemente faz seleção direta de covariáveis, uma vez que a predição do efeito de vários marcadores é próxima de zero, já o método RR-BLUP considera que a característica é governada por vários genes e todos com igual importância na explicação do caráter (Resende et al., 2012).

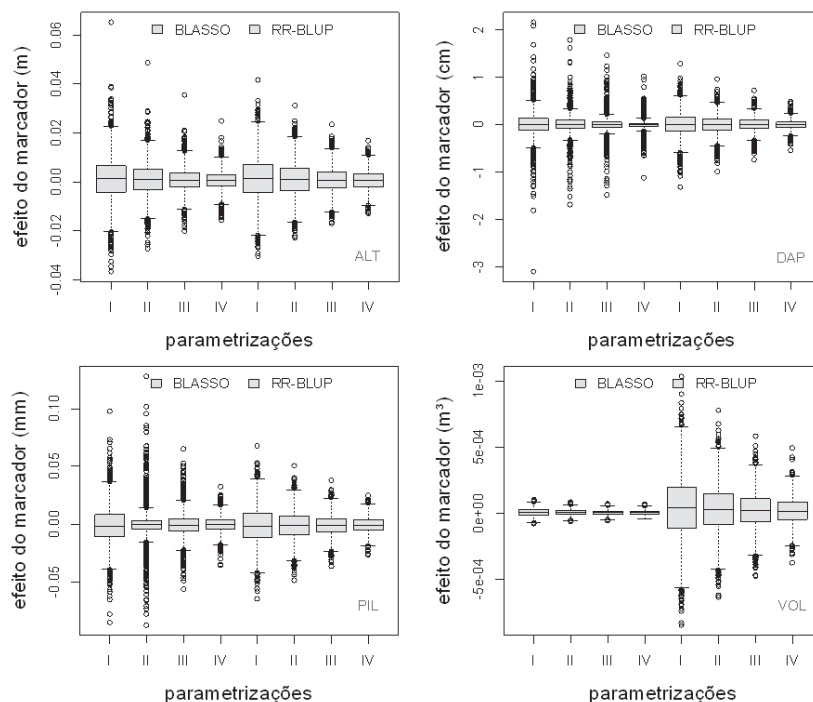


Figura 1. Boxplot dos efeitos de marcas considerando as parametrizações I 0 e 1 para ausência e presença da marca respectivamente; II 0 e 4/3 para ausência e presença da marca respectivamente; III 0 e 16/9 para ausência e presença da marca respectivamente e IV para valores da incidência de marcadores centrados na média e padronizados. Sendo as características analisadas Altura (ALT), Diâmetro na altura do peito (DAP), densidade básica pelo pilodyn® (PIL) e Volume (VOL).

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. E ao professor Leonardo Lopes Bhering, pelos recursos computacionais do laboratório de Biometria da UFV cedidos para realização das análises.

Referências

- de los Campos G and Rodriguez PP (2012). **BLR: Bayesian Linear Regression. R package version 1.3.** Disponível em: <http://CRAN.R-project.org/package=BLR>.
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome** 4:250-255.
- Meuwissen THE, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics** 57:1819-1829.
- Meuwissen THE, Luan T and Woolliams JA (2011) The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. **Journal of Animal Breeding and Genetics** 128:429-39.
- Resende MDV (2008) **Genômica quantitativa e seleção no melhoramento de plantas perenes e animais.** Embrapa Florestas, Colombo, 330p.
- Resende MDV, Resende Junior MFR, Aguiar AM, Abad JIM, Missiaggia AA, Sansaloni C, Petrolí C and Grattapaglia D (2010) **Computação da seleção genômica ampla.** Embrapa Florestas, Colombo, 79 p.
- Resende MDV, Silva FF, Lopes OS and Azevedo CF. (2012) **Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada (RRM) e estatística espacial.** Viçosa: UFV, 291p. Disponível em: http://www.det.ufv.br/ppestbio/corpo_docente.php.