

Projeto

Michel Marques Farah

27 de agosto de 2011

Sumário

1	Introdução e Justificativa	3
2	Objetivos Específicos	5
3	Material e Métodos	6
3.1	Dados	6
3.2	Matriz de parentesco pedigree-genômica	6
3.3	Matriz IBS	7
3.4	Matriz IBD	7
4	Cronograma	8

1 Introdução e Justificativa

Tradicionalmente, estudo para seleção de características de interesse são realizadas com base no fenótipo dos indivíduos e na informação do parentesco entre os animais. Esta seleção é eficiente, porém o processo é lento, principalmente para características que são medidas em apenas um sexo, como produção de leite, ou características medidas após o abate dos animais, como a qualidade da carne, ou ainda medidas mensuradas no final da vida do indivíduo, por exemplo, longevidade. Assim, para realizar programas de melhoramento para estas características, pesquisadores buscam identificar os genes para elas e a seleção de animais que carregam os alelos desejáveis (MEUWISSEN; GODDARD, 1996).

Os projetos de sequenciamento e geração de informações genômicas de alta qualidade estão cada vez mais sendo utilizados no melhoramento genético animal, a quantidade de nucleotídeos de polimorfismos simples (SNPs) identificados crescem rapidamente, por exemplo, Wong et al. (2004) relataram 2,8 milhões de SNPs no genoma de frangos, Andreassen et al. (2010) encontraram 112 SNPs em 88 genes diferentes de salmões e em bovinos, já estão disponíveis painéis para a identificação de mais de 777000 SNPs (Illumina Inc., San Diego, CA, 2011).

Com estas informações genômicas, permite, também, aos pesquisadores aumentar a precisão da avaliação genética dos animais, e segundo Meuwissen et al. (2001), esta seleção genômica aumenta a taxa de melhoramento genético e reduz o custo do teste de progênie, permitindo aos criadores pré-selecionar animais que tenham herdado segmentos cromossômicos de maior mérito.

Assim, vem crescendo cada vez mais a quantidade de pesquisadores interessados em utilizar as informações genômicas nos programas de melhoramento genético animal (MEUWISSEN; GODDARD, 1996; CHRISTENSEN; LUND, 2010; HEYES et al., 2010; GIANOLA et al., 2010), porém, estas previsões dos valores genéticos genômicos são obtidas por estimativa de um efeito causado por marcadores ou usando o modelo das Equações de Modelos Mistos com uma matriz de parentesco genômica, denominada Matriz G (VANRADEN, 2008), simples interpretação e de fácil manipulação (LEGARRA et al., 2009).

Contudo, as maioria das pesquisas visam a utilização das informações genômicas para a determinação dos parâmetros genéticos e para verificar o comportamento de determinadas características de interesse, porém, de acordo com Malhado et al. (2008) é imprescindível conhecer os diferentes fatores que interferem potencialmente na seleção e no progresso genético, como por exemplo, tamanho efetivo, intervalo de gerações, variabilidade genética e endogamia.

A endogamia pode ser definida como o acasalamento de indivíduos mais aparentados entre si, do que o parentesco médio esperado, se eles fossem escolhidos ao acaso, na população. Assim, uma consequência é o aumento da quantidade de homozigose em razão da diminuição da heterozigose, portanto, aumenta a probabilidade de dois alelos em um loco tomado ao acaso sejam idênticos por descendência (LOPES, 2005).

De acordo com Meuwissen e Luo (1992), existem diversos algoritmos computacionais que permitem o cálculo do coeficiente de endogamia de grandes populações, de forma rápida e eficiente. No entanto, estes algoritmos trabalham apenas com as informações de parentesco, utilizando uma matriz de parentesco (A) e quanto maior o número de gerações informadas maior a precisão desta estimativa da probabilidade de genes idênticos por descendência.

Porém, com o avanço da tecnologia e a possibilidade de genotipagem dos indivíduos possibilitou a utilização de informações mais precisas sobre os genes idênticos em estado que podem ser compartilhados através de ancestrais comuns ausentes no pedigree e assim tornou-se viável a utilização de uma matriz de parentesco genômica (G) que pode ser calculada por diversos métodos (FORNI et al., 2011).

Como é inviável a genotipagem de uma população inteira, Legarra et al. (2009) e Misztal et al. (2009) propuseram uma integração entre a matriz (A) e (G) em uma única matriz (H), assim, possibilitaram a avaliação genética via BLUP utilizando uma única etapa, esta aplicada com sucesso em gado de leite (AGUILAR et al., 2010).

Forni et al. (2011) avaliaram as diferentes formas de montar a matriz de parentesco genômica (G) de acordo com as frequências alélicas com posterior

integração com a matriz (A) sobre o impacto nas estimativas dos valores genéticos e componentes de variância e concluíram que estas estimativas utilizando uma matriz de parentesco (A) ou conjunta pedigree-genômica (H) foram similares e que se as frequências alélicas na população-base forem diferentes de 0,5, os alelos raros contribuem para uma maior semelhança genética entre os indivíduos.

Desta forma, o objetivo geral deste projeto será estudar os efeitos de diferentes matrizes genômicas, com informações de alelos idênticos por descendência (IBD), alelos idênticos por estado (IBS) e diferentes frequências alélicas, bem como a matriz de parentesco tradicional (A) e a matriz conjunta pedigree-genômica (H) e seus efeitos na estimação do coeficiente de endogamia em uma população de bovinos selecionada para características de carcaça e qualidade da carne.

2 Objetivos Específicos

Estudar os efeitos das matrizes IBD, IBS e A bem como suas metodologias de construção e integração entre as possíveis informações disponíveis.

Verificar o efeito da informação de SNPs na avaliação genética:

- Avaliação tradicional (A);
- Somente com a matriz IBS;
- Matriz e modelo com efeitos individuais.

```
«echo=TRUE»= x <- c(1,2,3,4,5,6,7,8,9,10) xm <- mean(x) xm @
```

3 Material e Métodos

3.1 Dados

Os dados utilizados serão provenientes de fazendas que integram o programa de melhoramento genético da Conexão Delta G, contando com 2000 animais machos da raça Nelore terminados em confinamento com idade próxima a dois anos, filhos de cerca de 100 touros, provenientes de seis rebanhos, todos com informações de desempenho e genealogia disponíveis. Destes, 980 animais serão genotipados utilizando um painel de aproximadamente 800000 SNP do BovineSNP BeadChip (High-Density Bovine BeadChip) (Illumina Inc., San Diego, CA, 2011).

Na análise de controle de qualidade, os SNPs serão submetidos a um teste de probabilidade, assumindo uma distribuição de 1ui-quadrado e nível de significância de $P < 0,05$.

Os SNPs com desvios significativos do equilíbrio de Hardy-Weinberg serão excluídos da análise, assim como SNPs com baixa frequência (*minor frequency alleles*: MAF) e aqueles que estejam genotipados em menos de 50% da população. Todas as análises serão realizadas utilizando o conjunto de dados completo e os registros serão analisados utilizando um modelo animal.

3.2 Matriz de parentesco pedigree-genômica

A matriz de parentesco pedigree-genômica (H) será utilizada para substituir a matriz de parentesco (A). A matrix H será obtida pela combinação das informações de pedigree e genômica (AGUILAR et al., 2010)

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}, \quad (1)$$

onde G^{-1} é a inversa da matriz de parentesco genômica e A_{22}^{-1} é a inversa da matriz de parentesco dos animais genotipados. As comparações irão envolver várias matrizes de parentesco genômica (G), de modo geral, para a obtenção da

matriz G foi usada a equação descrita por VanRaden (2008) :

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

em que M é uma matriz alélica com m colunas (m =número total de marcadores) e n linhas (n =número total de animais genotipados); e P é uma matriz que contém a frequência do segundo alelo (p_j), expresso como $2p_j$.

A matriz M será preenchida com -1 se o genótipo do animal i para o SNP j for homozigoto, pode ser preenchido com 0 se for heterozigoto ou preenchido com 1 se o genótipo for o outro homozigoto. As frequências alélicas utilizadas serão: 0,5 para todos os marcadores (G05), a frequência média do menor alelo (GMF) e a frequência observada do alelo de cada SNP (GOF).

3.3 Matriz IBS

Utilizando o método descrito anteriormente, para obter a matriz M , de acordo com VanRaden (2008), os elementos da diagonal de MM' representam o número de locus em homozigose para cada indivíduo e os elementos fora da diagonal representam o número de alelos compartilhados por parentes. Em contraste, os elementos da diagonal de $M'M$ demonstram o número de indivíduos homozigotos para cada locus e fora da diagonal indicam o número de vezes que os alelos em diferentes loci foram herdados pelo mesmo indivíduo.

3.4 Matriz IBD

Devido a complexidade para a obtenção da matriz genômica com informações de alelos idênticos por descendência, para a análise desta matriz será utilizada a simulação de dados.

4 Cronograma

2011	Programação
Setembro - Dezembro	Obtenção, edição e análises preliminares dos dados
2012	Programação
Janeiro - Junho	Preparação dos programas e algoritmos a serem utilizados
Julho - Dezembro	Ajustes, análises finais e redação do primeiro artigo sobre o uso das diferentes matrizes genômicas
2013	Programação
Janeiro - Junho	Ajustes, análises finais e redação do segundo artigo sobre ...
Julho - Setembro	Preparação da redação final e revisão para Qualificação
Outubro - Novembro	Qualificação e correções finais
Dezembro	Defesa

Referências

AGUILAR, I. et al. 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, n. 2, p. 743–752, Novembro 2010.

ANDREASSEN, R. et al. Targeted snp discovery in atlantic salmon (*salmo salar*) genes using a 3'UTR-primed snp detection approach. *BMC Genomics*, v. 11, n. 706, p. 1–10, Dezembro 2010.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, v. 42, n. 2, p. 1–8, 2010. [Http://www.gsejournal.org/content/42/1/2](http://www.gsejournal.org/content/42/1/2).

FORNI, S. et al. 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. [Http://www.gsejournal.org/content/43/1/1](http://www.gsejournal.org/content/43/1/1).

GIANOLA, D. et al. A two-step method for detecting selection signatures using genetic markers. *Genetics Research*, v. 92, p. 141–155, Março 2010.

HEYES, B. J. et al. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. *PLoS Genetics*, v. 6, n. 9, p. 1–11, Setembro 2010.

Illumina Inc., San Diego, CA. *Illumina Inc.* Agosto 2011. [Http://www.illumina.com/](http://www.illumina.com/).

LEGARRA, A. et al. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, v. 92, n. 9, p. 4656–4663, Abril 2009.

LOPES, P. S. *Teoria do Melhoramento Animal*. Belo Horizonte: FEPMVZ-Editora, 2005.

MALHADO, C. H. M. et al. Melhoramento e estrutura populacional em bubalinos da raça mediterrâneo no brasil. *Pesquisa Agropecuária Brasileira*, v. 43, n. 2, p. 215–220, Fevereiro 2008.

- MEUWISSEN, T. H. E.; GODDARD, M. E. The use of marker haplotypes in animal breeding schemes. *Genetics Selection Evolution*, v. 28, p. 161–176, 1996.
- MEUWISSEN, T. H. E. et al. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, v. 157, p. 1819–1829, Abril 2001.
- MEUWISSEN, T. H. E.; LUO, Z. Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*, v. 24, p. 305–313, Abril 1992.
- MISZTAL, I. et al. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, v. 92, n. 9, p. 4648–4655, Abril 2009.
- VANRADEN, P. M. Efficient methods to compute genomic predictions. *Journal of Dairy Science*, v. 91, n. 11, p. 4414 – 4423, Junho 2008.
- WONG, G. K. S. et al. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, v. 432, p. 717–722, dezembro 2004.