

Seleção genômica ampla para curvas de crescimento

[Genome Wide Selection for growth curves]

F.F. Silva¹, G.S. Rocha², M.D.V. Resende³, S.E.F. Guimarães¹,
L.A. Peternelli¹, D.A.S. Duarte^{4,5}, C. Azevedo¹

¹Universidade Federal de Viçosa – Viçosa, MG

²Aluno de pós-graduação – Universidade Federal de Viçosa – Viçosa, MG

³Empresa Brasileira de Pesquisa Agropecuária – Embrapa Florestas

⁴Aluno de graduação – Universidade Federal de Viçosa – Viçosa, MG

⁵Bolsista de Iniciação Científica

RESUMO

Foi proposta uma metodologia para avaliação genética de curvas de crescimento considerando-se informações de marcadores SNPs (*Single Nucleotide Polymorphisms*). Em um primeiro passo foram ajustados modelos de crescimento não lineares (logístico) aos dados de peso-idade de cada animal, e em um segundo passo as estimativas dos parâmetros de tais modelos foram consideradas como fenótipos em um modelo de regressão (LASSO Bayesiano – BL) cujas covariáveis foram os genótipos dos marcadores SNPs. Este enfoque possibilitou estimar os valores genéticos genômicos (GBV) para peso em qualquer tempo da trajetória de crescimento, refletindo na confecção de curvas de crescimento genômicas, as quais permitiram a identificação de grupos de indivíduos geneticamente superiores em relação à eficiência de crescimento. Os dados simulados utilizados neste estudo foram constituídos de 2000 indivíduos (1000 na população de treinamento e 1000 na população de validação) contendo 453 marcadores SNPs distribuídos sobre cinco cromossomos. Os resultados indicaram a alta eficiência do método BL em predizer GBVs da população de validação com base na população de treinamento (coeficientes de correlação variaram entre 0,79 e 0,93), bem como a alta eficiência na detecção de QTLs, uma vez que os marcadores com maiores efeitos estimados encontravam-se em posições dos cromossomos próximas àquelas nas quais se encontravam os verdadeiros QTLs postulados na simulação.

Palavras-chave: SNP, LASSO bayesiano, dados longitudinais

ABSTRACT

A methodology was proposed for the genetic evaluation of growth curves considering SNP (Single Nucleotide Polymorphisms) markers. At the first step, nonlinear regression growth models (Logistic) were fitted to the weight-age of each animal, and on second step the parameter estimates of the Logistic model were used as phenotype in a regression model (Bayesian LASSO - BL) which covariates were given by SNP genotypes. This approach allows the estimation of GBV (Genomic Breeding Values) for weight at either time of growth trajectory, allowing also the production of genomic growth curves, which selected groups of individuals with larger growth efficiency. The simulated data set was constituted of 2,000 individuals (being 1,000 in the training and 1,000 in the validation population) each one with 453 SNP markers distributed along 5 chromosomes. The results indicated high efficiency of the BL method to predict GBV in the validation population using information from the training population (correlation coefficients varying between 0.79 and 0.93). The BL also presented high efficiency to detect QTL, once the most expressive estimated SNP effects were located at positions closed to true QTL position fixed in the simulation.

Keywords: SNP, Bayesian LASSO, longitudinal data

Recebido em 5 de março de 2012

Aceito em 6 de março de 2013

E-mail: fabyanofonseca@ufv.br

INTRODUÇÃO

Atualmente, os avanços biotecnológicos na área de automação do processo de genotipagem, o qual passou a ser realizado em larga escala, permitiram o desenvolvimento de novas classes de marcadores, entre os quais se destacam os SNPs (*Single Nucleotide Polymorphisms*). Diante da abundância desses marcadores, Meuwissen *et al.* (2001) idealizaram a seleção genômica ampla (*Genome Wide Selection – GWS*), a qual consiste na análise de um grande número de marcadores amplamente distribuídos no genoma.

Considerando que polimorfismos do DNA são as fontes de variação de mérito genético, marcadores SNPs em desequilíbrio de ligação com QTL (*Quantitative Trait Loci*) podem ser utilizados como critério extra para identificação de indivíduos candidatos à seleção, o que, segundo Goddard e Hayes (2007), aumenta a acurácia da avaliação genética. Porém, o grande número de marcadores, geralmente dezenas ou centenas de milhares, proporciona problemas de multicolinearidade (diferentes marcadores com o mesmo perfil genotípico) e de dimensionalidade (número de marcadores muito maior que o número de animais genotipados). Assim, métodos estatísticos sofisticados são demandados com o objetivo de solucionar tais problemas.

Dentre os métodos estatísticos utilizados em GWS, destaca-se o LASSO Bayesiano (BL). Resumidamente, a técnica LASSO (*Least Absolute Shrinkage and Selection Operator*), a qual se caracteriza como um método de regressão penalizada, foi inicialmente proposta por Tibshirani (1996), e sua versão bayesiana foi apresentada por Park e Casella (2008) e posteriormente modificada e adaptada para GWS por de los Campos *et al.* (2009). Desde então, a eficiência de tal método tem sido reportada por vários autores (Cleveland *et al.*, 2010; de los Campos *et al.*, 2010; Silva *et al.*, 2011).

Embora recentemente vários métodos estatísticos tenham sido propostos para GWS, como o BL, em se tratando de programas de melhoramento para crescimento animal (ou desempenho), tais métodos vêm sendo aplicados apenas para características (fenótipos) pontuais, como pesos

em idades específicas e/ou taxas de crescimento em certos períodos. Assim, surge o interesse em aplicar esses métodos a trajetórias completas de crescimento, também denominadas de curvas de crescimento, as quais contemplam avaliações longitudinais de peso ao longo da vida produtiva do animal.

Diante do exposto, o presente trabalho tem como principal objetivo propor uma metodologia de seleção genômica ampla (GWS) para curvas de crescimento animal fundamentada no método BL. Para tanto, serão utilizados dados simulados e recursos computacionais (códigos do *software R*) de livre acesso.

MATERIAL E MÉTODOS

Os dados simulados utilizados neste estudo são provenientes do QTLMAS2009 (*Workshop of Quantitative Trait Loci Mapping and Marker Assisted Selection*). O conjunto de dados consiste de 2025 indivíduos de duas gerações. Todos os indivíduos têm informações completas de marcadores. Existem 453 marcadores SNPs que estão aleatoriamente distribuídos sobre cinco cromossomos. Os primeiros 25 indivíduos são pais, 20 fêmeas e cinco machos. Os 2000 indivíduos restantes são descendentes, 100 famílias de irmãos completos, uma de cada combinação entre machos e fêmeas. Cada família de irmãos completos tem 20 descendentes.

Das 100 famílias relatadas, 50 (população de treinamento contendo 1000 indivíduos) possuem registros fenotípicos de produção (pesos), as outras 50 (população de validação contendo 1000 indivíduos) não têm informações fenotípicas. Fenótipos foram registrados em cinco momentos distintos (zero, 132, 265, 397, 530 dias). As famílias fenotipadas foram escolhidas de tal forma que cada fêmea apresentasse pelo menos 40 descendentes fenotipados enquanto cada macho tivesse 100 descendentes fenotipados. Todo o conjunto de dados utilizado está disponível no seguinte endereço eletrônico: <http://www.qtlmas2009.wur.nl/UK/Dataset/>. A Fig. 1 é uma representação gráfica da estrutura de combinação utilizada para simular a geração 2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
22	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
23	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
24	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
25	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20

Figura 1. Representação gráfica da segunda geração simulada. Cada i,j representa uma família de irmãos completos simulada pela combinação de uma fêmea i e um macho j. Células pretas representam famílias de irmãos completos cujos dados fenotípicos foram simulados; células brancas representam famílias de irmãos completos cujos dados fenotípicos não foram simulados. Cada família de irmãos completos consiste de 20 descendentes.

As análises foram realizadas em dois passos subsequentes. Primeiramente, aos dados de peso-idade de cada indivíduo da população de treinamento, ajustou-se o seguinte modelo de crescimento logístico:

$$y_i = \frac{\varnothing_1}{1 + \exp[(\varnothing_2 - t_i)/\varnothing_3]} + e_i \quad (1)$$

em que: y_i é o peso no tempo t_i , $\varnothing_1$ é o peso adulto (peso assintótico), $\varnothing_2$ é a abscissa do ponto de inflexão da curva, $\varnothing_2 + \varnothing_3$ é a abscissa referente ao ponto no qual y_i corresponde a aproximadamente 73% do peso adulto e e_i é o efeito do erro aleatório. Em termos práticos, quanto menor os valores $\varnothing_2$ e $\varnothing_3$ e maior o valor de $\varnothing_1$, mais eficiente é o padrão de crescimento. Para ajustar o modelo (1) aos dados de crescimento de cada indivíduo, utilizou-se o método dos quadrados mínimos via processo iterativo de Gauss-Newton, por meio do PROC MODEL do software SAS® (Statistical..., 2011).

Em um segundo passo, as estimativas dos parâmetros, obtidas para cada indivíduo no passo anterior, foram consideradas como fenótipos em

análises utilizando o método BL. Nessas análises, a utilização dos parâmetros estimados como fenótipos permite compreender a arquitetura genética de toda a trajetória, uma vez que todas as pesagens estão condensadas em tais estimativas.

Considerou-se o seguinte modelo no segundo passo:

$$Y = 1\mu + X\beta + \varepsilon$$

em que: Y é o vetor de estimativas obtidas para $\varnothing_1$, $\varnothing_2$ e $\varnothing_3$ (porém cada variável analisada separadamente), μ é a média geral, β é o vetor de efeito dos marcadores e χ é a matriz de incidência dos marcadores SNPs codificados como 0 (aa), 1 (aA) ou 2 (AA), e o termo ε corresponde ao erro aleatório, $\varepsilon \sim N(0, \sigma_e^2)$.

O método BL consiste na obtenção das estimativas dos coeficientes de regressão, ou efeitos de marcadores (β_j , com $j = 1, 2, \dots, \rho$), que solucionam o seguinte problema de otimização:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ (Y - (1\mu + X\beta))'(Y - (1\mu + X\beta)) + \lambda \sum_{j=1}^{453} |\beta_j| \right\},$$

em que: $\sum_{j=1}^{453} |\beta_k|$ é a soma dos valores absolutos

dos coeficientes de regressão contidos no vetor β , e λ é um parâmetro de suavização que controla a intensidade da penalização. Quando este último parâmetro é igual a zero, não há penalização, caso contrário o método executa um encurtamento mais forte no sentido de que alguns coeficientes de regressão tenham valores iguais a zero. Dessa forma, ao conduzir tais valores a zero, tem-se simultaneamente um processo de seleção de variáveis, que, no presente contexto, representa a seleção de marcadores SNPs.

As estimativas dos efeitos dos marcadores (β_j , com $j = 1, 2, \dots, p$), e da média μ foram obtidas com a utilização do pacote BLR (*Bayesian Linear Regression*) disponível do

$$GBV = \begin{bmatrix} 30,4976 \\ 30,4976 \\ 30,4976 \\ \vdots \\ 30,4976 \end{bmatrix}_{1000 \times 1} + \begin{bmatrix} 1 & 0 & 2 & \dots & 1 \\ 1 & 0 & 1 & \dots & 2 \\ 0 & 0 & 1 & \dots & 2 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 2 & 1 & 1 & 0 & 0 \end{bmatrix}_{1000 \times 453} \times \begin{bmatrix} -0,1871 \\ -0,0990 \\ 0,0887 \\ \vdots \\ 0,1773 \end{bmatrix}_{453 \times 1} = \begin{bmatrix} 28,8778 \\ 33,2067 \\ 36,3633 \\ \vdots \\ 40,6342 \end{bmatrix}_{1000 \times 1}$$

A obtenção do valor genético genômico predito para cada indivíduo da população de validação (GBV_v) foi realizada da mesma forma, tomando-se o cuidado em substituir matriz X

programa R (R Development Core Team, 2010). A implementação bayesiana da regressão LASSO (Park e Casella, 2008) presente neste pacote foi adaptada para seleção genômica por de los Campos *et al.* (2009).

De posse do vetor estimado $\hat{\beta}$ e da média μ , prosseguiu-se com a obtenção do valor genético genômico predito (GBV) para cada indivíduo da população de treinamento:

$$GBV = \hat{\mu} + (X \times \hat{\beta}), \text{ ou seja:}$$

pela matriz X_v , que se refere à incidência para os efeitos dos SNPs para indivíduos da população de validação. Assim, $GBV_v = \hat{\mu} + (X_v \times \hat{\beta})$, ou seja;

$$GBV_v = \begin{bmatrix} 30,4976 \\ 30,4976 \\ 30,4976 \\ \vdots \\ 30,4976 \end{bmatrix}_{1000 \times 1} + \begin{bmatrix} 0 & 2 & 1 & \dots & 0 \\ 0 & 2 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 2 & 2 & 1 & 1 \end{bmatrix}_{1000 \times 453} \times \begin{bmatrix} -0,1871 \\ -0,0990 \\ 0,0887 \\ \vdots \\ 0,1773 \end{bmatrix}_{453 \times 1} = \begin{bmatrix} 33,7189 \\ 38,4528 \\ 36,9261 \\ \vdots \\ 35,9370 \end{bmatrix}_{1000 \times 1}$$

Os coeficientes de correlação (r) e regressão linear (b) envolvendo valores genéticos genômicos verdadeiros (GBV fixados na simulação) e preditos na população de treinamento (GBV) e validação (GBV_v) foram utilizados a fim de medir a capacidade do

método em realizar predições de forma acurada ($r=1$) e não viesada ($b=1$), respectivamente.

Com o intuito de identificar indivíduos de melhor desempenho, foi realizada uma análise de agrupamento considerando-se como variáveis todos os valores genômicos estimados nas duas populações. Para tanto, utilizou-se o PROC

CLUSTER (Statistical..., 2011), via método centroide e distância euclidiana. Por razões práticas, optou-se por dividir os indivíduos em 10 grupos distintos, sendo o PROC TREE (Statistical..., 2011) utilizado para verificar quais animais pertenciam a cada grupo. Dentro de cada grupo (indexado por k), calcularam-se as médias dos valores genéticos genômicos estimados para cada fenótipo ($\overline{GBV_{\phi_1 k}}$, $\overline{GBV_{\phi_2 k}}$ e $\overline{GBV_{\phi_3 k}}$), e essas foram,

$$\hat{y}_{ik} = \frac{(\mu_{\phi_1} + \overline{GBV_{\phi_1 k}})}{\left\{ \left[1 + \exp((\mu_{\phi_2} + \overline{GBV_{\phi_2 k}}) - t_{ik}) / (\mu_{\phi_3} + \overline{GBV_{\phi_3 k}}) \right] \right\}}. \quad (2)$$

Com objetivo de detectar QTLs de efeitos mais expressivos em nível genômico, foi obtido o quantil 95% da distribuição empírica dos módulos dos efeitos estimados dos SNPs para análises envolvendo, respectivamente, estimativas de ϕ_1 , ϕ_2 e ϕ_3 . Assim, foi possível identificar os SNPs com maiores efeitos em módulo e suas respectivas posições por meio do mapa de posição dos marcadores disponível no arquivo intitulado “Map File” fornecido pelo QTLMAS2009. As posições encontradas para marcadores com maiores efeitos nas análises envolvendo, respectivamente, estimativas de ϕ_1 , ϕ_2 e ϕ_3 , foram comparadas com as posições simuladas dos marcadores disponíveis no arquivo nomeado “Simulation Values”, também fornecido pelo QTLMAS2009. Isso foi realizado a fim de se verificar a correspondência entre posições encontradas e simuladas de QTLs mais expressivos.

RESULTADOS E DISCUSSÃO

Na Tab. 1, são apresentados os resultados referentes à avaliação da eficiência de predição dos valores genéticos genômicos (GBVs). Em relação aos valores de correlação obtidos nas duas populações (treinamento e validação) para os três fenótipos (estimativas de ϕ_1 , ϕ_2 e ϕ_3), observa-se que eles foram altos (entre 0,79 e 0,93), o que indica uma alta capacidade preditiva do método BL. Observa-se, ainda, que as

então, inseridas no modelo de crescimento logístico (1) para a construção de um gráfico contendo 10 curvas, uma para cada grupo. Estas podem ser denominadas “curvas de crescimento genômicas” médias, e a análise gráfica delas possibilita a identificação de grupos de indivíduos de maior eficiência de crescimento no intervalo de tempo considerado (zero aos 530 dias). As referidas curvas são definidas pela equação (2).

correlações referentes à população de validação (r_{GBV,GBV_v}) foram um pouco menores do que as da população de treinamento (r_{GBV,GBV_t}), como era de se esperar, uma vez que os GBVs dos indivíduos de validação são preditos com base nos efeitos dos marcadores estimados por meio da população de treinamento.

Tendo em vista os coeficientes de regressão linear (Tab. 1), nota-se que estes foram próximos de 1 (entre 0,89 e 1,14), mostrando que o método BL é efetivo em prever as reais magnitudes dos GBVs de indivíduos das populações de treinamento e validação. Também como era de se esperar, tais valores foram mais próximos da unidade na população de treinamento (b_{GBV,GBV_t}) que na população de validação (b_{GBV,GBV_v}).

A análise de agrupamento, considerando-se como variáveis os GBVs estimados em ambas as populações para cada parâmetro do modelo logístico (1), possibilitou condensar as 2.000 (1.000 indivíduos em cada população) curvas de crescimento genômicas em apenas 10 curvas representativas de cada um dos grupos considerados. Tais curvas foram confeccionadas de acordo com a equação (2) e são apresentadas na Fig. 2.

Tabela 1. Estimativas dos coeficientes de correlação (r) e regressão linear (b) envolvendo valores genéticos genômicos verdadeiros (GBV) e preditos nas populações de treinamento ($\hat{GBV}$) e validação ($\hat{GBV}_v$)

Populações	Coeficientes	Fenótipos (estimativas dos parâmetros do modelo logístico)		
		$\emptyset_1$	$\emptyset_2$	$\emptyset_3$
Treinamento	$r_{GBV, \hat{GBV}}$	0,93	0,80	0,91
	$b_{GBV, \hat{GBV}}$	1,01	0,96	0,98
Validação	$r_{GBV, \hat{GBV}_v}$	0,92	0,79	0,85
	$b_{GBV, \hat{GBV}_v}$	1,02	1,14	0,89

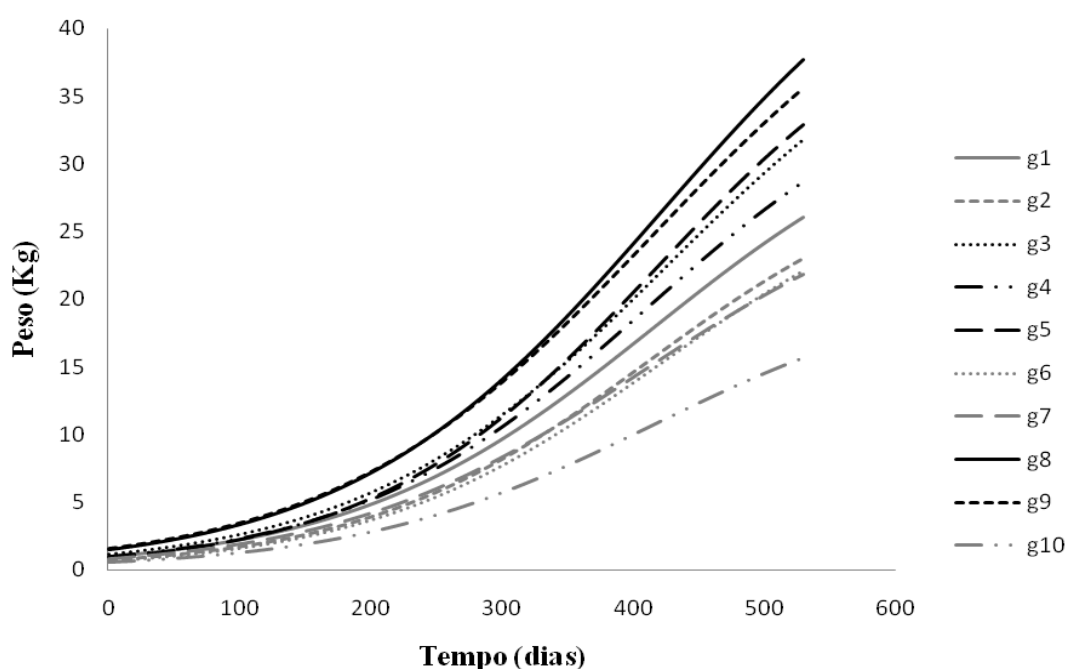


Figura 2. Curvas de crescimento genômicas médias para cada um dos 10 grupos de indivíduos.

De acordo com a Fig. 2, o grupo 8 foi o que apresentou maior crescimento dentro da amplitude de tempo considerada, portanto os indivíduos que compõem este grupo são, a princípio, aqueles destinados à seleção, tendo em vista toda a trajetória da curva de crescimento. Por outro lado, os indivíduos do grupo 10 são os menos indicados para seleção, uma vez que a baixa eficiência de crescimento deste grupo em relação aos demais é notória. Vale ressaltar que a estratégia usada permitiu selecionar indivíduos geneticamente superiores, tendo em vista o processo de crescimento como um todo, e não aqueles superiores apenas em relação a pesos em

tempos específicos da curva. Tal estratégia também foi usada com sucesso por Silva (2010), em estudos envolvendo curvas de crescimento de gado Nelore, e por Pong-Wong e Hadjipavlou (2010), em estudos de simulação envolvendo seleção genômica para curvas de crescimento.

Os resultados referentes à eficiência do método BL em detectar QTLs, que neste contexto pode ser caracterizado como sendo os marcadores de maiores efeitos, são apresentados na Tab. 2. Sob este enfoque, em cada análise foram encontrados 23 marcadores acima do quantil 95%, ou seja, 23 marcadores com efeitos mais expressivos em

nível genômico. Na tabela em questão também são apresentadas as posições reais assumidas na simulação para os seis marcadores de maiores efeitos e as posições dos seis marcadores de efeitos mais expressivos que mais se aproximaram das reais. A grande correspondência entre as posições em questão indica a eficiência do método LASSO em encontrar marcadores de grande efeito.

De forma geral, os marcadores de maiores efeitos identificados pelas posições encontradas (Tab. 2) podem ser usados diretamente na localização de QTLs. Isso pode ser realizado, pois a seleção genômica ampla (GWS), idealizada por Meuwissen *et al.* (2001), preconiza que a quantidade de marcadores é densa o suficiente para que eles estejam em desequilíbrio de ligação direto com o QTL.

Tabela 2. Posições (cM) simuladas e encontradas para os marcadores de maiores efeitos para os parâmetros analisados

Parâmetros					
$\emptyset_1$		$\emptyset_2$		$\emptyset_3$	
Posição simulada	Posição encontrada	Posição simulada	Posição encontrada	Posição simulada	Posição encontrada
0,4245	0,4447	0,5425	0,5309	0,8765	0,9137
1,0455	1,0356	1,3302	1,3445	1,4889	1,4849
1,8864	1,8831	2,0686	2,0515	2,2622	2,2202
2,8984	2,8827	2,5609	2,5826	3,0962	3,0663
3,6979	3,7004	3,3652	3,3582	3,8639	3,8537
4,7719	4,7396	4,5971	4,6264	4,3148	4,2936

CONCLUSÕES

A metodologia proposta mostrou-se eficiente para o estudo de curvas de crescimento considerando-se informações genômicas referentes a marcadores SNPs, pois esta possibilitou a identificação de grupos de indivíduos geneticamente superiores em relação ao padrão de crescimento, bem como a predição acurada de valores genéticos genômicos e a localização de QTLs tendo em vista o conjunto de dados simulados.

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), pelo apoio financeiro (APQ-00637-09) relacionado à aquisição de recursos computacionais e pela bolsa de Iniciação Científica. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro (473601/2010-9) relacionado à aquisição de recursos computacionais (servidores de última geração).

REFERÊNCIAS

- CLEVELAND, M.A.; FORNI, S.; DEEB, N.; MALTECCA, C. Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels. *BMC Proceedings*, v.4, Suppl 1, S6, 2010.
- de los CAMPOS, G.; NAYA, H.; GIANOLA, D. *et al.* Predicting quantitative traits with regression models for dense molecular markers. *Genetics*, v.182, p.375-385, 2009.
- de los CAMPOS, G.; GIANOLA, D.; ALLISON, D.B. Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat. Rev. Genet.* v.11, p.880-886, 2010.
- GODDARD, M.E.; HAYES, B.J. Genomic selection. *J. Anim. Breeding and Genetics*, v.124, p.323-330, 2007.
- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome wide dense marker maps. *Genetics*, v.157, p.1819-1829, 2001.

- PARK, T.; CASELLA, G. The Bayesian LASSO. *J. Am. Statist. Assoc.*, v.103, p.681-686, 2008.
- PONG-WONG, R.; HADJIPAVLOU, G. A two-step approach combining the Gompertz growth with genomic selection for longitudinal data. *BMC Proceedings*, v.4, Suppl 1:S4, 2010.
- R DEVELOPMENT CORE TEAM, 2010. R: A language and environment for statistical computing. Disponível em: <<http://www.R-project.org>> Acessado em: 5 dez. 2011.
- STATISTICAL Analysis Sistem - SAS Institute Inc. Statistical Analysis System user's guide. Version 9.1 ed. Cary: SAS Institute, USA, 2003.
- SILVA, N.A.M. *Seleção de modelos de regressão não linear e aplicação do algoritmo SAEM na avaliação genética de curvas de crescimento de bovinos Nelore*. 2010. 58f. Tese (Doutorado em Zootecnia) – Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte.
- SILVA, F.F.; VARONA, L.; RESENDE, M.D.V. et al. A note on accuracy of Bayesian LASSO regression in GWS. *Livest. Sci.*, v.142, p.310-314, 2011.
- TIBSHIRANI, R. Regression shrinkage and selection via the LASSO. *J. Royal Statist. Society*, v.58, p.267-288, 1996.