



43ª Reunião Anual da Sociedade Brasileira de Zootecnia
24 a 27 de Julho de 2006
João Pessoa - PB

INFLUÊNCIA DE GENES DE EFEITO MAIOR NA PREDIÇÃO BAYESIANA E CLÁSSICA DE VALORES GENÉTICOS EM POPULAÇÕES SELECIONADAS (1)

GISELLE MARIANO LESSA DE ASSIS (2), JOSÉ MARQUES CARNEIRO JÚNIOR (3), RICARDO FREDERICO EUCLYDES (4), ROBLEDO DE ALMEIDA TORRES (4), PAULO SÁVIO LOPES (4)

(1) Fonte Financiadora: CAPES

(2) Pesquisadora da Embrapa Acre, Rodovia BR 364, km 14, CP 321, CEP 69908-970, Rio Branco, AC, e-mail: giselle@cpafac.embrapa.br

(3) Pós-doutorando da Embrapa Acre, Rodovia BR 364, km 14, CP 321, CEP 69908-970, Rio Branco, AC

(4) Professor Adjunto do DZO da Universidade Federal de Viçosa, Departamento de Zootecnia, CEP 36571-000, Viçosa, MG

RESUMO

As metodologias usualmente empregadas na avaliação genética, inclusive quando os dados estão sob seleção, pressupõem que o modelo infinitesimal seja adequado para a característica de interesse. No entanto, características quantitativas podem ser governadas por poucos genes, o que pode tornar essas metodologias pouco apropriadas para prever valores genéticos. Com o objetivo de verificar a influência de genes de efeito maior e do tamanho da população na predição dos valores genéticos, em dados sob seleção, utilizando metodologia clássica (EBLUP) e Bayesiana, com dois níveis de informação 'a priori', quatro tipos de população foram simulados. Conforme os resultados obtidos concluiu-se que os valores genéticos são superestimados sob seleção fenotípica quando os registros de todos os indivíduos e a matriz completa de parentesco são incluídos nas análises, independentemente do tamanho da população. A queda na acurácia é ainda mais acentuada quando a característica é governada por reduzido número de genes, sendo a classificação correta dos indivíduos também prejudicada. A inclusão do registro de todos os indivíduos, assim como da matriz de parentesco completa beneficiam a classificação adequada dos indivíduos. Concluiu-se também que a acurácia na predição dos valores genéticos, assim como a classificação dos indivíduos não são alteradas pelo nível de informação 'a priori' das análises Bayesianas, cujos resultados também se assemelham aos da metodologia EBLUP.

PALAVRAS-CHAVE

avaliação genética, amostragem de Gibbs, modelo infinitesimal, modelos mistos, simulação computacional, viés da seleção

INFLUENCE OF MAJOR GENES IN BAYESIAN AND CLASSICAL PREDICTION OF BREEDING VALUES IN SELECTED POPULATIONS

ABSTRACT

Usually, the methodologies used in genetic evaluation, inclusive when data are under selection, assume that the infinitesimal model is adequate to the trait of interest. However, quantitative traits can be governed by few genes and this fact can turn these methodologies slightly appropriate to predict genetic

values. Four different types of populations were simulated, aiming to verify the influence of major genes and population size on breeding values prediction, under selection, using classical methodology (EBLUP) and Bayesian analyses, with two 'a priori' information levels. According to the results, it was concluded that breeding values are overestimated under phenotypic selection when data of all individuals and complete relationship matrix are included on analyses, independently of population size. Accuracy decrease is more accentuated when the trait is governed by reduced number of genes, being the correct classification of individuals also affected. The inclusion of all data, as well as complete relationship matrix benefit the appropriate classification of individuals. Finally, accuracy of breeding values prediction, as well as the rank of individuals are not changed by 'a priori' information level on Bayesian analyses, whose results are also similar to the EBLUP methodology.

KEYWORDS

computational simulation, Gibbs sampling, genetic evaluation, infinitesimal model, mixed models, selection bias

INTRODUÇÃO

As avaliações genéticas vêm sendo realizadas principalmente pelo emprego da Metodologia de Modelos Mistos para obtenção do melhor preditor linear não-viesado (BLUP), cuja predição depende normalmente da estimação dos componentes de variância, usualmente realizada por meio do Método da Máxima Verossimilhança Restrita (REML). Ambas as metodologias se baseiam na adequabilidade do modelo infinitesimal para a característica em questão.

Outra pressuposição para o uso apropriado da Metodologia de Modelos Mistos é que os dados não sejam oriundos de seleção. No entanto, os dados disponíveis para análise são obtidos a partir de rebanhos onde os acasalamentos não são realizados ao acaso pelos produtores. Com o intuito de eliminar ou minimizar o viés causado pela seleção, a análise deve incluir a matriz de parentesco completa até a população-base, além dos registros de todos os animais (Sorensen & Kennedy, 1984). No entanto, esses estudos com populações selecionadas são realizados considerando que as pressuposições implícitas ao modelo infinitesimal estão sendo atendidas.

O objetivo desse trabalho foi verificar a influência de genes de efeito maior e do tamanho da população na predição dos valores genéticos, utilizando metodologias Bayesiana e clássica, em populações sob seleção, por meio de simulação. Verificou-se também a influência de diferentes níveis de informação 'a priori' nos valores genéticos preditos.

MATERIAL E MÉTODOS

Os dados utilizados foram simulados em nível de gene, utilizando-se o pacote computacional Genesys (Euclides, 1996), composto por nove programas, escritos em linguagem 'Fortran'.

Quatro diferentes tipos de populações foram simulados: 1)GRAN 900: população grande, sendo cada geração constituída por 2.400 indivíduos com registros e 1.000 pais, em que a característica sob seleção é governada por 900 pares de locos; 2)GRAN 10: população grande, sendo cada geração constituída por 2.400 indivíduos com registros e 1.000 pais, em que a característica sob seleção é governada por 10 pares de locos; 3)PEQ 900: população pequena, sendo cada geração constituída por 120 indivíduos com registros e 50 pais, em que a característica sob seleção é governada por 900 pares de locos; 4)PEQ 10: população pequena, sendo cada geração constituída por 120 indivíduos com registros e 50 pais, em que a característica sob seleção é governada por 10 pares de locos.

Foram avançadas 10 gerações discretas para dois processos seletivos distintos: seleção ao acaso e seleção fenotípica. Para as populações GRAN 900 e GRAN 10 foram feitas 300 repetições e para as populações PEQ 900 e PEQ 10 foram realizadas 500 repetições para cada um dos processos seletivos na obtenção das 10 gerações.

Duas metodologias foram empregadas na avaliação genética dos dados simulados, visando estimar os

componentes de variância e prever os valores genéticos dos indivíduos: BLUP Empírico ou EBLUP, em que a Metodologia dos Modelos Mistos foi aplicada, utilizando-se as estimativas dos componentes de variância obtidas pelo REML e Análise Bayesiana, utilizando a Amostragem de Gibbs.

A análise Bayesiana foi realizada utilizando-se o programa MTGSAM – Multiple Trait Gibbs Sampling in Animal Models (Van Tassell & Van Vleck, 1995). O seguinte modelo linear misto unicaracterístico foi utilizado na análise dos dados simulados:

$y_{ijk} = m + A_i + B_j + a_{ijk} + e_{ijk}$, em que y_{ijk} é a observação fenotípica do k -ésimo indivíduo no i -ésimo efeito fixo A e no j -ésimo efeito fixo B; m é uma constante inerente a todas as observações; A_i é o efeito do i -ésimo efeito fixo A; B_j é o efeito do j -ésimo efeito fixo B; a_{ijk} é o efeito aleatório genético aditivo do k -ésimo indivíduo no i -ésimo efeito fixo A e no j -ésimo efeito fixo B; e_{ijk} é o efeito aleatório ambiental.

Dois níveis de informação a priori em relação aos componentes de variância foram considerados nessa metodologia: não-informativo (NI) e informativo (I).

O mesmo modelo utilizado na Análise Bayesiana foi considerado na estimação dos componentes de variância pelo REML e predição dos valores genéticos empregando-se a Metodologia de Modelos Mistos. Foram utilizados para este fim os programas do pacote MTDFREML – Multiple Trait Derivative-Free Restricted Maximum Likelihood (Boldman et al., 1995).

Com base nos valores genéticos preditos pela metodologia EBLUP e pela metodologia Bayesiana com priors não-informativo e informativo, foram calculados o Quadrado Médio do Desvio, a Porcentagem de Seleccionados em Comum e a Correlação de Ordem, visando comparar as metodologias empregadas, assim como as diferentes estruturas consideradas.

RESULTADOS E DISCUSSÃO

O Quadrado Médio do Desvio (QMD) dos valores genéticos preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação a priori, em relação aos valores genéticos reais, para as populações GRAN 900, GRAN 10, PEQ 900 e PEQ 10 encontram-se na Tabela 1. Verifica-se que, para todas as populações, os valores mais elevados do QMD se referem à situação em que ocorreu Seleção Fenotípica, sendo todos os registros e a informação de parentesco completa dos indivíduos utilizados para prever os valores genéticos (G10 SF 0-10). Esse resultado, provavelmente, ocorreu pelo fato dos valores genéticos preditos nesta situação serem, em média, superestimados (Assis, 2005). Ao comparar os resultados obtidos conforme o número de pares de locos que governa a característica, percebe-se que o QMD foi sempre maior para as populações com poucos pares de locos para todas as situações consideradas. Assim, ao violar a pressuposição do modelo infinitesimal de que a característica é determinada por infinitos pares de locos de efeitos aditivos infinitesimalmente pequenos, perde-se consideravelmente a acurácia na predição dos valores genéticos.

Na Tabela 2, encontram-se as Porcentagens de Seleccionados em Comum (presentes na classificação baseada nos valores genéticos preditos e na baseada nos valores genéticos reais) ao selecionar 15% dos melhores indivíduos em diferentes situações. De forma geral, a Porcentagem de Seleccionados em Comum aumentou ao incluir todos os registros e a informação completa de parentesco, tanto para Seleção ao Acaso quanto para Seleção Fenotípica para todas as populações. Verifica-se, portanto, que apesar do QMD e do desvio médio aumentarem ao incluir todos os registros e toda informação de parentesco na ocorrência de Seleção Fenotípica, a Porcentagem de Seleccionados em Comum também aumentou nessa situação, ou seja, apesar da diferença entre valores reais e preditos ter aumentado, a inclusão de informação permitiu, de forma geral, melhor classificação dos indivíduos. Em relação ao número de locos que governa a característica, nota-se que, para as populações grandes, a Porcentagem de Seleccionados em Comum diminuiu ao se analisar dados provenientes da população com 10 pares de locos na ocorrência de Seleção Fenotípica. Para as populações pequenas, ocorreu resultado inverso: ao diminuir o número de locos, a porcentagem de indivíduos em comum aumentou. Esse resultado pode estar ocorrendo pelo fato da inclusão de um único indivíduo na população

pequena alterar substancialmente a porcentagem de indivíduos em comum selecionados, enquanto que, para populações grandes, essa porcentagem é menor para cada indivíduo incluído. Portanto, para populações pequenas, esta forma de comparação não seria a mais adequada.

A Correlação de Ordem entre os valores genéticos reais e preditos pelo EBLUP e pela metodologia Bayesiana para dois níveis de informação 'a priori' encontram-se na Tabela 3 para as populações GRAN 900, GRAN 10, PEQ 900 e PEQ 10, respectivamente. A Correlação de Ordem aumentou com a inclusão de todos os registros e informação completa de parentesco para todas as populações, tanto para Seleção ao Acaso quanto para Seleção Fenotípica. A mesma observação feita anteriormente em relação à Porcentagem de Selecionados em Comum pode ser realizada. Em relação ao número de locos que governa a característica, nota-se que a Correlação de Ordem diminuiu ao se analisar dados provenientes das populações com poucos locos na ocorrência de Seleção Fenotípica, tanto para a população grande quanto para a pequena.

Com base nos resultados obtidos a partir dos valores genéticos preditos, nota-se que, de forma geral, a análise Bayesiana com 'prior' informativo não gerou predições mais acuradas nem mesmo trouxe melhorias no ordenamento dos melhores indivíduos. O mesmo não foi observado na estimação dos componentes de variância, em que, principalmente nas análises com pequena quantidade de dados, o aumento da informação 'a priori' foi capaz de produzir estimativas bastante acuradas (Assis, 2005). Portanto, a predição dos valores genéticos não sofreu influência expressiva das diferentes estimativas dos componentes de variâncias para o mesmo tipo de análise. Esse resultado pode ser explicado pelo fato de se ter realizado análises unicaracterísticas visando prever valores genéticos, que são relativamente insensíveis às mudanças nos valores dos componentes de variâncias (Kennedy, 1991).

CONCLUSÕES

Os valores genéticos são superestimados sob seleção fenotípica quando todos os registros e a matriz de parentesco são incluídos nas análises.

A classificação correta dos indivíduos é prejudicada quando genes de efeito maior governam a característica.

As metodologias Bayesianas e EBLUP geram resultados semelhantes em relação à acurácia da predição e à classificação dos indivíduos.

REFERÊNCIAS BIBLIOGRÁFICAS

ASSIS, G.M.L. Efeito do número de genes na avaliação genética utilizando dados simulados. Viçosa: Universidade Federal de Viçosa, 2005, 102p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2005.

BOLDMAN, K.G., KRIESE, L.A., VAN VLECK, L.D. et al. A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances. United States Department of Agriculture, Agriculture Research Service, 1995.

EUCLYDES, R. F. Uso do sistema para simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares. Viçosa: Universidade Federal de Viçosa, 1996, 150p. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, 1996.

KENNEDY, B.W. C.R. Henderson: the unfinished legacy. J. Dairy Sci., v.74, p.4067-4081, 1991.

SORENSEN, D.A., KENNEDY, B.W. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci., v.59, n.5, p.1213-1223, 1984.

VAN TASSEL, C.P., VAN VLECK, L.D. A manual for use of MTGSAM. A set of Fortran programs to apply Gibbs Sampling to animal models for variance components estimation. United States Department of Agriculture, Agriculture Research Service, 85p, 1995.