

VINÍCIUS SILVA JUNQUEIRA

**QUALIDADE DAS INFORMAÇÕES DE PARENTESCO NA  
AVALIAÇÃO GENÉTICA DE BOVINOS DE CORTE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister ScieEtiae*.

VIÇOSA  
MINAS GERAIS - BRASIL  
2014

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

J95q  
2014

Junqueira, Vinícius Silva, 1985-

Qualidade das informações de parentesco na avaliação  
genética de bovinos de corte. / Vinícius Silva Junqueira. –  
Viçosa, MG, 2014.

xi, 66f : il. (algumas color.) ; 29 cm.

Inclui anexo.

Orientador: Paulo Sávio Lopes.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Amostrador de Gibbs. 2. Marcadores moleculares.  
3. Reprodutores múltiplos. 4. Transferência de embriões.  
I. Universidade Federal de Viçosa. Departamento de Zootecnia.  
Mestrado em Genética e Melhoramento. II. Título.

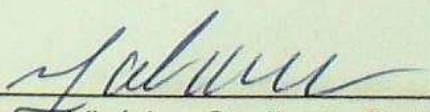
CDD. 636.082

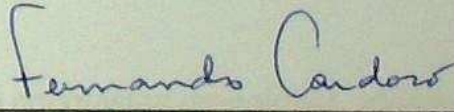
VINÍCIUS SILVA JUNQUEIRA

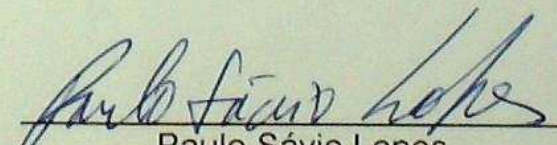
QUALIDADE DAS INFORMAÇÕES DE PARENTESCO NA AVALIAÇÃO  
GENÉTICA DE BOVINOS DE CORTE

Dissertação apresentada à Universidade  
Federal de Viçosa, como parte das exigências  
do Programa de Pós-Graduação em Genética  
e Melhoramento, para obtenção do título de  
*Magister Scientiae*

APROVADA: 25 de julho de 2014.

  
\_\_\_\_\_  
Marcos Vinícius Gualberto Barbosa da  
Silva

  
\_\_\_\_\_  
Fernando Flores Cardoso  
(Coorientador)

  
\_\_\_\_\_  
Paulo Sávio Lopes  
(Orientador)

*“É muito melhor lançar-se em busca de conquistas grandiosas, mesmo expondo-se ao fracasso, do que alinhar-se com os pobres de espírito, que nem gozam muito nem sofrem muito, porque vivem numa penumbra cinzenta, onde não conhecem nem vitória, nem derrota.”*

*Theodore Roosevelt*

**Aos meus pais Gilson e Lucilene,**

*“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”*

Marthin Luther King

**Dedico esse trabalho.**

## **AGRADECIMENTOS**

À Universidade Federal de Viçosa pela oportunidade da realização do mestrado no Programa de Pós-Graduação em Genética e Melhoramento e ao Departamento de Zootecnia por fornecer a estrutura de execução da pesquisa.

À Capes pela concessão da bolsa de estudos e financiamento de projetos correlacionados.

Ao meu orientador Professor Paulo Sávio Lopes, que me recebeu na minha chegada em Viçosa e, desde então, é disponível para ensinamentos, conversas e conselhos durante nesse processo de formação pessoal e profissional.

Ao PhD. Fernando Flores Cardoso, que no caminho do aprendizado foi fonte de inspiração e de ensinamentos, principalmente durante o período de trabalho no EMBRAPA Pecuária Sul, Bagé, RS.

Ao Professor Fabyano Fonseca e Silva, pela orientação, amizade, risadas, paciência e dedicação nos ensinamentos estatísticos.

Ao Professor Robledo de Almeida Torres, pelo incentivo e discussões sobre o melhoramento genético.

Ao Laboratório de Biometria, que forneceu a estrutura física para a execução das análises, e à FAPEMIG por proporcionar recurso financeiro para o estabelecimento da estrutura computacional.

Aos meus pais, Gilson e Lucilene, por toda educação e esforço para que o hoje fosse possível. Pela compreensão que eu precisava percorrer meu caminho para alcançar todos meus objetivos. Apesar de não estar presente, são eles o motivo de eu seguir na luta cotidiana.

À minha irmã, Andressa, que se esforça para cumprir o papel de filha e filho em minha ausência. Sem ela, eu não poderia estar ausente.

À minha namorada, Luciene, que, nesses quase seis anos de companheirismo, compreendeu o motivo de minha ausência e me deu suporte para finalizar mais essa etapa.

Aos amigos, Nadson Souza, André Mauric, Leonardo Peixoto, Leonardo Bhering e Humberto Fanelli pelas horas de companhia e risadas nesse processo de formação em que todos estamos longe de casa.

Aos amigos do Capril, Laís Costa, Hinayah Rojas, Delvan Silva, Alessandra Silva, Hugo Teixeira, Giovani Caetano, Bárbara Félix, William Mariano, Geraldo Iria e Otávio pela amizade e interesse na busca pelo conhecimento.

Aos amigos do melhoramento animal Edson Vinícius, Rodrigo Mota, Aline Crispim, Luciano Pinheiro, Jeferson Corrêa, Matilde Pessoa, Felipe Gomes, Renata Veroneze, Sirlene Lázaro, Leonardo Glória, Rodrigo Godinho, Ederson Camargo, Rogério Veloso, Raphael Rocha pelas trocas de conhecimento e ricas discussões.

Aos amigos da EMBRAPA, Bruna Pena, Leandro Lunardini, Marcos Yokoo, Rafael Martins, Ândrea Plotzki, Rodrigo Azambuja, Rodrigo Costa, Yuri, Bruno Teixeira, Helena Comin e Luíza Corrêa pela companhia e amizade durante a estadia em Bagé, RS.

## BIOGRAFIA

VINÍCIUS SILVA JUNQUEIRA, filho de Gilson Carlos Leite Junqueira e Lucilene Silva Junqueira, nasceu em 26 de novembro de 1985, na cidade de Brasília, Distrito Federal.

Iniciou a graduação em Medicina Veterinária na Faculdade de Agronomia e Medicina Veterinária (FAV) da Universidade de Brasília (UnB) no primeiro semestre de 2006.

Durante o curso de graduação se dedicou aos estudos das doenças infecciosas e das técnicas cirúrgicas. No segundo semestre de 2010 foi estagiário curricular na Clínica de Equínos da Escola de Veterinária da Universidade Federal de Minas Gerais (UFMG).

No primeiro semestre de 2011 recebeu o título de Médico Veterinário pela Universidade de Brasília.

Durante o ano de 2011 e janeiro de 2012 teve como foco profissional a orientação em planejamento rural e clínica cirúrgica na cidade de Curvelo e região, localizada em Minas Gerais.

Em março de 2012 se muda para Viçosa, MG, e em agosto de 2012 inicia o mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa.

Em 25 de julho de 2014, submeteu-se aos exames finais de defesa de dissertação para obtenção do título de *Magister Scientiae* em Genética e Melhoramento.

## SUMÁRIO

<b>RESUMO .....</b>	<b>viii</b>
<b>ABSTRACT .....</b>	<b>x</b>
<b>INTRODUÇÃO GERAL .....</b>	<b>1</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>3</b>
<b>CAPÍTULO I.....</b>	<b>6</b>
Uso de marcadores para aumentar a qualidade de informações de pedigree em avaliações genéticas de resistência a carrapatos em bovinos de corte....	6
<b>INTRODUÇÃO .....</b>	<b>8</b>
<b>MATERIAL E MÉTODOS .....</b>	<b>9</b>
<b>RESULTADOS.....</b>	<b>14</b>
<b>DISCUSSÃO .....</b>	<b>23</b>
<b>CONCLUSÃO .....</b>	<b>27</b>
<b>BIBLIOGRAFIA.....</b>	<b>27</b>
<b>CAPÍTULO II.....</b>	<b>31</b>
Avaliação genética de bovinos de corte considerando incerteza de paternidade e animais oriundos de biotécnicas da reprodução .....	31
<b>INTRODUÇÃO .....</b>	<b>33</b>
<b>MATERIAL E MÉTODOS .....</b>	<b>34</b>
<b>RESULTADOS.....</b>	<b>42</b>
<b>DISCUSSÃO .....</b>	<b>52</b>
<b>CONCLUSÃO .....</b>	<b>56</b>
<b>REFERÊNCIA BIBLIOGRÁFICA.....</b>	<b>57</b>
<b>ANEXO .....</b>	<b>60</b>

## RESUMO

JUNQUEIRA, Vinícius Silva, M.Sc., Universidade Federal de Viçosa, julho de 2014. **Qualidade das informações de parentesco na avaliação genética de bovinos de corte**. Orientador: Paulo Sávio Lopes. Coorientadores: Fernando Flores Cardoso e Fabyano Fonseca e Silva.

O objetivo desse estudo foi verificar o reflexo da qualidade das informações de parentesco sobre as estimativas de parâmetros genéticos e acurácias das predições de valor genético. Dessa forma, foi avaliada a qualidade dessas informações ao realizar correções no pedigree baseadas em marcadores do tipo SNPs. Os componentes de variância foram estimados sob enfoque Bayesiano por amostragem de Gibbs. A avaliação da correta definição de parentesco foi realizada utilizando-se informações de marcadores SNPs de 3.591 indivíduos em uma população constituída de 12.668 animais. Os conflitos mendelianos entre as marcas da progênie e dos pais foram utilizados como critério de avaliação de parentesco. Dessa forma, foram realizadas 460 mudanças no pedigree, dentre os quais 54% possuíam um touro ou vaca identificado. Foi observado que, em média, novos relacionamentos genéticos foram definidos com 75 marcas em conflito, enquanto que a rejeição do parentesco foi realizada com 2.700 marcas. O uso do programa Molecular Coancestry para inferência de parentesco a partir de informações de marcadores moleculares proporcionou a definição de 2.174 novos relacionamentos de meio-irmãos. Foi possível observar que as correções de parentesco proporcionaram aumento da acurácia média dos valores genéticos preditos. O aumento na qualidade das informações de pedigree proporcionou a estimação de herdabilidade de maior magnitude ( $0,22 \pm 0,0286$ ), sugerindo a possibilidade de seleção direta para a resistência a carrapatos. Foi utilizada uma estratégia de validação cruzada pelo método K-médias e também de forma aleatória, no qual cinco grupos de treinamento foram formados. Os resultados da validação cruzada indicam que maior qualidade nos relacionamentos do pedigree proporcionam maior valor de acurácia.

O peso padronizado aos 205 dias foi utilizado para avaliar a inclusão de informações de filhos de reprodutores múltiplos (RM) e de animais oriundos de biotécnicas da reprodução (TEF). Foram avaliadas três estratégias de inclusão

dessas informações: grupos genéticos (GG), método hierárquico bayesiano (HIER) e matriz da média dos numeradores dos coeficientes de parentesco (ANRM). O critério de informação da *deviance* (DIC) foi utilizado como avaliador da qualidade de ajuste e sugeriu que a estratégia HIER proporcionou melhor ajuste ao considerar as informações de RM. Entretanto, o método ANRM apresentou melhor ajuste ao incluir informações dos animais TEF. A inclusão das informações de animais TEF foi realizada pelo uso da informação da receptora para estimação do efeito genético materno e de ambiente permanente materno. Não foi observada diferença estatística nas estimativas de componentes de variância e de parâmetros genéticos ao considerar informações de animais TEF, porém os valores genéticos preditos apresentaram maior magnitude. As correlações de Spearman entre os valores genéticos foram de elevadas magnitudes para os animais fundadores, animais com certeza no parentesco e para os filhos de RM. Entretanto, o uso das informações dos animais TEF modificou de forma significativa as predições dos valores genéticos. O uso de metodologias adequadas para incluir a informação de animais com incerteza de paternidade e animais oriundos de transferência de embriões e fertilização *in vitro* proporcionam a predição de valores genéticos mais acurados e auxiliam em maiores taxas de ganho genético.

## ABSTRACT

JUNQUEIRA, Vinícius Silva, M.Sc., Universidade Federal de Viçosa, July of 2014. **Quality of relationship information on genetic evaluation of beef cattle**. Adviser: Paulo Sávio Lopes. Co-advisers: Fernando Flores Cardoso and Fabyano Fonseca e Silva.

The aim of this study was to evaluate the quality of relationship information over genetic parameters estimates and accuracies of breeding values. Thus, we evaluated the quality by making corrections based on SNPs markers. Variance components were estimated under Bayesian approach via Gibbs sampling. The evaluation of the correct relationship was performed using markers of 3,591 individuals in a population consisting of 12,668 animals. Mendelian conflicts were defined as conflicts between the progeny and parents markers. Thus, 460 changes were performed, which 54% had a bull or cow on pedigree. We observed, on average, match parent were defined with 75 markers, while no-match parent n was performed with 2,700 markers. Annealing algorithm Molecular coancestry program provided 2,174 new half-sibs relationships. We observed that higher quality on pedigree information provided increase in accuracy of breeding values and higher heritability estimates ( $0.22 \pm 0.0286$ ), suggesting the possibility of direct selection for tick resistance. We used a 5-fold cross-validation strategy using K-means and random methods to group the animals. The results of cross-validation indicate that higher quality of the pedigree relationships provide higher accuracy.

The weaning weight was used to evaluate multiple sires progeny (MS) and embryo transfer and *in vitro* fertilization animals (TEF). We used three strategies to include MS information: genetic groups (GG), bayesian hierarchical method (HIER) and the average numerator relationship matrix (ANRM). The deviance information criteria (DIC) was used as Bayesian measure of fit, which suggested that HIER strategy provided better fit when MS information is consider. However, ANRM provided best fit to include TEF animals. We use foster dam information to estimate the maternal genetic and maternal permanent environmental effects when considering TEF animals. We didn't observed statistical difference in variance components and genetic parameters considering information TEF animals, but breeding values showed greater values. Spearman correlations between breeding values were higher for

base, animals with certain paternity and MS progeny. However, the use of TEF animals information significantly changed the predicted breeding values. The use of appropriate methodologies to include the information of MS progeny and TEF animals provide most accurate prediction of breeding values and assist in higher rates of genetic gain.

## INTRODUÇÃO GERAL

A bovinocultura brasileira ocupa posição de destaque no cenário agropecuário mundial como maior exportador de carne bovina e possui um rebanho efetivo com cerca de 200 milhões de animais. O cenário econômico é favorável ao desenvolvimento da pecuária nacional, porém, tem sofrido forte influência negativa dos altos custos de produção e da falta de políticas públicas que valorizem o setor. Além disso, a competitividade com outros países tem causado grande variação nos preços da arroba do boi e isso causa baixa capacidade de planejamento de investimentos a curto e médio prazos. Problemas técnicos como variação anual de produção e qualidade de forragem, deficiências em manejo nutricional, reprodutivo e sanitário, sistemas inadequados e o baixo investimento em genética também são responsáveis pela baixa eficiência do sistema produtivo.

Neste sentido, o investimento em melhoramento genético do rebanho nacional tem um papel importante no aumento da quantidade e qualidade da oferta de animais para produção de alimento. Uma opção para atender demanda crescente de proteína de origem animal é fazer com que bovinos criados a pasto ou confinados tenham rápido crescimento (BULLOCK et al., 1993; BOLIGON et al., 2009).

### *Qualidade de informações do pedigree e sucesso no progresso genético*

Atualmente as metodologias utilizadas na avaliação genética de bovinos de corte e outras espécies comerciais utilizam informações fenotípicas e de pedigree para predizer valores genéticos (HENDERSON, 1973; SAATCHI et al., 2011). Ao utilizar os relacionamentos presentes no pedigree se assume que essas informações estão corretas, entretanto, estudos recentes (SANDERS et al., 2006; HABIER et al., 2010; MUNOZ et al., 2014) têm demonstrado que erros estão presentes e têm influenciado o ganho genético (ISRAEL e WELLER, 2000; SIMEONE et al., 2011).

Segundo Banos et al. (2001), aproximadamente 25% de erros são presentes na identificação de parentesco na população mundial de bovinos. Esses autores relatam que esses erros causam viés na estimação de parâmetros genéticos, na predição do valor genético dos animais e reduzem o

ganho genético. Segundo Van Vleck (1970), quanto maior for a quantidade de erros no pedigree maior será o viés na estimação dos parâmetros genéticos.

### *Reprodutores múltiplos*

A extensão territorial do Brasil determina diferentes climas, solos e manejos, por isso a pecuária extensiva é mais expressiva. Dessa forma, é prática comum em países com sistema de produção como do Brasil disponibilizar mais de um touro por grupo de fêmeas com o objetivo de se conseguir elevadas taxas de prenhez (CARDOSO e TEMPELMAN, 2003; 2004). Pelo ponto de vista reprodutivo e econômico, o manejo com reprodutores múltiplos é importante, porém, gera problemas aos programas de avaliação genética, pois o pai verdadeiro não é conhecido com certeza.

Grande parte dos programas de melhoramento animal predizem o valor genético dos candidatos utilizando as equações de modelos mistos, metodologia proposta por Henderson (1973). Essas equações utilizam o relacionamento genético entre todos os indivíduos da população como informação *a priori* ao incluir a matriz dos numeradores dos coeficientes de parentesco de Wright. Entretanto, existe incerteza na definição do relacionamento genético entre filhos de reprodutores múltiplos e seus possíveis pais. Como proposta para solucionar esse problema, metodologias (HENDERSON, 1988; CARDOSO e TEMPELMAN, 2003) foram propostas para incluir informações desses animais com objetivo de se alcançar valores genéticos com maior acurácia.

### *Melhoramento genético animal e uso de biotecnologias da reprodução*

O uso de técnicas como a transferência de embriões e fertilização *in vitro* tem sido cada vez mais comum na pecuária de corte com objetivo de rápida multiplicação dos animais geneticamente superiores. Segundo Alencar (2004) e Mota et al. (2013), a adoção dessas técnicas apresenta grandes vantagens ao melhoramento genético porque possibilita o aumento da intensidade de seleção das fêmeas nos programas de melhoramento, o que proporciona maior taxa de ganho genético.

As mães são responsáveis por proporcionar o ambiente de sobrevivência aos filhos durante os primeiros meses de vida. Por isso, parte do desempenho da progênie é influência direta da mãe, sendo conhecido como efeito materno. Dessa forma, para a estimação do efeito materno é necessário o registro do desempenho da progênie. Porém, nas situações em que animais são oriundos de técnicas como a transferência de embriões ou fertilização *in vitro*, são as receptoras que fornecem o ambiente materno em vez da mãe biológica, então, parte do desempenho da progênie tem origem da doadora e parte tem origem da receptora. Pensando nisso, Schaeffer e Kennedy (1989) e Van Vleck (1990) propuseram que para correta estimação dos efeitos genéticos aditivos, genético materno e de ambiente permanente materno fosse utilizada a informação da doadora dos embriões ou oócitos para estimação dos efeitos genéticos aditivos e a informação da receptora para a estimação dos efeitos genéticos materno e de ambiente permanente materno.

Portanto, esse estudo teve como objetivos: (a) avaliar a influência da certeza dos relacionamentos presentes no pedigree sobre a acurácia e predição do valor genético para a resistência a carrapatos em bovinos de corte e (b) avaliar parâmetros genéticos e tendência genética de filhos de reprodutores múltiplos juntamente aos animais oriundos de transferência de embriões ou fertilização *in vitro* em características de crescimento de bovinos de corte.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALENCAR, M. M. **Perspectivas para o melhoramento genético de bovinos de corte no Brasil** 2004. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/CPPSE/15172/1/PROCIMMA2004.00003.pdf%3E>. Acesso em

BANOS, G.; WIGGANS, G. R.; POWELL, R. L. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. **Journal of Dairy Science**, v. 84, n. 11, p. 2523-2529, 2001.

BOLIGON, A. A.; ALBUQUERQUE, L. G. D.; MERCADANTE, M. E. Z.; LÔBO, R. B. Herdabilidades e correlações entre pesos do nascimento à idade adulta em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia**, v. 38, p. 2320-2326, 2009.

BULLOCK, K. D.; BERTRAND, J. K.; BENYSHEK, L. L. Genetic and environmental parameters for mature weight and other growth measures in Polled Hereford cattle. **J Anim Sci**, v. 71, n. 7, p. 1737-1741, 1993.

CARDOSO, F. F.; TEMPELMAN, R. J. Bayesian inference on genetic merit under uncertain paternity. **Genetics Selection Evolution**, v. 35, n. 5, p. 469-488, 2003.

\_\_\_\_\_. Genetic evaluation of beef cattle accounting for uncertain paternity. **Livestock Production Science**, v. 89, n. 2, p. 109-120, 2004.

HABIER, D.; TETENS, J.; SEEFRIED, F.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution**, v. 42, n. 1, p. 5, 2010.

HENDERSON, C. R. Sire evaluation and genetic trends. In: PROCEEDINGS OF THE ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF DR. JAY L. LUSH, 1973, **Anais.**, 1973. p. 10-41.

\_\_\_\_\_. Use of an Average Numerator Relationship Matrix for multiple-sire joining. **Journal of Animal Science**, v. 66, p. 1614-1621, 1988.

ISRAEL, C.; WELLER, J. I. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. **Journal of Dairy Science**, v. 83, n. 1, p. 181-187, 2000.

MOTA, R. R.; LOPES, P. S.; MARQUES, L. F. A.; DA SILVA, L. P.; DE RESENDE, M. D. V.; TORRES, R. D. A. The influence of animals from embryo transfer on the genetic evaluation of growth in Simmental beef cattle by using multi-trait models. **Genetics and Molecular Biology**, v. 36, n. 1, p. 43-49, 2013.

MUNOZ, P. R.; RESENDE, M. F. R.; HUBER, D. A.; QUESADA, T.; RESENDE, M. D. V.; NEALE, D. B.; WEGRZYN, J. L.; KIRST, M.; PETER, G. F. Genomic relationship matrix for correcting pedigree errors in breeding populations: Impact on genetic parameters and genomic selection accuracy. **Crop Science**, v. 54, n. 3, p. 1115-1123, 2014.

SAATCHI, M.; MCCLURE, M. C.; MCKAY, S. D.; ROLF, M. M.; KIM, J.; DECKER, J. E.; TAXIS, T. M.; CHAPPLE, R. H.; RAMEY, H. R.; NORTHCUTT, S. L. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics Selection Evolution**, v. 43, p. 40, 2011.

SANDERS, K.; BENNEWITZ, J.; KALM, E. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. **Journal of Dairy Science**, v. 89, n. 1, p. 315-321, 2006.

SCHAEFFER, L. R.; KENNEDY, B. W. Effects of embryo transfer in beef cattle on genetic evaluation methodology. **Journal of Animal Science**, v. 67, n. 10, p. 2536-2543, 1989.

SIMEONE, R.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A. Evaluation of the utility of diagonal elements of the genomic relationship matrix as a diagnostic tool to detect mislabelled genotyped animals in a broiler chicken population. **Journal of Animal Breeding and Genetics**, v. 128, n. 5, p. 386-393, 2011.

VAN VLECK, L. D. Misidentification and sire evaluation. **Journal of Dairy Science**, v. 53, n. 12, p. 1697-1702, 1970.

\_\_\_\_\_. Alternative animal models with maternal effects and foster dams. **Journal of Animal Science**, v. 68, n. 12, p. 4026-4038, 1990.

## CAPÍTULO I

Uso de marcadores para aumentar a qualidade de informações de pedigree em avaliações genéticas de resistência a carrapatos em bovinos de corte

**RESUMO:** Este estudo teve como objetivo avaliar o reflexo de correções nos relacionamentos genéticos do pedigree sobre parâmetros genéticos, valores genéticos e acurácia. Para isso utilizaram-se registros de contagem de carrapatos de bovinos das raças Hereford e Braford. A genotipagem dos touros foi realizada utilizando-se painel de alta densidade (BovineHD - Illumina® bead chip – 777962 SNPs) e o painel Illumina BovineSNP50 (54609 SNPs) para a progênie. Equilíbrio de Hardy-Weinberg, taxa de genotipagem, alelo menos frequente, correlação de SNPs, SNPs na mesma posição, desvio de heterozigotidade, genótipos idênticos e erros de identificação de sexo foram utilizados como critérios de qualidade para SNPs e amostras. O parentesco entre genitores e progênie dos animais genotipados foi avaliado e incompatibilidades foram baseadas na contagem de conflitos mendelianos. Foram identificadas 460 modificações nas definições de parentesco, das quais 97,16% (447) foram relacionadas às informações de identificação do pai e 2,84% (13) às informações de identificação da mãe. Foram observados 199 (84,32%) conflitos em que o pai era genotipado e 13 (100%) conflitos com mães genotipadas. Foram realizadas 2.174 novas definições de meio-irmãos de acordo com o coeficiente de correlação entre a matriz de coancestralidade e a matriz de coancestralidade molecular. As estimativas de componentes de variância e de parâmetros genéticos foram obtidas por metodologia Bayesiana por amostragem Gibbs. Os valores genéticos foram preditos supondo um modelo com medidas repetidas. Foram utilizados três arquivos de pedigree com diferentes níveis de correções, os quais diferiam na estrutura de (co)variância genética. Foi observada que maior qualidade nos relacionamentos genéticos não estimou componentes de variância diferentes estatisticamente, porém resultou em valores genéticos mais acurados.

**Palavras-chave:** amostrador de gibbs, avaliação genética, pedigree, repetibilidade

Use of molecular markers to improve quality of pedigree information in genetic evaluation to tick resistance in beef cattle

**ABSTRACT:** This study aimed to evaluate the misidentification of paternity on pedigree and its influence in genetic parameters, predicted breeding values and accuracy. Tick count was used as response variable collected in Hereford and Braford repeated times. Genotyping of bulls (130) was performed using high-density panel (BovineHD - Illumina® bead chip – 777,962 SNPs) and the Illumina BovineSNP50 panel (54,609 SNPs) to their progeny (3461). Hardy-Weinberg, call rate, minor allele frequency, SNPs correlation, SNPs in the same position, heterozygosity deviance, identical genotypes and misidentification of sex were used as quality criteria for SNPs and samples. Variance components estimates and genetic parameters were obtained by Bayesian approach via Gibbs sampling algorithm. The parents and progeny relationship of genotyped animals was evaluated and incompatibilities were based on counts of mendelian conflicts. We identified 460 errors in the definitions relationships, of which 97.16% (446) were related to the sire identification and 2.84% (13) to dam's. We find 199 (84.32%) conflicts in which sire was genotyped and 13 (100%) conflicts with genotyped dam. We set 2,174 new definitions of half-siblings based on correlation coefficient between the coancestry matrix and molecular coancestry matrix. Breeding values and accuracy was performed according to animal model with repeated measures. Three pedigree files with different levels of correction were used, in which differ in genetic covariance structure. We observed that improved quality in genetics relationships didn't estimate variance components statistically different, however genetics values were found more accurate. The *deviance* information criteria (DIC) was used as a measure of fit. We used a *5-fold* strategy for cross-validation to evaluate the accuracy of methodology. It was defined 2,174 new relationships of half-siblings based on genotypes. Estimated variance components didn't differ between the models, but estimated breeding values and accuracies were predicted with lower predicted error of variance. Cross-validation was consistent to indicate that greater certainty in relationships predicts higher accuracies.

**Keywords:** genetic evaluation, gibbs sampling, pedigree, repetibility models

## INTRODUÇÃO

Bovinos são uma das principais espécies utilizadas para consumo de proteína animal. Por serem amplamente distribuídos em diversos ambientes pelo mundo são expostos a vários endo e ectoparasitas. Dentre os mais comuns estão os carrapatos da espécie *Rhipicephalus (Boophilus) microplus*, sendo responsáveis por prejuízos à saúde animal e humana, principalmente por transmitirem doenças, causar anemia, debilitação, perda de peso e morte (GASPARIN et al., 2007). Todos esses fatores podem reduzir a eficiência dos sistemas produtivos (JONSSON, 2006).

Durante anos, pecuaristas têm controlado endo e ectoparasitas principalmente pela aplicação de produtos químicos de uso tópico e/ou parenteral. Entretanto, a resistência natural dos animais aos ectoparasitas tem sido cada vez mais estudada visando a possibilidade de aumento da frequência dos alelos favoráveis. Neste sentido, o estudo da herança da resistência a carrapatos e a seleção de animais têm sido objetivo de diversas pesquisas (GASPARIN et al., 2007; PORTO NETO et al., 2011; OLIVEIRA et al., 2013).

A seleção dos indivíduos geneticamente superiores está relacionada com adequada estimação de parâmetros genéticos e predição acurada dos valores genéticos. Para isso, a matriz dos numeradores dos coeficientes de parentesco – calculada em função do pedigree – é utilizada como informação *a priori* nas equações de modelos mistos (MUNOZ et al., 2014).

Van Vleck (1970b) relata que erros no pedigree ou a falta de informação no parentesco afeta a taxa de ganho genético porque reduz a correlação entre o valor genético verdadeiro e o valor genético predito. São encontrados na literatura estudos de simulação de erros de pedigree (VAN VLECK, 1970b; 1970a; CARNEIRO et al., 1999; ISRAEL e WELLER, 2000; BANOS et al., 2001; HARDER et al., 2005), testes de paternidade em função do grupo sanguíneo (GELDERMANN et al., 1986), marcadores microssatélites (SANDERS et al., 2006), marcadores SNP's (HABIER et al., 2010; MUNOZ et al., 2014) e matriz de parentesco genômico (HABIER et al., 2010; SIMEONE et al., 2011).

Entretanto, ainda não existem estudos avaliando a correção das informações de parentesco pelo uso de marcadores SNPs em dados reais e

particularmente com relação a resistência de bovinos a carrapatos. Marcadores tipo SNPs podem ser utilizados com o objetivo de avaliar conflitos mendelianos entre genitores e progênes com conseqüente aumento de qualidade das relações de parentesco nos animais da população (FERNANDEZ e TORO, 2006; BARUCH e WELLER, 2008; WIGGANS et al., 2009; WIGGANS et al., 2010).

Portanto, o objetivo deste estudo foi avaliar se e como a correção das informações de parentesco pelo uso de marcadores SNPs incrementa a acurácia da predição do mérito genético para resistência de bovinos de corte a carrapatos.

## MATERIAL E MÉTODOS

Os registros de contagem de carrapatos de bovinos da raça Hereford e Braford utilizados neste estudo foram cedidos pela Conexão DeltaG, em parceria com o Laboratório de Bioinformática e Estatística Genômica (Labegen), localizado na Embrapa Pecuária Sul/Bagé-RS. Os animais participantes do programa de avaliação genética são avaliados com o objetivo de identificar animais resistentes a carrapatos. O sangue, pelo ou sêmen foram os materiais utilizados para extração do DNA e determinação do genótipo de 3.591 indivíduos. A genotipagem de 130 reprodutores foi realizada utilizando-se o painel de alta densidade (*BovineHD - Illumina® bead chip – 777.962 SNPs*), enquanto para os 3.461 animais restantes foi utilizado o painel *Illumina BovineSNP50* (54.609 SNP's).

Foram utilizados diferentes critérios de controle de qualidade para amostras e para SNPs. Os critérios adotados para exclusão de SNPs foram Equilíbrio de Hardy-Weinberg ( $P=10^{-7}$ ), taxa de genotipagem (CR) ( $< 98\%$ ), alelo menos frequente (MAF) ( $< 3\%$ ), correlação entre SNPs ( $r > 0,98$ ) e SNPs na mesma posição. Os critérios adotados para avaliação de amostras foram CR ( $< 90\%$ ), desvio de heterozigocidade acima de três desvios padrões, genótipos idênticos ( $> 99,5\%$ ) e erros de identificação de sexo, na situação de indivíduos identificados como fêmeas apresentarem genótipos heterozigotos para marcadores no cromossomo X.

A característica avaliada nesse estudo foi contagem de carrapatos transformada por  $\log_{10}(x_i + 1,0001)$ , em que  $x_i$  é a quantidade de carrapatos

contabilizada na lateral esquerda do animal. O arquivo de dados é composto por registros de 4.363 indivíduos criados a campo com infestação natural, em que 2.188 animais possuem 3 registros, 1.934 possuem 2 registros e 241 possuem somente 1 registro de contagem de carrapatos. A média e o desvio padrão para indivíduos com 1, 2 e 3 registros são  $1,3214 \pm 0,4344$ ,  $1,1941 \pm 0,4520$  e  $1,4504 \pm 0,4057$ , respectivamente.

O arquivo de dados é composto por 10.673 registros de contagem de carrapatos, dentre os quais 2.369 registros são de animais da raça Hereford e 8.304 de animais Braford com no máximo  $\frac{3}{4}$  de composição zebu. Os indivíduos mantidos no arquivo de dados apresentam entre 326 e 729 dias de idade na data da contagem. Os grupos contemporâneos (GC) foram formados pela combinação dos efeitos de fazenda, sexo, ano de nascimento, grupo de manejo e data da contagem. Foram eliminados do arquivo de dados registros que apresentavam menos de três animais e informações superiores ou inferiores a 3,5 desvios-padrão em relação ao GC. Após restrições, restaram 146 grupos contemporâneos. Os efeitos de heterozigosidade e perda por recombinação foram calculados como proposto por Cardoso e Tempelman (2004) e incluídos como covariável linear.

Foi avaliado o parentesco entre genitores-progênie dos animais genotipados. As incompatibilidades foram baseadas na contagem de conflitos mendelianos segundo a proposta de Wiggans et al. (2009). Nessa proposta o conflito é definido quando a progênie apresenta genótipo homozigoto e um dos pais também apresenta genótipo homozigoto, porém contrastante ao do filho. Quando a contagem dos conflitos excedia pelo menos 1% do total de SNPs o parentesco era rejeitado e um indivíduo alternativo foi procurado dentre a população genotipada. Assim, foi gerado um novo arquivo de pedigree (Pedigree 2). O programa SeekParentF90 (AGUILAR, 2014) foi utilizado para realizar essas avaliações.

Dados perdidos em genótipos (0,89% de todos os genótipos) foram imputados segundo a metodologia *Sliding Windows* pelo programa FImpute (SARGOLZAEI et al., 2011). A proposta de Fernandez e Toro (2006) foi utilizada para definir parentesco de meio-irmãos. A metodologia utiliza o coeficiente de correlação entre a matriz de coancestralidade e a matriz de coancestralidade molecular como critério para reconstrução do pedigree. Assim, pais fictícios foram atribuídos aos indivíduos genotipados que não

possuíam pai no Pedigree 2, o que promoveu a atribuição de novos relacionamentos genéticos de meio-irmãos. Dessa forma, foi criado o arquivo de Pedigree 3. A adoção dessa estratégia foi motivada porque os indivíduos genotipados da população são produtos de manejo reprodutivo com reprodutores múltiplos. Dessa forma, a reconstrução do pedigree pelas informações de meio-irmãos retrata a realidade do manejo reprodutivo adotado nas propriedades.

As estimativas de componentes de variância e de parâmetros genéticos foram obtidas com o uso do programa GIBBSF90, descrito por Misztal (2008), que utiliza a metodologia Bayesiana por amostragem de Gibbs. Foram considerados 500.000 ciclos, com intervalo de amostragem de 50 amostras e período de aquecimento das 100.000 primeiras amostras geradas. A convergência foi avaliada pelo Critério de Geweke (GEWEKE, 1991).

As equações de modelos mistos (HENDERSON, 1973) foram utilizadas para obtenção das predições dos valores genéticos para todos os indivíduos incluídos na análise em um modelo animal com medidas repetidas. O seguinte modelo em notação matricial foi adotado para a análise:

$$y = Xb + Zu + Wp + e,$$

em que  $y$  é o vetor de observações,  $X$ ,  $Z$  e  $W$  são matrizes de incidência que relacionam  $y$  aos vetores  $b$  de efeitos sistemáticos,  $u$  de efeito genético aditivo e  $p$  de efeito de ambiente permanente;  $e$  é o vetor dos erros aleatórios.

Os grupos contemporâneos foram incluídos como efeitos sistemáticos, enquanto proporção zebu na composição racial, heterozigosidade, perdas por recombinação gênica foram incluídas como covariáveis lineares, e a idade na data da contagem como covariável linear e quadrática.

As pressuposições assumidas para o primeiro e segundo momentos do modelo em questão foram:

$$E(y) = Xb, \quad E \begin{bmatrix} u \\ p \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad V \begin{bmatrix} u \\ p \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_u^2 & 0 & 0 \\ 0 & I\sigma_p^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

em que  $\sigma_u^2$  representa o componente de variância genética aditiva,  $\sigma_p^2$  o componente de variância de ambiente permanente,  $\sigma_e^2$  o componente de

variância residual,  $\mathbf{A}$  representa a matriz dos numeradores dos coeficientes de parentesco de Wright entre todos os indivíduos e  $\mathbf{I}$  uma matriz identidade. Sob enfoque Bayesiano, foi considerada a seguinte distribuição conjunta dos dados amostrais (função de verossimilhança) e as seguintes distribuições de probabilidade *a priori* para os parâmetros de interesse:

$$y|\mathbf{b}, \mathbf{u}, \mathbf{p}, \mathbf{e}, \sigma_u^2, \sigma_p^2, \sigma_e^2 \sim N(\mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp}, \mathbf{I}\sigma_e^2)$$

Efeito sistemático:  $\mathbf{b} \sim N(\mathbf{b}_0, \mathbf{V}_b)$ ;

Efeito genético aditivo:  $\mathbf{u}|\sigma_u^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$ ;

Efeito de ambiente permanente:  $\mathbf{p}|\sigma_p^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$ ;

A distribuição Qui-quadrado invertida escalada é assumida *a priori* para a variância residual ( $\sigma_e^2$ ), genética aditiva ( $\sigma_u^2$ ) e ambiente permanente ( $\sigma_p^2$ ):

$$\sigma_e^2 \sim S_e^2 \chi_{v_e}^{-2};$$

$$\sigma_u^2 \sim S_u^2 \chi_{v_u}^{-2};$$

$$\sigma_p^2 \sim S_p^2 \chi_{v_p}^{-2};$$

em que  $S_e^2$ ,  $S_u^2$  e  $S_p^2$  são os valores *a priori* das variâncias residual, genética aditiva e de ambiente permanente, respectivamente, e  $v_u = 5$ ,  $v_p = 5$  e  $v_e = 5$  são os graus de confiança nesses valores;  $\mathbf{V}_b$  é uma matriz diagonal da variância *a priori* de  $\mathbf{b}$ , assumindo-se  $\mathbf{V}_b < \infty$ .

Cada arquivo de pedigree apresenta estrutura de (co)variância genética específica entre os diferentes níveis da matriz dos numeradores dos coeficientes de parentesco, ou seja, diferentes estruturas de efeito genético aditivo. Neste sentido, foi considerada que cada análise representa uma modelagem específica – modelos diferentes – que foram comparadas pelo critério de informação da *deviance* (DIC). O DIC foi utilizado como critério de ajuste global e é definido em função da *deviance* (**DEV**) e pelo número efetivo de parâmetros ( $P_{D(r)}$ ) (SPIEGELHALTER et al., 2002).

$$DEV = \frac{1}{G} \sum_{i=1}^G -\log p(y|\theta^{(i)}, M_r)$$

$p_{D(r)} = DEV - D_r(\bar{\theta})$ , em que  $DEV = -2 \log p(y|\bar{\theta}, M_r)$ . Assim, o DIC para os modelos foi determinado por:

$$DIC = DEV + p_{D(r)}$$

Menores estimativas de DIC são indicativas de melhor ajuste do modelo aos dados. As probabilidades *a posteriori* foram calculadas em função do DIC utilizando a aproximação proposta por Wilberg e Bence (2008) e apresentada por Silva et al. (2011):

$$p(M_t|\hat{y}) = \frac{\exp(-\frac{\Delta_t}{2})}{\sum_{t=1}^3 \exp(-\frac{\Delta_t}{2})}$$

em que  $p(M_t|\hat{y})$  é a probabilidade *a posteriori* do modelo  $t$ ,  $\Delta_t$  é a diferença de DIC entre o modelo  $t$  e o melhor modelo (menor DIC). Para o melhor modelo essa diferença é igual a zero e a probabilidade *a posteriori* é igual a unidade.

Assumiu-se duas estratégias de cálculo para acurácia ( $\rho_{gg}^{\wedge}$ ) de predição de valor genético. A primeira estratégia foi utilizada para avaliar o ganho em acurácia ao realizar as correções no pedigree a partir de informações de SNPs e foi calculada como  $\rho_{gg}^{\wedge} = \sqrt{1 - \frac{PEV_i}{\hat{\sigma}_g^2}}$ , tal que  $\rho_{gg}^{\wedge}$  é a acurácia de predição do valor genético do animal  $i$ ,  $PEV_i$  é igual a variância do erro de predição do estimador  $\hat{g}_i$  do animal  $i$  e  $\hat{\sigma}_g^2$  é a variância genética aditiva estimada. A magnitude da acurácia média dos modelos foi avaliada da seguinte forma:

$$\Delta_{ij} = \frac{\bar{\rho}_{gg^i}}{\bar{\rho}_{gg^j}} \times 100,$$

tal que  $i=2,3$  e  $j=1,2$ , para  $i > j$ ,  $\bar{\rho}_{gg^i}$  e  $\bar{\rho}_{gg^j}$  são as acurácias médias dos arquivos de Pedigree  $i$  e  $j$ . Dessa forma, foi possível observar a superioridade, em porcentagem, da acurácia média predita do modelo  $i$  sobre o modelo  $j$ .

A segunda estratégia foi utilizada na avaliação da qualidade dos modelos por uma estratégia de validação cruzada considerando 5 grupos

distintos, considerando a proposta por Saatchi et al. (2011). Os autores propõem o agrupamento de animais de treinamento em função do parentesco pelo método K-médias, de forma que o coeficiente de parentesco dentro do grupo seja maior que o entre grupos. Uma segunda estratégia consiste em formar os grupos de treinamento selecionando os indivíduos aleatoriamente, com o objetivo de formar grupos de treinamento e validação independentes. As duas estratégias para formação dos grupos de treinamento foram utilizadas com o objetivo de avaliar diferenças na acurácia ao adotar diferentes agrupamentos para a validação cruzada. Os componentes de (co)variância estimados foram utilizados como informação *a priori* para população de treinamento na predição dos valores genéticos da população de validação. A acurácia da validação cruzada foi definida como a correlação entre o valor genético predito considerando a própria informação e predito em função do parentesco.

## RESULTADOS

Ao se utilizarem marcas SNPs, correções foram realizadas no Pedigree original (Pedigree 1) para formação do Pedigree 2. Foram identificados 460 modificações nas definições de parentesco (Tabela 1) referentes a 1.018 registros de contagem de carrapatos. Dessas modificações, 97,16% (447) foram relacionadas às informações de identificação do pai e 2,84% (13) às informações de identificação da mãe. Foram observados 199 (84,32%) conflitos em que o pai era genotipado e 13 (100%) conflitos com mães genotipadas. Foram realizadas 236 correções para definições inadequadas de touros, entretanto, apenas 37 dessas correções não apresentavam pai genotipado no pedigree original. Foi observado que 56,07% de conflitos de touro para as fêmeas e 53,38% para os machos, enquanto que 30,76% dos machos tinham conflito de vaca e nenhum para as fêmeas. O Pedigree 3 foi composto das informações presentes no Pedigree 2 e 2.174 novas definições de meio-irmãos.

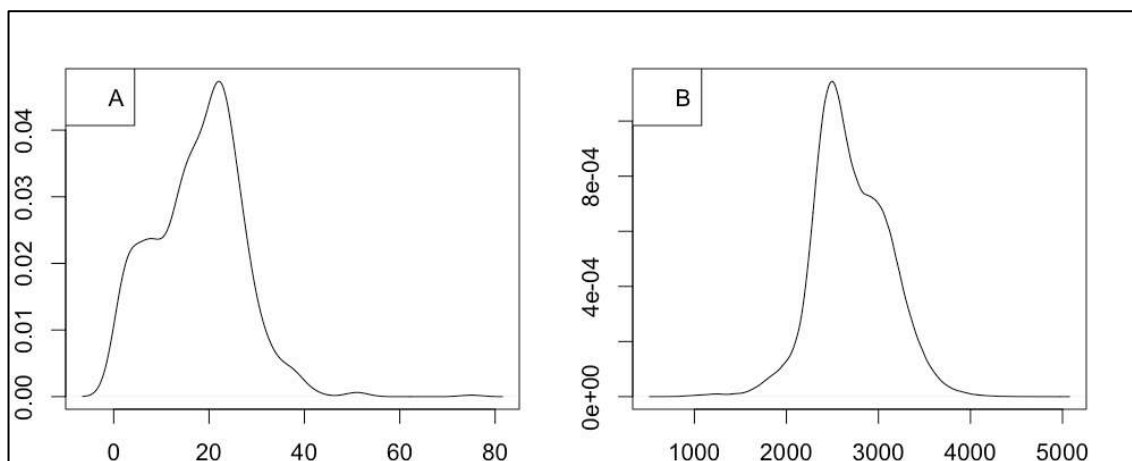
Os conflitos mendelianos identificados no Pedigree 1 foram definidos como a divergência do genótipo esperado da progênie em relação aos genótipos dos pais, caracterizado por distorção de segregação. Na Figura 1 pode ser observada a densidade dos conflitos mendelianos identificados a

partir de marcadores SNPs. A definição de parentesco foi baseada ao assumir no máximo 1% de marcas divergentes entre a progênie e um genitor por vez. A média de marcas divergentes para a confirmação do parentesco foi de 17,73 SNPs (0 a 75) e desvio-padrão de 9,28. A rejeição do parentesco apresentou em média 2.704,13 marcas (595 a 4993) em conflito e desvio-padrão de 408,68.

**Tabela 1.** Relação dos conflitos mendelianos entre genitores-progênie em função do sexo.

Parentesco	Progênie		Total
	Macho	Fêmea	
<i>Pai</i>			
Confirmado	750	293	1043
Identificado	185	26	211
Rejeitado	96	56	152
Alternativo	61	23	84
<i>Mãe</i>			
Confirmado	2127	1362	3489
Identificado	0	0	0
Rejeitado	8	0	8
Alternativo	5	0	5

Confirmado: indivíduo que manteve o mesmo pai/mãe após a avaliação de conflitos mendelianos, Identificado: indivíduo sem pai/mãe no arquivo de pedigree e identificado pai compatível, Rejeitado: indivíduo com pai/mãe não compatível, Alternativo: indivíduo com pai/mãe no arquivo de pedigree e encontrado um alternativo.



**Figura 1.** Distribuição das marcas com conflito mendeliano no Pedigree original (Pedigree 1) ao definir (A) ou rejeitar (B) o parentesco a partir de SNPs.

### *Cr terios de avalia o de modelos*

Os crit rios de sele o de modelos utilizados indicaram que o modelo que considerou o Pedigree 3 obteve melhor ajuste em rela o aos que consideraram o Pedigree 1 ou o Pedigree 2 (Tabela 2). Pode-se observar que o aumento na certeza do parentesco reduziu a DEV e o n mero efetivo de par metros, fator determinante para o DIC do modelo que utilizou o Pedigree 3 ser menor. A diferen a entre os DIC (calculado em escala log) dos modelos   ainda maior ao avaliar a m dia das probabilidades *a posteriori* (calculado em escala normal), o que comprova a superioridade em qualidade de ajuste do modelo que adotou o Pedigree 3. A diferen a entre os DIC calculados nesse estudo variou de 34 a 47 entre o modelo que utilizou o Pedigree 3 para os modelos que utilizaram o Pedigree 2 e 1, respectivamente.

O crit rio de converg ncia Geweke indicou converg ncia de todos os par metros nos tr s modelos ao gerar 500.000 cadeias de MCMC, 100.000 amostras para aquecimento e intervalo de amostragem de 50, totalizando 8.000 amostras efetivas utilizadas para estima o dos componentes de vari ncia (SILVA et al., 2013). O tamanho efetivo de amostras (TEA) estima o n mero de amostras independentes com informa es equivalentes  s contidas dentro da amostragem dependente. Dessa forma, verificou-se que o tamanho da cadeia gerada foi adequada, pois o menor TEA foi de 281 (Tabela 3).

**Tabela 2.** Estimativas dos avaliadores de ajuste bayesiano ao utilizar diferentes estruturas de (co)variância genética.

Pedigree	DEV	P <sub>D</sub>	DIC	$p(M_t \hat{c})$
1	2141,414	2224,001	4365,415	4,1345E-11
2	2136,276	2215,813	4352,089	3,2365E-08
3	2124,329	2193,268	4317,597	1,00000000

1 = arquivo de pedigree original, 2 = arquivo de pedigree com correções dos conflitos de genitor-progênie, 3 = arquivo de pedigree com correções dos conflitos de genitor-progênie e definições de meio-irmãos a partir de informações genotípicas; **DEV** = média da deviance, penalidade para número efetivo de parâmetros (**P<sub>D</sub>**), critério de informação da deviance (**DIC**), média das probabilidades a posteriori ( $p(M_t|\hat{c})$ ).

#### *Componentes de variância e parâmetros genéticos*

Pelos valores de média, moda e mediana *a posteriori* dos componentes de variância estimados para os três modelos, pode-se observar que a distribuição *a posteriori* marginal apresenta simetria, sugerindo formato campanular (Tabela 3). O modelo que utilizou o Pedigree 3 apresentou o menor intervalo de maior densidade *a posteriori* (HPD) para a variância genética aditiva e de ambiente permanente. Porém, o intervalo de todos os modelos se sobrepõem, o que indica que não há diferença estatística entre os modelos ao estimar esses parâmetros. Os desvios-padrão *a posteriori* dos parâmetros estimados foram semelhantes entre os diferentes modelos (Tabela 3).

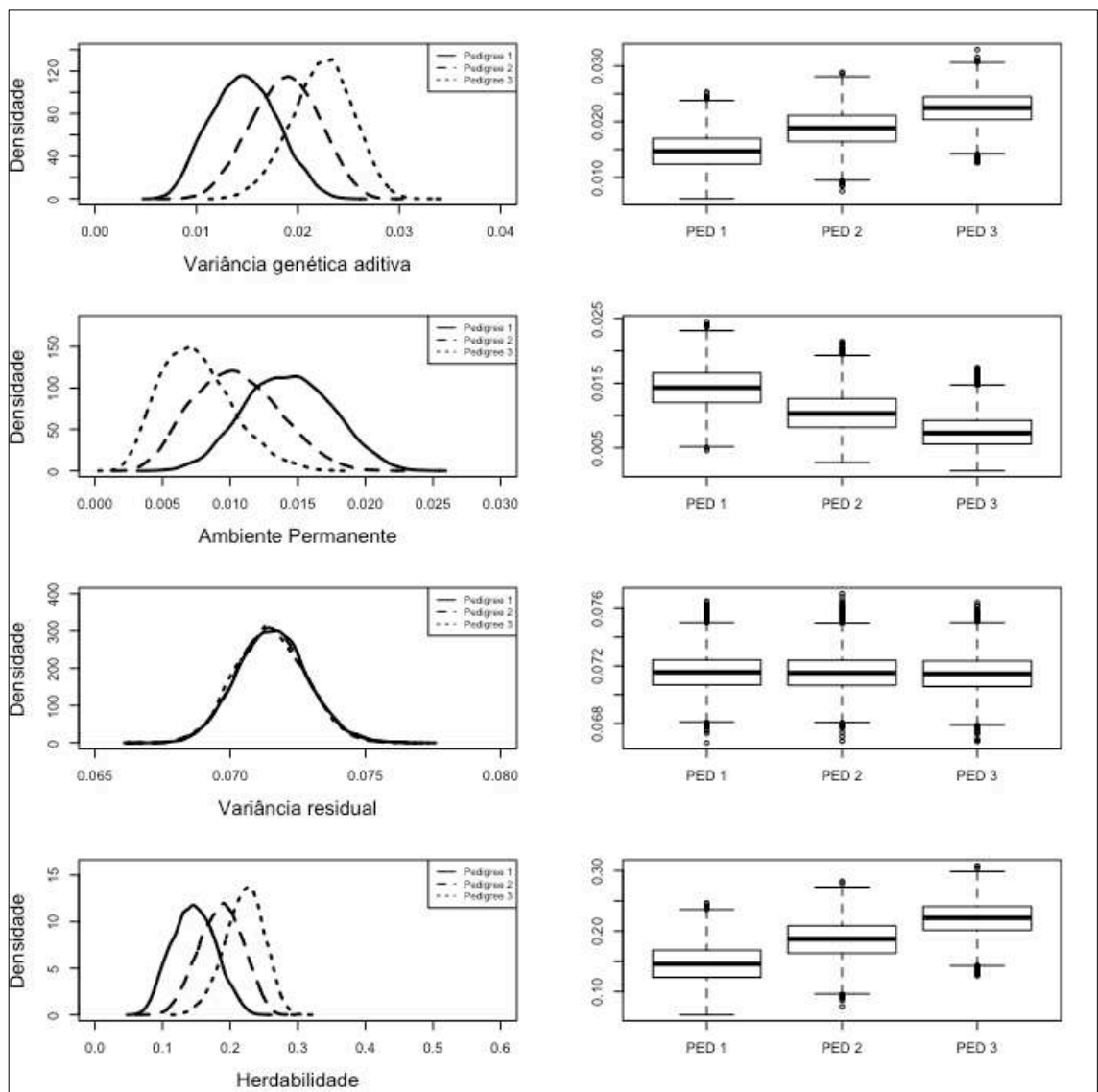
As densidades *a posteriori* das estimativas dos componentes de variância – genético aditivo, de ambiente permanente e residual – e da herdabilidade podem ser observadas na Figura 2. O aumento na qualidade das informações de parentesco proporcionou estimativas de variância genética aditiva de maior magnitude e menores para a variância de ambiente permanente. Efeitos capturados como de ambiente permanente pelo uso do Pedigree 1 foram corretamente capturados como efeito genético aditivo ao utilizarem os Pedigrees 2 e 3. Os aumentos nas estimativas de variância genética aditiva ao utilizarem os pedigrees 2 e 3 proporcionaram valores de

herdabilidade de maior magnitude. Pela Figura 2 pode-se observar que não houve efeito significativo nas estimativas de variância residual ao utilizar diferentes estruturas de (co)variância genética.

**Tabela 3.** Estatísticas observadas ao utilizar diferentes estruturas de (co)variância genética.

Pd	P	MP	MDP	MOP	PSD	HPD	Z	TEA
1	$\sigma_f^2$	0,100	0,100	0,100	0,0015	0,0977 – 0,1036	0,3209	6845
	$\sigma_u^2$	0,015	0,015	0,014	0,0032	0,0087 – 0,0209	0,1553	356
	$\sigma_p^2$	0,014	0,014	0,014	0,0032	0,0085 – 0,0207	-0,0108	368
	$\sigma_e^2$	0,072	0,071	0,071	0,0013	0,0691 – 0,0742	-1,1853	7476
	$h^2$	0,146	0,146	0,144	0,0316	0,0888 – 0,2085	0,1487	401
	$r$	0,288	0,288	0,288	0,0121	0,2647 – 0,3125	1,5523	6679
2	$\sigma_f^2$	0,100	0,101	0,100	0,0015	0,0984 – 0,1044	-0,0883	6499
	$\sigma_u^2$	0,019	0,019	0,019	0,0033	0,0122 – 0,0248	-0,5671	348
	$\sigma_p^2$	0,010	0,010	0,010	0,0031	0,0048 – 0,0166	0,7832	323
	$\sigma_e^2$	0,071	0,071	0,071	0,0013	0,0691 – 0,0741	-1,7512	7325
	$h^2$	0,186	0,187	0,189	0,0321	0,1230 – 0,2455	-0,5691	440
	$r$	0,289	0,289	0,290	0,0120	0,2661 – 0,3133	1,8360	6152
3	$\sigma_f^2$	0,101	0,101	0,101	0,0016	0,0985 – 0,1044	1,3220	5812
	$\sigma_u^2$	0,022	0,022	0,023	0,0030	0,0162 – 0,0279	0,3785	318
	$\sigma_p^2$	0,007	0,007	0,007	0,0027	0,0029 – 0,0129	-0,2182	281
	$\sigma_e^2$	0,072	0,071	0,071	0,0013	0,0690 – 0,0740	0,3076	7666
	$h^2$	0,220	0,222	0,227	0,0286	0,1604 – 0,2716	0,3482	437
	$r$	0,294	0,295	0,296	0,0123	0,2708 – 0,3192	0,9596	6126

1 = arquivo de pedigree original; 2 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie; 3 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie e definições de meio-irmãos a partir de informações genotípicas;  $\sigma_f^2$  = variância fenotípica;  $\sigma_u^2$  = variância genética aditiva,  $\sigma_p^2$  = variância de ambiente permanente;  $\sigma_e^2$  = variância residual,  $h^2$  = herdabilidade e  $r$  é a repetibilidade; Arquivo de pedigree (Pd), parâmetro (P), média a posteriori (MP), mediana a posteriori (MDP), moda a posteriori (MOP), desvio-padrão a posteriori (PSD), intervalo de maior densidade a posteriori (HPD), Z-Geweke (Z), tamanho efetivo de amostras (TEA).



**Figura 2.** Densidade (esquerda) e diagramas em caixa (direita) das estimativas *a posteriori* da variância genética aditiva, de ambiente permanente, residual e da herdabilidade considerando diferentes estruturas de (co)variância genética ao adotar diferentes Pedigrees (PED).

### *Predição de valor genético e acurácias*

Foram avaliadas as correlações de Spearman entre os valores genéticos preditos considerando diferentes proporções de candidatos à seleção (Tabela 4). A magnitude da correlação dos valores genéticos preditos foi maior entre os modelos que utilizaram os Pedigrees 1 e 2 quando comparado aos Pedigrees 1 e 3 ou 2 e 3. Quanto menor a proporção de indivíduos selecionados, menor é a correlação entre os modelos, principalmente entre 1 e 3. As correções realizadas para o Pedigree 3 foram responsáveis por modificar de forma

significativa a matriz de (co)variância genética e isto pode ser observado pela diferença na magnitude da correlação entre os valores genéticos.

A correlação de Spearman entre as acurácias dos valores genéticos preditos nas diferentes análises foram positivas e de alta magnitude (Tabela 5). Isso retrata que, de forma geral, as mudanças na magnitude da acurácia entre os modelos apresentaram tendência semelhante. Correlação positiva e de alta magnitude pôde ser observada em todas as proporções de seleção. Gráficos de dispersão da acurácia dos valores genéticos preditos entre os diferentes arquivos de pedigree dos 10% melhores indivíduos podem ser visualizados na Figura 3. É possível observar nessa figura que o aumento da qualidade nos relacionamentos genéticos proporcionou estimativas de acurácia de maior magnitude. Isso pode ser verificado no limite inferior das dispersões que relacionam indivíduos com estimativas de acurácia igual a zero pelo Pedigree 1, enquanto pelo Pedigree 2 (Figura 3A) e 3 (Figura 3B) relacionam valores maiores que 0,20.

**Tabela 4.** Correlação de Spearman entre os valores genéticos preditos considerando diferentes estruturas de (co)variância genética.

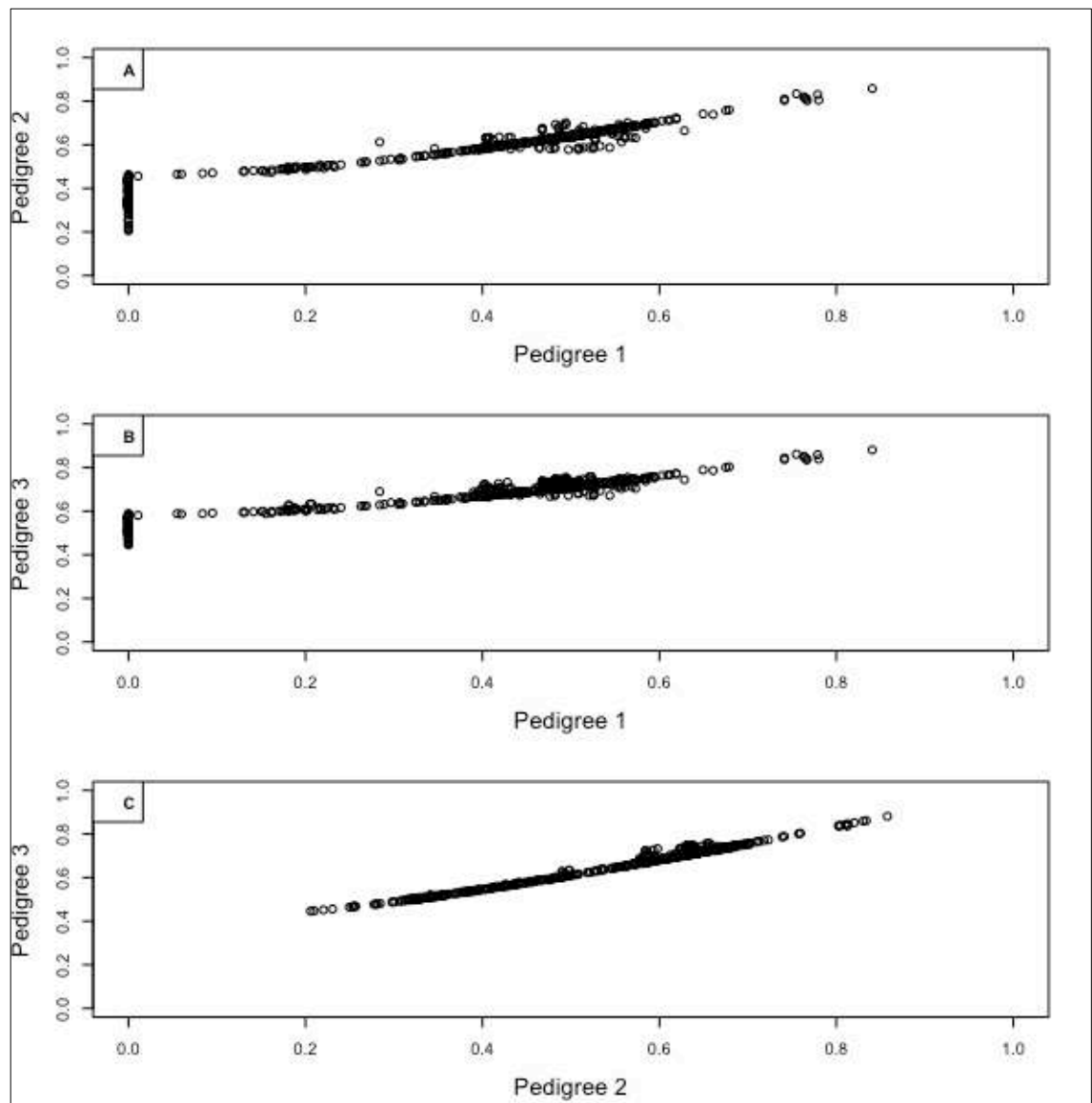
Pedigree	Porcentagem dos melhores indivíduos selecionados						
	1	5	10	20	30	50	100
1 – 2	0,953*	0,968*	0,974*	0,979*	0,980*	0,981*	0,987*
1 – 3	0,845*	0,911*	0,928*	0,946*	0,954*	0,960*	0,974*
2 – 3	0,890*	0,937*	0,950*	0,965*	0,972*	0,976*	0,987*

1 = arquivo de pedigree original, 2 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie, 3 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie e definições de meio-irmãos a partir de informações genotípicas. \*P < 0,0

**Tabela 5.** Correlação de Spearman entre acurácias dos valores genéticos preditos entre os três arquivos de pedigree.

Pedigree <sup>a</sup>	Porcentagem dos melhores indivíduos selecionados						
	1	5	10	20	30	50	100
1 – 2	0,971*	0,971*	0,971*	0,965*	0,953*	0,928*	0,924*
1 – 3	0,947*	0,964*	0,972*	0,972*	0,965*	0,954*	0,952*
2 – 3	0,986*	0,987*	0,992*	0,994*	0,994*	0,994*	0,993*

1 = arquivo de pedigree original, 2 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie, 3 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie e definições de meio-irmãos a partir de informações genotípicas. \*P < 0,05



**Figura 3.** Dispersão das estimativas de acurácia dos valores genéticos preditos ao utilizar o Pedigree 1 e 2 (**A**), Pedigree 1 e 3 (**B**) e Pedigree 2 e 3 (**C**).

O modelo que utilizou o Pedigree 3 estimou acurácia média maior em relação aos modelos que utilizaram os Pedigrees 1 ou 2 (Tabela 6). Ao aplicar maiores proporção de seleção (1% e 5%) as diferenças entre as acurácias são menores em magnitude. Entretanto, ao selecionar menor proporção de candidatos (1, 5 ou 10% de indivíduos selecionados), a diferença entre as acurácias torna-se mais evidente.

**Tabela 6.** Relação de superioridade ( $\Delta$ ) na magnitude da acurácia média entre os diferentes arquivos de pedigree utilizados para compor a matriz dos numeradores dos coeficientes de parentesco de Wright.

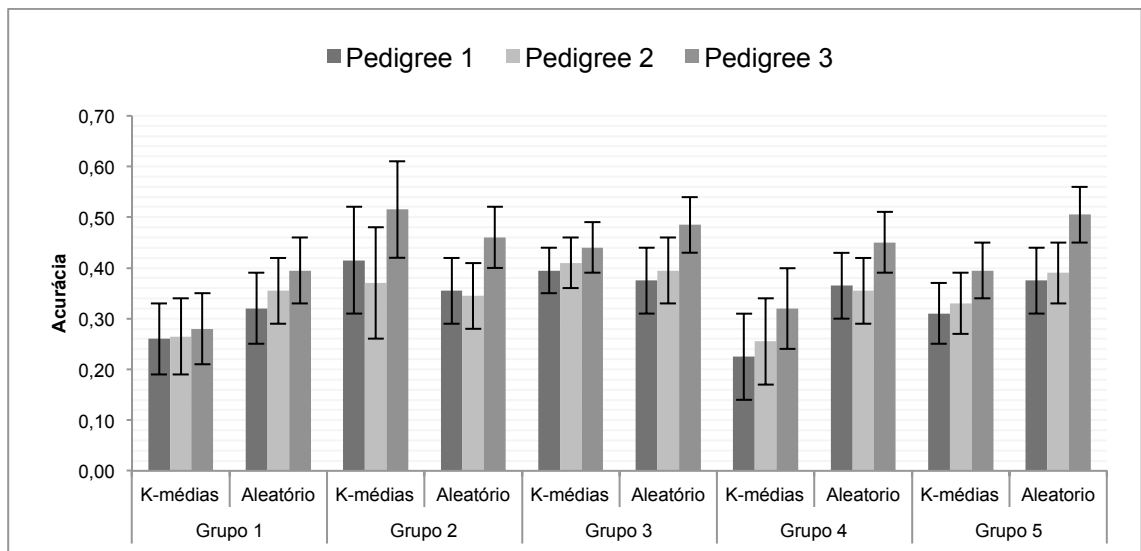
Pedigree <sup>a</sup>	Porcentagem dos melhores indivíduos selecionados						
	1	5	10	20	30	50	100
$\Delta_{21}$	34,8%	38,2%	46,3%	51,3%	50,3%	52,3%	49,5%
$\Delta_{31}$	53,4%	56,8%	69,2%	75,3%	74,9%	77,2%	72,4%
$\Delta_{32}$	14,1%	17,3%	20,9%	27,9%	35,1%	51,9%	56,6%

<sup>a</sup>1 = arquivo de pedigree original, 2 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie, 3 = arquivo de pedigree corrigido pelo conflito de parentesco entre genitor-progênie e definições de meio-irmãos a partir de informações genotípicas.

#### *Validação cruzada*

Foram utilizadas duas estratégias de agrupamento para a validação cruzada propostas por Saatchi et al. (2011). O método K-médias, primeira estratégia, agrupou os indivíduos em função do parentesco, de forma que o relacionamento genético dentro de cada grupo fosse maior que entre grupos. A segunda estratégia particionou o arquivo de dados selecionando-se os indivíduos ao acaso.

Nessa etapa as informações de contagem de carrapatos do grupo de validação foram eliminadas do arquivo de dados e o valor genético desses animais predito em função do parentesco. A acurácia da valor genético da validação foi calculada pela correlação entre o valor genético predito em função do parentesco e ao considerar a própria informação. A Figura 4 apresenta a acurácia média e o intervalo com 95% de confiança nos diferentes grupos de validação e estratégias de agrupamento. É possível observar que não houve diferença estatística nas diferentes estratégias de agrupamento apesar de a seleção aleatória dos animais para validação ter proporcionado acurácias de maior magnitude. De forma geral, maior qualidade nas informações de parentesco proporcionou acurácia média de maior magnitude nos diferentes grupos de validação e estratégias de agrupamento.



**Figura 4.** Acurácia média e intervalo de confiança nos diferentes grupos de validação e estratégias de agrupamento.

## DISCUSSÃO

### *Conflitos mendelianos*

Esse estudo avaliou o ganho em predição de valor genético e estimação da acurácia ao corrigir os relacionamentos genéticos definidos no pedigree de bovinos de corte a partir de informações de SNPs de forma avaliar a resistência desses animais a carrapatos.

Foi observado aumento na qualidade do relacionamento genético de 460 animais, sendo que a maior parte dos erros de identificação se referiram às informações de pai (97,16%). Os registros de contagem de carrapatos utilizados nesse estudo são provenientes de propriedades que utilizam manejo reprodutivo com reprodutores múltiplos, motivo pelo qual 210 animais não apresentavam pai definido no pedigree original. A definição de pai para esses animais reflete em expressivo ganho em relacionamentos genéticos na população. Entretanto, conflitos foram observados nas identificações de mães genotipadas, o que sugere baixo percentual de falha no controle reprodutivo das propriedades. De fato, verificou-se que 100% dos conflitos das mães eram provenientes de vacas genotipadas. Resultados diferentes foram relatados por Wiggans et al. (2012) ao avaliarem conflitos mendelianos entre genitor-progênie de bovinos Holandeses, Jersey e Pardo-Suiços usando SNPs com 3K

e encontraram conflitos de touro e vaca de igual proporção. A identificação e correção desses erros proporcionaram o aumento da qualidade na estimação dos parâmetros genéticos, principalmente em magnitude.

A definição do parentesco foi realizada com pequena divergência entre as marcas do indivíduo e de seus pais, o que demonstra elevada confiança ao considerar o relacionamento como verdadeiro. Apesar da identificação de conflitos mendelianos ter sido realizada avaliando-se um genitor por vez, a correção de erros no pedigree pôde ser efetuada com elevada acurácia com 41.045 SNPs, sugerindo elevada precisão das correções nos parentesco. Wiggans et al. (2009) encontraram resultados divergentes do presente estudo, ao utilizarem 37.811 marcadores SNPs, encontraram em média 2,3 (0 a 89) conflitos para a definição de parentesco e 2.411 (754 a 3.507) para parentesco rejeitado em uma população de bovinos da raça Holandesa.

#### *Avaliadores de ajuste dos modelos*

O DIC e o CPO do modelo que utilizou o Pedigree 3 foram menores do que aqueles encontrados quando utilizado o Pedigree 1 ou o Pedigree 2 (Tabela 2), o que comprova a superioridade de ajuste do modelo que adotou o Pedigree 3. O DIC foi proposto com objetivo de avaliar a qualidade de ajuste em modelos hierárquicos bayesianos (WILBERG e BENCE, 2008) e sugere como significativa diferença maior que sete unidades entre modelos (SPIEGELHALTER et al., 2002). Esse critério tem sido aplicado em diversos estudos e utilizado como critério de seleção de modelos (KIZILKAYA e TEMPELMAN, 2005; SHIOTSUKI et al., 2012; SHIOTSUKI et al., 2013).

#### *Componentes de variância e parâmetros genéticos*

O HPD é o intervalo que maximiza a densidade *a posteriori* de forma que quanto menor a amplitude entre os quantis 2,5 e 97,5% maior é a precisão dos parâmetros estimados. o HPD do Pedigree 3 para os efeitos genéticos aditivos e de ambiente permanente sugerem que as correções realizadas proporcionaram estimativas mais próximas da verdade que os Pedigrees 1 ou 2, apesar das estimativas entre modelos não serem estatisticamente diferentes (Figura 1). Apesar do HPD das estimativas do Pedigree 3 ter sido menor, não

observou-se diferença entre a média *a posteriori* dos desvios-padrão, sugerindo que apesar de encontradas estimativas de maior magnitude a dispersão desses dados não se modificou de forma significativa entre os modelos.

As correções no parentesco proporcionaram estimativas de herdabilidade de maior magnitude (Figura 2), sugerindo que a maior qualidade nos relacionamentos genéticos definidos no pedigree direcionam maior resposta a seleção. Estimativas de maior magnitude para esse parâmetro sugerem a possibilidade de se praticar a seleção direta para resistência de bovinos a carrapatos. A principal consequência de erros no pedigree é a redução na taxa de ganho genético anual, principalmente por herdabilidades subestimadas e incorreta predição de valor genético, com inadequada seleção dos candidatos (WELLER et al., 2004; HARDER et al., 2005; SANDERS et al., 2006; MUNOZ et al., 2014). A média *a posteriori* da herdabilidade do modelo que utilizou o Pedigree 3 apresentou maior magnitude. A estimação correta e acurada desse parâmetro apresenta implicação no sucesso dos programas de melhoramento genético, pois afeta a taxa de ganho genético.

#### *Implicações das correções no parentesco no processo de seleção*

As correlações entre os valores genéticos (Tabela 4) e entre acurácias dos valores genéticos (Tabela 5) relatam elevadas magnitudes em praticamente todas as proporções de seleção dos candidatos ao utilizar diferentes pedigrees. A correlação de elevada magnitude observada sugere relação linear entre os valores genéticos preditos e entre acurácias estimadas ao considerar diferentes níveis de correção no parentesco. Entretanto, as estimativas de correlação não mostram a diferença de escala entre as variáveis analisadas, sugerindo que não é indicado utilizar somente o coeficiente de correlação linear para concluir sobre as estimativas de valores genéticos preditos e acurácias estimadas.

Nessa situação, a avaliação da dispersão das acurácias (Figura 3) passa a apresentar maior importância, pois é possível observar o ganho em magnitude ao realizar correções no parentesco. Banos et al. (2001) relatam que erros na definição do parentesco modificam as médias de família e também afetam a estimação da endogamia. Assim, a predição do mérito

genético dos candidatos também é modificada, o que pode estimar valores genéticos de diferentes magnitudes. Resultados semelhantes ao do presente estudo foram obtidos por Munoz et al. (2014), que verificaram que correções no pedigree causam aumento nas estimativas das acurácias. Esses autores justificam que erros na definição de parentesco levam a estimativas de variância genética aditiva incorretas, o que, por sua vez, reduz a acurácia da predição dos valores genéticos e da avaliação genética.

### *Validação cruzada*

Os resultados obtidos pela formação do grupo de validação utilizando o método K-médias revelaram acurácias de menor magnitude nos grupos 1 e 4, sugerindo laços genéticos distantes com a população de treinamento. Com exceção do grupo de validação 2 (Figura 4), o agrupamento ao acaso apresentou acurácia média de maior magnitude em relação ao K-médias. Entretanto, a inspeção do intervalo com 95% de confiança relata não significância nas acurácias estimadas dentro e entre grupos de treinamento. Saatchi et al. (2011) encontraram resultados divergentes, pois a acurácia média do agrupamento aleatório apresentou valores médios de maior magnitude. Entretanto, os autores não associaram um teste estatístico que avaliasse a igualdade de médias nas diferentes estratégias adotadas. Segundo Pérez-Cabal et al. (2012), a avaliação da qualidade das predições de valor genético pode ser realizada utilizando-se técnicas de validação cruzada em que se elimina indivíduos geneticamente relacionados. Entretanto, esses mesmos autores sugerem que na prática isso raramente é possível, pois são encontrados diversos níveis de relacionamento genético na população. Dessa forma, a estratégia utilizada para montar os grupos de treinamento tem influência nos resultados da validação cruzada.

Esse estudo é o primeiro a apresentar o reflexo das correções de parentesco de bovinos de corte ao avaliar a herança de resistência a carrapatos com o uso de SNPs. Os carrapatos são responsáveis por prejuízos no sistema produtivo por debilitar a saúde dos bovinos. Nesse sentido, a identificação dos genótipos resistentes a esse ectoparasita deve ser considerada como objetivo de seleção bem como sua relação com outras características de interesse econômico.

A presença de erros na identificação de parentesco reduz a acurácia na seleção dos genótipos superiores e o uso de informações de SNPs para a composição dos relacionamentos genéticos verdadeiros passa a ter grande importância durante esse processo.

## CONCLUSÃO

A utilização de SNPs para avaliação de conflitos entre genitores e progênie aumenta a certeza nos relacionamentos de parentesco e conseqüentemente a acurácia nas predições dos valores genéticos dos animais candidatos à seleção.

## BIBLIOGRAFIA

AGUILAR, I. **SeekParentF90** 2014. Disponível em: <<http://nce.ads.uga.edu/wiki/doku.php?id=readme.seekparentf90%3E>>. Acesso em 10 de outubro de 2014.

BANOS, G.; WIGGANS, G. R.; POWELL, R. L. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. **Journal of Dairy Science**, v. 84, n. 11, p. 2523-2529, 2001.

BARUCH, E.; WELLER, J. I. Estimation of the number of SNP genetic markers required for parentage verification. **Animal Genetics**, v. 39, n. 5, p. 474-479, 2008.

CARDOSO, F. F.; TEMPELMAN, R. J. Hierarchical Bayes multiple-breed inference with an application to genetic evaluation of a Nelore-Hereford population. **Journal of Animal Science**, v. 82, n. 6, p. 1589-1601, 2004.

CARNEIRO, P. L. S.; EUCLYDES, R. F.; SILVA, M. A.; LOPES, P. S.; TORRES, R. A.; CARNEIRO, A. P. S.; TORRES, R. A. T. Efeito de Erros de Pedigree na Seleção. **Revista Brasileira de Zootecnia**, v. 28, n. 2, p. 269-274, 1999.

FERNANDEZ, J.; TORO, M. A. A new method to estimate relatedness from molecular markers. **Molecular Ecology**, v. 15, n. 6, p. 1657-1667, 2006.

GASPARIN, G.; MIYATA, M.; COUTINHO, L. L.; MARTINEZ, M. L.; TEODORO, R. L.; FURLONG, J.; MACHADO, M. A.; SILVA, M. V. G. B.; SONSTEGARD, T. S.; REGITANO, L. C. A. Mapping of quantitative trait loci controlling tick [*Rhipicephalus (Boophilus) microplus*] resistance on bovine chromosomes 5, 7 and 14. **Animal Genetics**, v. 38, n. 5, p. 453-459, 2007.

GELDERMANN, H.; PIEPER, U.; WEBER, W. E. Effect of misidentification on the estimation of breeding value and heritability in cattle. **Journal of Animal Science**, v. 63, n. 6, p. 1759-1768, 1986.

GEWEKE, J. **Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments**. Federal Reserve Bank of Minneapolis, Research Department, 1991.

HABIER, D.; TETENS, J.; SEEFRIED, F.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution**, v. 42, n. 1, p. 5, 2010.

HARDER, B.; BENNEWITZ, J.; REINSCH, N.; MAYER, M.; KALM, E. Effect of missing sire information on genetic evaluation. **Archiv Tierzucht**, v. 48, n. 1-3, p. 219-232, 2005.

HENDERSON, C. R. Sire evaluation and genetic trends. In: PROCEEDINGS OF THE ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF DR. JAY L. LUSH, 1973, **Anais.**, 1973. p. 10-41.

ISRAEL, C.; WELLER, J. I. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. **Journal of Dairy Science**, v. 83, n. 1, p. 181-187, 2000.

JONSSON, N. N. The productivity effects of cattle tick (*Boophilus microplus*) infestation on cattle, with particular reference to *Bos indicus* cattle and their crosses. **Veterinary Parasitology**, v. 137, n. 1, p. 1-10, 2006.

KIZILKAYA, K.; TEMPELMAN, R. J. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. **Genetics Selection Evolution**, v. 37, n. 1, p. 1-26, 2005.

MISZTAL, I. Reliable computing in estimation of variance components. **Journal of Animal Breeding and Genetics**, v. 125, n. 6, p. 363-370, 2008.

MUNOZ, P. R.; RESENDE, M. F. R.; HUBER, D. A.; QUESADA, T.; RESENDE, M. D. V.; NEALE, D. B.; WEGRZYN, J. L.; KIRST, M.; PETER, G. F. Genomic relationship matrix for correcting pedigree errors in breeding populations: Impact on genetic parameters and genomic selection accuracy. **Crop Science**, v. 54, n. 3, p. 1115-1123, 2014.

OLIVEIRA, M. C. S.; ALENCAR, M. M.; GIGLIOTI, R.; BERALDO, M. C. D.; ANÍBAL, F. F.; CORREIA, R. O.; BOSCHINI, L.; CHAGAS, A. C. S.; BILHASSI, T. B.; OLIVEIRA, H. N. Resistance of beef cattle of two genetic groups to ectoparasites and gastrointestinal nematodes in the state of São Paulo, Brazil. **Veterinary Parasitology**, v. 197, n. 1, p. 168-175, 2013.

PÉREZ-CABAL, M.; VAZQUEZ, A. I.; GIANOLA, D.; ROSA, G. J. M.; WEIGEL, K. A. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. **Livestock Genomics**, v. 3, p. 27, 2012.

PORTO NETO, L. R.; JONSSON, N. N.; D'OCCHIO, M. J.; BARENDSE, W. Molecular genetic approaches for identifying the basis of variation in resistance

to tick infestation in cattle. **Veterinary Parasitology**, v. 180, n. 3, p. 165-172, 2011.

SAATCHI, M.; MCCLURE, M. C.; MCKAY, S. D.; ROLF, M. M.; KIM, J.; DECKER, J. E.; TAXIS, T. M.; CHAPPLE, R. H.; RAMEY, H. R.; NORTHCUTT, S. L. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics Selection Evolution**, v. 43, p. 40, 2011.

SANDERS, K.; BENNEWITZ, J.; KALM, E. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. **Journal of Dairy Science**, v. 89, n. 1, p. 315-321, 2006.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. FImpute-An efficient imputation algorithm for dairy cattle populations. **Journal of Dairy Science**, v. 94, p. 421, 2011.

SHIOTSUKI, L.; CARDOSO, F. F.; SILVA, J. A.; ALBUQUERQUE, L. G. Comparison of a genetic group and unknown paternity models for growth traits in Nellore cattle. **Journal of Animal Science**, v. 91, n. 11, p. 5135-5143, 2013.

SHIOTSUKI, L.; CARDOSO, F. F.; SILVA, J. A.; ROSA, G. J. M.; ALBUQUERQUE, L. G. Evaluation of an average numerator relationship matrix model and a Bayesian hierarchical model for growth traits in Nellore cattle with uncertain paternity. **Livestock Science**, v. 144, n. 1, p. 89-95, 2012.

SILVA, F. F.; ROSA, G. J. M.; GUIMARÃES, S. E. F.; LOPES, P. S.; DE LOS CAMPOS, G. Three-step Bayesian factor analysis applied to QTL detection in crosses between outbred pig populations. **Livestock Science**, v. 142, n. 1, p. 210-215, 2011.

SILVA, F. F.; VIANA, J. M. S.; FARIA, V. R.; RESENDE, M. D. V. Bayesian inference of mixed models in quantitative genetics of crop species. **Theoretical and Applied Genetics**, v. 126, n. 7, p. 1749-1761, 2013.

SIMEONE, R.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A. Evaluation of the utility of diagonal elements of the genomic relationship matrix as a diagnostic tool to detect mislabelled genotyped animals in a broiler chicken population. **Journal of Animal Breeding and Genetics**, v. 128, n. 5, p. 386-393, 2011.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 64, n. 4, p. 583-639, 2002.

VAN VLECK, L. D. Misidentification and sire evaluation. **Journal of Dairy Science**, v. 53, n. 12, p. 1697-1702, 1970a.

\_\_\_\_\_. Misidentification in estimating the paternal sib correlation. **Journal of Dairy Science**, v. 53, n. 10, p. 1469-1474, 1970b.

WELLER, J. I.; FELDMESSER, E.; GOLIK, M.; TAGER-COHEN, I.; DOMOCHOVSKY, R.; ALUS, O.; EZRA, E.; RON, M. Factors affecting incorrect

paternity assignment in the Israeli Holstein population. **Journal of Dairy Science**, v. 87, n. 8, p. 2627-2640, 2004.

WIGGANS, G. R.; COOPER, T. A.; VANRADEN, P. M.; OLSON, K. M.; TOOKER, M. E. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. **Journal of Dairy Science**, v. 95, n. 3, p. 1552-1558, 2012.

WIGGANS, G. R.; SONSTEGARD, T. S.; VANRADEN, P. M.; MATUKUMALLI, L. K.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S.; VAN TASSELL, C. P. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. **Journal of Dairy Science**, v. 92, n. 7, p. 3431-3436, 2009.

WIGGANS, G. R.; VANRADEN, P. M.; BACHELLER, L. R.; TOOKER, M. E.; HUTCHISON, J. L.; COOPER, T. A.; SONSTEGARD, T. S. Selection and management of DNA markers for use in genomic evaluation. **Journal of Dairy Science**, v. 93, n. 5, p. 2287-2292, 2010.

WILBERG, M. J.; BENCE, J. R. Performance of deviance information criterion model selection in statistical catch-at-age analysis. **Fisheries Research**, v. 93, n. 1-2, p. 212-221, 2008.

## CAPÍTULO II

Avaliação genética de bovinos de corte considerando incerteza de paternidade e animais oriundos de biotécnicas da reprodução

**RESUMO:** Esse estudo teve como objetivo avaliar metodologias adequadas para inclusão de informações de filhos de reprodutores múltiplos e animais oriundos de transferência de embriões ou fertilização *in vitro* na avaliação genética para peso padronizado aos 205 dias. Foram utilizados registros de bovinos da raça Hereford, Braford, Nelore, Brahman e Tabapuã em análises unicaracterísticas. Utilizou-se a metodologia hierárquica bayesiana (HIER), matriz da média dos numeradores dos coeficientes de parentesco de Wright (ANRM) e grupos genéticos para avaliar a inclusão de informações de reprodutores múltiplos. A inclusão de informações de animais oriundos de biotecnologias da reprodução (TEF) foi realizada pelo uso da receptora para estimação dos efeitos genético materno e de ambiente permanente. O critério de informação da *deviance* (DIC) e a ordenada preditiva condicional (CPO) foram utilizados para avaliação de ajuste. O modelo GG apresentou pior ajuste ao incluir animais RM, segundo os critérios assumidos (CPO e DIC). Segundo o CPO, o modelo HIER apresentou melhor ajuste. Entretanto, pelo DIC o modelo ANRM apresentou melhor ajuste global. Ao incluir as informações dos animais TEF, o CPO indicou que o modelo ANRM apresentou melhor ajuste aos dados. Não foram observadas diferenças estatísticas nas estimativas dos componentes de variância e parâmetros genéticos entre os diferentes modelos, porém, as magnitudes dos valores genéticos preditos foram diferentes. Incluir informações de filhos de reprodutores múltiplos e animais oriundos de transferência de embriões e fertilização *in vitro* com metodologia apropriada altera de forma significativa a predição do valor genético e, conseqüentemente, modifica a seleção dos candidatos.

**Palavras-chave:** amostrador de Gibbs, tendência genética, transferência de embriões

Genetic evaluation under uncertainty of paternity and biotechnology  
reproduction techniques for weaning weights of beef cattle

**ABSTRACT:** This study aimed to evaluate appropriate methodologies for uncertain paternity in multiple sire schemes (MS), embryo transfer and *in vitro* fertilization (TEF) in genetic evaluation for weaning weight. We used records of Hereford, Braford, Nellore, Brahman and Tabapuã beef cattle. We fit hierarchical Bayesian methodology (HIER), average numerator relationship matrix (ANRM) and genetic groups (GG) to evaluate MS progeny. We used the foster dam information to estimate maternal genetic and permanent environmental effects under the TEF schemes. The *deviance* information criteria (DIC) and conditional predictive ordinate (CPO) were used as Bayesian measure of fit. The GG model presented the worse fit according to DIC and CPO. According to CPO, HIER model presented the best fit. However, DIC suggested that ANRM model showed better overall fit. By the use of TEF information, CPO suggests that ANRM model presented better fit. We didn't observed statistical differences on variance components and genetic parameters estimates, however, breeding values were different. The use of MS progeny and TEF information with appropriate methodology modifies the predicted breeding values and the selected candidates.

**Keywords:** embryo transfer, genetic trend, heritability

## INTRODUÇÃO

A demanda por proteína animal é cada vez maior e estratégias para maximizar a eficiência reprodutiva passam a ser importantes para a produção de maior quantidade de alimento. Dentre as estratégias utilizadas para se alcançar maior eficiência reprodutiva em bovinos de corte estão o uso de reprodutores múltiplos (RM) e biotecnologias da reprodução, tais como transferência de embriões (TE) e a fertilização *in vitro* (FIV).

O manejo reprodutivo com uso de reprodutores múltiplos (RM) ocorre pela disposição de mais de um touro em um grupo de vacas. Do ponto de vista reprodutivo, este tipo de manejo tem como objetivo aumentar a taxa de prenhez e os índices reprodutivos. Entretanto, geram problemas para a adoção de métodos de avaliação genética, pois impossibilita a correta identificação da paternidade dos animais nascidos neste sistema de manejo. Com o advento dos testes de compatibilidade genética, realizado pela comparação do DNA do indivíduo, da mãe e dos possíveis pais, a identificação do pai verdadeiro pode ser realizada, porém gera custos ao produtor. Neste sentido, diversas metodologias têm sido propostas para contornar esse problema na avaliação genética (HENDERSON, 1988; WESTELL et al., 1988; CARDOSO e TEMPELMAN, 2003; SAPP et al., 2007).

A metodologia proposta por Westell et al. (1988) agrupa indivíduos de acordo com critérios definidos pelo pesquisador e considera que o valor genético de cada indivíduo também está em função dos grupos formados. O principal questionamento sobre o uso dessa metodologia ao incluir animais produtos de RM é que se considera que os touros não são aparentados, o que pode não ser verdadeiro.

Henderson (1988) propôs o uso da matriz dos numeradores dos coeficientes de parentesco ajustada para incluir a informação de probabilidade *a priori* de um animal ser filho de determinado touro, essa matriz ficou conhecida como matriz da média dos numeradores dos coeficientes de parentesco (ANRM). Essa proposta permite a atribuição de probabilidades específicas para os indivíduos oriundos de RM quando se conhece o comportamento reprodutivo dos reprodutores. De forma geral, tem-se utilizado chances iguais para os reprodutores serem os potenciais pais. Cardoso e Tempelman (2003) propuseram a metodologia hierárquica bayesiana com

inclusão de indivíduos com incerteza de paternidade que utiliza Monte Carlo via Cadeias de Markov (MCMC) por amostragem de Gibbs.

Dentre as diversas vantagens que técnicas reprodutivas como a transferência de embriões (TE) e a fertilização *in vitro* (FIV) trouxeram para a pecuária, possivelmente a mais importante seja a rápida multiplicação de indivíduos geneticamente superiores. Nessas técnicas, os embriões gerados pela doadora – no caso da transferência de embriões – e *in vitro* – no caso da fertilização *in vitro* – são implantados em fêmeas receptoras.

Entretanto, não existem estudos que avaliam o reflexo da inclusão das informações de filhos de reprodutores múltiplos e de animais oriundos de biotécnicas da reprodução com metodologias apropriadas apresentadas na literatura (HENDERSON, 1988; SCHAEFFER e KENNEDY, 1989; CARDOSO e TEMPELMAN, 2003).

Este estudo teve como objetivo avaliar metodologias que possibilitam a inclusão da informação de indivíduos produtos de reprodutores múltiplos e de biotecnologias da reprodução (TEF), tais como a transferência de embrião e fertilização *in vitro*, na avaliação genética de características de crescimento de bovinos de corte.

## **MATERIAL E MÉTODOS**

Os registros de peso corporal de bovinos são provenientes do Programa de Melhoramento Genético PampaPlus, pertencente a Associação Brasileira de Hereford e Braford (ABHB), e cedidos pelo Laboratório de Bioinformática e Estatística Genômica (LabeGen), localizado na Embrapa Pecuária Sul, Bagé, RS. O arquivo de pedigree utilizado é composto por informações de bovinos da raça Hereford, Braford ( $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$ ,  $\frac{3}{4}$ ,  $\frac{5}{8}$ ), Brahman, Tabapuã e Nelore.

Os animais são mantidos à pasto com sal mineral à vontade. Os sistemas de acasalamento utilizados são monta natural, inseminação artificial e reprodutores múltiplos (RM). O manejo reprodutivo com mais de um touro no grupo é realizado disponibilizando para um mesmo grupo de fêmeas mais de um touro simultaneamente. Biotécnicas da reprodução como transferência de embriões (TE) e fertilização *in vitro* (FIV) são utilizadas. Os nascimentos ocorrem durante todo o ano, porém, são concentrados nos meses de agosto a novembro.

As edições, restrições e análises dos dados foram realizadas utilizando o pacote estatístico R (Versão 3.0.2) e o programa *Statistical Analysis System* (SAS) – versão 9.0. O programa Intergen – versão 1.2 – foi utilizado para estimação dos componentes de variância e predição dos valores genéticos (CARDOSO, 2010). A característica considerada neste estudo foi o peso padronizado a desmama (P205) com média e desvio-padrão de  $187,00 \pm 37,59$ .

Os grupos contemporâneos (GC) foram formados combinando efeitos de rebanho, ano de nascimento, época de nascimento, sexo, raça do pai, composição genética do pai, raça da mãe, composição genética da mãe, regime alimentar a desmama, grupo manejo a desmama, data do desmame e grupos de animais com diferença de idade de 90 dias. Durante a consistência dos dados foram descartados aqueles GC que apresentavam menos de três indivíduos e de animais cujas medidas foram superiores ou inferiores a mais ou menos 3,5 desvios-padrão em relação à média. Os efeitos sistemáticos considerados para P205 foram idade dos animais a desmama e classes de idade da mãe, ambos com efeito linear e quadrático. As classes de idades da mãe foram formadas pelos efeitos de idade da mãe em anos e sexo do animal.

A descrição do arquivo de dados e arquivo de pedigree estão apresentados na Tabela 1. Cada grupo de RM apresenta, em média, 3,07 reprodutores e o número máximo disponibilizado aos grupos foi cinco.

A metodologia proposta por Westell et al. (1988) foi utilizada para formar grupos genéticos com indivíduos que possuíam data de nascimento, mas não tinham mãe ou pai identificados. Adotou-se a proposta de Shiotsuki et al. (2013) e utilizado intervalo de geração dos machos para formação dos grupos genéticos, que no presente estudo é de seis anos.

Foram adotadas três abordagens para avaliar filhos de RM. Na primeira abordagem (GG) foi assumido que produtos oriundos de RM e animais TEF que apresentavam data de nascimento fossem incluídos como grupos genéticos. Na segunda abordagem foi considerado o uso da matriz da média dos numeradores dos coeficientes de parentesco (ANRM) (HENDERSON, 1988) para filhos de RM. Nessa metodologia, o objetivo é predizer o valor genético dos animais assumindo probabilidades conhecidas de cada animal ser produto de determinado touro dentro do grupo de reprodutores múltiplos. Quando a probabilidade de determinado touro ser o pai verdadeiro for

conhecida essa metodologia auxilia na correta especificação da matriz de (co)variâncias genéticas. Nessa abordagem, o coeficiente de parentesco é função das probabilidades adotadas de cada touro ser o pai verdadeiro para os animais com incerteza de paternidade e foi calculado como:

$$\begin{aligned} a_{ij} &= \frac{1}{2} \left[ a_{is_1}^{(j)} p_{k=1}^{(j)} + \dots + a_{is_g}^{(j)} p_{k=g}^{(j)} + \dots + a_{id_1}^{(j)} p_{d=1}^{(j)} + \dots + a_{id_m}^{(j)} p_{d=m}^{(j)} \right] \\ &= \frac{1}{2} \left[ \sum_{k=1}^g a_{is_k}^{(j)} p_k^{(j)} + \sum_{d=1}^m a_{id_m}^{(j)} p_{d_m}^{(j)} \right], \end{aligned}$$

em que  $a_{ij}$  é o parentesco aditivo entre os animais  $i$  e  $j$ ;  $a_{ij}^{(j)}$  é o parentesco aditivo entre o animal  $i$  e cada um dos possíveis pais (touro ou vaca) ( $k = 1, \dots, g$ ) do animal  $j$ ;  $p_k^{(j)}$  é a probabilidade do touro  $k$  ser o pai do animal  $j$ ;  $a_{id}^{(j)}$  é o parentesco aditivo entre o animal  $i$  e a mãe do animal  $j$  e  $p_{d_m}^{(j)}$  é a probabilidade da vaca  $m$  ser a mãe do animal  $j$ .

Quando a incerteza de paternidade está relacionada somente ao touro, a expressão acima pode ser simplificada como:

$$a_{ij} = \frac{1}{2} \left[ a_{is_1}^{(j)} p_1^{(j)} + \dots + a_{is_g}^{(j)} p_g^{(j)} + a_{id}^{(j)} \right] = \frac{1}{2} \left[ \sum_{k=1}^g a_{is_k}^{(j)} p_k^{(j)} + a_{id}^{(j)} \right],$$

em que o termo  $\sum_{d=1}^m a_{id_m}^{(j)} p_{d_m}^{(j)}$  é reduzido a  $a_{id}^{(j)}$ , pois  $p_{d_m}^{(j)}$  assume valor igual a unidade com a certeza da identificação da mãe. A diagonal da matriz ANRM pode ser calculada da seguinte maneira:

$$a_{ii} = \left[ 1 + \frac{1}{2} \sum_{j=1}^{s_i} \sum_{k=1}^{d_i} p_{ij} q_{ik} a_{m_j f_k} \right] = 1 + F_i$$

em que  $F_i$  é o coeficiente de endogamia do indivíduo  $i$  (PEREZ-ENCISO e FERNANDO, 1992).

**Tabela 1.** Descrição do arquivo de dados e de pedigree considerando reprodutores múltiplos (RM), produtos oriundos de TEF e grupos genéticos.

Número de animais no arquivo de pedigree	99.265
Número de touros	4.944
Número de vacas	40.589
Número de registros de P205	49.339
Animais com pai desconhecido	43.156
Animais com mãe desconhecida	29.180
Animais com pai e mãe desconhecidos	29.162
Animais com pai e mãe conhecidos	56.091
Animais filhos de reprodutores múltiplos	8.642
Animais produtos de TE/FIV	330
Número de grupos contemporâneos	4.043
Grupos genéticos	7

Na terceira abordagem (HIER) foi assumido o modelo hierárquico bayesiano (CARDOSO e TEMPELMAN, 2003) em que probabilidades *a priori* foram atribuídas a cada animal, enquanto as probabilidades *a posteriori* estavam em função das informações contidas nos dados. O modelo pode ser descrito em quatro estágios, quais sejam (a) definição do modelo linear de efeitos sistemáticos e aleatórios, (b) definição das distribuições *a priori* dos efeitos genéticos aditivos e não-aditivos, (c) definição da distribuição da variância genética aditiva e (d) as distribuições *a posteriori* para a probabilidade de paternidade.

O modelo unicaracterístico em notação matricial é dado por:

$$y = Xb + Za + Zm + Zmpe + Zgc + e$$

em que  $y$  é o vetor de informações das observações,  $b$  é o vetor de efeitos sistemáticos,  $a$  é o vetor de efeitos aleatórios genéticos aditivos;  $m$  é o vetor de efeitos aleatórios genéticos maternos,  $mpe$  é o vetor de efeito aleatório de ambiente permanente materno;  $gc$  é o vetor de efeito aleatório de grupos de contemporâneos;  $e$  é o vetor de efeito residual;  $X$ ,  $Z_a$ ,  $Z_m$ ,  $Z_{mpe}$  e  $Z_{gc}$  são

matrizes de incidência que relacionam as observações aos efeitos sistemáticos, efeito genético aditivo, efeito genético materno, efeito de ambiente permanente materno e de grupos de contemporâneos, respectivamente. Seguem abaixo os estágios propostos por Cardoso e Tempelman (2003):

1. Assume-se distribuição normal univariada para a distribuição condicional dos dados amostrais, dados os parâmetros.

$$y|b, a, m, mpe, gc, G, P, C, R \sim N(Xb + Z_a a + Z_m m + Z_{mpe} mpe + Z_{gc} gc, R),$$

A matriz de (co)variância residual é descrita da seguinte forma:

$$R = I_n \otimes \sigma_e^2$$

2. A distribuição *a priori* dos efeitos sistemáticos, genéticos aditivos e não genéticos são descritos a seguir:

Assume-se distribuição *a priori* uniforme para o efeito sistemático **b** com densidade  $p(\beta) \propto \text{constante}$ .

Efeito genético aditivo:  $a|s, A \sim N(0, A \otimes G_0)$ ;

$$\begin{bmatrix} a \\ m \end{bmatrix} | S, G_0 \sim N(0, G);$$

$$G = A \otimes G_0;$$

$$G_0 \sim N(0, G);$$

$$G_0 = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_m^2 \end{bmatrix};$$

E efeito de ambiente permanente materno:  $mpe \sim N(0, P)$ , tal que  $P = I \otimes \sigma_{mpe}^2$ ;

E efeito de grupo contemporâneo:  $gc \sim N(0, C)$ , tal que  $C = I \otimes \sigma_{gc}^2$ ;

em que **A** é a matriz dos numeradores dos coeficientes de parentesco de Wright, **s** são os touros utilizados como reprodutores múltiplos,  $\sigma_a^2$  é a variância

genética aditiva,  $\sigma_m^2$  é a variância genética materna,  $\sigma_{mpe}^2$  é a variância de ambiente permanente materno,  $\sigma_{gc}^2$  é a variância de grupos contemporâneos.

3. A distribuição Wishart invertida é definida *a priori* para as matrizes de (co)variâncias  $\mathbf{R}$ ,  $\mathbf{G}$ ,  $\mathbf{P}$  e  $\mathbf{C}$  com as respectivas densidades:

$$\sigma_e^2 \sim S_e^2 \chi_{w_R}^{-2},$$

$$\sigma_a^2 \sim S_a^2 \chi_{w_G}^{-2},$$

$$\sigma_{mpe}^2 \sim S_{mpe}^2 \chi_{w_P}^{-2},$$

$$\sigma_{gc}^2 \sim S_{gc}^2 \chi_{w_C}^{-2},$$

em que  $S_e^2$ ,  $S_a^2$ ,  $S_{mpe}^2$  e  $S_{gc}^2$  são os valores *a priori* das variâncias residuais, genética aditiva, de ambiente permanente materno e de grupo contemporâneo, respectivamente, e  $w_R = 4$ ,  $w_G = 5$ ,  $w_P = 5$  e  $w_C = 5$  são os graus de confiança nesses valores;

A probabilidade de  $s_j$  ser o pai verdadeiro do animal  $j$  é especificada nesse estágio e é dada por:

$\pi_j^{(k)} = P(s_j = s_j^{(k)})$  para os touros candidatos  $k = 1, 2, 3, \dots, v_j$ ; sendo o conjunto de probabilidades *a priori* de cada touro  $v_j$  ser o pai verdadeiro dado por:

$$\pi_j = \{\pi_j^{(1)}, \pi_j^{(2)}, \dots, \pi_j^{(k)}\} \text{ para os } v_j \text{ touros candidatos a pai do animal } j.$$

4. A distribuição *a posteriori* para a probabilidade de paternidade é descrita neste estágio. Por não se conhecer o completo conjunto de probabilidades para todos os animais pertencentes a população base, a distribuição de Dirichlet é assumida:

$$p(\pi|\alpha) = \prod_{j=q_b+1}^q \prod_{k=1}^{v_j} (\pi_j^{(k)})^{\alpha_j^{(k)}}$$

em que  $\alpha_j = \{\alpha_j^{(k)}\}_{k=1}^{V_j}$ ,  $\alpha_j^{(k)} > 0$  para  $k = 1, 2, \dots, V_j$  e  $\pi_j^{(V_j)} = 1 - \sum_{k=1}^{V_j-1} \pi_j^{(k)}$ .

Especificações do conjunto de hiperparâmetros  $\{\alpha_j^{(k)}\}_{k=1}^{V_j}$  podem ser baseadas na avaliação da confiabilidade de uma fonte de informação externa sobre as probabilidades *a priori* de cada touro designado. Na etapa de avaliação dos filhos de RM, os registros de P205 dos animais TEF foram eliminados do arquivo de dados, porém os animais TEF foram mantidos no arquivo de pedigree com a informação da mãe biológica. Nesse estágio os valores genéticos dos animais TEF foram preditos em função somente do parentesco.

A metodologia adotada para avaliar os indivíduos TEF foi proposta por Schaeffer e Kennedy (1989). Nessa proposta, é assumida a informação da receptora para a estimação dos efeitos maternos e de ambiente permanente materno. Considerou-se que as receptoras pertenciam a mesma base genética das doadoras para adequada distribuição dos efeitos genéticos e não genéticos.

Os componentes de variância foram estimados sob enfoque Bayesiano utilizando Métodos de Monte Carlo via Cadeias de Markov (MCMC) e os valores genéticos foram preditos segundo modelo animal. A cadeia gerada apresentou 500.000 ciclos com período de aquecimento de 10% da cadeia (50.000) e intervalo de amostragem de 50. A convergência foi avaliada pelo critério de Geweke (GEWEKE, 1991).

O critério de informação da *deviance* (DIC) e a ordenada preditiva condicional (CPO) foram utilizados como critérios de seleção de modelos. O DIC foi utilizado como critério de ajuste global e definido em função da *deviance* (**DEV**) e pelo número efetivo de parâmetros ( $pD_{(r)}$ ) (SPIEGELHALTER et al., 2002).

$$DEV = \frac{1}{G} \sum_{i=1}^G -\log p(y|\theta^{(i)}, M_r)$$

$p_{D(r)} = DEV - D_r(\bar{\theta})$ , em que  $DEV(\bar{\theta}) = -2 \log p(y|\bar{\theta}, M_r)$ . Assim, o DIC para os modelos foi determinado por:

$$DIC = DEV + p_{D(r)}$$

Para comparar dois modelos,  $M_1$  e  $M_2$ , o CPO é definido como aproximações de MCMC para o CPO dos modelos  $M_1$  e  $M_2$  com parâmetros  $\theta$ , os quais foram obtidos pela média harmônica de  $G$  ciclos de MCMC:

$$p(y_{ij}|y_{(-ij)}, M_r) \approx \frac{1}{\frac{1}{G} \sum_{l=1}^G p^{-1}(y_{ij}|\theta^{(l)}, M_r)}$$

em que  $p(y_{ij}|y_{(-ij)}, M_1)$  é a ordenada preditiva condicional do modelo 1 para observação  $y_{ij}$ , destinada a ser a densidade de validação cruzada, que sugere quais valores de  $y_{ij}$  são prováveis quando o modelo é ajustado a todas as outras observações  $y_{(-ij)}$  exceto  $y_{ij}$ . Enquanto  $p(y_{ij}|y_{(-ij)}, M_2)$  se refere à ordenada preditiva condicional do modelo 2.

Estimativas menores de DIC e CPO são indicativas de melhor ajuste do modelo aos dados. As probabilidades *a posteriori* foram calculadas em função do DIC utilizando a aproximação proposta por Wilberg e Bence (2008) e apresentada por Silva et al (2011):

$$p(M_t|\hat{c}) = \frac{\exp(-\frac{\Delta_t}{2})}{\sum_{t=1}^3 \exp(-\frac{\Delta_t}{2})}$$

em que  $p(M_t|\hat{c})$  é a probabilidade *a posteriori* do modelo  $t$ ,  $\Delta_t$  é a diferença de DIC entre o modelo  $t$  e o melhor modelo (menor DIC). Para o melhor modelo essa diferença é igual a zero e a probabilidade *a posteriori* é igual a unidade. Esse procedimento modifica a diferença entre os DIC dos modelos da escala log para a escala normal.

Correlações de Spearman foram obtidas entre os valores genéticos preditos dos animais nos diferentes modelos e diferentes estratégias de

inclusão de produtos de reprodutores múltiplos e TEF. Foram consideradas três abordagens para avaliar a correlação entre os valores genéticos dos animais, em que na abordagem (1) utilizou-se toda a população; na abordagem (2) foram avaliados somente o valor genético dos filhos de RM considerando os 10%, 20%, 50% e 100% melhores indivíduos; na abordagem (3) os valores genéticos dos animais TEF considerando os 10%, 20%, 50% e 100% melhores indivíduos.

As médias anuais dos valores genéticos dos animais filhos de RM e animais TEF foram utilizadas para obter estimativas de tendências genéticas utilizando-se o pacote estatístico R. A tendência genética foi estimada por uma regressão linear ponderada pela quantidade de animais nascidos em cada ano.

## **RESULTADOS**

A avaliação da inclusão de informações de reprodutores múltiplos (RM) e animais oriundos de biotecnologias da reprodução (TEF) foi realizada em duas etapas. A primeira etapa avaliou o melhor ajuste de animais RM nos modelos hierárquico bayesiano (HIER), matriz da média dos numeradores dos coeficientes de parentesco (ANRM) e grupos genéticos (GG). Pode-se verificar na Tabela 2 que o modelo GG apresentou pior ajuste ao incluir animais RM, segundo os critérios assumidos (CPO e DIC). Segundo o CPO, o modelo HIER apresentou melhor ajuste. Entretanto, o DIC, sugeriu que o modelo ANRM apresentou melhor ajuste global, apesar de ter apresentado maior penalização para número efetivo de parâmetros.

A avaliação do ajuste ao incluir informações de animais TEF, segunda etapa, foi realizada somente para as estratégias HIER e ANRM. Foi observado que a inclusão de animais TEF proporcionou aumento em magnitude dos critérios (Tabela 2), porém, os critérios DEV,  $P_D$  e DIC apresentaram padrão semelhante aos da primeira etapa. Na segunda etapa de avaliação, o CPO indicou que o modelo ANRM apresentou melhor ajuste aos dados.

**Tabela 2.** Critérios de ajuste bayesiano ao avaliar a inclusão de filhos de reprodutores múltiplos (primeira etapa de avaliação) e de animais oriundos de transferência de embriões e fertilização *in vitro* (segunda etapa de avaliação) para a característica peso padronizado aos 205 dias.

Modelo <sup>a</sup>	CPO	DEV	P <sub>D</sub>	DIC
<b>Primeira etapa de avaliação</b>				
GG	476.946,20	498.083,20	64.941,51	563.024,70
ANRM	476.605,70	494.108,80	66.786,18	560.895,00
HIER	476.595,40	494.517,10	66.584,91	561.102,00
<b>Segunda etapa de avaliação</b>				
ANRM-TEF	480.058,90	495.639,40	68.265,25	563.904,60
HIER-TEF	480.085,60	496.200,40	68.005,13	564.205,60

CPO = ordenada preditiva condicional, DEV = média da *deviance*, P<sub>D</sub> = penalização para número efetivo de parâmetros e DIC = critério de informação da *deviance*. <sup>a</sup>GG = grupos genéticos, ANRM = matriz da média dos numeradores dos coeficientes de parentesco, HIER = modelo hierárquico bayesiano.

#### *Componentes de variância e parâmetros genéticos*

Na Tabela 3 são apresentadas as estimativas dos componentes de variância, resultados de convergência pelo Teste Z-Geweke e intervalo de credibilidade com 95% de certeza. Todos os efeitos aleatórios (genético aditivo direto, genético materno, ambiente permanente materno, grupo contemporâneo e residual) e os parâmetros genéticos ( $h_a^2$  e  $h_m^2$ ) atingiram convergência pelo Teste Z-Geweke. É possível observar que os efeitos genético aditivo, de grupo contemporâneo e residual foram mais influenciados ao incluir animais TEF, de forma que o efeito genético aditivo apresentou aumento em magnitude e os efeitos de grupo contemporâneo e residual apresentaram redução. Apesar de ter sido observada diferença em magnitude, não foi observada diferença estatística nas estimativas dos componentes de variância e parâmetros genéticos entre os modelos (Figura 1). Não foi observada diferença na média dos desvios-padrão *a posteriori* entre os modelos.

A inclusão das informações dos animais TEF proporcionou estimativas de herdabilidade direta de maior magnitude, o que não foi observado para as estimativas de herdabilidade materna. Entretanto, existe sobreposição dos IC\_95% desses parâmetros genéticos, o que indica que apesar da diferença em magnitude na herdabilidade direta não foi observada diferença estatística significativa.

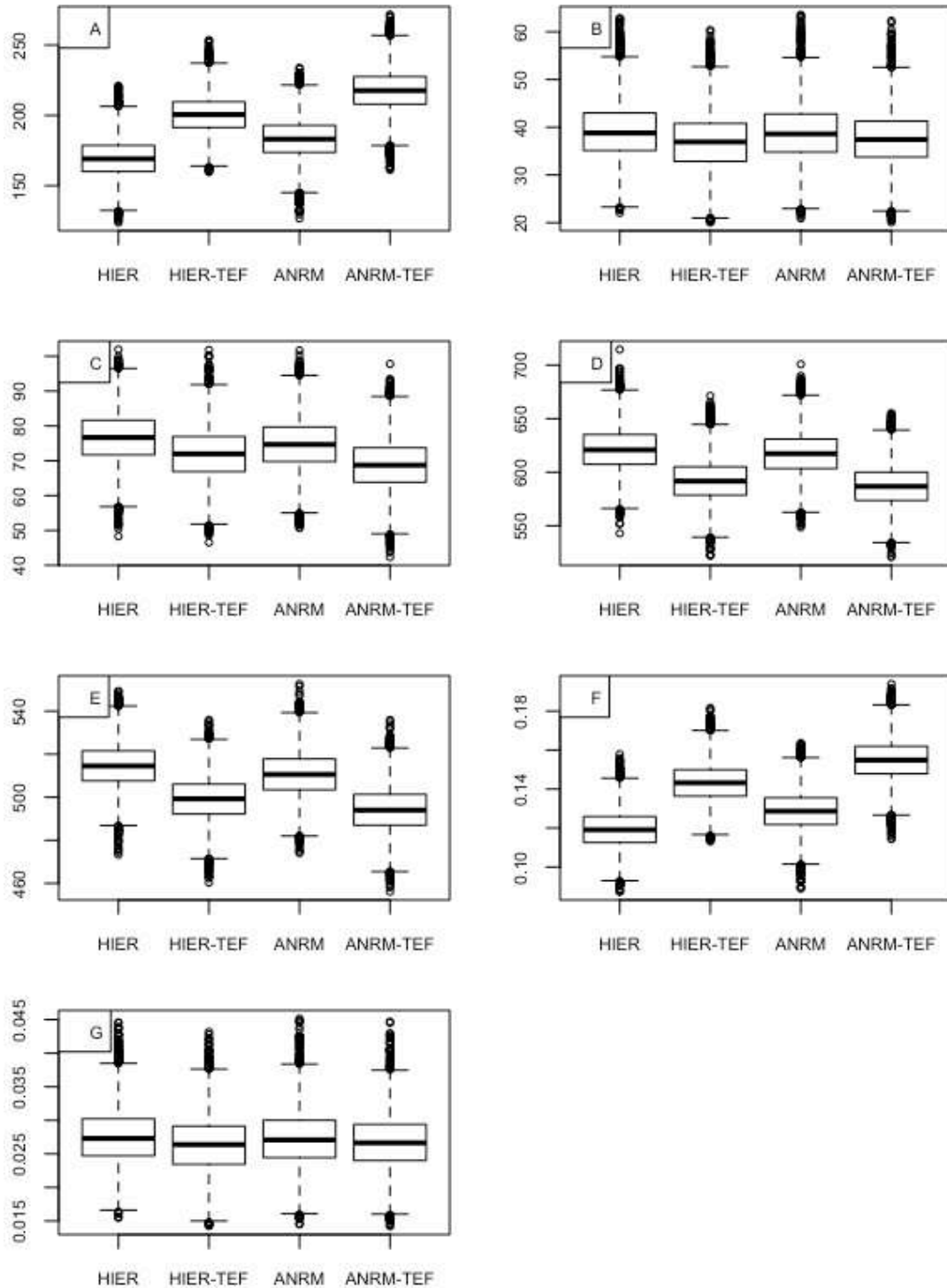
Foi observado que não existe diferença estatística, segundo o Teste F de Graybill, para os desvios-padrão *a posteriori* (DPP) dos animais com certeza de paternidade, fundadores e oriundos de monta natural ou inseminação artificial (dados não apresentados) ao utilizar os modelos HIER ou ANRM. De forma semelhante, também não foi observada diferença estatística dos DPP dos animais TEF entre os modelos HIER e ANRM ou HIER-TEF e ANRM-TEF (dados não apresentados), sendo que essas estimativas se apresentaram ao redor de 11kg. Entretanto, existe diferença estatística ao avaliar os DPP dos filhos de RM entre os modelos HIER e ANRM (Figura 2) ou HIER-TEF e ANRM-TEF (dados não apresentados). É possível observar que os valores de DPP maior que 11kg se desviam consideravelmente da hipótese nula (não existe diferença nas estimativas de DPP entre os modelos). Entretanto, pode ser observado que menores valores de DPP se localizam em cima do segmento com intercepto zero e inclinação um. Um total de nove animais apresentam DPP menor que 8kg, sendo que esses animais são touros que possuem no mínimo 47 filhos e no máximo 354. O menor valor de DPP (3,94kg) observado pertence a um touro com 275 filhos.

**Tabela 3.** Estatísticas dos componentes de variância, critério de convergência pelo Teste Z-Geweke e intervalo de credibilidade com 95% de certeza ao avaliar informações de peso padronizado aos 205 dias.

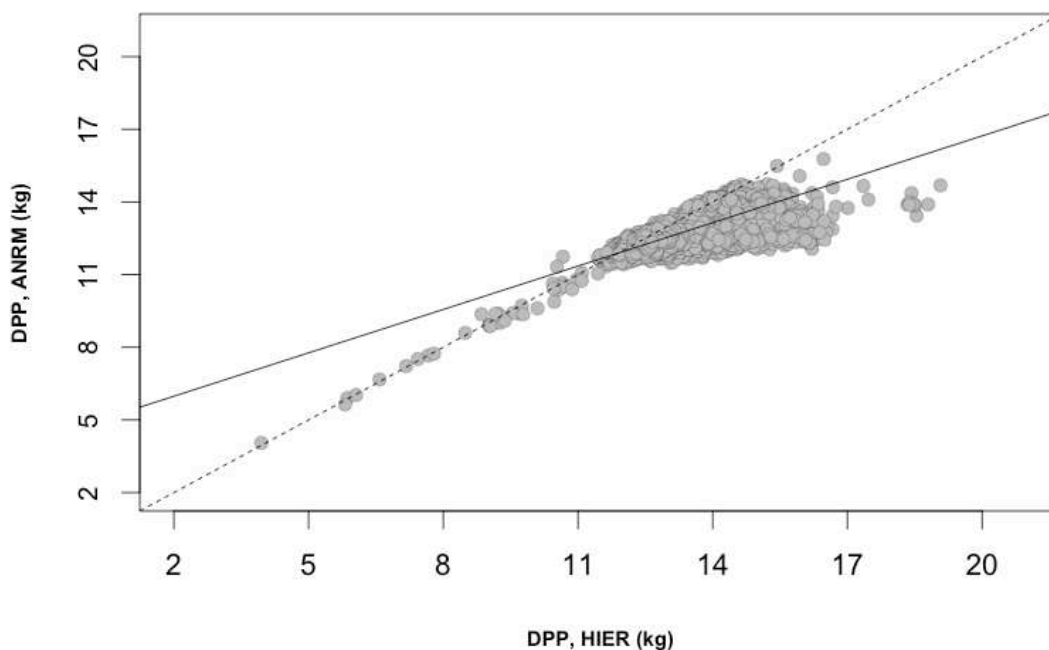
Modelo	Parâmetro	MP	DPP	Z	IC_95%	TEA
HIER	$\sigma_a^2$	169,40	13,88	0,5231	(141,90; 196,90)	439
	$\sigma_m^2$	39,25	5,81	-0,8401	(29,15; 51,51)	208
	$\sigma_{mpe}^2$	76,58	7,38	1,3439	(61,86; 90,82)	514
	$\sigma_{gc}^2$	621,38	20,54	-0,4824	(582,24; 661,75)	6984
	$\sigma_e^2$	514,55	10,36	-0,9175	(494,19; 534,80)	726
	$h_a^2$	0,1192	0,009	0,5598	(0,099; 0,139)	447
	$h_m^2$	0,0276	0,004	-0,8557	(0,020; 0,036)	210
HIER-TEF	$\sigma_a^2$	200,76	13,44	0,2653	(175,27; 227,70)	610
	$\sigma_m^2$	37,06	5,91	0,8398	(26,47; 49,51)	185
	$\sigma_{mpe}^2$	71,90	7,34	-0,5664	(57,58; 86,01)	511
	$\sigma_{gc}^2$	592,04	19,58	0,7097	(554,45; 631,21)	8142
	$\sigma_e^2$	499,25	10,19	-0,9378	(479,56; 519,16)	875
	$h_a^2$	0,1433	0,009	0,2092	(0,124; 0,162)	517
	$h_m^2$	0,0264	0,004	0,8162	(0,018; 0,035)	186
ANRM	$\sigma_a^2$	183,19	14,72	1,2492	(153,70; 212,58)	344
	$\sigma_m^2$	38,84	6,014	0,1910	(27,16; 50,87)	191
	$\sigma_{mpe}^2$	74,66	7,501	-0,6514	(59,7; 89,22)	400
	$\sigma_{gc}^2$	617,43	20,49	-1,4201	(577,58; 658,66)	7108
	$\sigma_e^2$	510,69	10,68	-1,4244	(490,01; 531,94)	694
	$h_a^2$	0,1286	0,010	1,2876	(0,107; 0,149)	352
	$h_m^2$	0,0272	0,004	0,2065	(0,019; 0,035)	193
ANRM-TEF	$\sigma_a^2$	217,62	15,02	1,4504	(187,66; 247,37)	407
	$\sigma_m^2$	37,55	5,73	-0,1180	(26,19; 48,95)	194
	$\sigma_{mpe}^2$	68,83	7,29	-0,4918	(54,71; 83,32)	518
	$\sigma_{gc}^2$	587,12	19,66	-0,8674	(549,23; 627,08)	7131
	$\sigma_e^2$	494,17	10,75	-1,7132	(473,00; 515,85)	806
	$h_a^2$	0,1548	0,010	1,5172	(0,133; 0,176)	417
	$h_m^2$	0,0267	0,004	-0,1100	(0,018; 0,034)	196

MP = média *a posteriori*, DPP = desvio-padrão *a posteriori*, Z = Z-Geweke, IC\_95% = intervalo de credibilidade, TEA = tamanho efetivo de amostras, HIER = modelo hierárquico bayesiano, HIER-TEF = modelo hierárquico bayesiano

com inclusão de indivíduos TEF; ANRM = matriz da média dos numeradores dos coeficientes de parentesco; ANRM-TE = matriz da média dos numeradores dos coeficientes de parentesco com inclusão de indivíduos TEF,  $\sigma_a^2$  = variância genética aditiva,  $\sigma_m^2$  = variância genética materna,  $\sigma_{mpe}^2$  = variância de ambiente permanente materno,  $\sigma_{gc}^2$  = variância de grupos contemporâneos,  $h_a^2$  = herdabilidade aditiva direta e  $h_m^2$  = herdabilidade materna.



**Figura 1.** Diagrama em caixa das estimativas *a posteriori* dos componentes de variância genético aditivo (A), genético materno (B), de ambiente permanente materno (C), de grupos contemporâneos (D), residual (E), herdabilidade direta (F) e herdabilidade materna (G).



**Figura 2.** Dispersão dos desvios-padrão *a posteriori* (DPP) do efeito genético aditivo para peso padronizado aos 205 dias (kg) dos filhos de reprodutores múltiplos obtido pelo modelo HIER e ANRM. A linha contínua representa o ajuste obtido pelo método dos quadrados mínimos ordinários e a linha tracejada possui inclinação um e intercepto zero.

#### *Animais coincidentes e correlação entre valores genéticos*

Foram observadas estimativas de correlação de Spearman de elevada magnitude ao selecionar diferentes proporções de candidatos entre os modelos (Tabelas 4, 5 e 6). De fato, foi observado que correlação de elevada magnitude proporcionou a seleção de elevada proporção de animais coincidentes entre os modelos, principalmente para os animais TEF.

A avaliação dos filhos de RM e animais TEF entre metodologias adotadas nesse estudo indicou que ao se utilizar o modelo hierárquico bayesiano foram selecionados mais indivíduos RM que a matriz da média dos coeficientes de parentesco (dados não apresentados) em qualquer proporção de seleção. Entretanto, o modelo ANRM estimou valores genéticos mais acurados (dados não apresentados). Os modelos HIER e ANRM selecionaram praticamente os mesmos animais TEF em qualquer proporção de seleção.

**Tabela 4.** Correlação de Spearman entre os valores genéticos preditos considerando toda a população.

<b>Modelo</b>	<b>Porcentagem de indivíduos selecionados</b>			
	<b>10%</b>	<b>20%</b>	<b>50%</b>	<b>100%</b>
HIER x ANRM	0,97*	0,98*	0,99*	0,99*
HIER x HIER-TEF	0,87*	0,87*	0,91*	0,98*
ANRM x ANRM-TEF	0,89*	0,89*	0,92*	0,98*
HIER-TEF x ANRM-TEF	0,96*	0,97*	0,99*	0,99*

HIER = modelo hierárquico bayesiano, HIER-TEF = modelo hierárquico bayesiano com inclusão de indivíduos TEF, ANRM = modelo da matriz da média dos numeradores dos coeficientes de parentesco, ANRM-TEF = modelo da matriz da média dos coeficientes dos numeradores dos numeradores dos coeficientes de parentesco com inclusão de indivíduos TEF. \*P < 0,05

**Tabela 5.** Número de filhos de reprodutores múltiplos (RM) coincidentes (entre parêntesis) e correlação de Spearman entre os valores genéticos preditos.

	<b>Porcentagem de filhos de RM selecionados</b>			
	<b>10%</b>	<b>20%</b>	<b>50%</b>	<b>100%</b>
<b>Número de filhos de RM</b>	991	1769	4272	8642
HIER x ANRM	0,92*	0,95*	0,98*	0,99*
	(827)	(1581)	(4053)	(8642)
<b>Número de filhos de RM</b>	1056	1802	4264	8642
HIER x HIER-TEF	0,94*	0,95*	0,97*	0,99*
	(926)	(1567)	(3845)	(8642)
<b>Número de filhos de RM</b>	886	1686	4257	8642
ANRM x ANRM-TEF	0,94*	0,95*	0,97*	0,99*
	(767)	(1446)	(3857)	(8642)
<b>Número de filhos de RM</b>	1001	1663	4052	8642
HIER-TEF x ANRM-TEF	0,93*	0,95*	0,97*	0,99*
	(816)	(1488)	(3846)	(8642)

HIER = modelo hierárquico bayesiano, HIER-TEF = modelo hierárquico bayesiano com inclusão de indivíduos TEF, ANRM = modelo da matriz da média dos numeradores dos coeficientes de parentesco, ANRM-TEF = modelo

da matriz da média dos numeradores dos coeficientes de parentesco com inclusão de indivíduos TEF. \*P < 0,05.

**Tabela 6.** Número de animais oriundos de transferência de embriões e fertilização *in vitro* coincidentes (entre parêntesis) e correlação de Spearman entre os valores genéticos preditos.

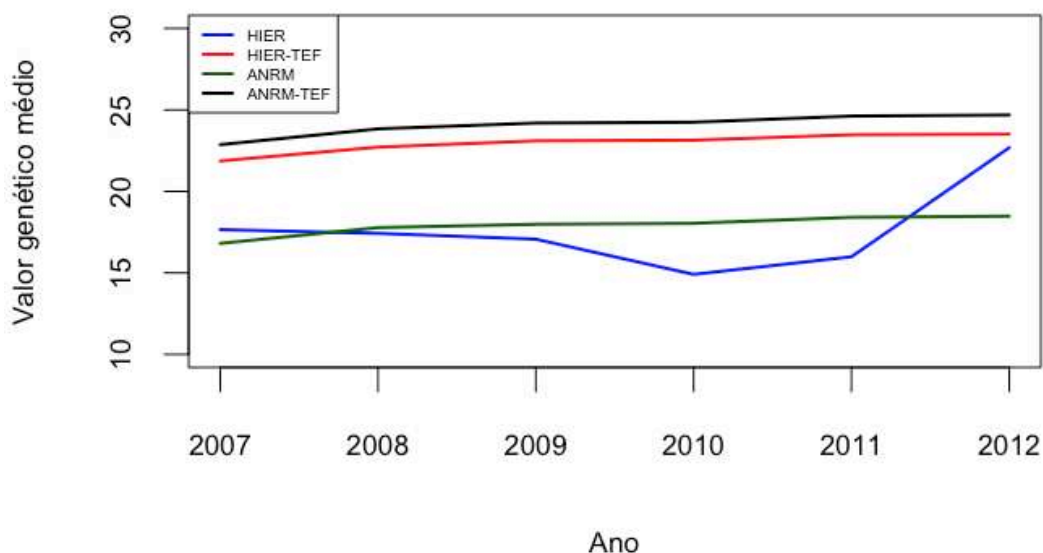
	Porcentagem de filhos de TEF selecionados			
	10%	20%	50%	100%
<b>Número de produtos TEF</b>	26	62	209	330
HIER x ANRM	0,91* (25)	0,98* (61)	0,99* (204)	0,99* (330)
<b>Número de produtos TEF</b>	136	184	271	330
HIER x HIER-TEF	0,28* (25)	0,34* (53)	0,25* (184)	0,39* (330)
<b>Número de produtos TEF</b>	134	185	272	330
ANRM x ANRM-TEF	0,43* (24)	0,35* (51)	0,22* (187)	0,38* (330)
<b>Número de produtos TEF</b>	135	176	251	330
HIER-TEF x ANRM-TEF	0,99* (133)	0,99* (174)	0,99* (250)	0,99* (330)

HIER = modelo hierárquico bayesiano, HIER-TEF = modelo hierárquico bayesiano com inclusão de indivíduos TEF, ANRM = modelo da matriz da média dos numeradores dos coeficientes de parentesco, ANRM-TEF = modelo da matriz da média dos numeradores dos coeficientes de parentesco com inclusão de indivíduos TEF. \*P < 0,05.

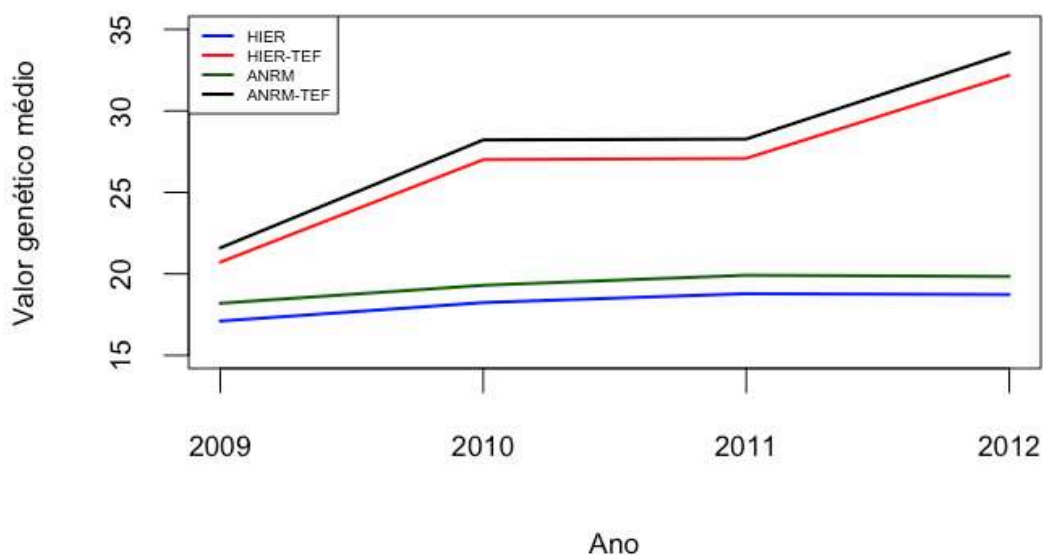
### *Tendência genética*

A tendência genética dos filhos de RM e animais oriundos de biotecnologias da reprodução pode ser observada nas Figuras 3 e 4, respectivamente. Foi observado que a inclusão das informações de RM pelo modelo HIER proporcionou o maior ganho anual (0,677kg/ano) em relação aos modelos HIER-TEF (0,277kg/ano), ANRM (0,270kg/ano) e ANRM-TEF (0,303kg/ano). Comportamento diferente foi observado para a tendência

genética dos animais TEF, em que os HIER-TEF e ANRM-TEF proporcionaram maiores ganhos anuais, 3,359kg/ano e 3,499kg/ano, respectivamente, que os modelos HIER (0,326 kg/ano) e ANRM (0,338 kg/ano).



**Figura 3.** Tendência genética dos filhos de reprodutores múltiplos ao considerar a metodologia HIER, HIER-TEF, ANRM e ANRM-TEF na avaliação de registros de pesos padronizados aos 205 dias.



**Figura 4.** Tendência genética dos animais oriundos de transferência de embriões e fertilização *in vitro* ao considerar as metodologias HIER, HIER-TEF, ANRM e ANRM-TEF na avaliação de registros de pesos padronizados aos 205 dias.

## DISCUSSÃO

A avaliação do desempenho dos animais de produção depende de vários fatores, dentre os quais a qualidade dos registros de fenótipo, da correta informação do pedigree e da adoção de metodologias estatísticas adequadas às singularidades presentes no sistema produtivo (CARDOSO et al., 2009). Segundo Van Vleck (1970), a incerteza de paternidade pode afetar o progresso genético ao reduzir a intensidade de seleção por excluir informações de animais com incerteza de paternidade e também por baixa acurácia da seleção. No presente estudo foram avaliadas metodologias adequadas ao utilizar informações de filhos de reprodutores múltiplos (RM) e de animais oriundos de transferência de embriões ou fertilização *in vitro* (TEF) em modelos unicaracterísticos para peso padronizado aos 205 dias de bovinos de corte.

Os resultados de DIC e CPO demonstram que o uso dos GG não se mostrou superior aos modelos HIER e ANRM, sugerindo que o uso dos registros dos grupos de reprodutores múltiplos e das receptoras com metodologia apropriada apresenta ajuste mais adequado. Uma possível justificativa é que os valores genéticos são corrigidos para efeito de grupos genéticos e para a (co)variância entre os pais fantasmas e indivíduos verdadeiros ao utilizar a metodologia proposta por Westell et al. (1988) (GG). Porém, a predição do valor genético pode não ser a mais adequada, pois é assumido que os pais fantasmas não são correlacionados (PEREZ-ENCISO e FERNANDO, 1992). Entretanto, o uso de grupos genéticos se apresenta como uma opção para ajustar as informações de filhos de RM e animais TEF quando os registros desses animais não são confiáveis. Shiotsuki et al. (2013) utilizaram as metodologias GG, ANRM e HIER em um rebanho Nelore para avaliar informações de RM e encontraram resultados que concordam com o do presente estudo. Shiotsuki et al. (2013) e Kerr et al. (1994) relatam que ajustar grupos genéticos auxilia na redução da variância do erro de predição em situações que os reprodutores dos grupos de reprodutores múltiplos não são conhecidos.

Foram observados menores valores no CPO pelo uso do modelo HIER, sugerindo que esse modelo apresentou estimativas *a posteriori* mais precisas. Segundo Pettit (1990), o critério CPO é utilizado como avaliador de dados discrepantes. Entretanto, os valores de DIC foram favoráveis ao modelo

ANRM, sugerindo que assumir probabilidades *a priori* conhecidas para os touros nos grupos de reprodutores múltiplos proporciona valores *a posteriori* menos discrepantes. De certa forma isso era esperado, pois a cada ciclo MCMC novas probabilidades *a posteriori* são calculadas para os touros participantes de grupos de reprodutores múltiplos no HIER. Cardoso e Tempelman (2004) e Shiotsuki et al. (2012) encontraram resultados discordantes do presente estudo ao avaliar o peso ao desmame e ganho pós-desmame em animais Hereford e Nelore, respectivamente, em que o modelo HIER apresentou melhor ajuste segundo o DIC. O DIC foi proposto com objetivo de avaliar a qualidade de ajuste em modelos hierárquicos bayesianos (WILBERG e BENCE, 2008) e sugere como significativa diferença maior que sete unidades entre modelos (SPIEGELHALTER et al., 2002). Esse critério tem sido aplicado em diversos estudos e utilizado como critério de seleção de modelos (KIZILKAYA e TEMPELMAN, 2005; SHIOTSUKI et al., 2012; SHIOTSUKI et al., 2013).

Cardoso e Tempelman (2004) relatam que diferenças entre os modelos HIER e ANRM são maiores quando as herdabilidades das características são de moderada magnitude e sugerem também que quanto menor a herdabilidade da característica, mais importante são as informações de registros fenotípicos na predição do valor genético dos touros candidatos.

Foi observado que a inclusão dos registros de pesagem de animais TEF causou aumento da média *a posteriori* da variância genética aditiva e redução das médias *a posteriori* das variâncias de grupo contemporâneo e residual. Isso sugere que negligenciar os registros dos animais TEF proporcionaria parâmetros genéticos viesados e, conseqüentemente, menor taxa de ganho genético seria alcançada. No presente estudo foi considerado que as receptoras são do mesmo agrupamento genético e que cada uma produza apenas uma vez. Entretanto, dificilmente isso é realizado na prática, pois se preconiza que as receptoras sejam F1 de raças que mesclam boa habilidade materna e rusticidade. Modelar todas essas informações não é tarefa simples, pois diferentes estruturas de (co)variância genética devem ser consideradas para grupos de indivíduos de raças diferentes. O uso de receptoras de diferentes raças exigiria a partição do vetor de soluções dos efeitos genético materno e de ambiente permanente materno. Dessa forma, seria ajustado (co)variância específica para essas informações. Além disso, devem ser

incluídos efeitos fixos para a correção de tendências relacionadas às receptoras, assim como covariáveis para ajustar a idade desses indivíduos. Entretanto, Van Vleck (1990) relata o mais importante é eliminar o efeito de seleção das receptoras entre os rebanhos.

De forma geral, a inclusão de informações de animais TEF proporcionou médias *a posteriori* para a herdabilidade direta de maior magnitude, sugerindo que a inclusão das informações desses indivíduos proporcionaria maior ganho genético. Ao considerar a acurácia da predição do valor genético como  $\sqrt{1 - \frac{DPP}{\sigma_a^2}}$ , foi observado incremento na acurácia pelo aumento da magnitude da variância genética aditiva, pois não foi observada redução no DPP. Os valores genéticos dos animais RM (Figura 3) e TEF (Figura 4) apresentaram maior magnitude, sugerindo que os valores genéticos desses animais estavam subestimados. O modelo ANRM apresentou tendência em estimar valores genéticos de maior magnitude que os do modelo HIER (KERR et al., 1994). Kerr et al. (1994) relatam que o viés na predição do mérito genético é mais evidente em características com herdabilidade de elevada magnitude. Além disso, o modelo ANRM pressupõe que o valor genético do animais RM é dado pelo somatório da metade dos valores genéticos dos touros do grupo de reprodutores múltiplos e da vaca mais um desvio referente a segregação mendeliana.

Os valores de herdabilidade encontrados podem ser considerados baixos em relação aos relatos da literatura (CARDOSO e TEMPELMAN, 2004; CHUD et al., 2014; GRION et al., 2014), entretanto, o efeito de grupo contemporâneo foi modelado como aleatório. Dessa forma, esse efeito foi incluído na composição da variância fenotípica, o que causou redução na herdabilidade. As estimativas de variância genética aditiva e genética materna encontradas nesse estudo estão de acordo com as encontradas por Mota et al. (2013) ao avaliarem a inclusão de animais TEF da raça Simental não observaram diferença significativa nas magnitudes de variância genética aditiva para P205 ao considerar ou não as informações dos animais TEF. Entretanto, diferentemente do presente estudo, os animais TEF não foram mantidos no pedigree quando os registros foram eliminados do arquivo de dados.

Os resultados apresentados na Tabela 5 relatam que o modelo ANRM seleciona menor quantidade de filhos de RM, apesar de apresentar correlação genética de elevada magnitude. O que sugere que a construção da matriz dos numeradores de parentesco com definição *a priori* da probabilidade de determinado touro ser o verdadeiro pai pode não ser adequada. Apesar desse modelo predizer valores genéticos de maior magnitude, essa predição é viesada. O modelo HIER permite que a cada iteração MCMC as probabilidades *a posteriori* de determinado touro ser o pai verdadeiro sejam influenciadas pela estrutura de dados e do pedigree, permitindo, assim, valores genéticos com menor viés na predição do mérito genético.

Os resultados apresentados na Tabela 6 demonstram que a inclusão dos registros de pesagem dos animais TEF modificou significativamente os valores genéticos. De fato isso era esperado, pois apesar das informações de família serem importantes, as próprias informações são as que mais interferem na predição do valor genético. Além disso, as informações desses animais foram incluídas com ajuste adequado para os efeitos genético materno e de ambiente permanente materno. Esses resultados concordam com os relatos feitos por Schaeffer e Kennedy (1989) e Van Vleck (1990) que ajustes adequados das informações de animais TEF proporcionam estimativas de valores genéticos menos viesadas.

A maior tendência de ganho observada dos animais TEF nos modelos HIER-TEF e ANRM-TEF possivelmente está relacionada com a habilidade materna superior das receptoras utilizadas nessa população. Possivelmente, as receptoras são selecionadas para apresentarem combinações gênicas favoráveis para características relacionadas à habilidade materna.

#### *Implicações e perspectivas do uso de reprodutores múltiplos e biotecnologias da reprodução*

Esse estudo é o primeiro que avalia os reflexos da inclusão das informações de filhos de reprodutores múltiplos e de animais oriundos de transferência de embriões e fertilização *in vitro* ao ajustar diferentes metodologias na avaliação genética de bovinos de corte.

O uso de métodos estatísticos na avaliação genética têm proporcionado grande avanço no setor pecuário. Grupos agropecuários têm realizado grandes

investimentos na seleção de touros melhoradores e, por facilidades de manejo, o uso de grupos de reprodutores múltiplos é adotado. Além disso, esse manejo é importante do ponto de vista econômico, porém, gera problemas aos programas de avaliação genética pelo desconhecimento do pai verdadeiro. Nessas situações, as informações de filhos de RM têm sido descartadas e o valor genético desses animais preditos em função somente do parentesco. Entretanto, apesar das informações de parentesco contribuírem para aumento da acurácia, a própria informação apresenta grande influência na estimação do mérito genético e isso pode ser observado nos presentes resultados.

O uso de marcadores moleculares seria suficiente para solucionar o problema de identificação do pai verdadeiro, entretanto, o custo ainda é elevado para genotipagem de grande quantidade de animais. Dessa forma, é importante a adoção de metodologias apropriadas para ajustar as informações desses animais.

O uso de biotécnicas da reprodução pelos pecuaristas é cada vez maior, sendo que o uso dessas tecnologias apresenta tendência de crescimento. Os custos para a adoção dessas técnicas normalmente estão relacionados à quantidade de animais nascidos. Entretanto, quanto maior a quantidade de animais nascidos, menor será o custo por animal. Logo, os avanços que vêm sendo alcançados tendem a causar a redução no custo por animal nascido, devido a maior quantidade de animais gerados. A grande vantagem do uso dessas tecnologias está relacionada à rápida multiplicação dos genótipos de interesse econômico em pequeno intervalo de tempo. De fato, a inclusão das informações desses animais sem considerar a correta modelagem pode influenciar negativamente as predições dos valores genéticos e suas acurácias. Por isso, os programas de avaliação genética têm optado pelo descarte dessas informações. Entretanto, a metodologia utilizada no presente estudo se apresenta como alternativa para o ajuste das informações desses animais, principalmente por proporcionar aumento na magnitude da predição de valor genético.

## **CONCLUSÃO**

O modelo hierárquico bayesiano se apresenta como alternativa para a inclusão das informações de filhos de reprodutores múltiplos. A inclusão das

informações dos animais oriundos de transferência de embriões e fertilização *in vitro* podem ser incluídas na avaliação genética, desde que os efeitos genéticos maternos e de ambiente permanente materno sejam modelados como efeitos das receptoras. A inclusão dessas informações com metodologia apropriada influencia de forma significativa os valores genéticos.

## REFERÊNCIA BIBLIOGRÁFICA

CARDOSO, F. F. **Aplicação da inferência bayesiana no melhoramento animal usando o programa Intergen: manual da versão 1.2.** Bagé: Embrapa Pecuária Sul: 74 p. 2010. Disponível em: <<http://arquivos.cppsul.embrapa.br/downloads/intergen/%3E>>. Acesso em 10 de setembro de 2013.

CARDOSO, F. F.; ROSA, G. J. M.; TEMPELMAN, R. J.; TORRES JUNIOR, R. A. A. Modelos hierárquicos bayesianos para estimação robusta e análise de dados censurados em melhoramento animal. **Revista Brasileira de Zootecnia**, v. 38, p. 72-80, 2009.

CARDOSO, F. F.; TEMPELMAN, R. J. Bayesian inference on genetic merit under uncertain paternity. **Genetics Selection Evolution**, v. 35, n. 5, p. 469-488, 2003.

\_\_\_\_\_. Genetic evaluation of beef cattle accounting for uncertain paternity. **Livestock Production Science**, v. 89, n. 2, p. 109-120, 2004.

CHUD, T. C. S.; CAETANO, S. L.; BUZANSKAS, M. E.; GROSSI, D. A.; GUIDOLIN, D. G. F.; NASCIMENTO, G. B.; ROSA, J. O.; LÔBO, R. B.; MUNARI, D. P. Genetic analysis for gestation length, birth weight, weaning weight, and accumulated productivity in Nellore beef cattle. **Livestock Science**, n. 0, 2014.

GEWEKE, J. **Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.** Federal Reserve Bank of Minneapolis, Research Department, 1991.

GRION, A. L.; MERCADANTE, M. E. Z.; CYRILLO, J.; BONILHA, S. F. M.; MAGNANI, E.; BRANCO, R. H. Selection for feed efficiency traits and correlated genetic responses in feed intake and weight gain of Nellore cattle. **Journal of Animal Science**, v. 92, n. 3, p. 955-965, 2014.

HENDERSON, C. R. Use of an Average Numerator Relationship Matrix for multiple-sire joining. **Journal of Animal Science**, v. 66, p. 1614-1621, 1988.

KERR, R. J.; HAMMOND, K.; KINGHORN, B. P. Effects of Multiple-Sire Matings on Genetic Evaluations, Selection Response and Rates of Inbreeding. **Livestock Production Science**, v. 38, n. 3, p. 161-168, 1994.

KIZILKAYA, K.; TEMPELMAN, R. J. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. **Genetics Selection Evolution**, v. 37, n. 1, p. 1-26, 2005.

MOTA, R. R.; LOPES, P. S.; MARQUES, L. F. A.; DA SILVA, L. P.; DE RESENDE, M. D. V.; TORRES, R. D. A. The influence of animals from embryo transfer on the genetic evaluation of growth in Simmental beef cattle by using multi-trait models. **Genetics and Molecular Biology**, v. 36, n. 1, p. 43-49, 2013.

PEREZ-ENCISO, M.; FERNANDO, R. L. Genetic evaluation with uncertain parentage: a comparison of methods. **Theoretical and Applied Genetics**, v. 84, n. 1-2, p. 173-179, 1992.

PETTIT, L. I. The Conditional Predictive Ordinate for the Normal-Distribution. **Journal of the Royal Statistical Society Series B-Methodological**, v. 52, n. 1, p. 175-184, 1990.

SAPP, R. L.; ZHANG, W.; BERTRAND, J. K.; REKAYA, R. Genetic evaluation in the presence of uncertain additive relationships. I. Use of phenotypic information to ascertain paternity. **Journal of Animal Science**, v. 85, n. 10, p. 2391-2400, 2007.

SCHAEFFER, L. R.; KENNEDY, B. W. Effects of embryo transfer in beef cattle on genetic evaluation methodology. **Journal of Animal Science**, v. 67, n. 10, p. 2536-2543, 1989.

SHIOTSUKI, L.; CARDOSO, F. F.; SILVA, J. A.; ALBUQUERQUE, L. G. Comparison of a genetic group and unknown paternity models for growth traits in Nellore cattle. **Journal of Animal Science**, v. 91, n. 11, p. 5135-5143, 2013.

SHIOTSUKI, L.; CARDOSO, F. F.; SILVA, J. A.; ROSA, G. J. M.; ALBUQUERQUE, L. G. Evaluation of an average numerator relationship matrix model and a Bayesian hierarchical model for growth traits in Nellore cattle with uncertain paternity. **Livestock Science**, v. 144, n. 1, p. 89-95, 2012.

SILVA, F. F.; ROSA, G. J. M.; GUIMARÃES, S. E. F.; LOPES, P. S.; DE LOS CAMPOS, G. Three-step Bayesian factor analysis applied to QTL detection in crosses between outbred pig populations. **Livestock Science**, v. 142, n. 1, p. 210-215, 2011.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 64, n. 4, p. 583-639, 2002.

VAN VLECK, L. D. Misidentification in estimating the paternal sib correlation. **Journal of Dairy Science**, v. 53, n. 10, p. 1469-1474, 1970.

\_\_\_\_\_. Alternative animal models with maternal effects and foster dams. **Journal of Animal Science**, v. 68, n. 12, p. 4026-4038, 1990.

WESTELL, R. A.; QUAAS, R. L.; VAN VLECK, L. D. Genetic groups in an animal model. **Journal of Dairy Science**, v. 71, n. 5, p. 1310-1318, 1988.

WILBERG, M. J.; BENCE, J. R. Performance of deviance information criterion model selection in statistical catch-at-age analysis. **Fisheries Research**, v. 93, n. 1-2, p. 212-221, 2008.

## ANEXO

É necessário a informação da proporção da variância genética que é atribuída à segregação mendeliana para cada animal ( $d$ ) para se utilizar metodologia hierárquica bayesiana. O código fonte apresentado a seguir foi desenvolvido em linguagem C++, e tem como objetivo calcular o parâmetro  $1/d$ , desde que informada a quantidade de touros nos grupos de reprodutores múltiplos e a identificação de cada touro participante dos grupos.

O programa abaixo ainda está em desenvolvimento para implementação de novas funções. Os algoritmos utilizados para o cálculo de  $1/d$  foram propostas por Sullivan, P. G. (1995) e Famula, T. R. (1992).

```
1. // -----
2. // Program: Mendelian_samples.cpp
3. // -----
4. //
5. // This program was developed as part of Master Science in Genetics and Breeding at
   // Universidade Federal de Viçosa (UFV).
6. //
7. // The program computes the diagonal elements of Henderson's L matrix (1988), as well as
   // the elements of D matrix necessary to compute the numerator relationship matrix
   // inverse.
8. //
9. // The program is in development.
10. //
11. // Created by Vinícius Junqueira on 08/02/14.
12. // Copyright (c) 2014 Vinicius Junqueira . All rights reserved.
13. //
14.
15. // Default libraries
16. #include <iostream>
17. #include <fstream>
18. #include <cstdlib>
19. #include <math.h>
20. #include <iomanip>
21. #include <string.h>
22. #include <cstring>
23. #include <stdio.h>
24. #include <stdlib.h>
25.
26. using namespace std;
27.
28. //Functions
```

```

29. void Inbred();
30. void Inbred(int lines, int Matrix[][9], string File);
31. void NoInbred(int lines, int Matrix[][9], string File);
32. void SaveFile(float dinv[], int lines, int Matrix[][9], string File);
33.
34. // Main function - import data file and pass the data to a new array
35. int main(){
36.     // Create and define variables
37.     string input, Animal, Dam, Sire, Progeny, decision, line;
38.     int l;
39.     ifstream Input;
40.
41.     /***** PROGRAM PARAMETERS *****/
42.     // Data file name
43.     cout << "\n*****\n";
44.     cout << setw(30)<< " " << "Software Mendelian Samples\n\n";
45.     cout << setw(28)<< " " << "Created by Vinicius Junqueira\n";
46.     cout << "*****\n\n";
47.     // Print the guide user
48.     cout << "\nThe columns in datafile must be ordered by:\n\n1. Animal id\n2. Dam id\n3
. Sire id\n4. Number of sires in the group\n5. Id of the first sire\n6. Id of the second
sire\n7. Id of the third sire\n8. Id of the fourth sire\n9. Id of the fifth sire\n\n";
49.     cout << "Data file name: ";
50.     cin >> input;
51.     cout << "Number of lines in data file: ";
52.     cin >> l;
53.
54.     /***** IMPORT FILE TO PROGRAM *****/
55.     // Import pedigree file
56.     Entrada.open(input.c_str());
57.     if(!Input){
58.         cerr<<"Data file could not be imported.\n";
59.         exit(1);
60.     }
61.
62.     /*****/
63.     // Set array dimension (number of animals) x (columns(id, dam, sire, progeny))
64.     // Data matrix
65.     int Matrix_input[l][9];
66.
67.     // Transfers input data to an array
68.     for(int i=0; i < l; i++){
69.         for(int j=0; j < 9; j++){
70.             Input>> Matrix_input[i][j];
71.         }
72.     }
73.
74.     // The user chooses if he wants to calculate Dinv considering inbreeding or not

```

```

75.     string inbreed;
76.     cout << "Would you like to consider inbreeding? [yes/no] ";
77.     cin >> inbreed;
78.
79.     // Split the program
80.     // Function that consider inbreeding
81.     if(inbreed == "yes"){cout<<"I am sorry, still in development...\n";}
82.     // Function that do not consider inbreeding
83.     else{
84.         cout << "Running..." << endl;
85.         NoInbred(1, Matrix_input, input);
86.     }
87.
88.     // Ends the program
89.     return 0;
90. }
91.
92. /***** Famula's algorithm (1992) *****/
93.
94. // Inbreeding function
95. void Inbred(int lines, int Matrix[][9], string File){
96.     // Vectors that receive values to compute mendelian samples
97.     // First iteration is zero
98.     float u[lines], v[lines], w[lines], ts, td, sum, Dinv[lines];
99.     // Integer variables
100.    int num_rm, column = 5, sire, dam;
101.    // Lines begins with zero
102.    for(int i=0;i<lines;i++) {
103.        Dinv[i] = 0;
104.    }
105.    // All vectors receive zero values
106.    for(int i=0; i<lines;i++){
107.        v[i] = 0.0;
108.        u[i] = 0.0;
109.        w[i] = 0.0;
110.    }
111.
112.    // u = vector (1 x 1) save sum of squares of Henderson's L matrix elements. At the
        end of process the vector "u" will be the diagonal elements of NRM.
113.    // v = vector (1 x 1) save the diagonal L elements (square root of D elements).
114.    // w = vector (1 x 1) save probabilities related with sum of products of sire and
        dams for all "1" animals.
115.
116.    // First iteration to fill the vectors "u","v" and "w"
117.    for(int i=0; i < lines; i++){
118.        // Step 2
119.        // If both parents are unknown
120.        if(Matrix[i][1]==0 && Matrix[i][2]==0){
121.            v[i] = 1;
122.            u[i] = u[i] + 1;
123.        }

```

```

124.     else{
125.         // At least one of the parents is known
126.         v[i] = sqrt(1 - u[i] + 0.5*w[i]);
127.         u[i] = u[i] + pow(v[i], 2.0);
128.     }
129.
130.     // Loop to fill the vectors u,v e w
131.     for(int k=i+1; k<lines; k++){
132.         // If both parents are known
133.         if(Matrix[k][1] > 0 && Matrix[k][2] > 0){
134.             dam = Matrix[k][1];
135.             td = v[dam-1];
136.             sire= Matrix[k][2];
137.             ts = v[sire-1];
138.         }
139.
140.         // If the dam is unknown and the sire known
141.         else if(Matrix[k][1] == 0 && Matrix[k][2] > 0){
142.             sire= Matrix[k][2];
143.             ts = v[sire-1];
144.             td = 0.0;
145.         }
146.
147.         // If the dam is known and the sire unknown
148.         else if(Matrix[k][1] > 0 && Matrix[k][2] == 0){
149.             dam = Matrix[k][1];
150.             td = v[dam-1];
151.             ts = 0.0;
152.         }
153.
154.         // If the sire is multiple sire (MS)
155.         else if(Matrix[k][2]<0){
156.             // Evaluate the number of sires on group
157.             num_rm = Matrix[k][3];
158.             // Set the probability to sire to be the truly father
159.             // Set the equal probability for group sires
160.             float p = 1/num_rm;
161.             // If the dam is known
162.             if(Matrix[k][1]>0){
163.                 // Compute ts based on how many sire are inside the MS group and
                yours "v" values
164.                 for(int n = 0; n < columns; n++){
165.                     if(Matrix[k][4+n] > 0){
166.                         sire= Matrix[k][4+n];
167.                         sum = sum + (v[sire-1]*p);
168.                     }
169.                 }
170.                 dam = Matrix[k][1];
171.                 td = v[dam-1];
172.                 ts = soma;
173.             }

```

```

174.         // If dam is unknown
175.         if(Matrix[k][1]==0){
176.             // Compute ts based on how many sire are inside the MS group and
           yours "v" values
177.             for(int n = 0; n < columns; n++){
178.                 if(Matrix[k][4+n] > 0){
179.                     sire = Matrix[k][4+n];
180.                     sum = sum + (v[sire-1]*p);
181.                 }
182.             }
183.             ts = sum;
184.             td = 0;
185.         }
186.     }
187.
188.     // Set 'ts' e 'td' when none of the last conditions are accepted
189.     else{
190.         ts=0.0;
191.         td=0.0;
192.     }
193.
194.     // Compute 'u', 'v' e 'w' values for "k" animal
195.     //u[k] = u[k] + 0.25 * pow((ts+td), 2.0);
196.     w[k] = w[k] + ts*td;
197.     v[k] = 0.5 * (ts+td);
198. }
199. }
200.
201. // Loop for compute Dinv value
202. for (int i=0;i<lines;i++){
203.     Dinv[i]= (1/pow(v[i],2));
204. }
205.
206. // Fill the Dinv vector with final data
207. for(int i=0; i < lines;i++){
208.     Dinv[i]= 1/pow(v[i],2);
209. }
210. // Export final datafile
211. SaveFile(Dinv, lines, Matrix, File);
212.}
213.
214. /***** Paul Sullivan algorithm (1995) *****/
215.
216. // Function that do not consider inbreeding
217.
218. void NoInbred(int lines, int Matrix[][9], string File){
219.
220.     float sire, dinv[lines];
221.     int dam;
222.     for(int i=0;i<lines;i++){
223.

```

```

224.     // Identify MS progenies
225.     if(Matrix[i][2]<0){
226.         sire = 0.25/Matrix[i][3];
227.         if(Matrix[i][1]>0){dam = 1;}
228.         else{dam = 0;}
229.     }
230.     // Identify animals with certain paternity
231.     else if(Matrix[i][2]>0){
232.         sire = 1;
233.         if(Matrix[i][1]>0){dam = 1;}
234.         else{dam = 0;}
235.     }
236.     // Identify animals with uncertain paternity
237.     else{
238.         sire = 0;
239.         if(Matrix[i][1]>0){dam = 1;}
240.         else{dam = 0;}
241.     }
242.
243.     // Compute Dinv for each animal
244.     dinv[i] = 1/(1-0.25*dam-0.25*sire);
245. }
246.
247. // Export final datafile
248. SaveFile(dinv, lines, Matrix, File);
249.}
250.
251.// Export function
252.void SaveFile(float dinv[], int lines, int Matrix[][9], string File){
253.
254.    // Output file name
255.    // Creates a new string variable
256.    string file = File.substr(0,File.length()-4);
257.    char * Files = new char[file.length()+1];
258.    strcpy(Files, file.c_str());
259.    strcat(Files, "_dinv.ped");
260.
261.    // Set output variables
262.    ofstream Output, Report;
263.    // Opens output file
264.    Output.open(Files);
265.    // Saves information on output file
266.    for(int i=0; i < lines; i++){
267.        Output<< Matrix[i][0] << " " << setprecision(4) << dinv[i] << "\n";
268.    }
269.    // Closes output file
270.    Output.close();
271.
272.    cout << "\nLog file: mendelian_sample.log\n" << endl;
273.
274.    // Loop for compute some users information

```

```

275.     float min=2, max=0, dmin=2, dmax=0;
276.     for(int i=0;i<lines;i++){
277.         if(dinv[i] < min){min = dinv[i];}
278.         if(dinv[i] > max){max = dinv[i];}
279.         if(1/dinv[i] < dmin){dmin = 1/dinv[i];}
280.         if(1/dinv[i] > dmax){dmax = 1/dinv[i];}
281.     }
282.
283.     // Opens report
284.     Report.open("mendelian_sample.log");
285.     Report << "Data file: " << File << endl;
286.     Report <<"Animals in data file: " << lines<<endl;
287.     Report << "\nOutput file created: "<< file <<"_dinv.ped\n";
288.     Report << "Minimun diagonal element of L matrix: " << dmin << " \n";
289.     Report << "Maximun diagonal element of L matrix: " << dmax << " \n";
290.     Report<< "Minimun element of inverse of the genetic variance due to mendelian sample
: " << min << " \n";
291.     Report<< "Maximun element of inverse of the genetic variance due to mendelian sample
: " << max << " \n";
292.}

```