

**UNIVERSIDADE FEDERAL DO PAMPA**

**Ândrea Plotzki Reis**

**DESCRIÇÃO METODOLÓGICA DO USO DE MARCADORES MOLECULARES  
ASSOCIADO A INFORMAÇÕES DE PEDIGREE NA PREDIÇÃO DO VALOR  
GENÉTICO DE ANIMAIS POR SIMULAÇÃO UTILIZANDO A METODOLOGIA  
BLUP**

**TRABALHO DE CONCLUSÃO DE CURSO**

**Dom Pedrito**

**2012**

R375d Reis, Ândrea Plotzki

Descrição metodológica do uso de marcadores moleculares associados a informações de pedigree na predição do valor genético de animais por simulação utilizando a metodologia BLUP / Ândrea Plotzki Reis ; orientador Prof. Dr. Eduardo Brum Schwengber. – Dom Pedrito : UNIPAMPA, Curso de Zootecnia, 2012.

1. Seleção tradicional 2. Seleção genômica 3. BLUP I.  
Título

CDD 636

**ÂNDREA PLOTZKI REIS**

**DESCRIÇÃO METODOLÓGICA DO USO DE MARCADORES MOLECULARES  
ASSOCIADO A INFORMAÇÕES DE PEDIGREE NA PREDIÇÃO DO VALOR  
GENÉTICO DE ANIMAIS POR SIMULAÇÃO UTILIZANDO A METODOLOGIA  
BLUP**

Trabalho de Conclusão de Curso apresentado a  
Universidade Federal do Pampa, como  
requisito parcial para obtenção do Título de  
Bacharel em Zootecnia.

Orientador(a): Eduardo Brum Schwengber  
Co-orientador: José Acélio Silveira da  
Fontoura Jr.  
Co-orientador: Fernando Flores Cardoso

**Dom Pedrito**

**2012**

**ÂNDREA PLOTZKI REIS**

**DESCRIÇÃO METODOLÓGICA PARA ANÁLISE DE UTILIZAÇÃO DE  
MARCADORES MOLECULARES ASSOCIADO A INFORMAÇÕES DE  
PEDIGREE PARA PREDIÇÃO DO VALOR GENÉTICO DE ANIMAIS  
ATRAVÉS DE SIMULAÇÕES UTILIZANDO A METODOLOGIA BLUP**

Trabalho de Conclusão de Curso apresentado a  
Universidade Federal do Pampa, como  
requisito parcial para obtenção do Título de  
Bacharel em Zootecnia.

Trabalho de Conclusão de Curso defendido e aprovado em: 11 de Julho de 2012

Banca examinadora:

---

Prof. Dr. Eduardo Brum Schwengber (Orientador)  
Universidade Federal do Pampa – UNIPAMPA  
Campus Dom Pedrito

---

Dr. Marcos Jun-Iti Yokoo  
Embrapa Pecuária Sul  
CPPSul

---

Dr. Mauricio Morgado de Oliveira

“Tenho agradecido por estar vivo e ter andado por todos os lugares onde andei e ter vivido tudo o que vivi e ser exatamente como sou.”

Caio F. Abreu

## AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida e por tudo o que sou e o que tenho até hoje.

Agradeço aos meus pais, Eduardo e Maria Claudete por todo amor, dedicação e principalmente por acreditarem no meu potencial sempre me instigando a seguir em frente.

Agradeço aos meus irmãos Rodrigo, Débora e Renan por serem meus companheiros e amigos. Agradeço a vó Maria por ser uma pessoa que sempre esteve presente tanto em amor quanto em presença.

Agradeço de coração aos meus mestres até aqui, porém não posso deixar de citar alguns:

- Aos professores e orientadores acadêmicos Eduardo Schwengber e o professor José Acélio Fontoura, que sempre encontrei apoio, incentivo, paciência com as minhas indecisões e dúvidas (quando queria mudar de tema de TCC, que queria um estágio, queria isso e fiz aquilo, etc), conhecimento e palavras amigas, só tenho a dizer muito obrigada por tudo ao longo desses quase cinco anos. Sei que causei algumas quedas de cabelos em vocês.

- Ao orientador de bolsa, Dr. Fernando Flores Cardoso, que colaborou na minha formação acadêmica e intelectual de forma ímpar. Levo ensinamentos além do intelectual, pois aprendi a não me subestimar, mesmo que muitas vezes não entenda no momento o que fazer.

- Ao Dr. Marcos Yokoo e ao Mauricio Morgado, que dispuseram de muitos momentos para me ajudar e paciência em me ensinar e explicar tudo o que era necessário e o que não compreendia, este trabalho não seria realidade sem a contribuição de vocês.

Não esquecendo os meus colegas do Laboratório de Bioinformática e Estatística Genômica, que proporcionaram momentos únicos, fossem de debates ou de descontração.

Aos meus amigos que me acompanham há algum tempo: Cissa, Bus, Dessa, Dorado, Leila e família. Vocês fizeram parte dos momentos mais memoráveis da minha vida, e nesse ocasião tão importante para mim não poderia ser diferente, obrigada por serem meus amigos.

Não poderia deixar de agradecer duas amigas mais do brete e boa que existem, Natalia Teixeira e Bruna Ponte:

-Natalia, como postou no face: lado a lado, do início ao fim. Se não fosse por ti, pela tua amizade e apoio talvez não estivéssemos tendo esse lado a lado, do início ao fim.

-Bruna, minha companheira que alegrava meus dias, preparava alimentinhos super du mau e fez uma falta depois que voltou pra casinha.

A minha família de Dom Pedrito, onde encontrei amizade, carinho e apoio: Tia Fátima e flia, vocês moram no meu coração e serei eternamente grata pelos momentos junto a vocês. Mas não se enganem: Sempre voltarei, nem que seja pelo mate e cupe cakes...hehe.

Natalião, Cris, Aninha, Carol, Ana, Bruna, Mudinho, Mozer, Nardino, Gabizinha e Natyta: alguns que fizeram esses cinco anos inesquecíveis com nossas conversas (produtivas ou não), sessões engordes, aguentando minhas crises existenciais, quase sendo expulsas do ap por brincar de mimica ou mostrando indevidos pela janela do carro... já sinto uma saudade indefinível. Mas chegaaaaaaa, digam adeus... mas hoje prefiro um até breve!

Muito obrigada a tudo e a todos que fizeram parte deste caminho, desde a universidade até as pessoas que de alguma forma contribuíram pro meu crescimento pessoal e profissional.

## RESUMO

### DESCRIÇÃO METODOLÓGICA DO USO DE MARCADORES MOLECULARES ASSOCIADO A INFORMAÇÕES DE PEDIGREE NA PREDIÇÃO DO VALOR GENÉTICO DE ANIMAIS POR SIMULAÇÃO UTILIZANDO A METODOLOGIA BLUP.

O presente trabalho teve por objetivo comparar estimativas de valores genéticos dos animais por meio de pedigree tradicional e o mesmo, porém associado a marcadores moleculares, verificando a similaridade genotípica e frequência gênica, além de descrever os passos para realizar essa análise. Foram gerados arquivos de pedigree tradicional (“pedigree.txt”) e de animais genotipados (“marcadores.txt”) com marcadores do tipo SNP para análises do BLUP tradicional e marcador, respectivamente, juntamente com um arquivo de fenótipos (“fenotipo.txt”) que também foi gerado. Para a seleção tradicional foi montada a matriz de parentesco, onde apresenta o valor genético médio dos animais, baseada no pedigree. Na seleção genômica obteve-se a matriz genômica, que apresentou o valor genético genômico dos animais, baseado nos marcadores moleculares do tipo SNP. Quando comparadas, a seleção genômica foi superior em 13% sobre a seleção tradicional, apresentando correlações de 0,95 e 0,83, respectivamente, com o valor genético verdadeiro. Os resultados deste estudo revelam um grande potencial da seleção genômica em aumentar a eficiência do melhoramento genético, uma vez que as predições genômicas aumentam a acurácia de seleção comparado com a seleção tradicional.

Palavras-chaves: **seleção tradicional, seleção genômica, BLUP, modelo misto.**

## **ABSTRACT**

### **METHODOLOGICAL DESCRIPTION OF THE USE OF MOLECULAR MARKERS LINKED TO PEDIGREE INFORMATION IN PREDICTING THE VALUE OF ANIMAL GENETIC BY USING THE SIMULATION METHODOLOGY BLUP.**

The present study aims to compare estimates of breeding values using traditional pedigree and the use of molecular markers in order to ascertain, the similarity checking genotypic and gene frequency, and describe the steps to perform such analysis. Files were generated from traditional pedigree ("pedigree.txt") and genotyped animals ("marcadores.txt") with the type SNP markers for analysis of traditional BLUP and marker, respectively, together with a phenotypes file ("fenotipo.txt ") that was also generated. For the traditional selection was assembled relationship matrix, which presents the average breeding value of animals, based on pedigree. In genomic selection to obtain the genomic matrix, that presents the genomic breeding value of animals, based on molecular markers SNP type. In comparison, the genomic selection was higher by 13% over traditional selection, with correlations of 0.95 and 0.83, respectively, with the true genetic value. The results of this study revealed a great potential of genomic selection in increasing the efficiency of breeding, since the genomic predictions increase the accuracy of selection compared to the traditional selection.

**Keywords: traditional selection, genomic selection, BLUP, mixed model.**

## SUMÁRIO

Tabela 1 – Pedigree .....	20
Tabela 2 – Animais genotipados para marcadores moleculares do tipo SNP .....	20
Tabela 3- Pedigree de marcadores moleculares do tipo SNP.....	21
Tabela 4- Valores dos marcadores .....	22
Tabela 5 – Fenótipo de característica não especificada .....	23
Tabela 6- Soluções do BLUP tradicional .....	28
Tabela 7- Soluções do BLUP genômico .....	32
Tabela 8- Valores genéticos estimados por métodos tradicionais e genômica usando a metodologia BLUP e valores genéticos verdadeiros.....	35
Tabela 9 - Correlação de <i>Spearman</i> entre os valores genéticos estimados por métodos tradicionais, genômico e valor genético verdadeiro dos indivíduos .....	37
Tabela 10- Correlação de <i>Pearson</i> entre os valores genéticos estimados por métodos tradicionais, genômico e valor genético verdadeiro dos indivíduos .....	38

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>9</b>
<b>2. REVISÃO BIBLIOGRÁFICA.....</b>	<b>10</b>
<b>2.1. Predição de Valores Genéticos pelo Modelo Animal.....</b>	<b>10</b>
<b>2.2. Matriz de Parentesco .....</b>	<b>13</b>
<b>2.3. Seleção Genômica .....</b>	<b>14</b>
<b>2.4. Marcadores Moleculares SNP (Single Nucleotide Polymorphism).....</b>	<b>16</b>
<b>3. MATERIAIS E MÉTODOS.....</b>	<b>19</b>
<b>3.1. Arquivos Gerados .....</b>	<b>19</b>
<b>3.2. Valores Genéticos Tradicionais .....</b>	<b>23</b>
<b>3.3. Valores Genéticos Genômicos.....</b>	<b>28</b>
<b>3.4. Matriz de Parentesco e Matriz G .....</b>	<b>32</b>
<b>4. RESULTADOS, DISCUSSÕES E IMPLICAÇÕES PRÁTICAS.....</b>	<b>35</b>
<b>5. REFERÊNCIAS .....</b>	<b>38</b>

## 1. INTRODUÇÃO

Uma das maiores conquistas no melhoramento animal foi a possibilidade de incorporação da inversa da matriz de parentesco (HENDERSON, 1976) nas equações de modelos mistos. O uso da matriz de parentesco deve melhorar a precisão das estimativas dos valores genéticos e quase sempre aumenta a correlação entre os valores genéticos dos indivíduos aparentados, dependendo da herdabilidade da característica estudada (HENDERSON, 1976). Diversos trabalhos comprovam que a inclusão da matriz de parentesco aditiva aumenta a acurácia das avaliações genéticas (KENNEDY e MOXLEY, 1975; CARLSON *et al.*, 1984).

Atualmente, com a detecção de milhares ou milhões de polimorfismo de base única (SNP), a seleção assistida por marcadores passou a ser uma realidade, dado o ganho em acurácia que os valores genéticos genômicos proporcionam, entre outras vantagens (HAYES *et al.*, 2009; VANRADEN *et al.*, 2009). Contudo, informações de pedigree continuam sendo muito importantes para avaliar a endogamia, o tamanho efetivo, a diversidade genética e diversos outros parâmetros populacionais importantes (FARIA *et al.*, 2001; FARIA *et al.*, 2002; MALHADO *et al.*, 2008; MARTÍNEZ *et al.*, 2008, *apud* CARNEIRO, 2009).

Nos arquivos de dados de animais os eventuais erros de pedigree prejudicam as avaliações genéticas, causando prejuízos no ganho genético. Além disso, quando se utiliza a tradicional matriz de parentesco, considera apenas uma proporção média de genes compartilhados entre os animais parentes, o que nem sempre é verdadeiro. Uma alternativa para a construção da matriz de parentesco aditiva de pedigrees é usar as informações de marcadores moleculares para inferir parentescos. Entretanto, a tentativa de fazer isso a partir de um número limitado de marcadores pode resultar em viés e estimativas imprecisas de parâmetros genéticos. No entanto, se os marcadores são suficientemente densos, essa abordagem poderia ser mais precisa do que usando as informações de pedigree coletadas no campo (tradicional matriz de parentesco). Assim, segundo Carneiro *et al.* (2006), recentemente, tem-se estudado a possibilidade da utilização de informações de similaridade genética, via marcadores moleculares, na aplicação da metodologia de modelos mistos para obtenção do BLUP.

Segundo Lôbo e Lôbo (2007), em diversas espécies, a seleção de animais com base no seu valor genético, tem sido praticada com sucesso. Exemplo marcante tem sido observado em bovinos de corte, entre outras espécies, em que os animais são selecionados para

características de importância econômica e são usados como reprodutores na geração seguinte, alterando-se a constituição genética do rebanho, obtendo-se assim um incremento na produtividade. Os mesmos autores ressaltam ainda que a avaliação e a seleção de indivíduos são realizadas, há muito tempo, com base em seus fenótipos, onde os caracteres de interesse, na sua grande maioria, controlados por muitos genes, apresentam baixa herdabilidade e são passíveis de alta influência ambiental. Diante disso, muitas das complicações da análise com base no fenótipo podem ser mitigadas através da identificação direta de genótipos, baseado em marcadores moleculares que segregam conjuntamente com os genes de interesse.

Visscher *et al.* (2000) afirmam que, com o uso da seleção assistida por marcadores, maiores ganhos genéticos seriam obtidos no melhoramento, quando comparados aos ganhos obtidos pelos métodos tradicionais. Entretanto, a seleção tradicional baseada no método BLUP por meio da mensuração dos fenótipos pode ser realizada simultaneamente, visando maximizar o ganho genético e evitando a perda de animais de alto mérito genético que possam ter sido eliminados por uma possível associação não-consistente de um marcador com um QTL. Assim, por causa desta associação não-consistente, a seleção assistida por marcadores utilizando-se alguns poucos marcadores, salvo algumas raras exceções, não foi muito eficiente na maioria dos programas de melhoramento animal. Por outro lado, com a detecção de milhares ou milhões de polimorfismo de base única (SNP), a seleção assistida por marcadores passou a ser uma realidade, dado o ganho em acurácia que os valores genéticos genômicos proporcionam, entre outras vantagens (HAYES *et al.* , 2009; VANRADEN *et al.*, 2009).

O objetivo deste trabalho foi comparar estimativas de valores genéticos verdadeiros dos animais com valores genéticos estimados por meio de um pedigree tradicional e também pela matriz genômica, baseada nos marcadores moleculares do tipo SNP, verificando a similaridade genotípica e frequência gênica, além de descrever os passos para realizar essa análise.

## **2. REVISÃO BIBLIOGRÁFICA**

### **2.1. Predição de Valores Genéticos pelo Modelo Animal**

O sucesso do melhoramento genético animal depende, essencialmente, da adoção de precisos métodos de seleção, os quais demandam a predição dos valores genéticos dos

animais candidatos à seleção (RESENDE e PEREZ, 1999). O procedimento de predição de valores genéticos denominado melhor predição linear não viciada (BLUP) foi desenvolvido por Henderson em 1949 e apresentado formalmente em 1973 (HENDERSON, 1973; 1984). Neste contexto, desenvolveu-se um algoritmo denominado de Equações dos Modelos Mistos que possibilita modelar simultaneamente os efeitos fixos (por exemplo, grupo de contemporâneos) e aleatórios (animal). Assim, pode-se obter estimativas para os efeitos fixos (BLUP) e simultaneamente a predição dos efeitos aleatórios, ou seja, os BLUP. Em um contexto de modelos mistos clássicos, com informação fenotípica, ou seja, as observações (medições) realizadas sobre determinada característica, como por exemplo, a produção de leite, peso vivo em uma determinada idade, as Equações dos Modelos Mistos podem ser explicadas pelo seguinte modelo matemático:

$$\tilde{y} = X\tilde{\beta} + Z\tilde{a} + e, \text{ em que:}$$

$\tilde{y}$  é o vetor dos dados observados nos vários (n) animais, de dimensão  $n \times 1$ ;

$\tilde{\beta}$  é o vetor contendo efeitos ambientais identificáveis ou efeitos fixos do modelo, como por exemplo, os efeitos de sexo, de grupo de manejo, de grupo contemporâneo, dentre outros. Este vetor tem dimensão  $p \times 1$ , onde  $p$  é o número de níveis de efeitos fixos;

$\tilde{a}$  é o vetor contendo os valores genéticos dos animais, de dimensão  $q \times 1$ ;

$e$  é o vetor de efeitos ambientais não identificáveis ou erros aleatórios, de dimensão  $n \times 1$ ;

$X$  e  $Z$  são matrizes de delineamento ou de incidência, contendo apenas valores 0 e 1, as quais associam as observações ( $\tilde{y}$ ) aos efeitos fixos ( $\tilde{\beta}$ ) e aos valores genéticos dos animais ( $\tilde{a}$ ), respectivamente.  $X$  tem dimensão  $(n \times p)$  e  $Z$  tem dimensão  $(n \times q)$ .

Por este modelo verifica-se que uma observação ( $\tilde{y}$ ) em determinado animal, pode ser desdobrada em termos de efeitos ambientais identificáveis ( $\tilde{\beta}$ ), efeitos genéticos ( $\tilde{a}$ ) e efeitos ambientais não identificáveis ou aleatórios ( $e$ ). O objetivo do BLUP é prever o valor genético ( $\tilde{a}$ ) dos animais, a partir de análises estatísticas dos valores medidos (fenotípicos) diretamente nos animais utilizando a matriz de parentesco tradicional, ou seja, considerando apenas uma proporção média de genes compartilhados entre os animais parentes (pedigree).

Para Resende e Perez (1999), a plena aplicação das equações dos modelos mistos, seja ela por meio do modelo animal no melhoramento de animais domésticos, torna-se dependente do conhecimento dos parâmetros genéticos associados aos caracteres de importância econômica, bem como a definição dos efeitos ambientais identificáveis (efeitos fixos) importantes a cada uma das espécies a serem melhoradas. Como exemplo, os mesmos autores utilizaram os caracteres economicamente importantes para o melhoramento genético do gado leiteiro destacam-se a produção de leite, a produção de gordura, a produção de proteína e a resistência à mastite (baseada na contagem de células somáticas ou CCS). Ainda no trabalho de Resende e Perez (1999), os parâmetros genéticos mais relevantes para a avaliação genética em gado leiteiro são a herdabilidade ( $h^2$ ) e a repetibilidade ( $r$ ) visto que os caracteres de interesse podem ser avaliados em várias lactações. As estimativas dos parâmetros a serem utilizados nas equações de modelo misto, podem ser obtidas pelo método da máxima verossimilhança restrita (REML), entre outros. Assim como dentre os efeitos ambientais identificáveis importantes em gado leiteiro, destacam-se: rebanho, ano de parto, estação de parto, frequência de ordenha, local ou região, grau de sangue e idade da vaca ao parto. Estes efeitos são os efeitos fixos a serem estimados no vetor  $\beta$ . De maneira geral, os efeitos de rebanho, ano e estação são agrupados em um único efeito (com vários níveis) denominado grupo de manejo (RESENDE e PEREZ, 1999).

Existem vários tipos de modelos que podem ser especificados para proceder as análises das equações dos modelos mistos. Dentre eles, pode-se destacar o modelo touro, o modelo touro-avô-materno, o modelo animal, o modelo animal reduzido, entre outros. A definição de qual modelo será utilizado para estimar os valores genéticos vai depender de vários fatores, como os recursos computacionais, do tipo e volume do arquivo de dados fenotípicos, entre outros. Segundo Gama et al. (2004), o modelo animal apresenta inúmeras vantagens em termos de avaliação genética, sendo uma das principais, o fato de, através da inclusão da matriz de parentesco, o registro de cada indivíduo contribuir para a avaliação de todos os seus parentes (com o peso adequado dado pela matriz de parentesco). Desta forma, podem-se obter estimativas dos valores genéticos de todos os indivíduos, mesmo aqueles que não possuem registros do fenótipo.

## 2.2. Matriz de Parentesco

A matriz de parentesco (A) é essencial na avaliação genética pelo BLUP, utilizando as equações do modelo misto 
$$\begin{vmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{vmatrix} \begin{vmatrix} \hat{\beta} \\ \hat{a} \end{vmatrix} = \begin{vmatrix} X'y \\ Z'y \end{vmatrix}$$
, onde  $\alpha$  é a razão entre a variância residual e a variância genética aditiva e  $A^{-1}$  é a inversa da matriz de parentesco, sendo que  $\hat{\beta}$  e  $\hat{a}$  são as soluções. Desta forma, utilizando-se a inversa da matriz de parentesco é permitido a obtenção de resultados mais acurados (pela consideração das informações de parentes) e não viciados (pela consideração da correlação entre os valores genéticos, na estimação dos efeitos fixos), levando em conta o progresso genético. Portanto, a incorporação da inversa da matriz de parentesco nas equações de modelos mistos é considerada uma das maiores conquistas no melhoramento animal (HENDERSON, 1976, apud RESENDE e PEREZ, 1999).

O uso dos coeficientes de parentesco e de endogamia tem sido frequente em análises genéticas das populações humanas e no gerenciamento de bancos de germoplasma associado aos programas de melhoramento de plantas e animais (PETERNELLI *et al.*, 2009). Uma característica inerente à estimação dos coeficientes de parentesco e de endogamia é a dependência sobre a informação do pedigree. Por exemplo, um coeficiente de endogamia não é uma característica intrínseca do indivíduo e sim uma informação genealógica dele mesmo (KHANG, 1989).

O uso da matriz de parentesco deve melhorar a precisão das estimativas dos valores genéticos e quase sempre aumenta a correlação entre os valores genéticos dos indivíduos aparentados (HENDERSON, 1976; DEMPFLER, 1990). Mais tarde, Quaas (1976) complementou o método com a inclusão da endogamia. Estes resultados tornaram esta metodologia computacionalmente aplicável a grandes conjuntos de dados, onde a robustez dos métodos de estimação de componentes de variância, baseados na verossimilhança, apoia-se, até certo ponto, na especificação completa e correta da matriz de parentesco.

De acordo com Hayes e Goddard (2007), programas de melhoramento de espécies animais dependem de estimativas precisas de parâmetros genéticos para predição dos reprodutores de valor, ou seja, para se ter um bom conhecimento do valor genético e proceder a seleção da melhor maneira possível. Em populações divergentes, a estimativa dos valores genéticos exige uma matriz que descreva a relação aditiva entre indivíduos na população (HENDERSON, 1984). Se a informação de pedigree foi coletada durante várias gerações, a matriz de parentesco aditiva pode ser construída a partir destas informações.

Diversos trabalhos comprovam que a inclusão da matriz de parentesco aditiva aumenta a acurácia das avaliações genéticas (KENNEDY e MOXLEY, 1975; POLLAK *et al.*, 1977; CARLSON *et al.*, 1984). Van Vleck & Hudson (1982), aplicando o método 3 de Henderson a um modelo de touro, demonstraram que a inclusão do parentesco entre os touros aumenta os valores das estimativas de herdabilidade. Assim, uma alternativa para a construção da matriz de parentesco aditiva de pedigrees sem considerar a proporção média de genes compartilhados entre os animais parentes, é utilizar as informações de marcadores moleculares do tipo SNP para inferir parentescos. Contudo, procede a discutir esta técnica sobre marcadores no item 2.3 (seleção genômica).

### **2.3. Seleção Genômica**

Em programas de melhoramento animal, a seleção assistida por marcadores utilizando-se alguns poucos marcadores, salvo algumas raras exceções, não foi muito eficiente, uma vez que, geralmente, as características de importância econômica são controladas por muitos pares de genes e, portanto, a informação destes poucos marcadores explica somente uma pequena parcela da variância genética total. Por outro lado, com a detecção de milhares ou milhões de polimorfismo de base única (SNP), a seleção assistida por marcadores passou a ser uma realidade, dado ao ganho em acurácia que os valores genéticos genômicos proporcionam, entre outras vantagens (HAYES *et al.*, 2009; VANRADEN *et al.*, 2009). Assim, a seleção assistida por marcadores em escala genômica, a qual é denominada seleção genômica (MEUWISSEN *et al.*, 2001), tem despertado grande interesse dos profissionais, pois permite a aceleração dos ganhos genéticos dos programas de melhoramento animal, que geralmente, praticam avaliações genéticas por meio de dados fenotípicos medidos a campo aliada a alta eficiência seletiva (RESENDE *et al.*, 2008). A seleção genômica é uma excelente ferramenta para ser aplicada, principalmente em caracteres de baixa herdabilidade, ao contrário da seleção assistida por marcadores, que não é útil para caracteres de baixa herdabilidade (MUIR, 2007, *apud* RESENDE, 2008). Esta metodologia se fundamenta no uso de marcadores do tipo SNP, os quais resultam da alteração de um único par de base, sendo esta a forma mais abundante de variação do DNA. Os SNP's podem ser usados para cobrir o genoma de um organismo, gerando marcas ou marcadores bem próximos uns dos outros, que estão no máximo a 1cM (um centimorgan) uns dos outros. Desta forma, foram geradas plataformas de DNA para a genotipagem de marcadores SNP de alta

densidade, denominadas de CHIP. Esta alta densidade significa uma plataforma que contém um número muito grande de marcadores, por exemplo, o CHIP de 54 mil marcadores (Bovine HapMap Consortium 2009). Muito embora os CHIP's de genotipagem disponíveis representam grandes avanços tecnológicos, os resultados dos primeiros estudos de aplicação dessas ferramentas em rebanhos bovinos de regiões de clima temperado demonstraram que a atual cobertura (1 marcador SNP a cada 50 mil pares de base) pode não ser suficiente para a implementação da seleção genômica em rebanhos bovinos brasileiros. Portanto, para o sucesso da seleção genômica, as abordagens desta metodologia devem ser estudadas a cada situação.

As principais abordagens para implementação da seleção genômica foram apresentadas por Meuwissen et al. (2001), e envolvem basicamente duas etapas, sendo que a primeira, envolve estimar os efeitos de cada marcador SNP distribuído pelo genoma do animal, e a segunda etapa, abrange a predição de valores genéticos (denominado de valor genético genômico) baseados nos genótipos dos marcadores e na estimativa de seus efeitos.

Atualmente este método tornou-se muito atrativo devido ao desenvolvimento dos marcadores tipo SNP, assim geneticistas e melhoristas renomados e adeptos de métodos tradicionais têm demonstrado e confirmado a superioridade e exequibilidade prática do método em benefício do melhoramento animal (SCHAEFFER, 2006; KOLBEHDARI *et al.*, 2007; MEUWISSEN, 2007; HAYES; GODDARD, 2007; LONG *et al.*, 2007; LEGARRA; MISZTAL, 2008, *apud* RESENDE et al, 2008) e vegetal (BERNARDO; YU, 2007, *apud* RESENDE *et al*, 2008). Esses trabalhos mostraram, definitivamente, que a seleção genômica terá grande utilidade no melhoramento genético, utilizando marcadores moleculares para obtenção da Melhor Predição Linear Não Viciada (BLUP) aplicando nos dados moleculares e permitindo a predição de valores genéticos genômicos.

Como mencionado anteriormente, a seleção genômica envolve um número bastante grande de marcadores do tipo SNP, os quais apresentam diferentes níveis de Desequilíbrio de Ligação com os QTL's ("Quantitative trait loci") que afetam o fenótipo de interesse (característica). Assim, se estima as associações de um grande número de marcadores SNP's em todo o genoma com os fenótipos, capitalizando esta associação no desequilíbrio de ligação entre os marcadores e QTL's proximamente ligados. Tudo isto, com base nas significâncias dos marcadores associados com o fenótipo. A princípio, sabe-se que os marcadores genéticos são herdados em blocos entre indivíduos aparentados. Esses blocos herdados, ou trechos de cromossomos idênticos por descendência, ou ainda haplótipos, são evidências da ligação genética que se propaga entre indivíduos de mesma família e de um ancestral comum.

Portanto, as predições são então obtidas para os efeitos dos haplótipos marcadores ou dos alelos em cada marcador. Essas predições derivadas de dados fenotípicos e de genótipos SNP's em alta densidade em uma geração são então usadas para obtenção dos valores genéticos genômicos dos indivíduos de qualquer geração subsequente, tendo por base os seus próprios genótipos marcadores (RESENDE *et al.*, 2008).

#### 2.4. Marcadores Moleculares SNP

Como sugerido pela sigla, um marcador SNP (polimorfismos de nucleotídeos simples) é apenas uma mudança de uma única base em uma sequência de DNA, comum a alternativa usual de dois nucleotídeos possíveis em uma determinada posição. Para este tipo de posição, para que uma variação seja considerada SNP, essa deve ocorrer em pelo menos 1% da população. Em princípio, em cada posição de um estiramento de sequência, qualquer uma das quatro possíveis bases nucleotídicas podem estar presentes. Os SNP's são a forma mais abundante de variação do DNA em genomas e são preferidos em relação a outros marcadores genéticos devido à sua baixa taxa de mutação e facilidade de genotipagem. Milhares de SNP's podem ser usados para cobrir o genoma de um organismo com marcadores que não estão a mais de 1 cM um do outro no genoma inteiro (RESENDE *et al.*, 2008). Os marcadores SNP tem como base as alterações mais elementares da molécula de DNA, ou seja, mutações em bases únicas da cadeia de bases nitrogenadas (Adenina (A), Citosina (C), Timina (T) e Guanina (G)). As mutações mais comuns são as transições, onde ocorrem trocas de uma purina por outra purina ( $A \leftrightarrow G$ ) ou de uma pirimidina por outra pirimidina ( $C \leftrightarrow T$ ). Menos frequentes, as transversões ocorrem quando há troca de uma purina por uma pirimidina, ou vice-versa ( $C/T \leftrightarrow A/G$ ) (CAETANO, 2009). Normalmente, os marcadores SNP's são bi-alélicos, ou seja, geralmente são encontrados apenas dois variantes em uma espécie (Ex: um alelo corresponde a um par de bases A/T e o outro a um G/C). Os SNP's podem ocorrer em regiões codificadoras ou com função regulatória, porém, na maior parte das vezes são encontrados em espaços intergênicos, sem função determinada.

Avanços tecnológicos recentes trouxeram metodologias de alto desempenho e acurácia, e um menor custo e mão-de-obra para prospecção, caracterização e genotipagem de marcadores SNP. Essas tecnologias trouxeram novas soluções para aplicações já solidificadas e também estão permitindo o desenvolvimento de novas aplicações, produtos e processos, com uso direto nas cadeias produtivas da pecuária mundial (CAETANO, 2009).

Segundo Carvalho (2009), as principais vantagens dos marcadores moleculares é o grande polimorfismo, o fato de não sofrerem influências do meio ambiente e de poderem ser analisados em qualquer estágio de desenvolvimento do indivíduo.

A utilização de marcadores moleculares SNP em estudos de associação e mapeamento genético, assim como em ensaios diagnósticos para confirmação de paternidade, identificação individual (rastreabilidade), detecção de doenças genéticas e/ou polimorfismos associados a características de produção, esteve limitada por muito tempo devido à restrições tecnológicas. Esse quadro mudou radicalmente nos últimos anos com a geração de novas tecnologias para descoberta e genotipagem em massa de SNP's (CAETANO, 2009).

O método mais básico para genotipagem de SNP's em regiões específicas do genoma é baseado na técnica de PCR-RFLP (MAEDA *et al.*, 1989). O fragmento contendo o SNP de interesse é amplificado com iniciadores específicos, purificado e submetido a um tratamento com uma enzima de restrição que reconheça apenas um dos alelos. Posteriormente, os fragmentos são separados por eletroforese para diferenciação dos alelos por tamanho.

Outro avanço tecnológico para genotipagem de SNP's surgiu com metodologias que utilizam sequenciadores automáticos para a realização da eletroforese de produtos de PCR marcados com corantes fluorescentes. Após a amplificação do fragmento desejado, realiza-se uma mineração do sequenciamento com a finalidade de se determinar o genótipo na base desejada (CAETANO, 2009). A metodologia permite a genotipagem de mais de um SNP em um mesmo fragmento, ou em fragmentos diferentes. Em termos práticos, é possível genotipar até 10 SNP's em um mesmo ensaio.

O desenvolvimento de metodologias para genotipar dezenas de milhares ou até um milhão de SNP's em um único ensaio trouxe novos benefícios e aplicações, levando a menores custos de geração de dados, contudo, quando se estuda rebanhos comerciais, os custos ainda são um entrave. CHIP's de genotipagem de alta densidade já foram gerados e validados para humanos, bovinos, ovinos, equinos, suínos e caninos, contendo em média mais de 50 mil SNP's. Em algumas espécies, novos CHIP's, contendo maiores números de SNP's já estão em desenvolvimento, refletindo o uso extenso que a tecnologia alcançou e as necessidades de aplicações específicas de cada setor. Atualmente, para bovinos, um CHIP de genotipagem de SNP's contendo aproximadamente 777, 54 e 9 mil marcadores custa em torno de US\$200, US\$120 e US\$30, respectivamente.

Metodologias para testes de exclusão de paternidade e identificação individual também foram desenvolvidas com marcadores SNP (HEATON *et al.*, 2002). Adicionalmente, painéis de baixa densidade foram gerados para rastreabilidade de amostras (HEATON *et al.*,

2005) e também para diagnosticar polimorfismos em genes associados a características de produção (CHESSA *et al.*, 2007). A medida que se reduz o custo do sequenciamento de nova geração de genomas (MARDIS *et al.*, 2008), estima-se que essa tecnologia possa substituir os CHIP's de DNA para uso nos estudos de associação global e de seleção genômica. Assim, haverá a disponibilização de um número imenso de genótipos de marcadores (de 9 a 12 milhões, em bovinos) de cada animal para seleção genômica a prospecção de genes associados a características de importância econômica. Porém, essas técnicas são economicamente inviáveis nos dias de hoje, principalmente em se tratando de estudos com rebanhos compostos por milhares de animais. Por outro lado, a genotipagem de somente um modesto número de marcadores SNP pode ser realizada e os genótipos faltantes podem ser imputados por técnicas computacionais. Esta é uma estratégia para reduzir os custos da genotipagem para a aplicação da seleção genômica, pois se pode sequenciar um pequeno número de indivíduos fundadores do rebanho, isto é, animais que mais contribuíram para a maioria dos genes que estão presentes na população e, em seguida genotipar os animais mais jovens com um CHIP de SNP, que contem poucos marcadores, para depois usar a técnica denominada de imputação de genótipos, no intuito de se obter os genótipos perdidos desses indivíduos a partir dos dados do sequenciamento dos animais fundadores (GODDARD & HAYES, 2009).

A técnica de imputação de genótipos é um procedimento não recente, no entanto com os estudos moleculares torna-se muito útil e está sendo largamente estudada na atualidade, sendo que esta técnica é uma ferramenta importante para os Estudos de Associação de Genômica Completa (GWAS ou Genome Wide Association Studies) (MOORE *et al.*, 2010) à fenótipos, doenças, performance de animais e, de uma forma geral, à seleção genômica (HABIER *et al.*, 2009). Esta técnica, como mencionado anteriormente, serve basicamente para estimar dados faltantes, por meio de sequenciadores de DNA de nova geração, e por meio de CHIP's de alta densidade de marcadores genéticos (geralmente bialélicos do tipo SNP, ou polimorfismos de base única), que permitem a genotipagem de centenas de milhares a até milhões de marcadores por animal. Geneticistas podem dessa forma, avaliar a associação de marcadores genéticos não genotipados, seja por problemas técnicos ou para combinar estudos que usam plataformas com diferentes densidades de marcadores (LI *et al.*, 2009). A imputação de genótipos é uma ferramenta que tem custo-benefício interessante para o estudo genético de rebanhos numerosos e principalmente para a implementação da seleção genômica.

A imputação de genótipos faltantes parte do princípio de que marcadores genéticos são herdados em blocos entre indivíduos relacionados. Esses blocos herdados, ou trechos de

cromossomos idênticos por descendência, ou ainda haplótipos, são evidências da ligação genética que se propaga entre indivíduos de mesma família e de um ancestral comum. Mesmo em indivíduos não-relacionados, essa técnica também pode ser usada, levando em consideração que os blocos herdados serão bem menores, devido à ancestrais comuns mais distantes (LI *et al.*, 2009). Ainda, em populações de animais comerciais a seleção artificial limita o número de pais o que acarreta ancestrais comuns em até poucas gerações (DAETWYLER *et al.*, 2011).

Grandes avanços devem ser incorporados nos programas de avaliação e melhoramento genético devido a geração de novas aplicações de CHIP's de genotipagem de alta densidade (CAETANO, 2009). A alta cobertura do genoma com marcadores (por exemplo: 1 marcador a cada 50.000 pares de bases) torna possível particionar grande parte da variabilidade genética aditiva de uma característica por todo genoma, permitindo estimar o valor de substituição do alelo em cada um dos loci envolvidos com a característica. Esse procedimento permite então estimar o valor genético de um indivíduo com base nos genótipos de todos os marcadores associados com a característica, ou seja, o valor genético genômico, sem a coleta da medida fenotípica do indivíduo (MEUWISSEN *et al.*, 2001). Desta forma, o valor genético genômico é estimado de uma forma mais objetiva comparada com o método tradicional, onde se tem uma estimativa do valor genético pela média dos genes em comum entre animais.

### **3. MATERIAIS E MÉTODOS**

#### **3.1. Arquivos Gerados**

Para o presente trabalho, foram determinados os seguintes arquivos:

- 1) Primeiramente foi criado um arquivo de pedigree, onde observa-se a genealogia dos animais em ordem cronológica e nomeado de “pedigree.txt” (Tabela 1). O animal, ou pai, ou mãe codificado como “0” (zero), significa ser desconhecido. Neste exemplo, o animal 5 é filho do animal 4 (pai) com uma mãe desconhecida.

Tabela 1 – Pedigree

<b>Animal</b>	<b>Pai</b>	<b>Mãe</b>
1	0	0
2	0	0
3	1	2
4	1	3
5	4	0
6	0	5
7	1	5
8	0	2

- 2) Na sequência, simulou-se um arquivo de marcadores para os animais os quais foram supostamente genotipados (Tabela 2), já codificados conforme o exemplo abaixo, onde recebeu o nome de “marcadores.gen”.

Tabela 2 – Animais genotipados para marcadores moleculares do tipo SNP

<b>Animais</b> i . j	<b>Genótipo</b>							
	<b>Marcador SNP (M)</b>							
	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>	<b>M5</b>	<b>M6</b>	<b>M7</b>	<b>M8</b>
1	AA	BB	AA	AB	AB	AA	BB	AA
2	AB	AB	BB	AA	BB	AB	AB	AA
3	AA	AB	AB	AA	BB	AA	AB	AA
4	AA	BB	AB	AB	BB	AA	BB	AA
5	AB	AB	BB	AA	AB	AB	AB	AA
6	AB	BB	AB	AA	BB	AA	BB	AB
7	AA	BB	AB	AB	AA	AB	BB	AA
8	AB	BB	AB	AA	AB	BB	AA	AB

Fonte: A autora

No entanto, deve-se codificar os marcadores moleculares como 0, ou 1, ou 2, para depois construir uma matriz  $X_{ij}$ , sendo  $i$  para animais e  $j$  para o marcador SNP, sendo que neste exemplo foram inferidos oito (8) marcadores (M1, ... , M8) e para todos os animais (8, neste exemplo).

$$X_{ij} = \begin{cases} 0 \text{ é homozigoto AA} \\ 1 \text{ é heterozigoto AB} \\ 2 \text{ é homozigoto BB} \end{cases}$$

Assim, o arquivo “marcadores.gen” (Tabela 3) de genótipos torna-se:

Tabela 3- Pedigree de marcadores moleculares do tipo SNP

<b>Animais</b>	<b>SNP's</b>
1	02011020
2	11202110
3	01102010
4	02112020
5	11201110
6	12102021
7	02110120
8	12101201

Fonte: A autora

- 3) No terceiro passo, simular-se-á um arquivo de fenótipo (desempenho), onde se utiliza a equação:

$$y_i = \mu_i + \sum_{j=1}^p m_{ij}g_j + \varepsilon_i,$$

onde  $\mu_i$ , representa os efeitos ambientais sistemáticos associados à observação  $y_i$  ou simplesmente uma média geral;  $m_{ij}$  representa o genótipo do indivíduo  $i$  no marcador  $j$  ( $j= 1,2, \dots , p$ );  $g_j$  é o efeito genético associado ao marcador  $j$  (marcador do tipo SNP); e  $\varepsilon_i$  é um termo residual de distribuição normal. Estes cálculos foram realizados

no programa excel (Microsoft Office 2010), estimando um valor simbólico para cada marcador, com média cinco ( $\mu_i=5$ ) e também um erro aleatório ( $\varepsilon_i$ ) de média zero e três desvios-padrão estimado para cada animal por uma função do excel (=INV.NORMAL(ALEATÓRIO);0;3). Os valores estimados para os marcadores, que recebe designação g estão na Tabela 4, neste exemplo abrange oito marcadores, logo g varia entre g1 e g8.

Tabela 4- Valores dos marcadores

<b>g1</b>	<b>g2</b>	<b>g3</b>	<b>g4</b>	<b>g5</b>	<b>g6</b>	<b>g7</b>	<b>g8</b>
0,001	-2	-0,001	0,05	3	-0,9	0,405	2

Como exemplo, utilizou-se o animal 1 que apresenta genótipo 02011020

$$y_i = 5 + ((0 \times 0,001) + (2 \times (-2)) + (0 \times (-0,001)) + (1 \times 0,05) + (1 \times 3) + (0 \times (-0,9)) + (2 \times 0,405) + (0 \times 2)) + (-1,91267)$$

$y_i = 2,9473332977$ ; e assim sucessivamente para os 8 animais. Para saber o valor genético verdadeiro, basta retirar a média ( $\mu_i=5$ ) e o erro ( $\varepsilon_i$ , neste exemplo = -1,91267).

Depois de calculado o valor fenotípico e o genético verdadeiro, se calcula a variância fenotípica (var\_fen) e a variância genotípica (var\_gen), sendo que a variância residual (var\_res) é:  $var\_res = var\_fen - var\_gen$ . Depois emprega-se estas variâncias para estimar os valores genéticos por meio das equações dos modelos mistos.

Os resultados fenotípicos geram o arquivo denominado de “fenotipo.txt” (Tabela 5), que contém três colunas (animal, um efeito da incidência da média que será 1(um) para todos os animais e valor fenotípico simulado). Assim, abrange o arquivo “fenotipo.txt”:

Tabela 5 – Fenótipo de característica não especificada

<b>Animal</b>	<b>Média</b>	<b>Fenótipo</b>
1	1	2,9473332977
2	1	7,0043310521
3	1	10,8523813691
4	1	9,2721251392
5	1	-1,4715132334
6	1	10,3607010922
7	1	-0,3564391925
8	1	2,8208193096

Fonte: A autora

Estes arquivos foram gerados para estimar os valores genéticos tradicionais e os valores genéticos genômicos pelo método da Máxima Verossimilhança Restrita, com uso do modelo animal. Foram utilizados os programas, “renumf90”, “preGSf90” e o “blupf90”, da família BLUP90 desenvolvido por Misztal et al. (2002), que foram escritos em linguagem FORTRAN 90. Além desses programas foram feitas análises estatísticas descritivas e de consistência utilizando-se o programa R (R a language..., 2006), que é um “software” livre com uma linguagem dirigida à análise gráfica e estatística. Depois de estimados os valores genéticos tradicionais e os valores genéticos genômicos, calcularam-se as correlações de *Pearson* e *Spearman* entre os valores genéticos tradicionais e os valores genéticos genômicos, entre valores genéticos tradicionais e os valores genéticos verdadeiros e entre valores genéticos genômicos e os valores genéticos verdadeiros.

### 3.2. Valores Genéticos Tradicionais

Para estimar os valores genéticos tradicionais, utilizam-se os arquivos: “pedigree.txt”, “fenotipo.txt” e também os programas executáveis “renumf90.exe” para preparar os arquivos e o “bluf90.exe” para nos mostrar as soluções dos valores genéticos tradicionais. Assim, segue os seguintes passos:

Fazer um arquivo de parâmetro para rodar o renumf90.exe. Este arquivo chama-se “parametros.txt”, que descreve o nosso modelo e prepara os arquivos para posteriormente rodar o “remlf90”.

- 1) Para rodar renumf90.exe pelo “Prompt de comando”, tem-se que ir ao diretório pelo comando “cd .. diretório”; depois é chamado o programa “renumf90” e se executa pelo arquivo denominado “parametros.txt”. Assim, o programa “renumf90” irá gerar outros novos arquivos como “renadd02.ped”, que é arquivo de pedigree, o “renf90.par” é arquivo de parâmetros do modelo, o “renf90.dat” é arquivo dos dados e o “renf90.tables”, que é um arquivo de tabelas.
- 2) Depois de rodar o renumf90.exe pelo “Prompt de comando”, tem-se que seguir no diretório e rodar o remlf90.exe, pelo novo arquivo de parâmetros denominado “renf90.par”. Assim, o programa “remlf90” irá gerar outros arquivos, entre eles o “solutions”, que é um arquivo das soluções do modelo, ou seja, os valores genéticos tradicionais e o efeito da média (neste exemplo).

A matriz de parentesco (A) é simétrica e seus elementos da diagonal ( $a_{ii}$ ) para cada indivíduo  $i$  é igual a  $(1+Fi)$ , onde  $Fi$  é o coeficiente de endogamia do indivíduo  $i$ . Os elementos fora ( $a_{ij}$ ) da diagonal referem-se ao coeficiente de parentesco entre os indivíduos  $i$  e  $j$ . O coeficiente de parentesco de Wright também pode ser definido como a correlação entre os valores genéticos de dois indivíduos devido exclusivamente ao parentesco. Pode-se esperar que, como indivíduos parentes têm genes e, portanto efeitos independentes daqueles genes em comum, eles devem possuir valores genéticos que são, em média, semelhantes, ou pelo menos mais semelhantes que valores genéticos de indivíduos não aparentados. O coeficiente de parentesco de Wright mede esta semelhança como uma correlação.

Para composição da matriz de parentesco (A), Henderson (1976) citado por Resende e Perez (1999), descreve os seguintes passos:

- Codificar os animais de 1 a N, ordenados de forma que os parentais precedam suas progênes;

- Se ambos os parentais (s do pai e d da mãe) do indivíduo  $i$  são conhecidos, fazer:

$$a_{ii} = 1 + 0,5 (a_{sd})$$

$$a_{ji} = a_{ij} = 0,5 (a_{js} + a_{jd}), j = 1 \text{ a } (i-1);$$

- Se somente um parental (s) é conhecido e supostamente não aparentado com a fêmea com que cruzou, fazer:

$$a_{ii} = 1$$

$$a_{ji} = a_{ij} = 0,5 (a_{js}), j = 1 \text{ a } (i-1)$$

- Se ambos os parentais são desconhecidos e supostamente não aparentados, fazer:

$$a_{ii} = 1$$

$$a_{ji} = a_{ij} = 0, j = 1 \text{ a } (i-1)$$

Foi considerado como pedigree o arquivo “pedigree.txt” (citado no passo 1 do item 3.1. Arquivos Gerados), onde se obtêm a seguinte matriz de parentesco:

$$A^{-1} = \begin{bmatrix} 1 & 0 & -0,5 & 0,75 & 0,375 & 0,18 & 0,68 & 0 \\ & 1 & 0,5 & 0,25 & 0,125 & 0,062 & 0,0625 & 0,5 \\ & & 1 & 0,75 & 0,375 & 0,187 & 0,437 & 0,250 \\ & & & 1,25 & 0,625 & 0,312 & 0,687 & 0,125 \\ & & & & 1 & 0,5 & 0,687 & 0,0625 \\ & \text{Simétrica} & & & & 1 & 0,343 & 0,031 \\ & & & & & & 1,187 & 0,031 \\ & & & & & & & 1 \end{bmatrix}$$

Os elementos foram obtidos da seguinte forma:

$$\begin{aligned} a_{11} &= 1 & a_{12} &= 0 \\ a_{22} &= 1 & a_{13} &= 0,5(a_{js} + a_{jd}) = 0,5 (a_{11} + a_{12}) = 0,5 \\ a_{33} &= 1 & a_{14} &= 0,5 (a_{js} + a_{jd}) = 0,5 (a_{11} + a_{13}) = 0,5(1+0,5) = 0,75 \\ a_{44} &= 1 & a_{15} &= 0,5 (a_{js}) = 0,5 (a_{14}) = 0,5 (0,75) = 0,375 \\ a_{55} &= 1 + 0 = 1 & a_{16} &= 0,5 (a_{15}) = 0,5 (0,375) = 0,1875 \\ a_{66} &= 1 + 0 = 1 & a_{17} &= 0,5 (a_{11} + a_{15}) = 0,5 (1+0,375) = 0,6875 \\ a_{77} &= 1 + 0 = 1 & a_{18} &= 0,5 (a_{12}) = 0,5 (0) = 0 \\ a_{88} &= 1 + 0 = 1 & & \\ \\ a_{23} &= 0,5(a_{21}+a_{21})=0,5(1+0) = 0,5 & a_{34} &= 0,5 (a_{31}+a_{33}) = 0,75 \\ a_{24} &= 0,5(a_{21}+a_{23}) = 0,5(0+0,5) = 0,25 & a_{35} &= 0,5 (a_{34}) = 0,375 \\ a_{25} &= 0,5 (a_{24}) = 0,5 (0,25) = 0,125 & a_{36} &= 0,5 (a_{35}) = 0,1875 \\ a_{26} &= 0,5 (a_{25}) = 0,5 (0,125) = 0,0625 & a_{37} &= 0,5 (a_{31} + a_{35}) = 0,4375 \\ a_{27} &= 0,5 (a_{21}+a_{25}) = 0,5 (0+0,125) = 0,0625 & a_{38} &= 0,5 (a_{32}) = 0,25 \\ a_{28} &= 0,5 (a_{22}) = 0,5 (1) = 0,5 & & \\ \\ a_{45} &= 0,5 (a_{43}+a_{42}) = 0,5 & a_{56} &= 0,5 (a_{55}) = 0,5 \\ a_{46} &= 0,5 (a_{45}) = 0,25 & a_{57} &= 0,5 (a_{51} + a_{55}) = 0,6875 \\ a_{47} &= 0,5 (a_{41}+a_{45}) = 0,1875 & a_{58} &= 0,5 (a_{52}) = 0,0625 \end{aligned}$$

$$a_{48} = 0,5 (a_{42}) = 0,5 (0,25) = 0,125$$

$$a_{67} = 0,5 (a_{61} + a_{65}) = 0,34375$$

$$a_{78} = 0,5 (a_{72}) = 0,03125$$

$$a_{68} = 0,5 (0,0625) = 0,03125$$

Como a matriz A é de elevada ordem (N×N), Henderson (1976) estabeleceu regras para se escrever diretamente  $A^{-1}$ , sem necessidade de inversão da A. Com parentais não endogâmicos, os seguintes passos devem ser adotados, partindo-se de uma matriz N×N, completamente zerada:

- Codificar os indivíduos de 1 a N, ordenados de forma que os parentais precedam suas progênes;

- Se ambos os parentais (s e d) do indivíduo são conhecidos adicionar:

2 ao elemento  $a^{ii}$ ;

-1 aos elementos  $a^{is}$ ,  $a^{si}$ ,  $a^{id}$  e  $a^{di}$ ;

1/2 aos elementos  $a^{ss}$ ,  $a^{sd}$ ,  $a^{ds}$  e  $a^{dd}$ ;

- Se apenas um dos parentais (s) do indivíduo i, for conhecido, adicionar:

4/3 ao elemento  $a^{ii}$ ;

-2/3 aos elementos  $a^{is}$  e  $a^{si}$ ;

1/3 ao elemento  $a^{ss}$ .

- Se ambos parentais do indivíduo i forem desconhecidos, adicionar 1 ao elemento  $a^{ii}$ .

Desta forma, tem-se então a inversa da matriz A para o presente exemplo:

$$A^{-1} = \begin{bmatrix} 2,5 & 0,5 & -0,5 & -1 & 0,5 & -1,0096 & -1 & 0 \\ & 1,8333 & -1 & 0 & 0 & 0 & 0 & -0,6666 \\ & & 2,5 & -1 & 0 & 0 & 0 & 0 \\ & & & 2,3636 & -0,7272 & 0 & 0 & 0 \\ & & & & 2,2878 & -0,666 & -1 & 0 \\ & & & & & 1,3333 & 0 & 0 \\ & & & & & & 2 & 0 \\ & & & & & & & 1,3333 \end{bmatrix}$$

Simétrica

Para exemplo de modelo animal, foi considerada a avaliação dos 8 animais do arquivo “pedigree.txt” (sem espécie especificada), com a observação fenotípica (mensuração aleatória

de uma característica qualquer, não específica chamada de fenótipo) do arquivo “fenotipo.txt” (tabela 4) descrito no passo 3 do item 3.1 Arquivos Gerados.

Utilizando-se as equações do modelo misto 
$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$
, onde  $\alpha$  é a

razão entre a variância residual e a variância genética aditiva, ou seja,  $\sigma_e^2 / \sigma_a^2 = (1-h^2)/h^2$ , e resolvendo estas equações pelo modelo matemático descrito no item 2.1. (Predição de Valores Genéticos pelo Modelo Animal), as pressuposições acerca da distribuição de  $y$ ,  $a$  e  $e$  podem ser descritas como:

$$\begin{bmatrix} y \\ \sim \\ a \\ \sim \\ e \\ \sim \end{bmatrix} \sim \left\{ \begin{bmatrix} X\beta \\ \sim \\ 0 \\ \sim \\ 0 \end{bmatrix}, \begin{bmatrix} ZGZ' + R & ZG & R \\ \sim & G & \Phi \\ \sim & R & \Phi & R \end{bmatrix} \right\},$$

em que  $G$  é a matriz de variâncias e covariâncias dos efeitos aleatórios do vetor  $a$ ;  $R$  é a matriz de variâncias e covariâncias residuais.

As matrizes  $G$  e  $R$  são descritas como:  $G = A \otimes G_0$ , em que  $A$  é a matriz que indica o grau de parentesco entre os indivíduos;  $G_0$  é a matriz de variâncias e covariâncias genéticas aditivas entre as características que compõem as observações;  $\otimes$  é o operador produto direto;  $R = I \otimes R_0$ , em que  $I$  é a matriz identidade de ordem igual à dimensão linha de  $y$ ;  $R_0$  é a matriz de variâncias e covariâncias residuais entre as características que compõem as observações;  $0$  é o vetor nulo e  $\Phi$  é a matriz nula.

Desta forma, encontraram-se as seguintes soluções (tabela 6):

Tabela 6- Soluções do BLUP tradicional

Nível	Soluções
Média ( $\hat{\beta}$ )	
1	5,1166
Animais (Valores Genéticos = $\hat{a}$ )	
1	-1,7029
2	1,8100
3	-0,3413
4	-0,4593
5	1,1469
6	0,9779
7	1,2239
8	-2,1588

Fonte: A autora

As soluções para as equações de modelo misto foram obtidas através do BLUP tradicional. Observa-se que neste caso,  $\hat{\beta}$  seria apenas a solução para a média e com apenas 1 (um) nível. Para os valores genéticos ( $\hat{a}$ ), como o modelo é o animal e neste exemplo apresenta oito (8) animais, a solução tem oito (8) níveis, ou seja, um valor para cada animal. A metodologia de modelos mistos para obtenção do BLUP utiliza a matriz do numerador do coeficiente de parentesco, que inclui todas as informações de parentes disponíveis. O uso da matriz completa do numerador do coeficiente de parentesco faz com que a seleção baseada nos valores genéticos preditos pelo BLUP apresente superioridade sobre outros métodos de seleção, especialmente quanto a características de baixa herdabilidade, em que o BLUP enfatiza as informações de parentes, em detrimento das informações do próprio indivíduo.

### 3.3. Valores Genéticos Genômicos

Para a estimação dos valores genéticos genômicos pelos marcadores moleculares tipo SNP utilizou-se os arquivos: “pedigree.txt”, “marcadores.gen” e “fenótipo.txt” (todos gerados no item 3.1 Arquivos Gerados). Neste caso, foram usados os programas executáveis “renumf90” para preparar os arquivos, o “preGSf90” para preparar as matrizes genômicas e o

“blupf90” para estimar as soluções dos valores genéticos genômicos. Assim, seguem-se os seguintes passos:

- 1) Fazer um arquivo de parâmetro para rodar renumf90.exe onde será chamado de arquivo de “parametros.txt”, que descreve o modelo e prepara os arquivos para posteriormente rodar o preGSf90.exe. Assim como se fez no passo 1 dos Valores Genéticos Tradicionais (3.2), e acrescentando apenas a opção “SNP\_FILE marcadores.gen” logo após a opção “FILE pedigree.txt”, tudo isto para acrescentar o arquivo de genótipos, além de poder usar a função “OPTION” para salvar arquivos desejados.
- 2) Depois de executado o renumf90.exe pelo “Prompt de comando”, foi executado o preGSf90.exe por meio do “renf90.par”, que é um novo arquivo de parâmetros do modelo. Tudo isto para gerar um nova matriz G (genômica).
- 3) Depois de rodar o preGSf90.exe pelo “Prompt de comando”, necessita seguir no diretório e rodar o blupf90.exe, pelo novo arquivo de parâmetros, também denominado “renf90.par”. Assim, o programa “blupf90” irá gerar outros arquivos, entre eles o “solutions”, que é um arquivo das soluções do modelo, ou seja, os valores genéticos genômicos.

A seguir, será mostrado um exemplo para se obter uma estimativa mais real da porcentagem de alelos em comum, entre os indivíduos de uma população. Foram utilizados marcadores moleculares na construção de uma matriz de similaridade genética, para substituição da matriz (G) estimada pelo método tradicional por uma nova matriz genômica ( $G^*$ ), por meio de uma nova matriz  $Z^*$ , através dos seguintes passos:

- Codificar os animais de 1 a N, ordenados de forma que os parentais precedam suas progênes (utilização dos mesmos 8 animais da matriz de parentesco tradicional) como mostrado no passo 1 do item 3.1 Arquivos Gerados;

- Codificar os marcadores moleculares de 1 a N, no caso 8 marcadores SNP com elementos do arquivo “marcadores.gen” (0 ou 1 ou 2) do passo 2 do item 3.1 Arquivos Gerados, construindo uma matriz  $X_{ij}$ : i para animais e j para SNP, sendo que neste exemplo apresenta-se oito (8) SNP e para todos os animais (8).

Assim, temos a matriz X:

$$X = \begin{vmatrix} AA & \cdots & BA \\ \vdots & \ddots & \vdots \\ BB & \cdots & AB \end{vmatrix}$$



Calculada a matriz P, obtem-se a nova matriz  $Z^*$ , pois  $Z^*_{ij} = X_{ij} - 2p_j$ , assim, a matriz

$Z^*$  fica:

$$Z^* = \begin{bmatrix} x_{11} - 2p_1 & \cdots & x_{1j} - 2p_j \\ \vdots & \ddots & \vdots \\ x_{ij} - 2p_1 & \cdots & x_{ij} - 2p_j \end{bmatrix}$$

Neste exemplo, a matriz  $Z^*$  fica:

$$Z^* = \begin{bmatrix} -0,5 & 0,375 & -1,125 & 0,625 & -0,375 & -0,625 & 0,625 & -0,25 \\ 0,5 & -0,625 & 0,875 & -0,375 & 0,625 & 0,375 & -0,375 & -0,25 \\ -0,5 & -0,625 & -0,125 & -0,375 & 0,625 & -0,625 & -0,375 & -0,25 \\ -0,5 & 0,375 & -0,125 & 0,625 & 0,625 & -0,625 & 0,625 & -0,25 \\ 0,5 & -0,625 & 0,875 & -0,375 & -0,375 & 0,375 & -0,375 & -0,25 \\ 0,5 & 0,375 & -0,125 & -0,375 & 0,625 & -0,625 & 0,625 & 0,75 \\ -0,5 & 0,375 & -0,125 & 0,625 & -1,375 & 0,375 & 0,625 & -0,25 \\ 0,5 & 0,375 & -0,125 & -0,375 & -0,375 & 1,375 & -1,375 & 0,75 \end{bmatrix}$$

Depois de computada a matriz  $Z^*$ , a matriz genômica  $G^*$  pode ser obtida por três métodos. Neste trabalho será usado o método proposto por VanRaden (2008), em que

$G^* = \frac{ZZ^*}{2\sum_1^j p_j(1-p_j)}$ . Assim, a matriz  $G^*$  neste exemplo, fica:

$$G^* = \begin{bmatrix} 1,0157 & -0,7853 & -0,0314 & 0,5130 & -0,6596 & 0,0523 & 0,5549 & -0,6596 \\ -0,7853 & 0,7643 & 0,1780 & -0,2827 & 0,5549 & -0,0732 & -0,5759 & 0,2198 \\ -0,0314 & 0,1780 & 0,5968 & 0,1361 & -0,0314 & 0,0104 & -0,4921 & -0,3664 \\ 0,5130 & -0,2827 & 0,1361 & 0,6806 & -0,4921 & 0,2198 & 0,0523 & -0,8272 \\ -0,6596 & 0,5549 & -0,0314 & -0,4921 & 0,6806 & -0,2827 & -0,1151 & 0,3455 \\ 0,0523 & -0,0732 & 0,0104 & 0,2198 & -0,2827 & 0,7643 & -0,4083 & -0,2827 \\ 0,5549 & -0,5759 & -0,4921 & 0,0523 & -0,1151 & -0,4083 & 1,0994 & -0,1151 \\ -0,6596 & 0,2198 & -0,3664 & -0,8272 & 0,3455 & -0,2827 & -0,1151 & 1,6858 \end{bmatrix}$$

Assim, utilizando o modelo matemático descrito no item 3.2. (Valores Genéticos Tradicionais), as pressuposições acerca da distribuição de  $\tilde{y}$ ,  $\tilde{a}$  e  $\tilde{e}$  podem ser descritas como:

$$\begin{bmatrix} \tilde{y} \\ \tilde{a} \\ \tilde{e} \end{bmatrix} \sim \left\{ \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} ZG^*Z' + R & ZG^* & R \\ G^*Z' & G^* & \Phi \\ R & \Phi & R \end{bmatrix} \right\},$$

em que  $G^*$  é a nova matriz de variâncias e covariâncias dos efeitos aleatórios do vetor. Desta forma, encontrou-se as seguintes soluções (tabela 7):

Tabela 7- Soluções do BLUP genômico

Nível	Soluções
Média ( $\hat{\beta}$ )	
1	5,1878
Animais (Valores Genéticos = $\hat{a}$ )	
1	-1,9741
2	2,3505
3	-2,2439
4	0,1266
5	2,7323
6	2,1658
7	0,5316
8	-3,7620

Fonte; A autora

### 3.4. Matriz de Parentesco e Matriz G

No programa R pode-se obter a matriz de parentesco, é chamada de matriz A, utilizando somente o arquivo “pedigree.txt” criado no passo 1 do item 3.1 Arquivos Gerados. Conforme os comandos a seguir:

```
## Matriz de parentesco
```

```
## Pacote necessário para ler como matriz de parentesco, no caso matriz A
```

```

library(kinship2)

# Primeiro passo para montar a matriz ler o arquivo de pedigree.txt

A<- read.table("C:/Users/Professor/tcc/simulacao/pedigree.txt", sep= "", header=
FALSE)

## Função necessária para montar a matriz de parentesco

A <- (2*kinship(A[,1], A[,2], A[,3]))

# inversa de A

solve (A)

```

Depois, no programa R, obtêm-se a matriz  $G^*$ , que neste trabalho usou-se a fórmula proposta por VanRaden (2008), onde  $G^* = \frac{ZZ'}{2\sum_j p_j(1-p_j)}$ . Sendo que,  $Z^*_{ij} = X_{ij} - 2p_j$  ou seja,  $Z^* = X - P$ , como descrito anteriormente. Assim, utilizado o arquivo “marcadores.gen” criado no passo 2 do item 3.1 Arquivos Gerados obteve-se algumas matrizes para chegar a matriz  $G^*$ , executando os comandos no programa R:

```

## MARCADORES

## Lendo arquivo de marcadores para formar a matriz X

X <- read.table ("C:/Users/Professor/tcc/simulacao/marcadores_R.gen", sep= "",
header= FALSE)

## Retira a coluna com a identificação do animal e já lê como matriz

X <- as.matrix(X [,-1])

## p é a frequência do marcador do alelo B, atribui a p uma matriz de dimensão da

## Matriz X para posterior cálculo das frequências dos marcadores do alelo B

```

```

p <- c(rep(0, ncol(X)))

## Para fazer a frequência do marcador em cada coluna do alelo B por coluna

for (i in 1:ncol(X)){

    freq <- sum(X[,i])/(2*nrow(X))

    p[i] <- freq

}

## Matriz de frequência dos marcadores

P2 <- (2*(matrix(p,ncol(X), nrow(X))))

## Transposta de P2, para organizar a frequência do marcador em colunas

P <- t(P2)

## Matriz Z que é o SNP- a frequência do SNP= efeito de substituição

Z <- X-P

## Fazendo o dividendo da equação  $G = K + (2*p[j]*(1-p[j]))$  para gerar G

K <- 0

## Função para formar o dividendo, o valor de K tem que coincidir com o valor do

## Arquivo “sum2pq” gerado após rodar preGSf90 do passo 3

## Coluna em comprimento de p (frequência do marcador)

for (j in 1:length(p)){

    ### Para gerar o dividendo de  $ZZ' / 2*p*(1-p_j)$ 

    K <- K + (2*p[j]*(1-p[j]))

```

```

}

## Gerar a matriz G* = ZZ'/k

G <- (Z%*%t(Z))/K

```

Os valores da matriz  $G^*$  obtidos no programa R tem que coincidir com os valores do arquivo gerado pelo “preGSf90.exe” “Gini” que é uma matriz gerada pelo programa bluf90 do item 3.3. (Valores Genéticos Genômicos).

#### 4. RESULTADOS, DISCUSSÕES E IMPLICAÇÕES PRÁTICAS

Na Tabela 8 estão apresentadas as soluções obtidas através do BLUP tradicional, BLUP genômico e valor genético verdadeiro.

Tabela 8- Valores genéticos estimados por métodos tradicionais e genômica usando a metodologia BLUP e valores genéticos verdadeiros.

<b>Animais (Valores Genéticos = <math>\hat{a}</math>)</b>	<b>Solução Tradicional</b>	<b>Solução Genômica</b>	<b>Valor Genético Verdadeiro</b>
1	-1,7029	-1,9741	-0,14
2	1,8100	2,3505	3,504
3	-0,3413	-2,2439	4,404
4	-0,4593	0,1266	2,859
5	1,1469	2,7323	0,504
6	0,9779	2,1658	4,81
7	1,2239	0,5316	-4,041
8	-2,1588	-3,7620	-0,8

Fonte: A autora

Ocorreu discrepância dos valores entre os dois métodos, por exemplo, quando utilizada seleção tradicional baseada no pedigree, pode-se observar que o animal 1 obteve

valor genético de -1,7029, quando utilizada seleção genômica baseada em marcadores o valor genético passou a ser de -1,9741. Essa mudança de valores genéticos obtidos pode ser devido ao ajuste que ocorre na matriz de parentesco pelo uso dos marcadores moleculares. De acordo com RESENDE *et al.*, (2008), a seleção genômica é o método que explora adequadamente a segregação de amostragem mendeliana, pois captura a matriz de parentesco realizada e não uma matriz de parentesco médio associada ao pedigree. Esses mesmos autores concluíram que a seleção genômica analisa diretamente o DNA associado (via marcadores moleculares) a cada loco de todo o caráter poligênico, e portanto avalia diretamente cada segregação em nível individual e não em nível médio. Isso leva à seleção de menos indivíduos aparentados do que faz a seleção tradicional, reduzindo assim o incremento da endogamia da população.

Analisando o animal de melhor classificação na seleção tradicional, o animal 2 fica em primeiro lugar em valor genético, por outro lado há uma inversão de classificação quando a avaliação é feita a partir da seleção genômica, passando a ser o animal 5 com maior valor genético. Uma explicação para essa troca de classificação pode estar relacionada com a falta de informações de pedigree para o animal 2, porém mesmo com esta incógnita apresenta fenótipo com valor superior e o animal 5 tem somente um dos pais conhecido. Ainda o animal 5 apresentou genótipo com maior índice de locos em heterozigose, logo, mais poligênico do que os demais animais, podendo isto estar relacionado a superioridade de alguma característica.

No arquivo de pedigree do presente trabalho, alguns animais não possuem informação completa, comprometendo a precisão da seleção tradicional o que pode ter incrementado a diferença entre a seleção tradicional e a seleção genômica das soluções. Quando se emprega seleção tradicional baseada no pedigree, monta-se uma matriz de parentesco médio, calculando o valor genético do indivíduo como uma probabilidade. Sabe-se, porém, que erros ou falta de informações no pedigree são corriqueiros, prejudicando as avaliações genéticas, causando prejuízos no ganho genético. Utilizando informações de marcadores moleculares ocorre um ajuste na matriz de parentesco baseada no pedigree, deixando de ser uma probabilidade média, passando a ser o valor genético preciso dos indivíduos. O procedimento BLUP genômico é superior ao BLUP tradicional, pois efetivamente captura a matriz de parentesco realizada e não uma matriz de parentesco médio associada ao pedigree.

As correlações de *Spearman e Pearson* servem como indicadores da acurácia das avaliações genéticas, pois, na classificação dos animais, informam a correspondência entre os valores genéticos verdadeiros e os preditos a partir das diferentes metodologias.

Tabela 9 - Correlação de *Spearman* entre os valores genéticos estimados por métodos tradicionais, genômico e valor genético verdadeiro dos indivíduos

	<b>Seleção Tradicional</b>	<b>Valor Genético Verdadeiro</b>
<b>Seleção Tradicional</b>	1	0,8333
<b>Seleção Genômica</b>	0,8095	0,9523

Fonte: A autora

No presente trabalho a correlação de *Spearman* (Tabela 9), entre seleção tradicional com o valor genético verdadeiro foi de 0,8333, já na correlação entre predição genômica com o valor genético verdadeiro foi maior (0,9523), indicando melhor associação entre o verdadeiro mérito genético do indivíduo. Isso pode ser reforçado conforme DAETWYLER *et al.* (2007) relataram que a seleção genômica aumenta em torno de 67 % a acurácia da predição de efeito genético aditivo predito do indivíduo dentro de família, em comparação com o BLUP tradicional e, conseqüentemente, eleva, em determinada situação, a acurácia da seleção individual de 71 % para 85 %. A correlação de *Spearman* entre seleção tradicional e genômica foi de 0,8095, onde alguns animais mudam de classificação quando comparados entre si pelos dois métodos, ou seja, eles poderão não ter o mesmo desempenho quando analisados pelas diferentes metodologias, essa mudança pode ser devido a alteração da variância genética aditiva estimada entre as duas soluções. Outra hipótese é que na seleção tradicional os pais desconhecidos causem uma discrepância maior no valor genético, o que não ocorre utilizando marcadores moleculares.

Utilizando a correlação de *Pearson* (Tabela 10), a seleção tradicional e o valor genético verdadeiro passou a ser de 0,9036, não diferindo do valor de correlação da seleção tradicional com a genômica, esse valor pode estar ligado ao pequeno número de amostragem tanto de animais quanto de marcadores moleculares utilizados na presente simulação. Quando correlacionada com o valor genético verdadeiro, a predição genômica passou ter valor 0,9232, apresentando um valor alto, porém inferior a correlação de *Sperman*. Pode-se inferir a partir destas correlações que ambas as seleções, tradicional e genômica, estão em concordância com o valor genético verdadeiro, mostrando serem eficientes para predizer o mérito genético dos animais domésticos.

Tabela 10- Correlação de Pearson entre os valores genéticos estimados por métodos tradicionais, genômico e valor genético verdadeiro dos indivíduos.

	<b>Seleção Tradicional</b>	<b>Valor Genético Verdadeiro</b>
<b>Seleção Tradicional</b>	1	0,9036
<b>Seleção Genômica</b>	0,9035	0,9232

Fonte: A autora

Observando os valores para as correlações, tanto de *Pearson* quanto de *Spearman* nas tabelas 8 e 9 (0,90 e 0,80, respectivamente) entre seleção genômica e seleção tradicional são superiores aos dados encontrados no trabalho de Hayes e Goddard (2008), onde encontraram correlação entre marcadores e pedigree aditivo com coeficiente médio de relação considerado alto, 0,69. Uma forma de avaliar a precisão relacionada aos marcadores moleculares é comparar a variabilidade da relação dos irmãos inteiros ajustados com a variabilidade esperada, devido a amostragem mendeliana. Se a variabilidade do marcador de relação estimado para irmãos inteiros está dentro do valor esperado, tem-se a confiança que as relações dos marcadores estimados são precisas. Uma explicação parcial para os resultados podem ser que a matriz de parentesco média derivada de pedigree não pode capturar efeitos de amostragem mendeliana.

Predições genômicas aumentam a acurácia de seleção comparado com a seleção tradicional, fazendo com que os valores genéticos dos indivíduos sejam precisos, auxiliando nos acasalamentos e seleção dos genótipos de interesse econômico do rebanho, ajudando a diminuir a endogamia. Assim, os resultados da simulação revelam um grande potencial da seleção genômica em aumentar a eficiência do melhoramento genético.

## 5. REFERÊNCIAS

AGUILA, I.; MISZTAL, I., JOHNSON, D. L. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal Dairy Science** :743–752 doi: 10.3168/jds.2009-2730 , v. 93, p. 743–752, 2009.

BERNARDO, R.; YU, J. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, v. 47, p. 1082-1090, 2007.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, p. 64-71, 2009.

CARLSON, J. P.; CHRISTIAN, L. L.; ROTHSCILD, M. F. et al. An evaluation of four procedures to rank centrally tested boars. **Journal of Animal Science**, v. 59, p. 934, 1984.

CARNEIRO, P. L. S., MALHADO, C. H. M., EUCLYDES, R. F. et al. Seleção tradicional e associada a marcadores moleculares na avaliação genética animal. **Pesquisa agropecuária brasileira**, Brasília, v. 41, n. 4, p. 615-621, abr. 2006.

CARNEIRO, P. L., S., MALHADO, C. H. M., EUCLYDES, R. F. et al. Endogamia, fixação de alelos e limite de seleção em populações selecionadas por métodos tradicionais e associados a marcadores moleculares. **Revista Brasileira Zootecnia.**, v. 36, n. 2, p. 369-375, 2007.

CARNEIRO, P. L. S., MALHADO, C. H. M., FILHO, R. M. et al. A raça Indubrasil no Nordeste brasileiro: melhoramento e estrutura populacional. , v.38, n.12, p., 2009. **Revista Brasileira de Zootecnia**, v. 38, n. 12, p. 2327-2334, 2009.

CARVALHO, T. D. de. **Marcadores moleculares aplicados à bovinocultura de corte.** Revisão de literatura apresentada como parte das exigências da disciplina Seminário I do Programa de Pós-Graduação em Ciência Animal, Faculdade de Medicina Veterinária e Zootecnia, Programa de Pós- Graduação em Ciência Animal, Universidade Federal de Mato Grosso do Sul. Campo Grande, MS. 2009.

CHESSA, S.; CHIATTI, F.; CERIOTTI, G. et al. Development of a single nucleotide polymorphism genotyping microarray platform for the identification of bovine milk protein genetic polymorphisms. **Journal of Dairy Science**, v.90, n.1, p.45164, 2007.

CONSORTIUM, B. H. A Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, v. 324, p. 528-532, 2009. ISSN 5926.

DEMPFLE, L. Problems in the use of the relationship matrix in animal breeding. **GIANOLA, D.; HAMMOND, K. (Eds.). Advances in statistical methods for genetic improvement of livestock**, Berlin: Springer-Verlag, p. 454-473, 1990.

FALCÃO, A. J. D. S., MARTINS, E. N.; COSTA, C. N. et al. Efeitos do número de animais na matriz de parentesco sobre as estimativas de componentes de variância para produção de leite usando os métodos de Máxima Verossimilhança Restrita e Bayesiano. **Revista Brasileira de Zootecnia**, v. 38, n. 8, p. 1478-1487, 2009. ISSN 1806-9290.

FARIA, F. J. C.; VERCESI FILHO, A.E.; MADALENA, F.E. et al. Parâmetros populacionais do rebanho Gir Mocho registrado no Brasil. **Revista Brasileira de Zootecnia**, v. 30, p. 1984-1988, 2001.

FARIA, F. J. C.; VERCESI FILHO, A.E.; MADALENA, F.E. et al. Estrutura populacional da raça Nelore Mocho. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 54, p. 501-509, 2002.

GAMA, L. T. D.; MATOS, C. P. D.; CAROLINO., N. Modelos Mistos em Melhoramento Animal. In: GAMA, L. T. D.; MATOS, C. P. D.; CAROLINO., N. **Modelos Mistos em Melhoramento Animal**. [S.l.]: DGV. Arquivos Veterinários., 2004. p. 281.

GIANOLA, D.; DE LOS CAMPOS,G.; HILL, W.G.; MANFREDI, E. et al. Additive genetic variability and the Bayesian alphabet. **Genetics**, v. 183, p. 347–363, 2009.

HAYES, B. J.; BOWMAN, P.J.; CHAMBERLAIN, A.J. et al. Invited review: genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, v. 92, p. 433-443, 2009.

HAYES, B. J.; GODDARD, M. E. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, p. 323-330, 2007.

HEATON, M. P.; HARHAY, G. P.; STONE, G. L. B. R. T. et al. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. **Mammalian Genome**, v. 13, n. 5, 2002.

HEATON, M. P.; KEEN, J.E.; CLAWSON, M.L. et al. Use of bovine single nucleotide polymorphism markers to verify sample tracking in beef processing. **Journal of the American Veterinary Medicine Associatio**, v. 226, n. 8, p. 1311-1314, 2005.

HENDERSON, C. R. Sire evaluation and genetic trends. **Animal Breeding and Genetics Symposium in Honor of J. Lush. American Society of Annimal Science**, , Champaign, v. III, p. 10-41, 1973.

HENDERSON, C. R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. **Biometrics**, v. 32, p. 69-83, 1976.

HENDERSON, C. R. Applications of linear models in animal breeding. In: HENDERSON, C. R. **Applications of Linear Models in Animal Breeding**. Guelph: University of Guelph, 1984, 1984. p. 462.

KAMIŃSKI, S.; AHMAN, A.; RUSC, A. E. A. MilkProtChip-a microarray of SNPs in candidate genes associated with milk protein biosynthesis — development and validation. **Journal of Applied Genetics**, v. 46, n. 1, p. 45-58, 2005.

KENNEDY, B. W.; MOXLEY, J. E. Comparison of genetic group and relationship methods for mixed model sire evaluation. **Journal of Dairy Science**, v. 58, p. 1507- 514, 1975.

KHANG, J. V. T. A FORTRAN subroutine to compute inbreeding and kinship coefficients according to the number of ancestral generations. **Computer Applications in the Biosciences**, v. 5, p. 199-204, 1989.

KOLBEHDARI, D.; SHAEFFER, L. R.; ROBINSON, J. A. B. Estimation of genome-wide haplotype effects in half-sib designs. **Journal of Animal Breeding and Genetics**, v. 124, p. 356-361, 2007.

LEGARRA, A.; MISZTAL, I. Computing strategies in genomewide selection. **Journal of Dairy Science**, v. 91, p. 360-366, 2008.

LI Y, Willer C, Sanna S, Abecasis G. Genotype imputation. **Annu Rev Genomics Hum Genet**. 2009;10:387-406. Review.

LÔBO, R. N. B.; LÔBO, A. M. B. O. Melhoramento genético como ferramenta para o crescimento e o desenvolvimento da ovinocultura de corte. **Revista Brasileira Reprodução Animal**, Belo Horizonte, v. 31, p. 247-253, 2007.

LONG, N.; GIANOLA, D.; ROSA, G. J. M. et al. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, n. 124, p. 377–389, 2007.

MAEDA, M.; MURAYAMA, N.; ISHII, H. et al. A simple and rapid method for HLA-DQA1 genotyping by digestion of PCR amplified DNA with allele specific restriction endonucleases. **Tissue Antigens**, v. 34, n. 5, p. 290-298, 1989.

MALHADO, C. H. M.; RAMOS, A.A.; CARNEIRO, P.L.S. et al. Melhoramento e estrutura populacional em bubalinos da raça Mediterrâneo no Brasil. **Pesquisa Agropecuária Brasileira**, v. 43, p. 215-220, 2008.

MARTÍNEZ, R.; GARCÍA, D.; GALLEGO, J.L. A. et al. Genetic variability in Colombian Creole cattle populations estimated by pedigree information. **Journal of Animal Science**, v. 86, p. 545-552, 2008.

MATUKUMALLI, L. K.; LAWLEY, C.T.; SCHNABEL, R.D. et al. Development and characterization of a high density SNP genotyping assay for cattle. **PLoS ONE**, v. 4, n.4, p. 5350, 2009.

MEUWISSEN, T. H. E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, v. 124, p. 321-322, 2007.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

MISZTAL, I.; TSURUTA, S.; STRABEL, T. et al. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS OF GENETIC APPLIED TO LIVESTOCK PRODUCTION, 7., Montpellier, France, 2002. **Proceedings...** Montpellier: Francis Minvielle, v.28, p.7, 2002.

MUIR., W. M. Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, v. 124, p. 342-355, 2007.

PETERNELLI, L. A., FERREIRA, F. M.; ROCHA, R. B. et al. ANÁLISE DOS COEFICIENTES DE ENDOGAMIA E DE PARENTESCO PARA QUALQUER NÍVEL DE PLOIDIA USANDO O PACOTE ESTATÍSTICO R. **Bragantia**, Campinas, v. 68, n. 4, p. 849-855, 2009.

POLLAK, E. J.; UFFORD, G. R.; GROSS, S. J. Comparison of alternative models for within-herd genetic evaluation fo beef cattle. **Journal of Animal Science**, v. 45, p. 1010, 1977.

QUAAS, R. L. Computing the diagonal elements and inverse of a large numerator relationship matrix. **Biometrics**, v. 32, p. 949-953, 1976.

R: a language and environment for statistical computing. Vienna, Austria: **R. Foundation for Statistical Computing**, 2006. Disponível em: <<http://www.R-project.org>, 2006> Acessado em: 06 out. 2006.

RESENDE, M. D. V. D.; LOPES, P. S.; SILVA, R. L. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, Colombo, n.56, p. 63-77, jan./jun 2008.

RESENDE, M. D. V. D.; PEREZ, J. R. H. R. Melhoramento animal: predição de valores genéticos pelo modelo animal – blup em bovinos de leite, bovinos de corte, ovinos e suínos. **Archives Veterinary Science**, v. 4, n. 1, p. 17-29, 1999.

SCHAEFFER., L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v. 123, p. 218-223, 2006.

SOLBERG, T. R., SONESSON, A.; WOOLIAMs, J.; MEUWISSEN, T. H. E. et al. Genomic selection using different marker types and density. **In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION**, Belo Horizonte: Ed. da UFMG, v. 8, 2006.

VAN VLECK, L. D.; HUDSON, G. F. S. Relationships among sires in estimating genetic variance. **Journal of Dairy Science**, v. 65, p. 1663, 1982.

VANRADEN, P. M. Efficient methods to Compute Genomic predictions. **Journal Dairy Science**, v. 91, p. 4414–4423, 2008.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R. et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, p. 16-24, 2009.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M. et al. A review on SNP and other types of molecular markers and their use in animal genetics. **Genet. Sel.**, v. 34, p. 275–305 , 2002.

VISSCHER, P., PONG-WONG, R.; WHITTEMORE, C. et al. Impact of biotechnology on (cross) breeding programs in pigs. **Liv. Prod Sci**, v. 65, p. 57-70, 2000.