

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP**

**CÂMPUS DE JABOTICABAL**

**AVALIAÇÃO DE ESTRATÉGIAS DE GENOTIPAGEM EM  
SITUAÇÕES DE INCERTEZA DE PATERNIDADE E SEU  
IMPACTO SOBRE AS AVALIAÇÕES GENÔMICAS EM  
BOVINOS DE CORTE**

**Rafael Lara Tonussi**

**Zootecnista**

**2016**

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP**

**CÂMPUS DE JABOTICABAL**

**AVALIAÇÃO DE ESTRATÉGIAS DE GENOTIPAGEM EM  
SITUAÇÕES DE INCERTEZA DE PATERNIDADE E SEU  
IMPACTO SOBRE AS AVALIAÇÕES GENÔMICAS EM  
BOVINOS DE CORTE**

**Rafael Lara Tonussi**

**Orientador: Prof. Dr. Fernando Sebastián Baldi Rey**

**Coorientador: Dr. Rafael Medeiros de Oliveira Silva**

**Tese apresentada à Faculdade de  
Ciências Agrárias e Veterinárias – Unesp,  
Câmpus de Jaboticabal, como parte das  
exigências para a obtenção do título de  
Doutor em Genética e Melhoramento  
Animal**

**2016**

T667a Tonussi, Rafael Lara  
Avaliação de estratégias de genotipagem em situações de incerteza de paternidade e seu impacto sobre as avaliações genômicas em bovinos de corte / Rafael Lara Tonussi. – Jaboticabal, 2016  
xii, 51 p. : il. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2016  
Orientador: Fernando Sebastian Baldi Rey  
Banca examinadora: Raysildo Barbosa Lôbo, Cláudio de Ulhoa Magnabosco, Daniel Gustavo Manzan Gordo, Daniel Jordan de Abreu Santos

Bibliografia

1. Acurácia. 2. Avaliação genética. 3. Pedigree incompleto. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.2

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação  
– Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA TESE: AVALIAÇÃO DE ESTRATÉGIAS DE GENOTIPAGEM EM SITUAÇÕES DE INCERTEZA DE PATERNIDADE E SEU IMPACTO SOBRE AS AVALIAÇÕES GENÔMICAS EM BOVINOS DE CORTE

**AUTOR: RAFAEL LARA TONUSSI**


**ORIENTADOR: FERNANDO SEBASTIAN BALDI REY**

**COORIENTADOR: RAFAEL MEDEIROS DE OLIVEIRA SILVA**

Aprovado como parte das exigências para obtenção do Título de Doutor em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

  
Prof. Dr. FERNANDO SEBASTIAN BALDI REY  
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

  
Prof. Dr. RAYSILDO BARBOSA LOBO  
Departamento de Genética / USP - Ribeirão Preto, SP

  
Prof. Dr. CLAUDIO DE ULHOA MAGNABOSCO  
EMBRAPA / Brasília, DF

  
Pós-doutorando DANIEL GUSTAVO MANSAN GORDO  
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

  
Pós-doutorando DANIEL JORDAN DE ABREU SANTOS  
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Jaboticabal, 14 de dezembro de 2016

## **DADOS CURRICULARES DO AUTOR**

Rafael Lara Tonussi, nascido em Laranjal Paulista – São Paulo, no dia 05 de Outubro de 1.985, filho de João Batista Tonussi e Rosa Maria de Arruda Lara Tonussi. Iniciou o curso de Zootecnia na Universidade Estadual de Maringá, PR. Em dezembro de 2009 concluiu a graduação, obtendo-se o título de Zootecnista. Em março de 2011 ingressou no curso de Pós-Graduação em Genética e Melhoramento Animal, na faculdade de Ciências Agrárias e Veterinárias (FCAV) – Câmpus Jaboticabal, como bolsista, inicialmente CNPq e posteriormente da Fundação de Amparo à Pesquisa, FAPESP, sob orientação da professora Dra. Lucia Galvão de Albuquerque, obtendo o título de mestre em 26 de fevereiro de 2013. Em março de 2013 ingressou no curso de doutorado no mesmo programa de Pós-Graduação sob orientação do Prof. Dr. Fernando Sebastián Baldi Rey, inicialmente bolsista CAPES e posteriormente FAPESP. Obteve o título de doutor em 14 de dezembro de 2016.

“Os nossos sonhos renascem, pois é preciso a gente lutar

Mil coisas pra pensar, mil coisas pra gente fazer

De repente o medo, o medo de vencer, esqueço que o mais importante da vida é  
viver

É preciso ter fé, é preciso ter força é preciso a gente lutar

Mas se um dia eu errar, não posso me sentir tão derrotado

Sei que o importante é tentar, não posso ter pressa tenho que ser calmo

A vida é uma estrada que temos que caminhar, posso até ser fraco, mas tenho uma  
força dentro de mim

Está no que eu faço, está em minha vida, enfim é preciso lutar

É preciso lutar, os sonhos renascem a cada instante, vale a pena olhar o horizonte e  
lá estão os nossos sonhos, as nossas lutas, a nossa força e a nossa fé

É preciso ter fé, é preciso ter força, é preciso a gente poder lutar”

(movimento MINI e T.L.C.)

Aos meus pais pelo exemplo de vida, pelas oportunidades a mim  
oferecidas, dedicação, carinho e amor

Aos meus irmãos pelo apoio e companhia

À Ana Fabrícia pelo companheirismo, compreensão e amor

Aos meus avós Joaquim, Deolinda, Batistino e Thereza (*in memoriam*)  
pelos valiosos ensinamentos

**DEDICO**

## AGRADECIMENTOS

À Deus, pelo dom da vida, pela oportunidade de continuar a vida e pelo infinito amor.

Ao meu orientador Fernando Baldi, pela oportunidade concedida, ensinamentos, colaboração e sugestões sem os quais não seria possível a realização dessa tese. Agradeço também pelo companheirismo e por sempre estar presente e disposto a ajudar em todas as etapas desse trabalho.

Ao meu coorientador e amigo Rafael Medeiros, o qual também sempre esteve presente e disposto a ajudar a qualquer dia e horário. O meu muito obrigado pelas sugestões, paciência e ensinamentos.

Aos membros da banca examinadora, Prof. Dr. Cláudio, Prof. Dr. Raysildo, Dr. Daniel Gordo e Dr. Daniel Jordan, pelas valiosas sugestões.

Aos docentes do programa em Genética e Melhoramento Animal pelos ensinamentos.

À CAPES e a Fapesp (processo 2013/25910-0) pelo apoio financeiro, sem o qual, não seria possível a dedicação exclusiva às atividades da pós-graduação.

À ANCP (Associação Nacional dos Criadores e Pesquisadores) por disponibilizar os dados.

Aos meus pais, João e Rosa, por todo amor, dedicação, incentivo e educação. Exemplos de vida, perseverança, garra e caridade. Eternamente grato por tê-los como pais.

Aos meus irmãos Renato e Rosana pelo carinho, amor e apoio. Por todas as mensagens e ligações trocadas. São minhas melhores lembranças de infância. E serão para sempre meu elo, parte de mim. Afinal somos frutos do mesmo amor.

Aos meus novos irmãozinhos, os gêmeos João e Pedro, os quais vieram para renovar e alegrar a vida de toda a família.



Ao amor da minha vida, minha querida Ana Fabrícia, por todo apoio, dedicação e ajuda, a qual tornou esse trabalho mais leve. Pelo amor, paciência e compreensão. A minha eterna gratidão.

Aos meus queridos afilhados Felipe, Ethan e Gustavo pela grande alegria de poder contribuir e fazer parte das suas vidas.

Às minhas tias Alice, Ana, Iracema, Maria e Regina por serem um exemplo de humildade, temor à Deus e de ajuda ao próximo desde minhas primeiras lembranças até os dias de hoje. Mulheres guerreiras que têm toda a minha admiração.

À Larissa Gabriela, Vinícius, Ana Laura, Guilherme, Clara, Beatriz e André por me fazerem sentir crianças novamente e tornar as idas para Laranjal/Chácara tão especiais. As brincadeiras com eles são tão simples, puras e divertidas.

Aos meus grandes amigos Tiago, Mayra, Guilherme, Israel, Sheila, Bruno, Janaína, Paola, Gustavo, Pamela, Fabiana, Kattu II, Mara, Célia, Lígia, Maria Inês, Renan, Carlos Eduardo, Carina, Arthur, Pessin, Thaís, Nicolas, Tharsila, Karina, Felipe Fernando, Lucas, Luiz Cláudio, Tiaguinho, Vanessa, Rodrigo, Jovane, Juliane, Grazielle, Fernanda Capucho, Fernanda Balarini, Fred, Lucilene, Paulinho, Mariane, Alessandra e Bruno pelas agradáveis companhias, reuniões e infinitas conversas. Fazem parte da minha história e crescimento pessoal. Amigos(as), irmãos(as) para a vida toda.

Aos meus queridos “hermanitos” Elisa, Bianca, Fabi, Marcos, Mari e Hermenegildo por tornarem o ambiente de trabalho, as intermináveis tardes na salinha muito mais divertidas. Enfim pela companhia e convivência diária.

Aos amigos da salinha Fabrícia, Espigolan, Willian, Malane, Ana Cristina, Dani Barreto, Dani Beraldo, Larissa, Luciana, Lúcio, Lucas, Samuel, Gabriela, Guilherme, Anderson, Andrés, André, Daiane, Diércles, Laiza, Giovana, Thaise, Henrique José por tornarem as pausas para o café mais agradáveis. Além de serem os melhores companheiros (as) para churras. Também não podemos deixar de agradecer as agregadas Bárbara e Amanda, as quais tornam as saídas e conversas mais divertidas.

Aos meus amigos da UEM, os quais estão presentes todos os dias nas mais variadas e engraçadas mensagens pelo whatsapp. Obrigado Barão, Che, Renan, Birigui, Chico, Rubem, Fernandão, Buiú, Jaum Preto e Jaum Branco por me fazerem rir à toa.

À minha eterna amiga Lilian (*in memoriam*), pelas boas lembranças deixadas enquanto permaneceu entre nós.

À minha família VPA Henrique, Renata, Thais, André, Neia, Valmir, Cláudia, Bianca, Júlia, Kátia, Cris, Dito, Isa, Neto e Valentina os quais tive a honra de conhecer por uma causa tão nobre. Obrigado pelas companhias, resgates, reuniões e comilanças hahaha.

À minha grande família Gislene, Adelino, Luís Carlos, Simone, Maria Luzia, Pedro, Eduardo, Vanda, Camila, Vinícius, Zezinho, Maria, Donizete, Lucas, André, Neusa, Danilo, Daniel, Djalma, Ana, Toninho, Etiane, Saulo, Ivana, Kaká, Salomé, Bruno, Carlinhos, Juliana, Rogério, Carlos, Ana, Tiago, Débora, Dila, Irineu (*in memoriam*), Maria, Luciana, Pedro, Paulo, Karol, Joaquim, Alexandre e Larissa pela grande torcida e reuniões de família.

Aos meus padrinhos Célia e Ricardo por se fazerem sempre tão presentes e serem companhias agradáveis. Também as suas filhas Patrícia, Débora e Larissa que são muito queridas e especiais.

À família da Fabrícia (Patrícia, Karla, Tiago, Rafael, Maria Luzia, Francisco) pela acolhida nas festas de fim de ano, por me receber tão bem e me fazer sentir em casa.

À minha família Maringaense Adelina, Israel, Andrea, Gil, Israel, Sheila e Ethan por me receberem sempre tão bem, com muita festa, alegria e hospitalidade. Família do coração que sempre estarão comigo aonde quer que eu vá.

À Suria, Catita, Sushi e Berenice que são mais que apenas cachorros. São exemplos de amor incondicional. A alegria de nos receber todos os dias quando chegamos em casa, é mais que gratificante.

## SUMÁRIO

Resumo .....	i
Abstract .....	iii
Lista de abreviaturas .....	v
CAPÍTULO 1 – Considerações Gerais.....	1
1.Introdução.....	1
2. Objetivo geral .....	3
2.1. Objetivos específicos .....	3
3. Revisão de Literatura.....	4
3.1. Características de interesse econômico.....	4
3.2. Métodos para a predição de valores genômicos .....	5
3.3. Modelos para predição dos valores genéticos em situações de incerteza de paternidade.....	7
4. Referências .....	9
CAPÍTULO 2 - Application of single step genomic BLUP method under different uncertain paternity scenarios using simulated data .....	14
Abstract .....	14
1. Introduction .....	16
2. Material and Methods .....	17
2.1. Simulated Population .....	17
2.2. Simulated Genome.....	18
2.3. BLUP and ssGBLUP methods.....	20
2.4. Scenarios .....	21
3. Results and Discussion .....	21
4. Conclusions .....	30

5. References .....	30
CAPÍTULO 3 - Impact of multiple sire mating system on the accuracy of genomic breeding value prediction in a beef cattle population under selection .....	35
Abstract .....	35
1.Introduction .....	36
2. Material and Methods.....	38
2.1. Data.....	38
2.2. BLUP and ssGBLUP method .....	39
2.3. Tested scenarios .....	40
3. Results and Discussion .....	40
4. Conclusions .....	46
5. Implications .....	46
6. References .....	47

## AVALIAÇÃO DE ESTRATÉGIAS DE GENOTIPAGEM EM SITUAÇÕES DE INCERTEZA DE PATERNIDADE E SEU IMPACTO SOBRE AS AVALIAÇÕES GENÔMICAS EM BOVINOS DE CORTE

**Resumo** – Reprodutores múltiplos (RM) é o sistema de manejo mais comum em produção de bovinos de corte principalmente devido a facilidade e baixo custo de manejo. No entanto, RM não permite a identificação de paternidade, tornando o pedigree incompleto um dos principais obstáculos para avaliações genéticas precisas. Existe um crescente interesse na investigação do uso de dados genômicos em modelos com paternidade incerta, visando aumentar a acurácia e diminuir o viés nas avaliações genéticas. Portanto, o objetivo desse estudo foi investigar a aplicação do BLUP e single step genomic BLUP (ssGBLUP) em diferentes cenários com paternidade incerta utilizando dados simulados e reais em bovinos de corte. Foram simulados genótipos, pedigree e fenótipos para idade ao primeiro parto (IPP) e peso aos 550 dias (P550) utilizando herdabilidades baseadas em dados reais (0,12 para IPP e 0,34 para P550). O genoma foi simulado com um comprimento total de 2.333 cM, com 735.293 marcadores bialélicos e 7.000 QTL distribuídos aleatoriamente ao longo dos 29 cromossomos. Foi assumido que os QTLs explicaram 100% da variância genética. Para QTL, a quantidade de alelos variou aleatoriamente de dois a quatro por loci. Foram estudados cenários com 0%, 25%, 50%, 75% e 100% de RM e foram testados quatro tipos de escalas entre as matrizes **G** e **A<sub>22</sub>**. A acurácia e viés foram calculados para cinco grupos: ALL = todos os animais; BULL = apenas touros; GEN = animais genotipados, FEM = fêmeas e YOUNG = machos jovens. O uso da informação genômica no modelo (ssGBLUP) apresentou acurácia maior (variando de 0,31 a 0,97) do que o BLUP tradicional (variando de 0,05 a 0,97), especialmente no grupo YOUNG. No estudo com dados reais, todos os modelos incluíram grupos de contemporâneos e classe de idade da vaca como efeitos fixos. A acurácia do valor genético (EBV / GEBV) foi calculada em cada cenário para oito grupos de animais: ALL = todos os animais, BULL = apenas touros com dez ou mais progênes; GEN = animais genotipados, GENwithPHEN = animais genotipados com fenótipos, GENwithoutPHEN = animais genotipados sem fenótipos, YOUNG = machos e fêmeas jovens sem fenótipos, YwithoutGEN =

animais jovens sem fenótipos e genótipos e YwithGEN = animais jovens sem fenótipos e com genótipos. As acurácias do EBV (método BLUP) variaram de 0,02 a 0,46 para P455 e 0,04 a 0,18 para IPP, enquanto que as acurácias do GEBV (ssGBLUP) variaram de 0,13 a 0,48 para P455 e 0,16 a 0,33 para IPP. Os resultados obtidos com dados simulados e reais mostraram que acurácias do EBV e GEBV diminuíram à medida que as proporções de RM aumentaram. Além disso, o uso da informação genômica na avaliação genética pelo ssGBLUP aumentou a acurácia da avaliação, especialmente para animais com menos informações e para animais jovens.

**Palavras-chave:** acurácia, avaliação genética, pedigree incompleto, reprodutores múltiplos, viés

## EVALUATION OF GENOTYPING STRATEGIES IN SITUATIONS OF UNCERTAINTY PATERNITY AND IMPACT ON GENOMIC EVALUATION IN BEEF CATTLE

**Abstract** - Multiple service sire (MS) is the most common mating system in extensive beef production systems mainly due the facility and low management cost. However, MS does not allow the paternity identification, which makes the incompleteness of pedigree one of the main obstacles for accurate genetic evaluations. There is a grown interest to investigate the use of genomic data in uncertain paternity models aiming to increase the accuracy and to decrease bias in genetic evaluations. Therefore, the objective of this study was to investigate the application of BLUP and single step genomic BLUP (ssGBLUP) under different scenarios of uncertain paternity, using simulated and real data in beef cattle population. Genotypes, pedigree, and phenotypes for age at first calving (AFC) and weight at 550 days (W550) were simulated using heritabilities based on real data (0.12 for AFC and 0.34 for W550). The simulated genome had a total length of 2,333 cM, with 735,293 biallelic markers and 7,000 QTLs randomly distributed over 29 BTA. It was assumed that QTLs explained 100% of the genetic variance. For QTL, the amount of alleles per loci ranged randomly from two to four. Uncertain paternity scenarios using 0%, 25%, 50%, 75%, and 100% were studied. Four ways of scaling the mean of the genomic matrix ( $\mathbf{G}$ ) to match the mean of the pedigree relationship matrix among genotyped animals ( $\mathbf{A}_{22}$ ) were tested. Accuracy, bias and inflation were investigated for five groups of animals: ALL = all animals; BULL = only bulls; GEN = genotyped animals; FEM = females; YOUNG = young males. The use of genomic information in the model (ssGBLUP) provided more accurated prediction (ranging from 0.31 to 0.97) than traditional BLUP (ranging from 0.05 to 0.97), especially in the YOUNG group. In real data, all models included contemporary groups and age at calving in classes as fixed effects. The accuracy of the estimated breeding value (EBV/GEBV) prediction was calculated in each scenario with eight groups of animals: ALL = all animals in the population, BULL = only bulls with ten or more progenies; GEN = genotyped animals, GENwithPHEN = genotyped animals with phenotypes, GENwithoutPHEN = genotyped animals without phenotypes, YOUNG = male and

female young animals without phenotypes,  $Y_{\text{withoutGEN}}$  = young animals without phenotypes and genotypes, and  $Y_{\text{withGEN}}$  = young animals with phenotypes and with genotypes. Accuracies of EBV (BLUP method) ranged from 0.02 to 0.46 for W450 and 0.04 to 0.18 for AFC, while the accuracies of GEBV (ssGBLUP) ranged from 0.13 to 0.48 for W450 and 0.16 to 0.33 for AFC. The results obtained in simulated and real data showed that EBV and GEBV accuracy decreased as the proportion of MS increased. Additionally, the use of genomic information in the genetic evaluation by ssGBLUP increases the accuracy of evaluation, especially for animals with few number of information, such as young animals.

**Keywords:** accuracy, bias, genetic evaluation, missing pedigree, multiple sires



## Lista de abreviaturas

### Capítulo 1

Matriz **A**: matriz de parentesco

Matriz **A<sub>22</sub>**: matriz coeficientes de parentesco tradicional entre os indivíduos genotipados

Matriz **G**: matriz de dados genômicos

Matriz **H**: matriz que combina pedigree e informações genômicas

RM: reprodutores múltiplos

IPP: idade o primeiro parto

HIER: modelo animal hierárquico

### Capítulo 2

**A** matrix: numerator relationship matrix

**A<sub>22</sub>** matrix: genomic relationship matrix

**G** matrix: pedigree-based relationship matrix for genotyped animals

**H** matrix: matrix that combines the pedigree and genomic information

MS: multiple sires

SNP: *Single Nucleotide Polymorphism*

QTLs: *quantitative trait loci*

BLUP: best linear unbiased prediction

ssGBLUP: single step genomic BLUP

AFC: age at first calving

W550: weight at 550 days

BTA: *Bos Taurus* autosomes

ALL: all animals in the population

BULL: only bulls with at least one progeny

GEN: genotyped animals

FEM: females

YOUNG: young males without phenotype

S<sub>1</sub>: scaling for all animals considering mean diagonal  $\mathbf{A}_{22}$  = mean diagonal  $\mathbf{G}$  and mean off diagonal  $\mathbf{A}_{22}$  = mean off diagonal  $\mathbf{G}$

S<sub>2</sub>: no scaling between  $\mathbf{A}_{22}$  and  $\mathbf{G}$  matrix

S<sub>3</sub>: scaling between  $\mathbf{A}_{22}$  and  $\mathbf{G}$  matrix only the animals which have known sire and dam

S<sub>4</sub>: scaling between  $\mathbf{A}_{22}$  and  $\mathbf{G}$  matrix only those animals which have one known parent.

UPG: Unknown parent groups

HIER: hierarchical animal model

GEBV: genomic estimated breeding value

EBV: estimated breeding value

TBV: true breeding value

LD: linkage disequilibrium

MAF: Minor allele frequencies

REML: restricted maximum likelihood

GREML: genomic restricted maximum likelihood

### **Capítulo 3**

W450: weight adjusted at 450 days of age

AFC: age at first calving

ALL: all animals with phenotypic records

BULL: only bulls with ten or more progenies

GEN = genotyped animals

GENwithPHEN = genotyped animals with phenotypes

GENwithoutPHEN = genotyped animals without phenotypes

YOUNG = male and female young animals without phenotypes and progenies

YwithoutGEN = young animals without phenotypes and genotypes

YwithGEN = young animals without phenotypes and with genotypes

AIREML= average information restricted maximum likelihood estimation

## **CAPÍTULO 1 – Considerações Gerais**

### **1.Introdução**

A bovinocultura de corte destaca-se no agronegócio brasileiro, uma vez que o país possui o segundo maior rebanho bovino mundial com mais de 213 milhões de cabeças de gado (MAPA, 2015), destacando-se como um dos maiores exportadores de carne bovina. A extensão territorial do Brasil é um dos principais fatores que contribuem para esses resultados, permitindo que a maioria dos animais sejam criados em sistema de pastagem, que favorece o baixo custo de criação em comparação com outros países produtores de carne.

O Brasil também possui uma ampla variedade de sistemas de produção de carne que reflete na diversificação dos produtos. Assim, pode-se atender diversos mercados, desde cortes com menor valor comercial até nichos específicos com carnes mais nobres (ABIEC, 2013). Diante da crescente demanda mundial por alimentos e proteínas de origem animal, o desenvolvimento e implementação de novas tecnologias têm sido necessários em sistemas de produção de gado de corte para atender aos padrões do mercado consumidor. A fim de aumentar a produtividade em bovinos de corte uma das alternativas é o uso da seleção, a qual altera as frequências alélicas, selecionando animais geneticamente superiores para determinadas características de interesse econômico.

Uma das ferramentas mais eficientes para realizar a identificação de animais geneticamente superiores é a seleção, a qual é realizada em todos os programas de melhoramento animal visando maior eficiência na produção. Animais com maior velocidade de crescimento ou precocidade sexual implicam diretamente no encurtamento do ciclo de produção, possibilitando maior retorno econômico ao criador (BOLIGON et al., 2009).

A informação parcial de pedigree, limita a possibilidade de realizar avaliações genéticas confiáveis em rebanhos de bovinos de corte e, conseqüentemente, afeta a seleção de animais geneticamente superiores, além de afetar diretamente a acurácia dessas avaliações e o diferencial de seleção (CARDOSO; TEMPELMAN, 2003). No Brasil, o sistema de acasalamento mais comum é o uso de reprodutores múltiplos

(RM) em regime extensivo, o qual consiste na utilização de vários touros em um mesmo lote de matrizes durante a estação de monta, dessa forma, não é possível a identificação da paternidade das crias, dificultando a avaliação dos animais.

Em rebanhos comerciais, a porcentagem de animais com paternidade incerta (filhos de reprodutores múltiplos) pode chegar até 60%, e em muitas situações os touros são utilizados apenas na estação de monta e posteriormente descartados, dificultando a seleção de animais com características de interesse. A preferência por esse sistema de acasalamento justifica-se, principalmente, por não requerer mão de obra qualificada, baixo custo comparado a outros sistemas, além de minimizar as dificuldades quanto ao manejo pela monta controlada (VALLE; ANDREOTTI, THIAGO, 1998).

Atualmente, uma ferramenta utilizada para auxiliar a seleção de animais é o uso de marcadores do tipo SNP (*Single Nucleotide Polymorphism*), amplamente distribuídos ao longo do genoma, pressupondo que tais marcadores estão em desequilíbrio de ligação com os QTLs (*quantitative trait loci*). Os SNPs são adequados para genotipagem dos animais e podem ser utilizados nas avaliações genéticas, facilitando a identificação de animais superiores e candidatos à seleção para determinada característica, mesmo que não haja informação de pedigree. Outra vantagem do uso desses marcadores é em relação ao sucesso na identificação de bovinos em testes de paternidade, além de serem úteis na reconstrução de pedigree e para estabelecer relações entre dois ou mais indivíduos (CLARKE et al., 2014).

Em geral avaliações genômicas são mais acuradas em relação à seleção tradicional (VANRADEN, 2008), uma vez que os marcadores são espalhados por todo o genoma, dessa forma, é possível explicar uma grande porção da variância genética entre os animais. Em rebanhos comerciais é comum pedigrees incompletos (paternidade incerta) na predição das avaliações genéticas afetando diretamente a estrutura da matriz de parentesco. Logo, deve haver uma conectabilidade entre animais da matriz de parentesco (**A**) e a matriz de dados genômicos (**G**) a fim de prever valores genéticos não viesados (FORNI et al., 2011). No entanto, ainda são escassos na literatura trabalhos que consideram inclusão de paternidade incerta nas avaliações (BERRY et al., 2016).

A matriz com informações genômicas pode ser calculada de diferentes maneiras (VANRADEN 2008; GIANOLA; VANKAAM 2008), não havendo um consenso, porém, o ideal seria os elementos da diagonal e fora da diagonal da matriz **G** terem a mesma média dos da **A**<sub>22</sub> (MISZTAL et al., 2013a). Aguilár et al. (2010) utilizando uma matriz **G** observaram previsões diferentes para EBV, provavelmente essas diferenças são relacionadas as escalas entre as matrizes **A** e **G**.

A formação de pedigrees incompletos também pode afetar a acurácia do valor genômico, a qual é de fundamental importância para a aplicação da seleção genômica no melhoramento animal (BODDHIREDDY et al., 2014). No entanto, existem diversos fatores que influenciam essa acurácia, como tamanho da população (DAETWYLER et al., 2010), o grau de desequilíbrio de ligação entre marcador e o QTL, herdabilidade da característica, e o método utilizado para previsão genômica (DE LOS CAMPOS et al., 2013).

Modelos estatísticos que consideram incerteza de paternidade utilizando dados genômicos não são utilizados em programas de melhoramento genético de bovinos de corte. Considerando que pedigrees incompletos são comuns na maioria dos sistemas de produção, e devido ao pequeno número de estudos sobre eles, é de grande importância o uso de dados genômicos na avaliação com o objetivo de aumentar a precisão e diminuir viés da previsão.

## **2. Objetivo geral**

Objetivou-se com este trabalho avaliar o impacto da aplicação de informações genômicas em situações de incerteza de paternidade sobre as avaliações genéticas em bovinos de corte utilizando dados simulados e reais.

### **2.1. Objetivos específicos**

- Estimar parâmetros e valores genéticos utilizando o método BLUP e ssGBLUP utilizando dados simulados e reais para duas características, de herdabilidade baixa e moderada;
- Comparar os valores genéticos preditos pelos modelos avaliados (BLUP e ssGBLUP);

- Avaliar o impacto da inclusão das informações genômicas sobre as acurácias e viés dos valores genéticos nos diferentes cenários propostos;

### **3. Revisão de Literatura**

#### **3.1. Características de interesse econômico**

A seleção de animais geneticamente superiores para as características de interesse econômico tem sido uma das ferramentas mais importantes aplicadas pelos programas de melhoramento genético para aumentar a produção de carne visando maior lucro para o criador.

Em bovinos de corte, medidas de peso e ganho em peso em diferentes idades se destacam como critério de seleção em programas de melhoramento, uma vez que estas características são relativamente de fácil medição e apresentam herdabilidade de moderada à alta magnitude (YOKOO et al., 2007; BOLIGON et al., 2008). Isto indica que essas características podem responder rápido à seleção, uma vez que grande parte da variação fenotípica total é devido aos efeitos aditivos dos genes. Além disso, possuem correlações genéticas altas e positivas entre si (BOLIGON et al., 2009), indicando que a seleção para peso em qualquer idade poderá promover progresso genético nos pesos nas demais idades, inclusive no peso ao abate (DIAZ, 2009).

As características reprodutivas ligadas à precocidade sexual são determinantes para a eficiência econômica do sistema de produção (BOLIGON et al., 2010). Dentre estas, a idade ao primeiro parto (IPP) é utilizada em programas de melhoramento da raça Nelore como indicadora da precocidade sexual em fêmeas. Essa característica influencia diretamente a eficiência da produção, pois a redução da idade ao primeiro parto está associada à antecipação do início da idade produtiva da vaca, com isso reduz os custos de produção (PEROTTO et al., 2006). Além disso, a IPP afeta a produtividade do rebanho pois influencia a produção de bezerros durante a vida útil da vaca, aumentando intensidade e seleção e reduzindo intervalo de gerações (YOKOO et al., 2012).

Geralmente, as estimativas de herdabilidade para características reprodutivas mensuradas em fêmeas da raça Nelore são de baixa magnitude (BOLIGON et al., 2008) indicando que a característica não deve responder rapidamente à seleção. No

entanto, a IPP possui correlações genéticas positivas com características de crescimento, como o peso em diferentes idades (BOLIGON et al., 2010), dessa forma, a seleção para uma característica poderá afetar outra favoravelmente, por resposta correlacionada.

### **3.2. Métodos para a predição de valores genômicos**

Algumas metodologias têm sido propostas para estimação dos valores genômicos, resumindo-se basicamente em procedimentos que utilizam diversos (*multi-step*) ou um único passo (*single-step*) (AGUILAR et al., 2010; CHEN et al., 2011). Para prever os valores genômicos utilizando o *multi-step*, os valores genéticos que serão utilizados como pseudo-fenótipos, são estimados pelo BLUP tradicional, o qual considera a matriz de parentesco baseada apenas no pedigree.

Em seguida, estes pseudo-fenótipos são utilizados em modelo de predição genômica para estimar os efeitos dos marcadores na população de referência, nesse caso, todos os animais devem estar genotipados e, por último, prever os valores genômicos por um índice de seleção (VANRADEN et al., 2009; HAYES et al., 2009) que é formado por uma média da habilidade de transmissão genética gerada a partir das informações dos passos anteriores. As ponderações do índice são realizadas com base na herdabilidade e acurácia das características analisadas (MISZTAL et al., 2009).

As etapas do método de múltiplos passos dependem de muitos parâmetros e suposições, os quais são extremamente difíceis de verificar, sobretudo na seleção de animais. Os principais problemas estão relacionados a má qualidade e definição dos pseudo-fenótipos, por exemplo, para animais com pequeno número de progênes, como acontece em alguns casos de bovinos de corte. Assim, quando apenas uma parcela da população for genotipada, pode-se utilizar o método de único passo (AGUILAR et al., 2010), o qual apresenta fácil aplicação.

Os efeitos dos marcadores podem ser estimados com diferentes pressuposições quanto a sua distribuição, assim como podem ser estimados pelo uso de um modelo simples que inclui uma matriz com dados genômicos (AGUILAR et al., 2010). No entanto, o uso de um grande número de marcadores com variância homogênea para todos é apropriado para a maioria das características seguindo



alguns estudos com dados reais em gado de leite (HAYES et al., 2009; VANRADEN et al., 2009).

Proposto por Legarra et al. (2009) e Misztal et al. (2009) a avaliação genômica em um único passo (ssGBLUP), consiste na combinação da matriz de parentesco tradicional (**A**) com a matriz de parentesco genômico (**G**), em uma única matriz (**H**), utilizada para prever o valor genético genômico (GEBV). Assim, no modelo animal, a inversa da matriz de parentesco (**A**<sup>-1</sup>) é substituída pela matriz **H**<sup>-1</sup>, que combina pedigree e informações genômicas permitindo a predição de valores genéticos por qualquer modelo (WANG et al., 2012).

Com o ssGBLUP, todos os marcadores do tipo SNP são considerados simultaneamente com as informações fenotípicas dos animais genotipados e não genotipados. Tal método também é adequado para análises multi características (CHEN et al., 2011). Esse método também permite a predição de valores genéticos indireta para animais jovens através de efeitos dos SNP com confiabilidade de uma avaliação completa (LOURENÇO et al., 2014).

Forni et al. (2011) também demonstraram que a inclusão de dados genômicos na matriz de parentesco (**A**) permite prever e estimar valores genéticos utilizando simultaneamente informações de dados fenotípicos, pedigree e genômicos. No entanto, animais da matriz **A** devem estar em escala próxima aos da matriz **G**, para que haja conectabilidade entre esses animais. Nesse sentido, esses mesmos autores relataram que se a matriz **A** for mais esparsa do que a matriz **G** pode ocorrer superestimação da variância genética, e conseqüentemente os valores de predição genômica podem ser viesados.

Uma das limitações em relação a avaliação genética pelo método ssGBLUP é em relação a quantidade de animais genotipados no conjunto de dados. No entanto, recentemente, Fragomeni et al. (2015) apresentaram uma alternativa para esse problema, onde a inversa da matriz **G** foi derivada por um “algoritmo de animais jovens e provados (*APY – proven and young animals*), o qual considerada apenas a inversão da matriz com esses animais. Nesse mesmo estudo, foi aplicado o método ssGBLUP em grupos com 2, 5, 10, 15 e 20 mil animais genotipados e a resposta da acurácia com a utilização do algoritmo foi adequada para modelos com qualquer número de animais genotipados.

### 3.3. Modelos para predição dos valores genéticos em situações de incerteza de paternidade

Considerando animais com paternidade desconhecida Scarth e Graser (1987) relataram grandes aumentos no erro de predição e viés nas avaliações genéticas. No entanto, a utilização de equações de modelos mistos pelo método BLUP podem minimizar esses efeitos (HENDERSON, 1963). Posteriormente, o modelo com base na matriz de parentesco médio, foi proposto por Henderson (1988), o qual pressupõe o conhecimento da probabilidade de cada touro ser o verdadeiro pai de um determinado animal. Isso permite que seja usado em equações de modelos mistos para avaliações genéticas em grupos de animais com paternidade incerta. Assim, diferentes modelos estatísticos têm sido utilizados a fim de minimizar o viés nas predições dos valores genéticos de animais com paternidade desconhecida (WESTELL et al., 1988; FAMULA 1992).

O modelo animal hierárquico (HIER) também é um método para avaliação genética de animais com paternidade desconhecida, proposto por Cardoso e Tempelman (2003). Este método considera a incerteza nas atribuições de paternidade incerta, em que probabilidades *a priori* fixas são atribuídas a cada possível pai dos animais com paternidade incerta, sendo que as probabilidades *a posteriori* são alteradas de acordo com a informação contida nos dados. Desta forma, o modelo infere o mérito genético dos animais com paternidade incerta e seus possíveis pais. A principal vantagem do modelo HIER é a possibilidade de considerar adequadamente a incerteza sobre as atribuições de touros na inferência do mérito genético. Com base nos critérios bayesianos de escolha do modelo, Cardoso e Tempelman (2004) demonstraram que o HIER foi consistentemente melhor para prever o valor genético de animais com paternidade incerta quando comparado ao modelo com base na matriz de parentesco médio.

A inclusão de informações fenotípicas para inferir sobre a incerteza da paternidade pode aumentar a acurácia do valor genético predito (CARDOSO; TEMPELMAN, 2003; 2004; SAPP et al., 2007). Além disso, Famula (1993) observou que a inclusão de animais com paternidade desconhecida é mais vantajosa para touros jovens com poucos descendentes identificados. Dessa forma, os touros jovens podem se beneficiar com a inclusão de descendência com paternidade

incerta nas avaliações genéticas. Em geral, à medida que o nível de incerteza de paternidade é substituído pela inclusão da progênie com paternidade conhecida, a precisão da avaliação aumenta.

Outra maneira de suprir a ausência de informação de pai ausente sobre as progênies é a inclusão de grupos genéticos (WESTELL et al., 1988) os quais representariam a contribuição genética média sobre as progênies sem identificação de paternidade com registros de produção. Dessa forma, o mérito genético predito do animal seria um valor corrigido em função do grupo genético a que este pertença (THERON; KANFER; RAUTENBACH, 2003).

Os grupos genéticos foram inicialmente desenvolvidos para os animais fundadores e representam os efeitos referentes ao parentesco, não contabilizado na análise. Henderson (1974) e Kennedy (1981) observaram que ignorar esses grupos genéticos poderia aumentar o viés do EBV. Assim, a formação de grupos de pais desconhecidos (*UPG - Unknown Parents Group*) podem ser incluídos nas equações do modelo misto para explicar as diferenças genéticas entre os grupos de animais definidos em avaliações genéticas (TSURUTA et al., 2014) apenas tomando cuidado para não serem confundidos com grupos de contemporâneos.

Recentemente, o ssGBLUP tem sido utilizado para a avaliação genética de animais com genealogia desconhecida através da incorporação de UPG nas equações dos modelos mistos (TSURUTA et al., 2014). Neste sentido, Misztal et al. (2013a) e Tsuruta et al. (2014), reportaram problemas de convergência com métodos iterativos e incompatibilidade entre matriz  $\mathbf{A}_{22}$  e  $\mathbf{G}$ . Segundo os autores, estes problemas podem ser devido a pedigrees pequenos ou incompletos, erros no pedigree, falta de qualidade dos genótipos, além da presença de múltiplas raças. Em situações de animais genotipados com paternidade desconhecida, problemas de conexão e escala entre a matriz de pedigree ( $\mathbf{A}_{22}$ ) e matriz  $\mathbf{G}$  são esperados, porque a matriz  $\mathbf{A}_{22}$  é menos informativa, devido à presença de pedigree ausente.

Em avaliações genéticas utilizando o método ssGBLUP, quando nos modelos forem inclusos UPGs, as soluções podem ser tendenciosas principalmente se esses grupos forem considerados apenas na criação da matriz  $\mathbf{A}$ . Nesse caso, o viés pode ser pequeno para grandes populações com poucos animais genotipados, caso contrário, pode ser maior (MISZTAL et al., 2013b). Sendo assim, a utilização de

modelos mistos, de pais desconhecidos como efeitos separados e as possibilidades da obtenção da matriz **A** deverão minimizar o viés (MISZTAL et al., 2013b).

Nesse mesmo trabalho, Misztal et al. (2013b), aplicando o método ssGBLUP em população com pedigree incompleto, consideraram várias opções para minimizar a inclusão dos UPG. Dentre as opções destacam-se: modificar as equações de modelo misto para incluir os UPG nas relações genômicas; considerá-los como efeitos separados ou até ignorá-los caso sejam relativamente pequenos. Além disso, pode ser considerado um efeito poligênico que incluem os UPG ao modelo. Assim, todas essas opções podem ser utilizadas para diminuir o viés com utilização dos UPG, no entanto, devem avaliadas a origem desse UPG (pedigree incompleto na população de seleção, múltiplas raças, diferentes linhagens).

#### 4. Referências

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, p. 743–752, 2010.

ABIEC – Associação Brasileira das Indústrias Exportações de Carne. Disponível em: [http://www.abiec.com.br/3\\_pecuaria.asp](http://www.abiec.com.br/3_pecuaria.asp). Acesso em 31/08/2016.

BERRY, D. P.; GARCIA, J. F.; GARRICK, D. J. Development and implementation of genomic predictions in beef cattle. **Animal Frontiers**, v. 6, p. 32-38, 2016.

BODDHIREDDY, P.; KELLY, M. J.; NORTHCUTT, S.; PRAYAGA, K. C.; RUMPH, J.; DENISE, S. Genomic predictions in Angus cattle: Comparisons of sample size, response variables, and clustering methods for cross-validation. **Journal of Animal Science**, v. 92, p. 485-497, 2014.

BOLIGON, A. A.; ALBUQUERQUE, L. G.; RORATO, P. R. N.; Associações genéticas entre pesos e características reprodutivas em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia**, v. 37, n. 41, p. 596-601, 2008.

BOLIGON, A. A.; ALBUQUERQUE, L. G.; 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.

BOLIGON, A. A.; ALBUQUERQUE, L. G.; MERCADANTE, M. E. Z.; LOBO, R. B. Study to relations between age at first calving, average weight gains and weights

from weaning to maturity in Nelore cattle. **Revista Brasileira de Zootecnia**, v. 39, p.746- 751, 2010.

CARDOSO, F. F.; TEMPELMAN, R. J. Bayesian inference on genetic merit under uncertain paternity. **Genetic Selection Evolution**, v. 35, p. 469-487, 2003.

CARDOSO, F. F.; TEMPELMAN, R. J. Genetic evaluation of beef cattle accounting for uncertain paternity. **Livestock Production Science**, v. 89, p. 109-120, 2004.

CHEN, C. Y.; MISZTAL, I.; AGUILAR, I., TSURUTA, S.; MEUWISSEN, T. H. E.; AGGREY, S. E.; WING, T.; MUIR, W. M. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. **Journal of Animal Science**, v. 89, p. 23–28, 2011.

CLARKE, S. M.; HENRY, H. M.; DODDS, K. G.; JOWETT, T. W. D.; MANLEY, T. R.; ANDERSON, R. M.; MCEWAN, J. C. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep, **PLOS ONE**, v. 9, p. 1-11, 2014.

DAETWYLER HD, PONG-WONG R, VILLANUEVA B, WOOLLIAMS JA: The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021-1031. 2010.

DE LOS CAMPOS, HICKEY, J. M. PONG-WONG, R. DAETWYLER, H. D. CALUS, M. P. L. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. **Genetics**, v. 193, p. 327–345. 2013.

DIAZ, I. D. P.S. **Interação Genótipo – Ambiente no peso ao sobreano na raça Nelore**. 2009. 61f. Dissertação (Mestrado em Genética e Melhoramento Animal) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho, Jaboticabal, 2009.

FAMULA, T. R. Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty. **Journal of Animal Science**, v. 70, p.1045-1058, 1992.

FAMULA, T. R. The Contribution of Progeny of Uncertain Paternity to the Accuracy of Sire Evaluation. **Journal of Animal Science**, v. 71, p.1136-1141, 1993.

FRAGOMENI, B. O.; LOURENÇO, D. A. L.; TSURUTA, S.; MASUDA, Y.; AGUILAR, I.; LEGARRA, A.; LAWLOR, T. J.; MISZTAL. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. **Journal of Dairy Science**, v. 98, p. 4090-4094, 2015.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetic Selection Evolution**, v. 43, n. 1, p. 1-7, 2011.

GAO, H.; CHRISTENSEN, O. F.; MADSEN, P.; NIELSEN, U. S.; ZHANG, Y.; LUND, M. S.; SU, G. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. **Genetics Selection Evolution**, v. 44, p. 1-8, 2012.

GIANOLA, D.; VANKAAM, J. B. C. H. M. Reproducing kernel Hilbert spaces regression methods for genomic prediction of quantitative traits. **Genetics**, v. 178, n. 4, p. 2289-2303, 2008.

HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v. 92, p. 433-443, 2009.

HENDERSON, C. R., Selection index and expected genetic advance, Publication. **Statistical Genetics and Plant Breeding**, p. 982, pp. 141-163, edited by HANSON, W. D.; ROBINSON, H. F. National Academy of Science, National Research Council, Washington, DC, 1963.

HENDERSON, C. R. General flexibility of linear model techniques for sire evaluation. **Journal of Dairy Science**, v. 57, p. 963-972, 1974.

HENDERSON, C. R. Use of an average numerator relationship matrix for multiple-sire joining, **Journal of Animal Science**, v. 66, p. 1614-1621, 1988.

KENNEDY, B.W. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. **Journal of Dairy Science**, v. 64, p. 689-697, 1981.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656-4663, 2009.

LOURENÇO, D. A. L.; MISZTAL, I.; TSURUTA, S.; AGUILAR, I.; LAWLOR, T. J.; FORNI, S.; WELLER, J. I. Are evaluations on Young genotyped animals benefiting from the past generations? **Journal of Dairy Science**, v. 97, p.3930-3942, 2014.

MAPA – Ministério de Agricultura, Pecuária e Abastecimento. Disponível em: <http://www.agricultura.gov.br/animal/especies/bovinos-e-bubalinos>. Acesso em 31/08/2016.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, p. 4648-4655, 2009.

MISZTAL, I.; AGGREY, S. E.; MUIR, W. M. Experiences with a single step genome evaluation. **Poultry Science**, v. 92, p. 2530-2534, 2013a.

MISZTAL, I.; VITEZICA, Z. G.; LEGARRA, A.; AGUILAR, I.; SWAN, A. A. Unknown-parent groups in single-step genomic evaluation. **Journal of Animal Breeding and Genetics**, v. 130, p. 252-258, 2013b.

PEROTTO, D.; MIYAGI, A. P.; SOUZA, J. C.; MOLETTA, J. L.; FREITAS, J. A. Estudos de características reprodutivas de animais da raça Canchim, criados a pasto, no Estado do Paraná, Brasil. **Archives of Veterinary Science**, v. 11, n. 2, p. 1-6, 2006.

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

SCARTH, R. D.; GRASER, H. U. Comparisons of genetic evaluation from animal models for single and multiple sire joinings. Proc. 6th Conf. Aust. Assoc. **Animal Breeding and Genetics**, p. 369, 1987.

SU, G.; MADSEN, P.; NIELSEN, U. S.; MÄNTYSAARI, E. A.; AAMAND, G. P.; CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction for Nordic Red Cattle using one-step and selection index blending. **Journal of Dairy Science**, v. 95, p. 909-917, 2012.

THERON, H. E.; KANFEN, F. H. J.; RAUTENBACH, L. The effect of phantom parent groups on genetic trend estimation. **South African Journal of Animal Science**, v. 32, n. 2, p. 130-135, 2003.

TSURUTA, S.; MISZTAL, I.; LOURENÇO, D. A. L.; LAWLOR, T. J. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. **Journal of Dairy Science**, v. 97, p. 5814–5821, 2014.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, p. 4414-4423, 2008.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v.92, p. 16-24, 2009.

VALLE, E. R.; ANDREOTTI, R.; THIAGO, L. R. S. Estratégias para aumento da eficiência reprodutiva e produtiva em bovinos de corte. Campo Grande- MS, **EMBRAPA**, 80p., 1998.

WANG, H.; MISZTAL, I.; AGUILAR, I., LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v. 94, p. 73-83, 2012.

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

YOKOO, M. J.; ALBUQUERQUE, L. G.; LOBO, R. B.; SAINZ, R. D.; CARNEIRO JUNIOR, J. M.; BEZERRA, L. A. F.; ARAUJO, F. R. C. Estimativas de parâmetros genéticos para altura do posterior, peso e circunferência escrotal em bovinos da raça Nelore. **Revista Brasileira de Zootecnia**, v.36, n.6, p.1761-1768, 2007.

YOKOO, M. J.; MAGNABOSCO, C. U.; ROSA, G. J. M.; LOBO, R. B.; ALBUQUERQUE, L. G. Características reprodutivas e suas associações com outras características de importância econômicas na raça Nelo. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 64, n. 1, p. 91-100, 2012.



## **CAPÍTULO 2 - Application of single step genomic BLUP method under different uncertain paternity scenarios using simulated data**

**Abstract** - In extensive beef cow-calf production systems, multiple service sire (MS) is the most common mating system, which consists in keeping several bulls in the same paddock with cows during the breeding season. Despite the management advantages of this mating system, it restricts the possibility of progeny paternity identification and increases the occurrence of missing pedigree, reducing the reliability of beef cattle genetic evaluations. The objective of this study was to investigate the application of BLUP and single step genomic BLUP (ssGBLUP) in different scenarios of uncertain paternity in a simulated beef cattle population. Genotypes, pedigree, and phenotypes for age at first calving (AFC) and weight at 550 days (W550) were simulated using heritabilities based on real data (0.12 for AFC and 0.34 for W550). Uncertain paternity scenarios using 0, 25, 50, 75, and 100% of multiple service sires (MS) were studied. The simulated genome had a total length of 2,333 cM, with 735,293 biallelic markers and 7,000 QTLs randomly distributed over 29 BTA. It was assumed that QTLs explained 100% of the genetic variance. For QTL, the amount of alleles per loci ranged randomly from two to four. The BLUP method that considers phenotypic and pedigree data, and ssGBLUP method that combines phenotypic, pedigree, and genomic information were used for genetic evaluations. Four ways of scaling the mean of the genomic matrix ( $\mathbf{G}$ ) to match the mean of the pedigree relationship matrix among genotyped animals ( $\mathbf{A}_{22}$ ) were tested. Accuracy, bias and inflation were investigated for five groups of animals: ALL = all animals; BULL = only bulls; GEN = genotyped animals; FEM = females; YOUNG = young males. The use of genomic information in the model (ssGBLUP) provided more accurate predictions (ranging from 0.31 to 0.97) than traditional BLUP (ranging from 0.05 to 0.97), especially in the YOUNG group. The results obtained in this study showed that it is advantageous to include genomic information in the genetic evaluation, especially for animals with less information, such as young animals. It is

important to scale the **G** matrix to be compatible with the numerator relationship matrix, even in situations where the **A<sub>22</sub>** is less informative due to the presence of missing pedigree.

**Keyword:** accuracy, bias, beef cattle, genomic evaluation, missing pedigree, multiple sires

## 1. Introduction

In extensive beef cow-calf production systems, multiple service sire (MS) is the most common mating system, which consists in keeping several bulls in the same paddock of cows during the breeding season. Despite the management advantages of this mating system, it restricts the possibility of progeny paternity identification and increases the occurrence of missing pedigree, reducing the reliability of beef cattle genetic evaluations.

Henderson (1974) and Kennedy (1981) noted that ignoring genetic groups or defining poor genetic groups could introduce bias to evaluations. Westell et al. (1988) developed rules to easily setting up genetic groups in the transformed mixed model equations. Several methods have been suggested for genetic evaluations of animals with uncertain paternity. Unknown parent groups (UPG) can be included in the mixed model equations to account for genetic differences among defined animal groups in genetic evaluations (TSURUTA et al., 2014).

The hierarchical animal model (HIER), proposed by Cardoso and Tempelman (2003), is used for the genetic evaluation of animals with uncertain paternity. This method combines records and a priori information to determine a posteriori probabilities of candidate sire, inferring the genetic merit of individuals with uncertain paternity and their respective sire. Recently, DNA markers can be used to assign calves to their individual sires based on inheritance rules. However, in most of the extensive beef cow-calf system, it is not common to identify or record the sires utilized in the breeding season, especially when they are used in only one breeding season (cleanup bulls).

Legarra et al. (2009) and Misztal et al. (2009) proposed a single-step procedure (ssGBLUP) that consists of integrating the pedigree (**A** matrix) and genomic information (**G** matrix) into a single matrix (**H**) to predict the genomic estimated breeding value (GEBV). Several studies reported that ssGBLUP is computationally efficient and accurate for genomic evaluation purposes (AGUILAR et al., 2010; TSURUTA et al., 2011; CHEN et al., 2011a; CHRISTENSEN et al., 2012), however, there are no reports regarding the consequences of using ssGBLUP in situations with uncertain paternity under MS system. In situation of genotyped

animals with uncertain paternity, problems of connect ability and scale between the pedigree relationship matrix among genotyped animals ( $\mathbf{A}_{22}$ ) and  $\mathbf{G}$  matrix are expected, because the  $\mathbf{A}_{22}$  matrix is less informative due to the presence of missing pedigree. Forni et al., (2011) reported overestimated variances and biased GEBV when  $\mathbf{A}_{22}$  was sparser than  $\mathbf{G}$ . For accurate and unbiased genomic predictions, it is necessary to assure that  $\mathbf{G}$  is compatible with  $\mathbf{A}_{22}$  when using MS information.

Currently, models that consider the uncertain paternity using genomic data are not used in animal breeding programs. Considering that the MS is the most common mating system in extensive beef cattle production systems, it is important to investigate the use of genomic data in uncertain paternity models aiming to increase the accuracy and decrease bias in genetic evaluations. Therefore, the objective of this study was to investigate the application of BLUP and ssGBLUP in different scenarios of uncertain paternity, using beef cattle simulated data.

## **2. Material and Methods**

Phenotypes, pedigree, and genotypes were simulated using the software QMSim version 1.00 (SARGOLZAEI; SCHENKEL, 2009). Regarding to phenotypes, two traits assuming different heritabilities were age at first calving (AFC;  $h^2 = 0.12$ ; only in females) and weight at 550 days (W550;  $h^2 = 0.34$ ). Heritabilities were based on real data estimates (DIAS et al., 2004; BOLIGON et al., 2008; BOLIGON et al., 2010; KOURY FILHO et al., 2010; LAUREANO et al., 2011 and YOKOO et al., 2012), and for simplicity, the phenotypic variance was assumed to be 1.0. Ten replicates were performed for each trait and results were averaged among replicates.

### **2.1. Simulated Population**

A historical population was created from generation zero to 2,020, with a constant size of 2,000 animals from generation zero to 1,000. A gradual reduction in the number of animals (from 2,000 to 600) produced a “bottleneck effect” and consequently, genetic drift and linkage disequilibrium (LD) from generation 1,001 to 2,020.

Two hundred out of the 600 animals from the latest generation of the historical population were randomly selected (males and females equally distributed) for the

expanded population, which had its effective size simulated based on the real population (BRITO et al., 2011). To simulate the expanded population, a mating system based on random union of gametes, absence of selection, an exponential growth of the number of females, and an average of five progenies per dam were considered.

After the expansion process, 240 males and 6,000 females from the last generation were randomly selected, including the founder animals of selection population. This population was spanned over 10 generations and the selected males and females from each generation were randomly mated, generating a single progeny with equal probability of being a male or a female. The replacement rate of dams and sires was kept constant over generations at 20% and 60%, respectively. The genotypes of 10,000 animals of the last three generations (8, 9, and, 10) were randomly selected. The estimated LD between adjacent markers in the 8, 9 and 10 generations were 0.17, 0.18 and 0.18, respectively. These results were similar to those reported by Espigolan et al. [21] using a Nellore cattle population genotyped with the BovineHD BeadChip (Illumina®).

## **2.2. Simulated Genome**

The simulated genome presented a total length of 2,333 cM, 735,293 markers and 7,000 QTLs randomly distributed over 29 *Bos Taurus* autosomes (BTA). The length of the bovine genome was based on Base\_4.6.1 (SNELLING et al., 2007) and was assumed that QTLs explain 100% of genetic variance. The number of markers and QTLs per chromosome ranged from 12,931 to 46,495 and from 121 to 438, respectively. All markers were bi-allelic, mimicking SNPs present in the bovine commercial panels. For QTL, the amount of alleles per loci randomly ranged from 2 to 4. Minor allele frequencies (MAF) were assumed equally for markers and QTL alleles. QTL allele effects were sampled from a gamma distribution with a shape parameter equal to 0.4 (HAYES; GODDARD, 2001).

The rate of missing marker genotypes was 0.01 and the rate of marker genotyping error was 0.005. Mutation rate of  $10^{-5}$  for markers and QTLs in the historical populations was considered. A total of 335,000 markers (with MAF greater or equal to 0.02) and 1,000 QTL were randomly selected from the last generation of

the historical population to generate genotypic data for the selection population. The phenotypes of the animals were computed as the sum of the QTL effects and an error term sampled from a normal distribution with zero mean and variance equal to 0.88 for AFC and 0.66 for W550. The parameters of the simulation are showed in the Table 1.

Table 1. Parameters used for the simulation

<b>Historical generations (HG)</b>	
Number of generations (size) – phase 1	1000
Number of generations (size) – phase 2	1020
<b>Expanded generations (EG)</b>	
Number of founder males from HG	100
Number of founder females from HG	100
Number of generations	10
Number of offspring per dam	5
<b>Recent generations</b>	
Number of founder males from EG	240
Number of founder females from EG	6000
Number of generations	10
Number of offspring per dam	1
Ratio of male	50%
Mating system	random
Replacement ratio for males	60%
Replacement ratio for females	20%
Selection/culling	EBV
BV estimation method	BLUP animal method
Ratio of missing sire and dam	0
Heritability of the trait	0.12 or 0.34
Phenotypic variance	1.0
<b>Genome</b>	
Number of chromosomes	29
Total length	2333cM
Number of markers	735,293
Marker distribution	Evenly spaced
Number of QTL	7000
QTL distribution	Random
MAF for markers	0.10

MAF for QTL	0.10
Additive allelic effects for markers	neutral
Additive allelic effects for QTL	Gamma distribution (shape = 0.4)
Rate of missing marker genotypes	0.01
Rate of marker genotyping error	0.005
Rate of recurrent mutation	0.00001

### 2.3. BLUP and ssGBLUP methods

In the BLUP method, a traditional genetic evaluation was applied using pedigree and phenotypic information. The model can be represented as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of phenotypes,  $\mathbf{b}$  is the vector of fixed effects,  $\mathbf{u}$  is the vector of additive genetic effect,  $\mathbf{X}$  e  $\mathbf{Z}$  are incidence matrices and  $\mathbf{e}$  is the vector of random residuals effects. Considering an infinitesimal model,  $\text{var}(\mathbf{u}) = \mathbf{A}\sigma_u^2$ , where  $\mathbf{A}$  is the numerator relationship matrix obtained from pedigree information and  $\sigma_u^2$  is the variance of genetic effect.

In the single-step genomic BLUP (ssGBLUP) proposed by Misztal et al., (2009), the inverse of the numerator relationship matrix ( $\mathbf{A}^{-1}$ ) was replaced by  $\mathbf{H}^{-1}$  that combines pedigree and genomic information. Aguilar et al. (2010) showed that  $\mathbf{H}^{-1}$  is constructed as:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where  $\mathbf{H}^{-1}$  is the inverse of the realized relationship matrix that incorporates the inverse of the genomic relationship matrix ( $\mathbf{G}^{-1}$ ) and the inverse of the numerator relationship matrix for genotyped animals  $\mathbf{A}_{22}^{-1}$ . The  $\mathbf{G}$  matrix was created according to VanRaden (2008):

$$\mathbf{G} = \frac{(\mathbf{M} - \mathbf{P})(\mathbf{M} - \mathbf{P})'}{2 \sum_{j=1}^m p_j (1 - p_j)}$$

where  $\mathbf{M}$  is a matrix of marker alleles with  $m$  columns ( $m$  = total number of markers) and  $n$  rows ( $n$  = total number of genotyped individuals), and  $\mathbf{P}$  is a matrix containing the frequency of the second allele ( $p_j$ ), expressed as  $2p_j$ .  $\mathbf{M}_{ij}$  was 0 if the genotype of individual  $i$  for SNP  $j$  was homozygous for the first allele, was 1 if heterozygous, or 2 if the genotype was homozygous for the second allele.

## 2.4. Scenarios

The **A** matrix was created assuming different proportions of MS matings (0, 25, 50, 75, and 100%). Four ways of scaling **G** to match **A**<sub>22</sub> were tested: S<sub>1</sub> – scaling for all animals considering mean diagonal **A**<sub>22</sub> = mean diagonal **G** and mean off diagonal **A**<sub>22</sub> = mean off diagonal **G** (*default in blupf90*; Chen et al., 2011b); S<sub>2</sub> - no scaling between **A**<sub>22</sub> and **G** matrix; S<sub>3</sub> - scaling between **A**<sub>22</sub> and **G** matrix only the animals which have known sire and dam; and S<sub>4</sub> - scaling between **A**<sub>22</sub> and **G** matrix only those animals which have one known parent. A total of 5 scenarios were tested for each trait (10 scenarios in total) considering the BLUP and ssGBLUP, five proportions of known paternity animals and four situations of scaling on relationship matrix animals.

Accuracy of prediction and bias were calculated for five groups of animals in each scenario: ALL = all animals in the population; BULL = only bulls with at least one progeny; GEN = genotyped animals; FEM = females; and YOUNG = young males without phenotype. The accuracy of prediction was the correlation between true breeding value (TBV) and EBV or genomic EBV (GEBV). In addition, bias was calculated as the regression coefficient of TBV on EBV/GEBV. Results were the mean of the 10 replicates of each scenario. For AFC, a total of 10,000; 91; 10,000; 5008 and 4901 genotyped animals for ALL, BULL, GEN, FEM and YOUNG respectively, were used. For W550, a total of 10,000; 98; 10,000; 5012 and 4890 genotyped animals for ALL, BULL, GEN, FEM and YOUNG respectively, were used. The variance component estimation and solutions were obtained by BLUPF90 family programs (MISZTAL et al. 2002).

## 3. Results and Discussion

The variance component estimates obtained for the two traits using traditional pedigree (REML) and genomic information (GREML) with different proportions of multiple sires (MS) are presented in Table 2. It is important to highlight that the software utilized for simulation does not estimate variance components, but uses the heritability values provided by the user (0.12 for AFC and 0.34 for W550).

In the BLUP method, additive genetic variance ranged from 0.13 to 0.20 and from 0.35 to 0.63 for AFC and W550, respectively. The highest additive genetic



variance was observed for the scenario with 100% of MS for both traits. Thus, the increase in the percentage of MS (0, 25, 50, 75 and 100%) in the population overestimated the additive genetic variance, being more evident in the trait with moderate heritability. Considering 50% of MS, the additive genetic variance increased 18.75 and 27% for AFC and W550 when compared to scenario 0%, respectively. Nevertheless, with 100% of MS this increase was 35 and 44.4%, respectively.

These results can be explained by the reduction of the number of inbred animals according to MS proportion included in the pedigree (Table 3), therefore, there was an increase in the variance between families, which consequently increased the additive genetic variance. Nietlisbach et al. (2016) stated that the inbreeding effect on the heritability estimate depends on the considered population, but increasing the level of inbreeding in the population reduces the heritability estimates. Applying the ssGBLUP ( $S_1$ ) for the same scenarios (0 to 50% and 0 to 100% of MS) the additive genetic variance increased (37.5 and 41.2%) and (36.7 and 43.6%) for AFC and W550, respectively.

The results in Table 2 showed that the inclusion of genomic relationship matrix caused a decrease in estimates of additive genetic variance for both traits, compared to the variance obtained using traditional REML for practically in all evaluated scenarios.

Table 2. Estimation of variance components, heritability and standard errors for the studied traits using traditional and genomic REML with different proportions of multiple sires

Trait	MS	REML			GREML		
		$\sigma_a^2$	$\sigma_e^2$	$h^2 \pm SE$	$\sigma_a^2$	$\sigma_e^2$	$h^2 \pm SE$
AFC	0%	0.13±0.004	0.88±0.002	0.13±0.003	0.10±0.002	0.89±0.003	0.10±0.002
	25%	0.14±0.004	0.88±0.002	0.14±0.004	0.13±0.004	0.87±0.003	0.13±0.003
	50%	0.16±0.006	0.88±0.002	0.15±0.005	0.16±0.005	0.86±0.003	0.15±0.004
	75%	0.18±0.007	0.87±0.003	0.17±0.006	0.18±0.006	0.86±0.003	0.17±0.005
	100%	0.20±0.009	0.86±0.003	0.19±0.007	0.17±0.005	0.88±0.003	0.16±0.004
W550	0%	0.35±0.008	0.66±0.002	0.35±0.006	0.31±0.008	0.67±0.001	0.32±0.006
	25%	0.41±0.012	0.66±0.002	0.38±0.007	0.40±0.001	0.64±0.002	0.39±0.007
	50%	0.48±0.014	0.67±0.002	0.42±0.008	0.49±0.014	0.63±0.002	0.44±0.007
	75%	0.55±0.017	0.64±0.002	0.46±0.008	0.55±0.015	0.63±0.002	0.46±0.007
	100%	0.63±0.020	0.62±0.003	0.50±0.009	0.55±0.017	0.68±0.002	0.45±0.008

AFC=age at first calving; W550=weight at 550 days; MS=percentage of multiple sires; REML=restricted maximum likelihood estimation; GREML= genomic restricted maximum likelihood estimation  $\sigma_a^2$ = additive genetic variance;  $\sigma_e^2$ = residual variance;  $h^2$ = heritability; SE= standard errors

Accuracies of genetic evaluation and bias for all studied groups using BLUP and ssGBLUP ( $S_1$ ) are shown in Table 4 and 5. For BLUP model, accuracies of genetic evaluations for both traits decreased when the proportion of unknown sires in the population was increased, and EBV accuracy reduction was higher for GEN and YOUNG. Comparing the scenarios for YOUNG (from 0 to 100% of MS) the loss of accuracy was 87.8 and 86% for AFC and W550, respectively. These results indicated that select young animals based on their EBV in a scenario with missing pedigree information could be unreliable.

Table 3. Pedigree structures with different proportions of multiple sires

	Percentage of multiple sires				
	0%	25%	50%	75%	100%
All animals	66,240	66,240	66,240	66,240	66,240
Inbred animals of AFC	23,326	20,927	18,534	16,153	13,785
Inbred animals of W550	19,134	16,830	14,542	12,292	10,062
Bulls	1,536	1,536	1,536	1,534	1,297
Dams	16,800	16,800	16,800	16,800	16,800
Progenies with only known sire	0	0	0	0	0
Progenies with only known dam	0	2,500	5,000	7,500	10,000
Progenies with known sire and dam	60,000	57,500	55,000	52,500	50,000

Applying the ssGBLUP ( $S_1$ ) method, the accuracies of genetic evaluation also decreased when increasing the MS in the pedigree for both traits. However, the accuracy reduction was smaller than those for BLUP model. Using the same comparison (scenario 0 to 100% of MS), the accuracy decreased 38 and 44.6% for AFC and W550, respectively. These results showed that ssGBLUP ( $S_1$ ) method was more accurate for evaluation of young animals in situations with missing pedigree information.

For both traits, the accuracies for ALL, BULL, and FEM were similar for BLUP and ssGBLUP ( $S_1$ ) methods, and lower accuracies were obtained with an increase in the proportion of MS. It is expected that genomic information contributes less for a group of animals that have enough phenotypic or progeny information, because this contribute to accuracy (LOURENÇO et al., 2015). Forni et al. (2011) working with real pig data showed that the inclusion of genomic relationship matrix had little impact on the accuracy of sires and higher impact on females with few phenotypic information.

Considering the scenario with 0% of MS, the accuracies of genetic evaluation practically remained constant with the inclusion of genomic relationship matrix in the population for ALL, BULL, and FEM for both traits studied. However, for YOUNG when the ssGBLUP ( $S_1$ ) was applied in a scenario with non-missing pedigree, the accuracies increased 22 and 30% for AFC and W550, respectively. Wiggans et al. (2011) reported gains in reliability above parent average ranging in young bulls from 2.7 to 47.6 percentage units for Holsteins, 9.6 to 29.2 percentage units for Jerseys, and 3.0 to 25.8 percentage units for Brown Swiss.

In beef cattle, Garrick (2011) stated that genomic prediction offers accuracies that exceed those of pedigree-based parent average of young selection candidates, and it can be equivalent to progeny tests based on up to 10 offspring. It is important to highlight that for YOUNG, as the proportion of MS increased in the pedigree, the ssGBLUP ( $S_1$ ) compensated the accuracy reduction obtained with the BLUP method. In the scenario that considers all unknown sires (100% of MS), the accuracy increased from 0.05 to 0.31 for AFC and from 0.07 to 0.36 for W550, being 6 and 5 times higher for AFC and W550, respectively.

According to Olson et al. (2011) the regression coefficient of TBV on EBV/GEBV is an alternative to evaluate the genetic evaluation bias, which indicates overestimation of the variance of genetic evaluation when is less than 1 (inflation), and underestimation when it is larger than 1 (deflation). In general, the regression coefficients for ALL, BULL, and FEM were close to 1 (Table 5), indicating that EBV predictions were less biased, although for GEN and YOUNG the regression coefficients were inflated as the number of unknown sires in the population increased. Large differences in regression coefficients between BLUP and ssGBLUP ( $S_1$ ) for YOUNG were observed in all the scenarios. Biases in GEBV have been reported and discussed in several studies (AGUILAR et al., 2010; WIGGANS et al., 2011; TSURUTA et al., 2013) and could be due to the difference in scale between pedigree and genomic relationships, especially for young genotyped animals.

Table 4. Accuracies of evaluation and standard errors using BLUP and ssGBLUP with different proportions of multiple sires

		Percentage of multiple sires						
		0%	25%	50%	75%	100%		
AFC	ALL	BLUP	0.79±0.006	0.76±0.005	0.73±0.005	0.70±0.005	0.68±0.004	
		ssGBLUP	0.80±0.006	0.78±0.005	0.76±0.005	0.73±0.005	0.69±0.004	
	BULL	BLUP	0.87±0.005	0.86±0.005	0.84±0.006	0.83±0.006	0.85±0.005	
		ssGBLUP	0.87±0.005	0.85±0.005	0.81±0.004	0.79±0.004	0.85±0.005	
	GEN	BLUP	0.46±0.007	0.20±0.005	0.16±0.004	0.16±0.005	0.30±0.006	
		ssGBLUP	0.55±0.009	0.57±0.009	0.56±0.009	0.56±0.009	0.56±0.010	
	FEM	BLUP	0.81±0.005	0.78±0.005	0.76±0.005	0.74±0.005	0.72±0.004	
		ssGBLUP	0.81±0.005	0.80±0.005	0.78±0.005	0.76±0.005	0.73±0.004	
	YOUNG	BLUP	0.41±0.007	0.14±0.005	0.08±0.005	0.06±0.007	0.05±0.008	
		ssGBLUP	0.50±0.009	0.50±0.008	0.46±0.008	0.42±0.007	0.31±0.007	
	W550	ALL	BLUP	0.92±0.001	0.89±0.001	0.86±0.002	0.84±0.002	0.82±0.002
			ssGBLUP	0.92±0.001	0.92±0.001	0.90±0.001	0.88±0.001	0.84±0.001
BULL		BLUP	0.97±0.001	0.96±0.001	0.95±0.001	0.93±0.001	0.97±0.001	
		ssGBLUP	0.97±0.001	0.97±0.001	0.95±0.001	0.94±0.001	0.97±0.001	
GEN		BLUP	0.60±0.005	0.29±0.003	0.25±0.003	0.28±0.003	0.41±0.003	
		ssGBLUP	0.72±0.006	0.70±0.005	0.68±0.004	0.67±0.004	0.66±0.004	
FEM		BLUP	0.93±0.001	0.91±0.001	0.89±0.001	0.88±0.001	0.87±0.002	
		ssGBLUP	0.93±0.001	0.92±0.001	0.91±0.001	0.89±0.001	0.87±0.002	
YOUNG		BLUP	0.50±0.004	0.16±0.003	0.09±0.005	0.06±0.005	0.07±0.006	
		ssGBLUP	0.65±0.006	0.60±0.005	0.53±0.003	0.46±0.003	0.36±0.003	

AFC=age at first calving; W550=weight at 550 days; ALL=all animals; BULL=bulls; GEN=genotyped animals; FEM= females; YOUNG=young males; BLUP=best linear unbiased prediction; ssGBLUP= single step genomic BLUP

As described by Vitezica et al. (2011), the comparison of average TBV and EBV/GEBV was also used to assess the bias of genetic evaluation with different proportions of multiple sires for ALL and YOUNG (Table 6). As expected, for the scenario with 0% of MS in the pedigree, the average of EBV/GEBV were very close to the average of TBV in ALL and YOUNG for both traits. As the percentage of missing pedigree increased, the BLUP and ssGBLUP ( $S_1$ ) methods overestimated the TBV mean for both traits, mainly for young animals. However, the ssGBLUP ( $S_1$ ) method predicted less biased TBV average than the BLUP method in situations with missing pedigree.

Table 5. Bias and standard errors using BLUP and ssGBLUP with different proportions of multiple sires

		Percentage of multiple sires					
		0%	25%	50%	75%	100%	
AFC	ALL	BLUP	1.00±0.005	1.01±0.006	1.01±0.007	1.02±0.007	1.02±0.007
		ssGBLUP	0.95±0.005	0.98±0.005	0.99±0.006	1.01±0.006	0.91±0.006
	BULL	BLUP	1.01±0.005	0.99±0.006	0.97±0.007	0.95±0.008	0.93±0.006
		ssGBLUP	1.01±0.006	1.04±0.006	1.04±0.007	1.03±0.007	0.93±0.007
	GEN	BLUP	1.03±0.014	0.24±0.006	0.18±0.003	0.21±0.005	0.78±0.013
		ssGBLUP	0.94±0.020	0.85±0.015	0.80±0.013	0.81±0.012	0.91±0.017
	FEM	BLUP	1.01±0.005	1.02±0.006	1.04±0.007	1.05±0.007	1.07±0.007
		ssGBLUP	0.95±0.005	0.96±0.006	0.97±0.006	0.10±0.007	1.01±0.007
	YOUNG	BLUP	1.03±0.018	0.18±0.006	0.09±0.006	0.07±0.009	0.08±0.012
		ssGBLUP	0.91±0.020	0.81±0.015	0.68±0.013	0.58±0.009	0.38±0.009
W550	ALL	BLUP	1.00±0.002	1.01±0.003	1.02±0.002	1.04±0.002	1.07±0.003
		ssGBLUP	0.99±0.002	1.03±0.003	1.06±0.003	1.07±0.003	1.07±0.003
	BULL	BLUP	1.01±0.002	1.00±0.002	0.98±0.002	0.97±0.002	0.94±0.003
		ssGBLUP	1.03±0.002	1.06±0.002	1.08±0.002	1.08±0.003	0.95±0.003
	GEN	BLUP	0.97±0.009	0.26±0.002	0.21±0.003	0.26±0.003	0.60±0.005
		ssGBLUP	1.06±0.009	0.89±0.006	0.75±0.005	0.73±0.006	0.78±0.006
	FEM	BLUP	1.00±0.003	1.02±0.003	1.05±0.003	1.07±0.003	1.11±0.003
		ssGBLUP	0.99±0.003	1.01±0.002	1.04±0.003	1.06±0.004	1.08±0.004
	YOUNG	BLUP	0.97±0.006	0.14±0.003	0.07±0.004	0.05±0.004	0.07±0.007
		ssGBLUP	1.05±0.009	0.80±0.006	0.60±0.005	0.48±0.003	0.34±0.004

AFC=age at first calving; W550=weight at 550 days; ALL=all animals; BULL=bulls; GEN=genotyped animals; FEM= females; YOUNG=young males; BLUP=best linear unbiased prediction; ssGBLUP= single step genomic BLUP

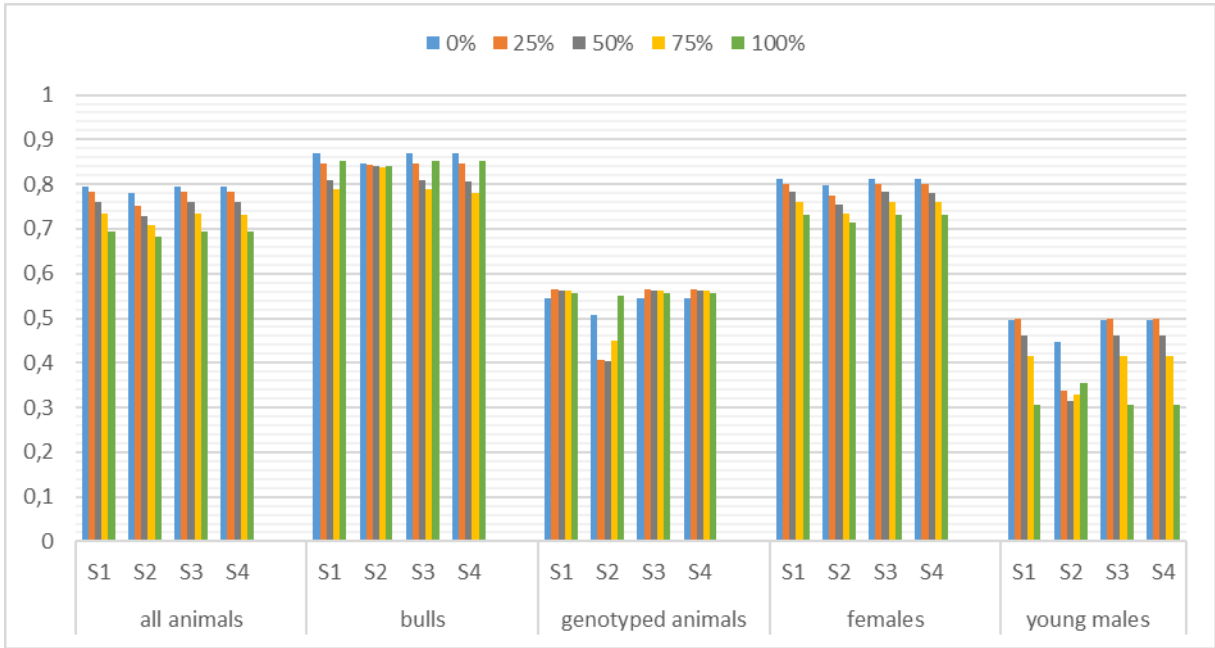
Accuracy of prediction for all groups studied using different scaling for **G** to match **A**<sub>22</sub> are presented in Figures 1 and 2 for AFC and W550, respectively. In general, for both traits the GEBV accuracy was invariant to the scaling method used. The GEN and YOUNG showed less accurate GEBV in situations with missing pedigree and with no scaling for **G** (**S**<sub>2</sub>). However, there is not differences between the scaling methods existed when all the genotyped animals had unknown sires. There was not also difference between the strategies for scaling the **G** matrix for ALL, BULL and FEM groups under the different scenarios with missing pedigree, showing that the uninformative part of the **A**<sub>22</sub> matrix did not influence the scaling of **G**. The results of this study showed that it is important to have a **G** matrix in the same scale of the **A**<sub>22</sub> matrix for GEBV especially in the situation with missing pedigree information for young animals evaluations.

Table 6. Average and standard deviations (SDs) for true breeding values (TBV) and breeding values from different traits using BLUP and ssGBLUP with different proportions of multiple sires

	Trait	% of MS in the pedigree	TBV	BLUP	ssGBLUP
ALL	AFC	0		-0.51 ± 0.40 <sup>a</sup>	-0.60 ± 0.42 <sup>b</sup>
		25		-0.47 ± 0.38 <sup>a</sup>	-0.58 ± 0.40 <sup>b</sup>
		50	-0.49 ± 0.50 <sup>a</sup>	-0.43 ± 0.36 <sup>b</sup>	-0.55 ± 0.38 <sup>b</sup>
		75		-0.39 ± 0.35 <sup>b</sup>	-0.50 ± 0.37 <sup>a</sup>
		100		-0.36 ± 0.33 <sup>b</sup>	-0.43 ± 0.35 <sup>b</sup>
	W550	0		-1.27 ± 0.96 <sup>a</sup>	-1.34 ± 0.97 <sup>a</sup>
		25		-1.12 ± 0.92 <sup>b</sup>	-1.27 ± 0.93 <sup>a</sup>
		50	-1.30 ± 1.04 <sup>a</sup>	-0.99 ± 0.88 <sup>b</sup>	-1.14 ± 0.89 <sup>b</sup>
		75		-0.86 ± 0.84 <sup>b</sup>	-0.99 ± 0.85 <sup>b</sup>
		100		-0.75 ± 0.80 <sup>b</sup>	-0.82 ± 0.82 <sup>b</sup>
YOUNG	AFC	0		-1.03 ± 0.15 <sup>a</sup>	-1.12 ± 0.18 <sup>b</sup>
		25		-0.87 ± 0.27 <sup>b</sup>	-0.99 ± 0.21 <sup>a</sup>
		50	-1.01 ± 0.34 <sup>a</sup>	-0.73 ± 0.30 <sup>b</sup>	-0.85 ± 0.23 <sup>b</sup>
		75		-0.46 ± 0.28 <sup>b</sup>	-0.58 ± 0.24 <sup>b</sup>
		100		-0.48 ± 0.23 <sup>b</sup>	-0.56 ± 0.28 <sup>b</sup>
	W550	0		-2.50 ± 0.28 <sup>a</sup>	-2.58 ± 0.33 <sup>a</sup>
		25		-2.08 ± 0.62 <sup>b</sup>	-2.30 ± 0.41 <sup>b</sup>
		50	-2.54 ± 0.54 <sup>a</sup>	-1.70 ± 0.69 <sup>b</sup>	-1.95 ± 0.47 <sup>b</sup>
		75		-1.35 ± 0.64 <sup>b</sup>	-1.61 ± 0.52 <sup>b</sup>
		100		-1.06 ± 0.51 <sup>b</sup>	-1.24 ± 0.58 <sup>b</sup>

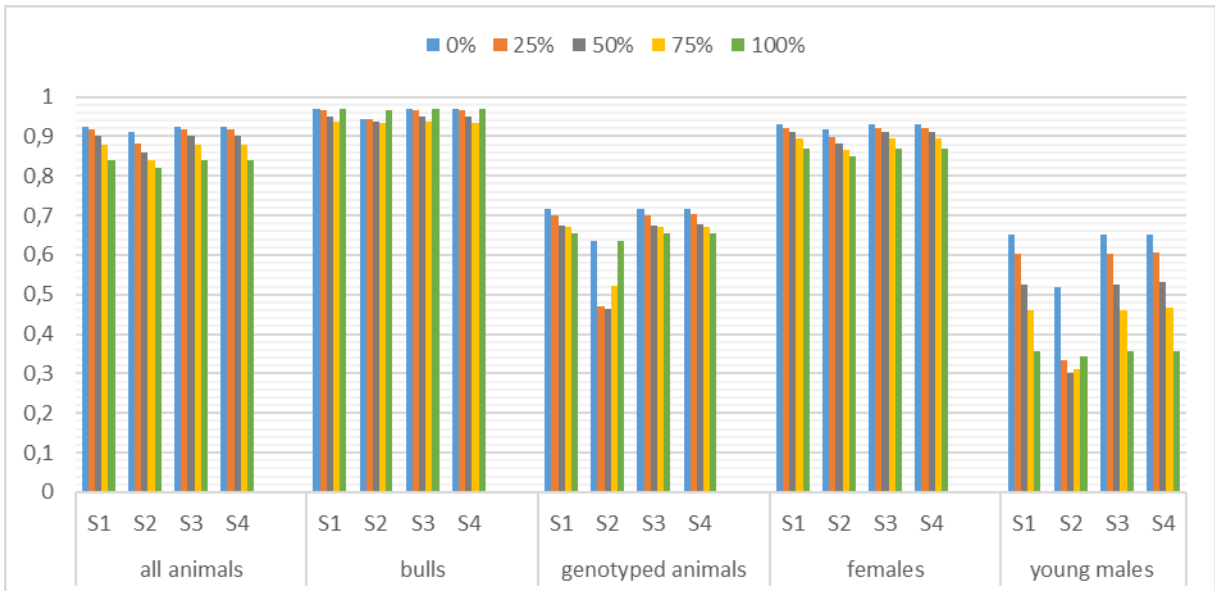
AFC=age at first calving; W550=weight at 550 days; MS=multiple sires; BLUP=best linear unbiased prediction; ssGBLUP= single step genomic BLUP; ALL=all animals, YOUNG=young males; <sup>a,b</sup> Means with different superscripts within the same row differ ( $P < 0.05$ ) according with the Dunnett's test (Control="TBV")

The theory for constructing **H** matrix makes many assumptions that may not hold in practice (MISZTAL et al., 2013a). Those assumptions include the same genetic parameters in the genotyped sample as in the complete population, and the existence of complete data for all traits for which selection occurred to account for selection bias (MISZTAL et al., 2013a). Chen et al. (2011b) reported that the scale of **G** influences ranking of genotyped versus non-genotyped animals. The optimal **G** should have the same average of diagonals and off-diagonals as **A**<sub>22</sub> (MISZTAL et al., 2013a). Vitezica et al. (2011) derived a formal proof and showed that a well-constructed **G**, in ssGBLUP, gives more accurate and less biased GEBV than a multistep approach.



AFC – age at first calving;  $S_1$  – scaling for all animals considering mean diagonal  $A_{22}$  = mean diagonal  $G$  and mean off diagonal  $A_{22}$  = mean off diagonal  $G$ ;  $S_2$  - no scaling between  $A_{22}$  and  $G$  matrix;  $S_3$  - scaling between  $A_{22}$  and  $G$  matrix only for animals which have known sire and dam;  $S_4$  - scaling between  $A_{22}$  and  $G$  matrix only for animals which have one known parent.

Figure 1: Accuracies of evaluation for AFC for the ALL, BULL, GEN, FEM, and YOUNG using different scaling for the genomic matrix ( $G$ ) to match the numerator relationship matrix for genotyped animals ( $A_{22}$ ).



W550 - weight at 550 days;  $S_1$  – scaling for all animals considering mean diagonal  $A_{22}$  = mean diagonal  $G$  and mean off diagonal  $A_{22}$  = mean off diagonal  $G$ ;  $S_2$  - no scaling between  $A_{22}$  and  $G$  matrix;  $S_3$  - scaling between  $A_{22}$  and  $G$  matrix only for animals which have known sire and dam;  $S_4$  - scaling between  $A_{22}$  and  $G$  matrix only for animals which have one known parent.

Figure 2: Accuracies of evaluation for W550 for the ALL, BULL, GEN, FEM, and YOUNG using different scaling for the genomic matrix ( $G$ ) numerator relationship matrix for genotyped animals ( $A_{22}$ ).

The ssGBLUP has been used for several large-scale analyses including dairy cattle (TSURUTA et al., 2011; AGUILAR et al., 2011), pigs (FORNI et al., 2011; CHRISTENSEN et al., 2012), and chickens (CHEN et al., 2011a). These studies showed that the ssGBLUP generally is reliable or more than the multistep procedure and the GEBVs were less biased. There are large differences among beef cattle production compared to dairy cattle, swine, and chicken. The beef cattle production is often cited to be inferior to poultry and swine production (WILKISON et al., 2011), and most of beef cattle production is in harsh environments and with low input and investment levels. In dairy cattle, a large proportion of calves in most populations are offspring of few artificial insemination sires. The lack of artificial insemination in beef contributes to poor genetic connectedness and sire identification, compromising the reliability of genetic evaluations compared to dairy cattle.

Several studies have been developed to apply the ssGBLUP in situations with missing pedigree using UPG (MISZTAL et al., 2013b; TSURUTA et al. 2014). Misztal et al. (2013b) explained that potential bias could occur in genomic EBV (GEBV) using ssGBLUP with UPG. They also reported convergence problems with iterative methods and incompatibility between  $\mathbf{G}$  and  $\mathbf{A}_{22}$  due to short or incomplete pedigrees, pedigree mistakes, incorrect assignment of genotypes, poor quality of genotypes, and the unaccounted presence of multiple/lines breeds. In our study there was not convergence problems, but it is important to emphasize that low correlations between the off-diagonal elements of  $\mathbf{G}$  and  $\mathbf{A}_{22}$  were observed due to incomplete pedigrees, mainly when the percentage of MS was higher than 50%. Tsuruta et al. (2014) assigned UPG in mixed-model equations using ssGBLUP, which reduced bias and increased accuracy of GEBV. In the present study no convergence problems were identified, even in situation with large proportion of missing pedigree. However, it is expected a higher occurrence of convergence problems with real data, in which the population structure and phenotypic records are unbalanced and model complexity is higher. Lourenço et al. (2014) showed that removing old phenotypes and pedigree helped to improve convergence without decreasing accuracy for selection candidates.

According to Berry et al. (2016), the development of accurate genomic evaluations in beef populations is more difficult than in dairy populations. The



reasons include the presence of multiple breeds, poor extent of phenotyping, lack of artificial insemination, and because beef systems are generally a lower-profit business that fails in adopting new technologies. However, the results of this study showed that when large proportion of the pedigree is missing and BLUP is not reliable, it is possible to increase prediction accuracy for selection candidates using ssGBLUP. In beef cattle systems, young animals with unreliable genetic evaluation are frequently discarded. Through the application of ssGBLUP, it is possible to obtain more reliable genetic evaluations for young animals with missing pedigree, increasing their market value, since they will be sold as young bulls in the market. Moreover, with the implementation of ssGBLUP, the breeder has more animals to selection from, allowing an increased selection intensity. The results of this study could support breeders to reduce the risk of selecting young animals with missing pedigree information, allowing higher genetic gains in the beef cattle systems.

#### 4. Conclusions

Despite the ssGBLUP procedure was not developed to deal with uncertainty paternity situations, the results of this study shown that the application of the ssGBLUP was appropriate for the simulated scenarios. In situations of missing pedigree, the ssGBLUP method is an appropriate alternative to obtain a more reliable and less biased breeding values, especially for young animals with few or no phenotypic information.

It is important to scale the genomic relationship matrix to be compatible with the numerator relationship matrix for genotyped animals, even in situations where the latter is less informative due to the presence of missing pedigree. For accurate and unbiased genomic predictions with ssGBLUP, it is necessary to assure that the G matrix is compatible with  $A_{22}$  matrix even in situations with uncertainty paternity.

#### 5. References

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, p. 743-752, 2010.

AGUILAR, I.; MISZTAL, I.; LEGARRA, A.; TSURUTA, S. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. **Journal of Animal Breeding and Genetics**, v. 128, p. 422-428, 2011.

BERRY, D. P.; GARCIA, J. F.; GARRICK, D. J. Development and implementation of genomic predictions in beef cattle. **Animal Frontiers**, v. 6, p. 32-38, 2016.

BOLIGON, A. A.; ALBUQUERQUE, L. G.; RORATO, P. R. N.; Associações genéticas entre pesos e características reprodutivas em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia**, v. 37, n. 41, p. 596-601, 2008.

BOLIGON, A. A.; ALBUQUERQUE, L. G.; MERCADANTE, M. E. Z.; LOBO, R. B. Study to relations between age at first calving, average weight gains and weights from weaning to maturity in Nelore cattle. **Revista Brasileira de Zootecnia**, v. 39, p.746- 751, 2010.

BRITO, F. V.; NETO, J. B.; SARGOLZAEI, M.; COBUCI, J. A.; SCHENKEL, F. S. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. **BMC Genetics**, v. 12, p. 80-89, 2011.

CARDOSO, F. F.; TEMPELMAN, R. J. Bayesian inference on genetic merit under uncertain paternity. **Genetic Selection Evolution**, v. 35, p. 469-487, 2003.

CHEN, C. Y.; MISZTAL, I.; AGUILAR, I., TSURUTA, S.; MEUWISSEN, T. H. E.; AGGREY, S. E.; WING, T.; MUIR, W. M. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. **Journal of Animal Science**, v. 89, p. 23–28, 2011a.

CHEN, C. Y.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Effect of different genomic relationship matrices on accuracy and scale. **Journal of Animal Science**, v. 89, p. 2673–2679, 2011b.

CHRISTENSEN, O. F.; MADSEN, P.; NIELSEN, B.; OSTERSEN, T; SU, G. Single-step methods for genomic evaluation in pigs. **Animal**, v. 6, p. 1565–1571, 2012.

DIAS, L. T.; EL FARO, L.; ALBUQUERQUE, L. G. Estimativas de herdabilidade para idade ao primeiro parto na raça Nelore. **Revista Brasileira de Zootecnia**, v.33, p.97-102, 2004.

ESPIGOLAN, R.; BALDI, F.; BOLIGON, A. A.; SOUZA, F. R.; GORDO, D. G.; TONUSSI, R. L.; CARDOSO, D. F.; OLIVEIRA, H. N.; TONHATI, H.; SARGOLZAEI, M.; SCHENKEL, F. S.; CARVALHEIRO, R.; FERRO, J. A.; ALBUQUERQUE, L. G. Study of whole genome linkage disequilibrium in Nelore cattle. **BMC Genomics**. V. 14, p. 305-312, 2013.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v. 43, n. 1, p. 1-7, 2011.

GARRICK, J. D. The nature, scope and impact of genomic prediction in beef cattle in the United States – Review. **Genetics Selection Evolution**, v. 47, p. 1-11, 2011.

HAYES, B.; GODDARD, M. E. The distribution of the effects of genes affecting quantitative traits in livestock. **Genetics Selection Evolution**, v. 33, n.3, p.209-229, 2001.

HENDERSON, C. R. General flexibility of linear model techniques for sire evaluation. **Journal of Dairy Science**, v. 57, p. 963-972, 1974.

KENNEDY, B.W. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. **Journal of Dairy Science**, v. 64, p. 689-697, 1981.

KOURY FILHO, W.; ALBUQUERQUE, L. G.; FORNI, S.; SILVA, J. A. V.; YOKOO, M. J.; ALENCAR, M. M. Estimativas de parâmetros genéticos para os escores visuais e suas associações com peso corporal em bovinos de corte. **Revista Brasileira de Zootecnia**, v. 39, p. 1015-1022, 2010.

LAUREANO, M. M. M.; BOLIGON, A. A.; COSTA, R. B.; FORNI, S.; SEVERO, J. L. P.; ALBUQUERQUE, L. G. Estimativas de herdabilidade e tendências genéticas para características de crescimento e reprodutivas em bovinos da raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 63, p. 143-152, 2011.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656-4663, 2009.

LOURENÇO, D. A. L.; MISZTAL, I.; TSURUTA, S.; AGUILAR, I.; LAWLOR, T. J.; FORNI, S.; WELLER, J. I. Are evaluations on Young genotyped animals benefiting from the past generations? **Journal of Dairy Science**, v. 97, p.3930-3942, 2014.

LOURENÇO, D. A. L.; FRAGOMENI, B. O.; TSURUTA, S.; AGUILAR, I.; ZUMBACH, B.; HAWKEN, R. J.; LEGARRA, A.; MISZTAL, I. Accuracy of estimated breeding values with genomic information on males, females, or both: an example in broiler chicken. **Genetic Selection Evolution**, v. 47, p. 56, 2015.

MISZTAL I.; TSURUTA S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90). **Proceedings** of the 7th World Congress on Genetics Applied to Livestock Production; 2002 August 19-23; Montpellier, France. Communication No 28-07, 2002.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, p. 4648-4655, 2009.

MISZTAL, I.; AGGREY, S. E.; MUIR, W. M. Experiences with a single step genome evaluation. **Poultry Science**, v. 92, p. 2530-2534, 2013a.

MISZTAL, I.; VITEZICA, Z. G.; LEGARRA, A.; AGUILAR, I.; SWAN, A. A. Unknown-parent groups in single-step genomic evaluation. **Journal of Animal Breeding and Genetics**, v. 130, p. 252-258, 2013b.

NIETLISBACH, P. KELLER, L.; POSTMA, E. Genetic variance components and heritability of multiallelic heterozygosity under inbreeding. **Heredity**, v. 116, p. 1–11, 2016.

OLSON, K. M.; VANRADEN, P. M.; TOOKER, M. E.; COOPER, T. A. Differences among methods to validate genomics evaluations for dairy cattle. **Journal of Dairy Science**, v. 94, p. 2613-2620, 2011.

SARGOLZAEI, M.; SCHENKEL, F. S.: QMSim: a large-scale genome simulator for livestock. **Bioinformatics**, v. 25, p. 680-681, 2009.

SNELLING, W. M.; CHIU, R.; SCHEIN, J. E.; HOBBS, M.; ABBEY, C. A.; ADELSON, D. L.; AERTS, J.; BENNETT, G. L.; BOSDET, I. E.; BOUSSAHA, M.; BRAUNING, R.; CAETANO, A. R.; COSTA, M. M.; CRAWFORD, A. M.; DALRYMPLE, B. P.; EGGEN A.; VAN DER WIND, A. E.; FLORIOT, S.; GAUTIER, M.; GILL, C. A.; GREEN, R. D.; HOLT, R.; JANN, O.; JONES, S. J. M.; KAPPES, S. M.; KEELE, J. W.; PONG, P. J.; LARKIN, M.; LEWIN, H. A.; MCEWAN, J. C.; MCKAY, S.; MARRA, M. A.; MATHEWSON, C. A.; MATUKUMALLI, L. K.; MOORE, S. S.; MURDOCH, B.; NICHOLAS, F. W.; OSOEGAWA, R.; ROY, A.; SALIH, H.; SCHIBLER, L.; SCHNABEL, R. D.; L.; SILVERI, L.; SKOW, L. C.; SMITH, T. P. L.; SONSTEGARD, T. S.; TAYLOR, J. F.; TELLAM, R.; VAN TASSEL, C. P.; WILLIAMS, J. L.; WOMACK, J. E.; WYE, N. H.; YANG, G.; ZHAO, S. for the International Bovine BAC Mapping Consortium: A physical map of the bovine genome. **Genome Biology**, 8:R165, 2007.

TSURUTA, S.; AGUILAR, I.; MISZTAL, I.; LAWLOR, T. J. Multiple trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. **Journal of Dairy Science**, v. 94, p. 4198–4204, 2011.

TSURUTA, S.; MISZTAL, I.; LAWLOR, T. J. Short communication: Genomic evaluations of final score for US Holsteins benefit from the inclusion of genotypes on cows. **Journal of Dairy Science**, v. 96, p. 3332–3335, 2013.

TSURUTA, S.; MISZTAL, I.; LOURENÇO, D. A. L.; LAWLOR, T. J. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. **Journal of Dairy Science**, v. 97, p. 5814–5821, 2014.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, p. 4414-4423, 2008.

VITEZICA, Z. G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions of populations under selection. **Genetics Research**, v. 93, p. 357-366, 2011.

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

WIGGANS, G. R.; VANRADEN, P. M.; COOPER, T. A. The genomic evaluation system in the United States: Past, present, future. **Journal of Dairy Science**, v. 94, p.3202-3211, 2011.

WILKINSON, S.; WIENER, P.; ARCHIBALD, A. L.; LAW, A.; SCHNABEL, R. D.; MCKAY, S. D.; TAYLOR, J. F.; OGDEN, R. Evaluation of approaches for identifying population informative markers from high density SNP Chips. **BMC Genetics**, v. 12, p. 1-14, 2011.

YOKOO, M. J.; MAGNABOSCO, C. U.; ROSA, G. J. M.; LÔBO, R. B.; ALBUQUERQUE, L. G. Características reprodutivas e suas associações com outras características de importância econômica na raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 64, p. 91-100, 2012.

### **CAPÍTULO 3 - Impact of multiple sire mating system on the accuracy of genomic breeding value prediction in a beef cattle population under selection**

**Abstract** - The objective of this study was to investigate the application of BLUP and ssGBLUP in different scenarios of uncertain paternity using data from a Nellore cattle population. Data from 60,325 and 18,526 records for weight adjusted at 450 days of age (W450) and age at first calving (AFC), respectively, were studied. The variance components were estimated using BLUP and ssGBLUP methods. The relationship matrix (**A**) was created with different proportions of animals with unknown sires (0, 25, 50, 75, and 100%). All models included contemporary groups and cow age at calving in classes as fixed effects. The accuracy of the estimated breeding value (EBV/GEBV) was evaluated in each scenario with eight groups of animals: ALL = all animals in the population, BULL = only bulls with ten or more progenies; GEN = genotyped animals, GENwithPHEN = genotyped animals with phenotypes, GENwithoutPHEN = genotyped animals without phenotypes, YOUNG = male and female young animals without phenotypes, YwithoutGEN = young animals without phenotypes and genotypes, and YwithGEN = young animals without phenotypes and with genotypes. Prediction accuracies obtained with BLUP method ranged from 0.02 to 0.46 for W450 and 0.04 to 0.18 for AFC. For ssGBLUP method ranged from 0.13 to 0.48 for W450 and 0.16 to 0.33 for AFC. For AFC, additive genetic variance remained practically constant as the proportion of MS increased; however, for W450 additive genetic variance decreases as the proportion of MS increased in the population. The accuracies decreased according to the increase of MS in the population, for both traits using BLUP and ssGBLUP.

**Keywords** - genetic evaluation, missing pedigree, Nellore cattle, uncertain paternity

## 1. Introduction

The growth and reproductive traits are determinant for the profit of beef cattle industry. The weights and weight gains at different ages stand out as the selection criteria, since showed medium to high heritability, can be measured directly in the selection candidates and they are correlated with carcass weight (FERRIANI et al., 2013; TONUSSI et al., 2015). The growth precocity is one of the most desirable trait, because it can increase the weight gain, reduce the time to attained the slaughter of animals in the pasture or feedlot, reducing the cost and increasing the feed conversion (PAULINO et al., 2002).

The age at first calving is used to evaluate the sexual precocity and fertility in breeding programs of beef cattle and it can be easily obtained (AZEVEDO et al., 2006). Selection for sexual precocity decreases the generation interval, increases the replacement rate and improves profitability, as well as the economic value of the breeding animals. Although this trait generally has low heritability, not responding quickly to selection (REGATIERI et al., 2012), genetic and phenotypic associations with growth traits are favorable (BOLIGON et al., 2010a).

Multiple service sires (MS) is the most common used mating system in Brazilian livestock, which consists in clustering a group of cows with several bulls in the same pen during the mating season (CARDOSO et al., 2003). This mating system has the advantage of improving the conception rate, however it results in an uncertainty paternity scenario. Consequently, MS causes incomplete pedigrees, which could negatively affect the accuracy of genetic evaluations and, subsequently, the genetic progress by selection (CARDOSO; TEMPELMAN, 2003).

Genetic evaluations can be performed using the traditional BLUP (best linear unbiased prediction) model, proposed by Henderson (1988), which is based on relationship matrix (**A**), and presupposes the knowledge of probabilities of each candidate male being the real sire of a particular animal. Famula (1992) also presented a strategy for rapidly computing the inverse of the numerator relationship matrix under parental uncertainty. The method is simple to apply and analogous to the usual case, in which parents are known with certainty. This algorithm allows uncertainty on both sides of the pedigree (paternal and maternal), also allows the

possibility of inbreeding and offers knowledge into the genetic evaluation of sires with uncertain paternity. Posteriorly, Famula (1993) evaluated the effect on the accuracy in the sire evaluation, increasing progeny records and concluded that this inclusion can improve the accuracy of evaluation with uncertainly paternity.

Genomic selection is a developed tool to predict breeding values for quantitative traits through of the use of dense markers covering the whole genome (LUAN et al., 2009). Legarra et al. (2009) and Misztal et al. (2009) proposed an approach to integrate the pedigree (**A** matrix) and the genomic information (**G** matrix) into a single matrix (**H**) to predict the GEBV. Thus, in the animal model, the inverse of the numerator of relationship matrix ( $\mathbf{A}^{-1}$ ) is replaced by matrix  $\mathbf{H}^{-1}$ , which combines the pedigree and genomic information. Recently, several studies using ssGBLUP have been done using genomic information to predict genetic breeding values in many species as dairy cattle (FRAGOMENI et al., 2015), beef cattle (SILVA et al., 2016), pig (OSTERSEN et al., 2016) and chicken (LOURENÇO et al., 2015).

In situations of missing pedigree, problems in the scales between **A** and **G** matrices are expected, since the **A** matrix is less informative. Nevertheless, animals from **A** and **G** matrix must be in nearby scales, having connect ability among these animals. If the **A** matrix is sparser than the **G** matrix, an overestimation of variance may occur (FORNI et al., 2011), and consequently, biased genomic prediction values are obtained. However, according to Misztal et al. (2013a), besides of the incomplete pedigree, others factors may affect the incompatibility between **A** and **G** matrix as incorrect assignment of genotypes, poor quality of genotypes, and the unaccounted presence of multiple/lines breeds.

Use of statistical models for unknown parent's groups (UPG) has been applied in genetic evaluations, mainly with unstructured pedigrees (QUAAS et al., 1988) and with multiracial populations (LEGARRA et al., 2007). Misztal et al. (2013b) working with inclusion of (UPG) in the animal effect, considered some approaches in order to reduce or even eliminate these bias about UPG solutions: mixed model equations were modified including UPG information; unknown parents groups were considered as separate effects, however, it requires a high computational demand or contributions of UPG were ignored when they were relatively small. Recently, Tsuruta et al. (2014) examined how to define the UPG assigned in mixed-model equations



(MME) and concluded that if poorly defined may to reduce bias and possibly increase accuracy in genomic evaluations for young Holstein bulls using ssGBLUP.

Genomic data is not used in genetic breeding programs in scenarios with uncertain paternity, however, there is not consensus on how the use of ssGBLUP can to affect evaluation genetic in this situation. The objective of this work was to investigate the application of BLUP and ssGBLUP for weight adjusted at 455 days of age and age at first calving under different scenarios with uncertain paternity due to incomplete pedigrees in Nelore cattle.

## **2. Material and Methods**

### **2.1. Data**

The analyzed data set was provided by the National Association of Farmers and Researchers (*Associação Nacional de Criadores e Pesquisadores - ANCP*). The data set contained information from 18 Nelore herds located in the southeast and mid-west regions of Brazil, which participate in the ANCP breeding program.

A total of 60,325 and 18,526 records for weight adjusted at 450 days (W450) and age at first calving (AFC) were used, respectively. The mean values  $\pm$  standard deviations were as follows:  $290.20 \pm 50.26$  kg for W450 and  $35.08 \pm 4.8$  months for AFC.

The contemporary groups (CG) were defined according to each trait: for W450 were formed 2,312 CGs and defined as farm, year of birth, season of birth, sex and management group. For AFC were formed 276 CGs defined as farm, year and season of birth. Records with values above or below the range of 3.5 standard deviations from the CG mean were excluded, as well as CGs with less than five animals.

A total of 3,809 animals were genotyped using low-density panel (Clarified Nelore 2.0), and subsequently imputed to arrays with 54k and 777k, using the Fimpute software (SARGOLZAEI et al., 2012), which has high efficiency for imputation of genotypes (SUN et al., 2012). As reference were used 963 animals in high density.

## 2.2. BLUP and ssGBLUP method

In the BLUP method, a traditional genetic evaluation was implemented using pedigree and phenotypic information. The model can be represented as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of phenotype,  $\mathbf{b}$  is the vector of fixed effects (CG for both traits and cow age at calving in classes for W450),  $\mathbf{u}$  is the vector of additive genetic effect,  $\mathbf{X}$  e  $\mathbf{Z}$  are incidence matrices and  $\mathbf{e}$  is the vector of random residuals. Considering an infinitesimal model,  $\text{var}(\mathbf{u}) = \mathbf{A}\sigma_u^2$ , where  $\mathbf{A}$  is the numerator relationship matrix obtained from pedigree information and  $\sigma_u^2$  is the variance of genetic effect.

In the single-step genomic BLUP (ssGBLUP) proposed by Misztal et al., (2009), the inverse of the numerator relationship matrix ( $\mathbf{A}^{-1}$ ) was replaced by  $\mathbf{H}^{-1}$  that combines pedigree and genomic information. Aguilar et al., (2010) showed that  $\mathbf{H}^{-1}$  is constructed as:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where  $\mathbf{H}^{-1}$  is the inverse of the realized relationship matrix that incorporates the inverse of the genomic relationship matrix ( $\mathbf{G}^{-1}$ ) and the inverse of the numerator relationship matrix for genotyped animals  $\mathbf{A}_{22}^{-1}$ . The  $\mathbf{G}$  matrix was created according to VanRaden (2008):

$$\mathbf{G} = \frac{(\mathbf{M} - \mathbf{P})(\mathbf{M} - \mathbf{P})'}{2 \sum_{j=1}^m p_j (1 - p_j)}$$

where  $\mathbf{M}$  is a matrix of marker alleles with  $m$  columns ( $m$  = total number of markers) and  $n$  rows ( $n$  = total number of genotyped individuals), and  $\mathbf{P}$  is a matrix containing the frequency of the second allele ( $p_j$ ), expressed as  $2p_j$ .  $\mathbf{M}_{ij}$  was 0 if the genotype of individual  $i$  for SNP  $j$  was homozygous for the first allele, was 1 if heterozygous, or 2 if the genotype was homozygous for the second allele. The variance component estimation (AIREMLF90 software) and solutions (BLUPF90 software) were obtained by BLUPF90 family programs (Misztal et al., 2002, Aguilar et al., 2014). The relationship matrix has 23 generations and included data of 129,369 animals with 5,257 sires and 50,742 cows.

### 2.3. Tested scenarios

The **A** matrix was created with different proportions of known paternity for genotyped animals (0, 25, 50, 75 and 100%). The genotyped animals with missing sire were selected or chosen at random. These proportions of missing sire for genotyped animals were randomly. Thus, a total of five scenarios were tested for each trait (10 scenarios in total) considering the BLUP and ssGBLUP.

Accuracy of prediction was calculated for eight groups of animals in each scenario: ALL = all animals with phenotypic records (60,325 and 18,526 animals for W450 and AFC, respectively); BULL = only bulls with at least ten progenies (1,363 bulls for both traits); GEN = genotyped animals (3,809 animals); GENwithPHEN = genotyped animals with phenotypes (1,973 and 658 animals for W450 and AFC, respectively); GENwithoutPHEN = genotyped animals without phenotypes (1,836 and 3,151 animals for W450 and AFC, respectively); YOUNG = male and female young animals without phenotypes and progenies (13,529 and 29,824 animals for W450 and AFC, respectively); YwithoutGEN = young animals without phenotypes and genotypes (12,014 and 26,931 animals for W450 and AFC, respectively) and YwithGEN = young animals without phenotypes and with genotypes (1,515 and 2,893 records for W450 and AFC, respectively). The accuracies of predictions were calculated according to BIF (beef improvement federation), represented below:

$$ACC_{BIF} = 1 - \sqrt{\frac{VEP}{\sigma_a^2}}$$

where VEP is the variance of the prediction error and  $\sigma_a^2$  is the additive variance. To evaluate the impact of missing pedigree on the rank of genotyped animals and young genotyped animals, the EBV and GEBV spearman rank correlations for AFC and W450 were estimated for each scenario.

### 3. Results and Discussion

Variance component estimates, heritability and standard errors for AFC and W450 with different proportions of uncertain paternity were showed in Table 1. For AFC, additive genetic variance remained practically unchanged with the different

proportion of MS; however, for W450 the additive genetic variance decreases as the proportion of MS increased in the population. Harder et al. (2005) working with dairy cattle also found a decrease in additive genetic variance according to increasing missing sire information in the pedigree.

Dong et al. (1988) estimated heritability for three traits (first lactation, fat and protein) with different pedigree (full pedigree, full less base – pedigree without base animals, and sire only – only relationships through sires) and observed a decrease in heritability estimation for all traits according missing pedigree. In another study, Cantet et al. (2000), evaluated four alternative pedigree structures: complete pedigree information, 50% of phenotypes with sire identities missing, 50% of phenotypes with dam identities missing, and 50% of phenotypes with sire and dams identities missing for different heritability's (0.10, 0.40 and, 0.70). In all cases with pedigree missing, the heritability decreases compared to complete pedigree.

Table 1. Estimation of variance components, heritability and standard errors for AFC and W450 using traditional and genomic AIREML with different proportions of multiple sires.

Trait	MS	AIREML			Genomic AIREML		
		$\sigma_a^2 \pm SE$	$\sigma_e^2 \pm SE$	$h^2$	$\sigma_a^2 \pm SE$	$\sigma_e^2 \pm SE$	$h^2$
AFC	0%	1.47±0.18	11.19±0.19	0.12	1.90±0.21	10.90±0.20	0.15
	25%	1.47±0.19	11.18±0.19	0.12	1.87±0.21	10.92±0.20	0.15
	50%	1.48±0.19	11.17±0.19	0.12	1.84±0.21	10.93±0.19	0.14
	75%	1.49±0.19	11.17±0.19	0.12	1.82±0.21	10.94±0.19	0.14
	100%	1.47±0.19	11.16±0.19	0.12	1.74±0.20	10.99±0.19	0.14
W450	0%	291.31±9.25	380.47±6.22	0.43	307.45±9.45	371.99±6.23	0.45
	25%	278.47±9.13	390.48±6.22	0.42	301.65±9.38	376.48±6.24	0.44
	50%	274.88±9.08	393.53±6.23	0.41	296.92±9.33	379.96±6.24	0.44
	75%	273.63±9.06	394.66±6.23	0.41	291.96±9.27	383.12±6.25	0.43
	100%	272.67±9.03	395.14±6.24	0.41	285.15±9.21	387.70±6.26	0.42

AFC=age at first calving; W450=weight adjusted for 450 days of age; MS=percentage of multiple sires; AIREML= average information restricted maximum likelihood estimation;  $\sigma_a^2$ = additive genetic variance;  $\sigma_e^2$ = residual variance;  $h^2$ = heritability; SE= standard errors

The heritability estimates obtained for AFC and W450 were according to literature for Nelore breed animals and reared in similar regions of the country, ranging from 0.10 to 0.16 for AFC (CAETANO et al., 2013; CAVANI et al., 2015) and ranging from 0.37 to 0.51 for W450 (BOLIGON et al., 2010b; DIAZ et al., 2011).

It is important to highlight that the correlation between the relationship matrix ( $\mathbf{A}_{22}$ ) and the genomic matrix ( $\mathbf{G}$ ) decreased when the proportion of MS in the

population increases. This is expected because in situations of missing pedigree the **A** matrix is sparser in relation to the **G** matrix. Forni et al. (2011) demonstrated that genomic and pedigree matrices should be compatible in scale, so that there is connect ability between these animals, and consequently, more accurate genomic prediction values.

Breeding value accuracies for W450 obtained by BLUP and ssGBLUP methods for different percentage of MS were showed in the Table 2. For the BLUP method, the prediction accuracy decreased for all groups as the proportion of MS in the population increased. The EBV accuracy reduction was 8.8%, 4.3%, 46.2%, 27.3%, 82.4%, 25.0%, 18.8% and 87.5% for ALL, BULL, GEN, GENwithPHE, GENwithoutPHE, YOUNG, YwithoutGEN and YwithGEN groups, respectively.

Table 2. Breeding value accuracy prediction for W450 using the BLUP and ssGBLUP with different proportions of multiple sires in the genotyped animals

		Percentage of multiple sires				
		0%	25%	50%	75%	100%
ALL	BLUP	0.34	0.32	0.31	0.31	0.31
	ssGBLUP	0.35	0.35	0.34	0.34	0.33
BULL	BLUP	0.46	0.44	0.44	0.44	0.44
	ssGBLUP	0.48	0.47	0.47	0.46	0.45
GEN	BLUP	0.26	0.22	0.19	0.17	0.14
	ssGBLUP	0.33	0.31	0.28	0.25	0.22
GENwithPHE	BLUP	0.33	0.30	0.28	0.26	0.24
	ssGBLUP	0.39	0.37	0.35	0.33	0.30
GENwithoutPHE	BLUP	0.17	0.13	0.09	0.06	0.03
	ssGBLUP	0.26	0.24	0.21	0.17	0.13
YOUNG	BLUP	0.16	0.14	0.13	0.13	0.12
	ssGBLUP	0.20	0.19	0.18	0.16	0.15
YwithoutGEN	BLUP	0.16	0.14	0.14	0.14	0.13
	ssGBLUP	0.18	0.18	0.17	0.16	0.15
YwithGEN	BLUP	0.16	0.12	0.08	0.05	0.02
	ssGBLUP	0.25	0.23	0.20	0.16	0.12

BLUP = best linear unbiased predictor; ssGBLUP = single step genomic BLUP; ALL = all animals in the population; BULL = only bulls with ten or more progenies; GEN = genotyped animals; GENwithPHEN = genotyped animals with phenotypes; GENwithoutPHEN = genotyped animals without phenotypes; YOUNG = male and female young animals without phenotypes; YwithGEN = young animals with genotypes and YwithoutGEN = young animals without genotypes

Accuracy prediction of EBV and GEBV for AFC using BLUP and ssGBLUP were presented in the Table 3. Similar to W450, EBV accuracies with BLUP method decreased as the percentage of MS increased in the population, with accuracies reduction of 6.3%, 0.0%, 54.5%, 37.5%, 50.0%, 10.0%, 0.0% e 60.0% for ALL,

BULL, GEN, GENwithPHE, GENwithoutPHE, YOUNG, YwithoutGEN e YwithGEN, respectively.

In general, for all scenarios using the ssGBLUP, the GEBV prediction accuracy (Table 2 and 3) decreased as the proportion of MS in the population increased, however, this reduction was different for each trait in the studied groups. Probably, this is due to different number of phenotypes used for each trait, because for AFC, the number of phenotypic records used were lower than for W450. In addition, another reason for this difference can be associated with the lower heritability estimate of AFC compared to W450. Thus, the estimation of breeding values for AFC is more affected by the population structure than for W450.

Table 3. Breeding value accuracy prediction for AFC using the BLUP and ssGBLUP methods with different proportions of multiple sires in the genotyped animals.

		Percentage of multiple sires				
		0%	25%	50%	75%	100%
ALL	BLUP	0.16	0.16	0.16	0.16	0.15
	ssGBLUP	0.26	0.25	0.24	0.24	0.22
BULL	BLUP	0.18	0.18	0.18	0.18	0.18
	ssGBLUP	0.29	0.29	0.27	0.26	0.24
GEN	BLUP	0.11	0.10	0.09	0.07	0.05
	ssGBLUP	0.30	0.27	0.25	0.22	0.18
GENwithPHE	BLUP	0.16	0.14	0.13	0.12	0.10
	ssGBLUP	0.33	0.31	0.28	0.26	0.22
GENwithoutPHE	BLUP	0.10	0.09	0.08	0.07	0.05
	ssGBLUP	0.29	0.27	0.24	0.22	0.18
YOUNG	BLUP	0.10	0.10	0.10	0.10	0.09
	ssGBLUP	0.22	0.22	0.20	0.19	0.17
YwithoutGEN	BLUP	0.10	0.10	0.11	0.11	0.10
	ssGBLUP	0.21	0.21	0.20	0.19	0.16
YwithGEN	BLUP	0.10	0.09	0.08	0.06	0.04
	ssGBLUP	0.29	0.27	0.24	0.22	0.18

BLUP = best linear unbiased predictor; ssGBLUP = single step genomic BLUP; ALL = all animals in the population; BULL = only bulls with ten or more progenies; GEN = genotyped animals; GENwithPHEN = genotyped animals with phenotypes; GENwithoutPHEN = genotyped animals without phenotypes; YOUNG = male and female young animals without phenotypes; YwithoutGEN = young animals without genotypes and YwithGEN = young animals with genotypes.

In the scenario with 0% of MS (Table 2), comparing the EBV and GEBV accuracy predictions for W450 obtained with BLUP and ssGBLUP methods, it was observed an accuracy gain of 2.9; 4.3; 26.9; 18.2; 52.9; 25.0; 12.5 and 56.3% for ALL, BULL, GEN, GENwithPHE, GENwithoutPHE, YOUNG, YwithoutGEN and YwithGEN, respectively. For AFC, the accuracy gains were higher when the ssGBLUP method

was applied, in some cases (young animals) being higher than 100%. Gains in accuracy also were observed by Lourenço et al. (2013), when the ssGBLUP was compared with BLUP for simulated data of weaning weight, using an animal model with direct and maternal effects. For all genotyped animals, the gains were 1.4 and 4.76% for the direct and maternal effect, respectively. While for genotyped young animals without phenotypes the gains were 7.27% and 5.6%, respectively for the same effects.

Recently, Lourenço et al. (2016) evaluated the application of the ssGBLUP using records from growth traits from the American Angus, and showed an increased in prediction accuracy of 25% compared with traditional evaluations. In Nellore cattle, Gordo et al. (2016) using ssGBLUP found 76.47%, 13.04% and 11.90% accuracy gain for longissimus muscle area, backfat thickness and hot carcass weight, respectively, in Nellore cattle. In addition, several studies showed that the ssGBLUP was more reliable than BLUP model in other species, like dairy cattle (AGUILAR et al., 2010) that obtained accuracy gain of 14,29% for final score; in chicken (CHEN et al., 2011) the accuracy gain was of 44,44% in breast meat; in pigs (CHRISTENSEN et al., 2012) was of 48,57% for daily gain (weight gain/days) and in dairy sheep (BALOCHE et al., 2014) the accuracy gain was of 31,92% for milk yield.

Predictions of EBV and GEBV for AFC and W450 using BLUP and ssGBLUP methods for all genotyped animals under different proportions of MS in the population were showed in Table 4. The EBV and GEBV for AFC and W450 decreased as the proportion of MS in the population increased, probably due to poor pedigree structure, which causes EBV/GEBV of these animals to be regressed to zero. Similar results were reported by Harder et al. (2005), where breeding values declined according to the missing pedigree for two traits with different heritability estimates (0.10 and 0.25).

Table 4. Averages and standard deviations for breeding values estimation for AFC and W450 using BLUP and ssGBLUP methods

	AFC		W450	
	BLUP	ssGBLUP	BLUP	ssGBLUP
0%	0.17 ± 0.47	0.04 ± 0.37	12.57 ± 12.97	2.19 ± 9.99
25%	0.17 ± 0.47	0.04 ± 0.37	7.53 ± 11.28	0.55 ± 9.72
50%	0.17 ± 0.47	0.03 ± 0.38	4.79 ± 10.56	-0.76 ± 9.58
75%	0.16 ± 0.47	0.03 ± 0.39	3.09 ± 10.18	-1.49 ± 9.53
100%	0.16 ± 0.47	0.02 ± 0.41	2.52 ± 10.07	-1.66 ± 9.55

AFC = age at first calving; W450 = weight at 450 days; BLUP = best linear unbiased predictor; ssGBLUP = single step genomic BLUP

The EBV and GEBV rank correlations for genotyped animals and young genotyped animals without phenotypic records for different proportion of MS in the population were presented in Table 5. In general, as the percentage of MS increases, the rank correlation using EBV and GEBV for each scenario decreased in relation to the scenario with 0% of MS. Nevertheless, using the ssGBLUP model the decrease of rank correlation was lower, showing that this method was more reliable than the BLUP. The ssGBLUP method performed more reliable genetic evaluation than the BLUP model, mainly for young animals with missing pedigree.

Table 5 - Rank correlation of EBV of AFC and W450 of genotyped animals and young genotyped animals for different percentage of multiple sire

	AFC - GEN		W450 - GEN		AFC - YwithGEN		W450 - YwithGEN	
	BLUP	ssGBLUP	BLUP	ssGBLUP	BLUP	ssGBLUP	BLUP	ssGBLUP
0 vs 25%	0.91	0.96	0.86	0.89	0.90	0.96	0.74	0.84
0 vs 50%	0.79	0.92	0.82	0.87	0.78	0.91	0.64	0.79
0 vs 75%	0.67	0.86	0.80	0.86	0.64	0.85	0.64	0.77
0 vs 100%	0.56	0.74	0.80	0.84	0.51	0.72	0.65	0.74

EBV = estimated breeding values; AFC = age at first calving; W450 = weight at 450 days; BLUP = best linear unbiased prediction, ssGBLUP = single step genomic BLUP; GEN = genotyped animals; YwithGEN = young animals without phenotypes and with genotypes



#### **4. Conclusions**

The additive genetic variance, reliability of genetic evaluation and breeding values for AFC and W450 were influenced by presence of the paternity uncertainty in the pedigree. The presence of paternity uncertainty affects more intensively the breeding value of young animals. The genotyped young animals were benefited from the application of ssGBLUP, particularly in situations with missing pedigree.

The ssGBLUP method could be applied in situations of uncertainty paternity, especially for selection of young animals, although it was not developed to work with paternity uncertainty, was shown to be accurate, mainly in scenarios with high percentage of multiple sires, but it is important and necessary to scale the genomic relationship matrix to be compatible with the numerator relationship matrix for genotyped animals.

#### **5. Implications**

In Brazil, there are many studs or herds that belong to breeding program that frequently use the multiple sire serving and it doesn't recognize paternity, and in some herds is common observe close to 60% of the progeny with unknown sire. This fact decreases the reliability of genetic evaluations. Additionally, there is a growing interest in commercial beef cattle herds to start genetic evaluation in herds without registered animals by the breeder association, in order to commercialize animals with genetic evaluation information.

In this context, this study allowed to evaluate the technical feasibility of the ssGBLUP in situation with paternity uncertainty, since there are many doubts and concerns about the most adequate G matrix scaling method under paternity uncertainty scenarios and, their impact on the genomic evaluation (accuracy and bias). It is important to highlight, that the ssGBLUP procedure was not developed to deal with paternity uncertainty scenarios, since it increase the problems of scaling between the G and A22 matrix. However, the results of this study shown that the application of the ssGBLUP was appropriate, and is important to scale the G matrix, even in situations with paternity uncertainty (simulated and real data).

Considering real data is not possible to access the TBV or have animals with accurate EBV, mainly in beef cattle, to really evaluate the impact of multiple sires. Also, in some circumstances, is not possible to have all the scenarios with real data, and adequate population structure to draw reliable conclusions. In this way, the simulation solved such problems. However, the study with real data is important because it is the practice scenarios and it possible to evaluate the strategy of more adequate scale in populations with absent pedigrees. Both studies showed the importance of structure pedigree, once the accuracies decreased according increase multiple sires in the population.

## 6. References

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, p. 743-752, 2010.

AGUILAR, I.; MISZTAL, I.; TSURUTA, S.; LEGARRA, A. PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. In: Internacional, 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada. **Proceedings**, 2014.

AZEVEDO, D. M. M. R.; MARTINS FILHO, R.; LÔBO, R. N. B.; MALHADO, C. H. M.; LÔBO, R. B.; MOURA, A. A. A.; PIMENTA FILHO, E. C. Desempenho reprodutivo de vacas Nelore no Norte e Nordeste do Brasil. **Revista Brasileira de Zootecnia**, v.35, n.3, p.988- 996, 2006.

BALOCHE, A.; LEGARRA, A.; SALLÉ, G.; LARROQUE, H.; ASTRUC, J. M.; ROBERT-GRANIÉ, C.; BARILLET, F. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. **Journal of Dairy Science**, v. 97, p. 1107-1116, 2014.

BOLIGON, A. A.; ALBUQUERQUE, L. G.; MERCADANTE, M. E. Z.; LÔBO, R. B. Study of relations among age at first calving, average weight gains and weights from weaning to maturity in Nelore cattle. **Revista Brasileira de Zootecnia**, v.39, n.4, p.746-751, 2010a.

BOLIGON, A. A.; SILVA, J. A. V.; SESANA, R. C.; SESANA, J. C.; JUNQUEIRA, J. B.; ALBUQUERQUE, L. G. Estimation of genetic parameters for body weights,

scrotal circumference, and testicular volume measured at different ages in Nelore cattle. **Journal of animal science**, v. 88, p. 1215-1219, 2010b.

CAETANO, S. L.; SAVEGNAGO, R. P.; BOLIGON, A. A.; RAMOS, S. B.; CHUD, T. C. S.; LÔBO, R. B.; MUNARI D. P. Estimates of genetic parameters for carcass, growth and reproductive traits in Nelore cattle. **Livestock Science**, v. 155, n.1, p. 1-7, 2013.

CANTET, R. J. C.; BIRCHMEIER, A. N.; SANTOS-CRISTAL, M. G.; AVILA, V. S. Comparison of restricted maximum likelihood and method R for estimating heritability and predicting breeding value under selection. **Journal of Animal Science**, v. 78, p. 2554-2560, 2000.

CARDOSO, V.; ROSO, V. M.; SEVERO, J. L. P.; QUEIROZ, S. D.; FRIES, L. A. Formando lotes uniformes de reprodutores múltiplos e usando-os em acasalamentos dirigidos, em populações Nelore. **Revista Brasileira de Zootecnia**, v. 32, n.4, p. 834-842, 2003.

CARDOSO, F. F.; TEMPELMAN, R. J. Bayesian inference on genetic merit under uncertain paternity. **Genetics Selection Evolution**, v. 35, p. 469-487, 2003.

CAVANI, L.; GARCIA, D. A.; CARREÑO, L. O. D.; ONO, R. K.; PIRES, M. P.; FARAH, M. M.; VENTURA, H. T.; MILLEN, D. D.; FONSECA, R. Estimates of genetic parameters for reproductive traits in Brahman cattle breed. **Journal of Animal Science**, v. 93, p.3287-3291, 2015.

CHEN, C.; MISZTAL, I.; AGUILAR, I.; TSURUTA, S.; AGGREY, S.; WING, T.; MUIR, W. Genome-wide marker-assisted selection combining all pedigree phenotypic information with phenotypic data in one step: an example using broiler chickens. **Journal of Animal Science**, v. 89, p. 23-28, 2011.

CHRISTENSEN, O.; MADSEN, P.; NIELSEN, B.; OSTERSEN, T.; SU, G. Single-step methods for genomic evaluation in pigs. **Animal**, v. 6, p. 1565-1571, 2012.

DIAZ, I. D. P. S.; OLIVEIRA, H. N. D.; BEZERRA, L. A. F.; LÔBO, R. B. Genotype by environment interaction in Nelore cattle from five Brazilian states. **Genetics and Molecular Biology**, v. 34, n.3, p. 435-442, 2011.

DONG, M. C.; VAN VLECK, L. D.; WIGGANS, G. R. Effect of relationships on estimation of variance components with an animal model and restricted maximum likelihood. **Journal of Dairy Science**, v. 71, p. 3047-3052, 1988.

FAMULA, T. R. Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty. **Journal of Animal Science**, v. 70, p. 1045-1048, 1992.

FAMULA, T. R. The Contribution of Progeny of Uncertain Paternity to the Accuracy of Sire Evaluation, **Journal of Animal Science**, v. 71, p. 1136-1141, 1993.

FERRIANI, L.; ALBUQUERQUE, L. G.; BALDI, F.; VENTURINI, G. C.; BIGNARDI, A. B.; SILVA, J. A. V.; CHUD, T. C. S.; MUNARI, D. P.; OLIVEIRA, J. A. Parâmetros genéticos de características de carcaça e de crescimento de bovinos da raça Nelore. **Archivos de Zootecnia**, v. 62, p. 123-129, 2013.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v. 43, n. 1, p. 1-7, 2011.

FRAGOMENI, B. O.; LOURENÇO, D. A. L.; TSURUTA, S.; MASUDA, Y.; AGUILAR, I.; LEGARRA, A. LAWLOR, T. J.; MISZTAL, I. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. **Journal of Dairy Science**, v. 98, p. 4090-4094, 2015.

GORDO, D. G. M.; ESPIGOLAN, R.; TONUSSI, R. L.; JÚNIOR, G. A. F.; BRESOLIN, T.; MAGALHÃES, A. F. B.; FEITOSA, F. L. B.; BALDI, F.; CARVALHEIRO, R.; TONHATI, H.; OLIVEIRA, H. N.; CHARDULA, L. A. L.; ALBUQUERQUE, L. G. Genetic parameter estimates for carcass traits and visual scores including or not genomic information. **Journal of Animal Science**, v. 94, p. 1821-1826, 2016.

HARDER, B. I. R. T. E.; BENNEWITZ, J. Ö. R. N., REINSCH, N. O. R. B. E. R. T.; MAYER, M. A. N. F. R. E. D.; KALM, E. R. N. S. T. Effect of missing sire information on genetic evaluation. **Arch. Tierz.**, v. 48, p. 219-232, 2005.

HENDERSON, C. R. Use of an average numerator relationship matrix for multiple-sire joining. **Journal of Animal Science**, v. 66, p. 1614-1621, 1988.

LEGARRA, A.; BERTRAND, J. K.; STRABEL, T.; SAPP, R. L.; SANCHEZ, J. P.; MISZTAL, I. Multi-breed genetic evaluation in a Gelbvieh population. **Journal of Animal Breeding and Genetics**, v. 124, p. 286– 295, 2007.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, p. 4656-4663, 2009.

LOURENÇO, D.; MISZTAL, I.; WANG, H.; AGUILAR, I.; TSURUTA, S.; BERTRAND, J. Prediction accuracy for a simulated maternally affected trait of beef cattle using different genomic evaluation models. **Journal of Animal Science**, v. 91, p. 4090-4098, 2013.

LOURENÇO, D. A.; FRAGOMENI, B. O.; TSURUTA, S.; AGUILAR, I.; ZUMBACH, B.; HAWKEN, R. J.; LEGARRA, A.; MISZTAL, I. Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. **Genetics Selection Evolution**, v. 47, p. 1-9, 2015.

LOURENÇO, D. A. L.; TSURUTA, S.; FRAGOMENI, B. D.; MASUDA, Y.; POCRNIC, I.; AGUILAR, I.; BERTRAND, J. K.; MOSER, D. W.; MISZTAL, I. Issues in

commercial application of single-step genomic BLUP for genetic evaluation in American Angus. **Journal of Animal Science**, v. 94, p. 144-145, 2016.

LUAN, T.; WOOLLIAMS, J. A.; LIEN, S.; KENT, M.; SVENDSEN, M.; MEUWISSEN, T. H. E. The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. **Genetics**, v. 183, p. 1119-1126, 2009.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D.H. **BLUPF90 and related programs (BGF90)**. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, 2002.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, p. 4648-4655, 2009.

MISZTAL, I.; AGGREY, S. E.; MUIR, W. M. Experiences with a single step genome evaluation. **Poultry Science**, v. 92, p. 2530-2534, 2013a.

MISZTAL, I.; VITEZICA, Z. G.; LEGARRA, A.; AGUILAR, I.; SWAN, A. A. Unknown-parent groups in single-step genomic evaluation. **Journal of Animal Breeding and Genetics**, v. 130, p. 252-258, 2013b.

OSTERSEN, T.; CHRISTENSEN, O. F.; MADSEN, P.; HENRYON, M. Sparse single-step method for genomic evaluation in pigs. **Genetics Selection Evolution**, v. 48, p. 1-10, 2016.

PAULINO, M. F., ZERVOUDAKIS, J. T., MORAES, E. H. B. K., DETMANN, E., Valadares Filho, S. C. Bovinocultura de ciclo curto em pastagens. **Simpósio de Produção de Gado de Corte**, p. 153-196, 2002.

QUAAS, R. L. Additive genetic model with groups and relationships. **Journal of Dairy Science**. v 71. p. 1338–1345. 1988.

REGATIERI, I. C.; BOLIGON, A. A.; BALDI, F.; ALBUQUERQUE, L. G. Genetic correlations between mature cow weight and productive and reproductive traits in Nellore cattle. **Genetics and Molecular Research**, v.11, n.3, p.2979-2986, 2012.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. Accuracy of imputed 50k genotypes from 3k and 6k chips in dairy cattle breeds using FImpute. In: International Plant and Animal Genome. **Conference**, San Diego, 2012.

SILVA, R. M. O.; FRAGOMENI, B. O.; LOURENÇO, D. A. L.; MAGALHÃES, A. F. B.; IRANO, N.; CARVALHEIRO, R.; CANESIN, R. C.; MERCADANTE, M. E. Z.; BOLIGON, A. A.; BALDI, F.; MISZTAL, I.; ALBUQUERQUE, L. G. Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. **Journal of Animal Science**, v. 94, p. 3613-3623, 2016.

SUN, C.; WU, X. L.; WEIGEL, K. A.; ROSA, G. J. M.; BAUCK, S.; WOODWARD, B. W.; SCHNABEL, R. D.; TAYLOR, J. F.; GIANOLA, D. An ensemble-based approach to imputation of moderate-density genotypes for genomic selection with application to Angus cattle. **Genetics research (Cambridge)**, v. 94, n.3, p.133–50, 2012.

TONUSSI, R. L.; ESPIGOLAN, R.; GORDO, D. G. M.; MAGALHÃES, A. F. B.; VENTURINI, G. C.; BALDI, F.; OLIVEIRA, H. N.; CHARDULO, L. A. L.; TONHATI, H.; ALBUQUERQUE, L. G. Genetic association of growth traits with carcass and meat traits in Nellore cattle. **Genetics and Molecular Research**, v. 14, n. 4, p. 18713-18719, 2015.

TSURUTA, S.; MISZTAL, I.; LOURENÇO, D. A. L.; LAWLOR, T. J. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. **Journal of Dairy Science**, v. 97, p. 5814–5821, 2014.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, v. 91, p. 4414-4423, 2008.