

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

**DIVERSIDADE GENÉTICA E ACURÁCIA DA INFORMAÇÃO GENÔMICA EM  
BOVINOS DE CORTE**

FERNANDA VARNIERI BRITO  
Engenheira Agrônoma – UFRGS  
Mestre em Zootecnia – UFRGS

Tese apresentada como um dos requisitos à obtenção do Grau de Doutor em  
Zootecnia  
Área de Concentração Produção Animal

Porto Alegre (RS), Brasil  
Fevereiro de 2011

## DEDICATÓRIA

*Às minhas filhas, Julia e Bruna, fontes de vida e inspiração,  
e ao meu esposo e companheiro Harold,  
dedico este trabalho.*

## AGRADECIMENTOS

Particularmente agradeço ao meu orientador José Braccini por seus conselhos e por ter depositado sua confiança em mim em todo este período;

De forma especial, meus agradecimentos ao meu co-orientador no exterior, Flávio Schramm Schenkel, por sua grande contribuição neste trabalho através de sua experiência, seus ensinamentos e conselhos que me fizeram crescer;

À CAPES, que me concedeu a bolsa de estudos no Brasil e no exterior;

À Pró-Reitoria de Pós-Graduação, na pessoa de Ana Lucia Nunes, que me orientou e assessorou, tanto no Brasil como no exterior, nos trâmites necessários para a realização do estágio no exterior;

Ao *Centre for Genetic Improvement of Livestock (CGIL)* do *Depto of Animal and Poultry Science* da *University of Guelph* pela estrutura disponibilizada para a realização deste estudo, principalmente a utilização de seus computadores de alto desempenho;

À Mehdi Sargolzaei, pela cedência dos programas utilizados neste trabalho e por todo o seu auxílio, sempre de maneira muito gentil;

À família Schenkel, Flávio, Sandra, Mariana e Daniel que receberam a mim e minha família no Canadá, tornando o período no exterior uma vivência inesquecível e prazerosa;

Aos colegas do GenSys, Piccoli, Severo, Vanerlei, Roberto e Vânia, pela confiança e pela sua disposição em cobrir a minha ausência durante o curso de doutorado;

À Léia, pela amizade e pela disposição em tomar qualquer providência que se fizesse necessária;

À Ione, por sua experiência no Programa de Pós-Graduação, sempre auxiliando no cumprimento das normas;

À Agropecuária Jacarezinho, por ter cedido seus dados para a execução de parte deste trabalho e à ABCZ, por ter cedido o seu banco de dados genealógico;

À minha família, Harold, Julia e Bruna pelo amor, apoio e compreensão;

In memoriam, com especial gratidão, ao amigo, sócio e inspirador, Luiz Alberto Fries, pedra fundamental da minha construção profissional e de vida.

# DIVERSIDADE GENÉTICA E ACURÁCIA DA INFORMAÇÃO GENÔMICA EM BOVINOS DE CORTE<sup>1</sup>

Autor: Fernanda Varnieri Brito

Orientador: José Braccini Neto

Coorientador no exterior: Flávio Schramm Schenkel

## RESUMO

Avanços na área da genômica têm propiciado a inclusão da informação genômica no melhoramento genético animal. O sucesso de sua utilização depende da extensão do desequilíbrio de ligação (DL) entre marcadores e QTL, do número de animais genotipados no grupo de treinamento (GT) e da  $h^2$  da característica. O conhecimento de parâmetros de diversidade genética tais como endogamia e tamanho efetivo da população ( $N_e$ ), associados ao aumento dos níveis de DL, são úteis no estudo da viabilidade da seleção genômica. Com o objetivo de estimar estes parâmetros em uma população de gado de corte, bovinos Nelore foram estudados através de uma análise de pedigree. O  $N_e$  estimado foi de 114, 245 e 101 para os anos de 95-99, 99-03 e 03-07, respectivamente. Para estimar a acurácia do valor genômico direto (VGD) foram simuladas duas populações com estrutura de DL semelhante àquela em gado de corte com 40k (POP1) e 800k (POP2) marcadores SNP. O DL entre os SNP adjacentes, medido pelo  $r^2$ , foi de 0,25 e 0,32 e o  $N_e$  foi de 250 e 120 para POP1 e POP2, respectivamente. Para POP1, GT foram formados por 1920, 960 e 480 touros com acurácia média do valor genético estimado (VGE) de 0,79, 0,90 e 0,94 para  $h^2$  0,10, 0,25 e 0,40, respectivamente, e para POP2, 480 touros com a mesma acurácia média dos VGE. Outros dois GT (7680 e 1920 para POP1 e POP2) foram formados por animais com fenótipo disponível para estimar os efeitos dos marcadores ao invés do VGE. A acurácia do VGD aumentou significativamente ( $p < 0,05$ ) com o aumento do número de touros no GT,  $h^2$  e densidade dos marcadores. O número de animais no GT teve um efeito quadrático sobre as acurácias do VGD para todas as  $h^2$  ( $p < 0,0001$ ). Para a POP1, a acurácia mais alta (0,64) foi observada para  $h^2 = 0,40$  e  $n = 1920$ . As acurácias do VGD na POP2 foram maiores (de 16% a 24%) do que aquelas para 40k SNP e  $n = 480$  e similares àquelas para 40k SNP e  $n = 960$ . Aumentar em 4 vezes o GT e usar fenótipos para calcular os efeitos dos marcadores não foi suficiente para manter ou aumentar a acurácia do VGD para  $h^2 < 0,40$ , tanto para POP1, como POP2. Os resultados obtidos quando o número de SNP foi aumentado de 40k para 800k SNP com  $n = 480$  indicou que o painel mais denso requer 2 vezes menos touros no GT do que 40k SNP para produzir acurácias do VGD similares. Os resultados sugerem que a seleção genômica pode trazer ganhos em acurácia na ordem de 27, 13 e 10%, considerando  $n = 1920$  e  $h^2 = 0,10$ , 0,25 e 0,40, respectivamente e 7% para  $n = 960$  e  $h^2 = 0,10$  com relação aos VGE de animais jovens (média dos pais). Estes níveis de ganho foram menores do que aqueles relatados para gado de leite. Portanto, justificam-se mais trabalhos de pesquisa nesta área.

---

<sup>1</sup> Tese de Doutorado em Zootecnia – Produção Animal, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. (134 p.). Fevereiro de 2011.

# GENETIC DIVERSITY AND ACCURACY OF GENOMIC INFORMATION IN BEEF CATTLE<sup>1</sup>

Author: Fernanda Varnieri Brito

Advisor: José Braccini Neto

Co-advisor in the foreign institution: Flávio Schramm Schenkel

## ABSTRACT

Advances in genomics have led to the inclusion of genomic information in livestock breeding. The success of its use depends mainly on the extent of linkage disequilibrium (LD) between markers and QTL, the number of genotyped animals in the training set (TS), and the heritability of the trait. The knowledge of genetic diversity parameters such as rate of inbreeding and the effective population size ( $N_e$ ) which are associated with the increase in LD levels is useful for studying the feasibility of genomic selection. Aiming to estimate these parameters in beef cattle, a population of Nellore cattle was studied by pedigree analysis. The  $N_e$  was estimated to be 114, 245 and 101 for the years 95-99, 99-03 and 03-07, respectively. Aiming to estimate the accuracy of genomic selection, two populations were simulated with a LD structure similar to that in beef cattle with 40k (POP1) and 800k (POP2) SNP markers. The LD between adjacent markers, as measured by  $r^2$ , was 0.25 and 0.32 and  $N_e$  was 250 and 120 for POP1 and POP2, respectively. TS were formed by 1920, 960 and 480 bulls with EBV with mean accuracy of 0.79, 0.90 and 0.94 for  $h^2$  of 0.10, 0.25 and 0.40, respectively, for POP1, and 480 bulls with the same average accuracies of EBV for POP2. Two large TS (7680 and 1920 for POP1 and POP2) were composed by animals with phenotypic information to estimate the marker effects, rather than EBV. The accuracy of direct genomic EBV (DGEVB) increased significantly ( $p < 0.05$ ) with the increasing of number of bulls in the TS,  $h^2$  and density of the markers. The number of animals in the TS showed a quadratic effect on the accuracy of DGEVB ( $p < 0.0001$ ) for all  $h^2$ . For POP1, the highest accuracy (0.64) was observed for  $h^2$  of 0.40 and 1920 bulls in TS. The accuracies of DGEVB in POP2 were higher (from 16% to 24%) than those from 40K SNP and 480 bulls in the TS and similar to those from 40K SNP and 960 bulls in TS. Increasing the number of animals in TS by 4-fold and using phenotypes to calculate the marker effects was not enough to maintain or increase the accuracy of DGEVB estimated using EBV as response variable to estimate the marker effects, when the  $h^2$  was lower than 0.40 for both POP1 as POP2. Results of increasing the number of markers from 40k to 800k with a TS of 480 genotyped bulls indicate that the high density marker panel would require 2 times less bulls in the TS than 40K for yielding similar accuracies of DGEVB. The results suggest that genomic selection can bring gains in accuracy of 27, 13 and 10%, considering 1920 bulls in the TS for  $h^2 = 0.10, 0.25$  and  $0.40$ , respectively, and 7% for TS of 960 bulls and  $h^2 = 0.10$  when compared to the parent average of young animals. These gains were lower than those reported in dairy cattle. Therefore, more investigations are warranted.

---

<sup>1</sup> Doctoral thesis in Animal Science, Faculdade de Agronomia. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. (134 p.). February, 2011.

## SUMÁRIO

	<b>Página</b>
<b>CAPÍTULO I</b>	
<b>1.INTRODUÇÃO .....</b>	<b>2</b>
<b>2.REVISÃO BIBLIOGRÁFICA .....</b>	<b>6</b>
2.1. Diversidade Genética.....	6
2.1.1. A origem genética dos bovinos domesticados.....	9
2.1.2. Tamanho efetivo da população .....	10
2.1.3. Endogamia .....	13
2.1.4. Parâmetros com base na probabilidade da origem do gene .....	15
2.2. O uso da genômica no melhoramento genético animal.....	17
2.2.1. Marcadores moleculares.....	17
2.2.2. Desequilíbrio de ligação.....	19
2.2.3. Seleção Assistida por Marcadores.....	23
2.2.4. Seleção Genômica.....	26
<b>CAPÍTULO II</b>	
<b>Pedigree analysis in a large Brazilian Nellore herd .....</b>	<b>38</b>
Abstract .....	39
Introduction .....	39
Material and Methods .....	41
Results and Discussion .....	44
Conclusions .....	48
Acknowledgements.....	48
References .....	49
<b>CAPÍTULO III</b>	
<b>Accuracy of genomic selection in a simulated population mimicking the extent of linkage disequilibrium in beef cattle .....</b>	<b>57</b>
Abstract .....	58
Background .....	59
Methods .....	61
Results and Discussion .....	67
Conclusions .....	74
Authors' Contributions.....	75
Acknowledgements.....	75
References.....	76
<b>3. CONSIDERAÇÕES FINAIS .....</b>	<b>94</b>
<b>4. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>98</b>
<b>5. APÊNDICES .....</b>	<b>109</b>
<b>6. VITA .....</b>	<b>134</b>

## LISTA DE TABELAS

<b>CAPÍTULO II</b>	<b>Página</b>
Table 1 - Generation interval (Gen Int), inbreeding calculated by VanRaden's method ( $\Delta F$ ) and effective numbers in the Nellore herd ....	53
Table 2 - The eight ancestors with highest genetic contributions to the gene pool of animals born between 2003 and 2007.....	54
<b>CAPÍTULO III</b>	
Table 1 - Parameters of the simulation process.....	82
Table 2 - Average Linkage Disequilibrium ( $r^2$ ) between adjacent SNP markers .....	83
Table 3 - Average Linkage Disequilibrium ( $r^2$ ) for different distances between closely located SNP pairs .....	84
Table 4 - Average Linkage Disequilibrium ( $r^2$ ) between adjacent SNPs pairs and distribution across different $r^2$ ranges.....	85
Table 5 - Accuracy of Direct Genomic Estimated Breeding Value.....	86

## LISTA DE FIGURAS

<b>CAPÍTULO II</b>	<b>Página</b>
Figure 1a - Regular average inbreeding coefficient and average inbreeding coefficient calculated by VanRaden's method .....	55
Figure 1b - Regular average inbreeding coefficient, inbreeding expected under random mating (F.ST) and deviation from random mating (f.is) .....	55
 <b>CAPÍTULO III</b>	
Figure 1 - Schematic representation of the simulation steps .....	87
Figure 2 - Schematic representation of the simulated scenarios.....	88
Figure 3 - Extent of Linkage Disequilibrium (LD).....	89
Figure 4 - Effective Population Size.....	90
Figure 5 - Accuracy of Direct Genomic Estimated Breeding Value (DGE BV) for 40k Scenario.....	91
Figure 6 - Accuracy of Direct Genomic Estimated Breeding Value (DGE BV) for 40k Scenario.....	92
Figure 7 - Accuracy of Direct Genomic Estimated Breeding Value (DGE BV) for 800k Scenario.....	93



## LISTA DE ABREVIATURAS

AI	<i>Artificial Insemination</i>
CMP	<i>Carcass Merit Project</i>
corr <sup>2</sup>	Quadrado do coeficiente de correlação
DEP	Diferença Esperada na Progenie
DGEBV	<i>Direct Genomic Estimated Breeding Value</i>
DL	Desequilíbrio de Ligação
E <sub>a</sub>	<i>Effective number of ancestors</i>
EBV	<i>Estimated Breeding Value</i>
E <sub>f</sub>	<i>Effective number of founders</i>
E <sub>fg</sub>	<i>Effective number of founder genomes</i>
E <sub>nf</sub>	<i>Effective number of non-founders</i>
F	Coeficiente de endogamia
f.is	<i>Deviation from random mating</i>
F.ST	<i>Inbreeding expected under random mating</i>
GEBV	<i>Genomic Estimated Breeding Value</i>
GPA	<i>Genomic Parentage Average</i>
GT	Grupo de treinamento
h <sup>2</sup>	Herdabilidade
LD	<i>Linkage disequilibrium</i>
MACE	<i>Multi-trait Across-Country Evaluation</i>
MAF	<i>Minor allele frequency</i>
MARC	<i>Meat Animal Research Center</i>
MS	<i>Multiple Sire</i>
NBCEC	<i>National Beef Cattle Evaluation Consortium</i>
N <sub>e</sub>	Tamanho efetivo da população
PA	<i>Parentage average</i>
POP1	População 1
POP2	População 2
QTL	<i>Quantitative Trait Locus</i>
r <sup>2</sup>	Quadrado da correlação entre os alelos de dois <i>loci</i>
SAM	Seleção Assistida por Marcadores
SNP	<i>Single Nucleotide Polymorphism</i>
TBV	<i>True Breeding Value</i>
TS	<i>Training Set</i>
USDA	<i>United States Department of Agriculture</i>
VGD	Valor Genético Direto
VGE	Valor Genético Estimado
VGG	Valor Genético Genômico
VGP	Valor Genético médio dos Pais
VRM	VanRaden's method of calculate inbreeding coefficient

## RELAÇÃO DE APÊNDICES

	<b>Página</b>
Apêndice 1 – Normas para publicação – Genetics and Molecular Biology .....	109
Apêndice 2 – Normas para publicação – BMC Genetics .....	115
Apêndice 3 – Resultados originais do Capítulo III .....	131

## **CAPÍTULO I**

## 1.INTRODUÇÃO

O melhoramento genético dos animais domésticos para características de importância econômica tem sido tradicionalmente realizado por meio da seleção com base no fenótipo do indivíduo e de seus parentes e na informação de parentesco disponível entre os indivíduos (pedigree). Os maiores avanços nesta área vinham ocorrendo no campo da informática e estatística onde foram desenvolvidos estimadores não viciados dos valores genéticos de indivíduos criados em diferentes ambientes e, concomitantemente, a possibilidade de estimação de uma grande quantidade de parâmetros, utilizando grandes volumes de dados.

Nas últimas décadas, uma série de importantes descobertas e avanços na área de análise do DNA propiciou o surgimento da Genômica, ciência que trata do estudo do genoma completo dos diferentes organismos. Além do seqüenciamento completo do DNA de várias espécies, incluindo os bovinos, o mapeamento dos SNPs (*Single Nucleotide Polymorphism*) em conjunto com plataformas modernas de genotipagem capazes de processar muitas amostras para muitos marcadores numa única análise, tem propiciado a inclusão da informação genômica em esquemas de melhoramento genético.

Os resultados de estudos envolvendo gado leiteiro, inicialmente desenvolvidos por uma parceria entre instituições de pesquisa e indústria dos

Estados Unidos e Canadá, entusiasmaram os pesquisadores e, atualmente, alguns países já adotaram esta importante ferramenta nestes rebanhos.

Em gado de corte a situação ainda é bem diferente com relação aos estudos e à viabilidade da seleção genômica. Os motivos são diversos, porém destaca-se a estrutura genética dos rebanhos que envolvem um número maior de raças em relação ao gado leiteiro e que resulta em diferenças significativas para parâmetros como extensão do desequilíbrio de ligação (DL) e tamanho efetivo da população ( $N_e$ ).

Na seleção genômica, o número e as posições dos *loci* de características quantitativas (QTL) não são conhecidos, portanto é necessário cobrir todo o genoma. Desta forma busca-se capturar o efeito de todos os genes responsáveis pela característica de interesse por assumir que os marcadores estejam próximos aos QTLs e em desequilíbrio de ligação, ou seja, que haja uma associação entre eles persistindo através das gerações sendo, esta associação, uma propriedade da população. A densidade de SNPs necessária depende da extensão do desequilíbrio de ligação que, por sua vez, depende do tamanho efetivo da população.

A seleção artificial tem sido um importante fator que afeta o desequilíbrio de ligação em rebanhos comerciais sujeitos a programas intensivos de melhoramento genético. A metodologia de modelos mistos é atualmente utilizada para obter-se a melhor predição não viesada dos valores genéticos dos animais para características produtivas, sendo que as avaliações genéticas realizadas no mundo inteiro são usualmente baseadas no modelo animal. Tem sido demonstrado que esta metodologia tem ótimas propriedades

estatísticas o que garante resultados muito satisfatórios com relação aos ganhos genéticos a curto e médio prazo. No entanto, uma vez que todas as relações de parentesco existentes na população são levadas em conta, animais mais aparentados têm mais probabilidade de serem selecionados, levando a uma maior taxa de endogamia e perda de variância genética.

A preocupação com relação a perdas econômicas relacionadas com o fenômeno da depressão endogâmica e a diminuição da variabilidade genética, além do interesse em conhecer mais detalhadamente a estrutura genética no que diz respeito ao tamanho efetivo e extensão do DL, têm sido tema de inúmeros estudos. Uma maneira de analisar a diversidade genética de uma população é baseada na informação fornecida pelo pedigree. Além do coeficiente de endogamia e tamanho efetivo da população, estatísticas baseadas na probabilidade do gene de origem fornecem informações importantes para estudar populações sob vários anos de seleção.

Os padrões de DL, por serem decisivos na viabilidade da utilização de dados genômicos no melhoramento genético, têm sido objeto de vários estudos envolvendo diferentes raças no mundo inteiro. Contudo, a extensão do DL ainda não foi bem estudada em populações de gado de corte de países tropicais onde as raças e seus cruzamentos bem como os sistemas de produção diferem de outros países. Com a utilização de parâmetros de DL disponíveis na literatura e do tamanho efetivo estimado nas populações de interesse, é possível simular dados de genotipagem para estudos de viabilidade da seleção genômica em gado de corte.

Este trabalho considerou as seguintes hipóteses: 1) a seleção

intensa para características de valor econômico afeta a diversidade genética da população; 2) o uso da informação genômica nos programas de melhoramento genético de gado de corte trará ganhos em acurácia do Valor Genético Estimado (VGE), porém, o número de animais a serem genotipados necessários para a estimação dos efeitos dos marcadores será alto, se comparado com populações de gado leiteiro, devido a diferentes estruturas de DL. Espera-se que a maior densidade dos painéis disponíveis para genotipagem, aumente a acurácia do VGE.

### **Objetivo geral**

Estimar parâmetros que dêem suporte a estudos de viabilidade da aplicação da seleção genômica em rebanhos de bovinos de corte.

### **Objetivos específicos**

- Estimar parâmetros de diversidade genética de uma população de bovinos Nelore sujeita há vários anos de seleção;
- Estimar a acurácia da seleção genômica utilizando diferentes parâmetros de densidade de marcadores, herdabilidade e número de animais, através da simulação de uma população com estrutura de DL semelhante à encontrada em sistemas de produção de gado de corte.

## 2.REVISÃO BIBLIOGRÁFICA

### 2.1. Diversidade genética

De acordo com UNEP (2010) diversidade biológica ou biodiversidade é o termo usado para descrever o número e variedade de organismos vivos do planeta. A diversidade pode ocorrer em três níveis: diversidade genética (variação em genes e genótipos), diversidade de espécies (riqueza de espécies) e diversidade de ecossistemas (comunidades de espécies e seu ambiente) (Rao and Hodgkin, 2002).

A importância da biodiversidade para a humanidade tem sido bem reconhecida nas últimas décadas. Segundo UNEP (2010), é estimado que 40% da economia global são baseadas em produtos e processos biológicos, o que permite o desenvolvimento sustentável de várias atividades humanas. No campo da agricultura, a biodiversidade permitiu que os sistemas agrícolas evoluíssem e resultassem nas espécies de plantas e animais domésticos altamente produtivos existentes atualmente após milhares de anos de intervenção humana. Porém, a constante diminuição do

número de espécies cultivadas/criadas e a perda de variabilidade dentro destas espécies é alvo de atenção por parte dos pesquisadores.

A diversidade genética é a variabilidade genética existente entre indivíduos de uma variedade/raça ou entre populações de uma espécie (Brown,



1983). As diferenças genéticas podem se manifestar em diferenças na sequência do DNA, em características bioquímicas, fisiológicas ou morfológicas, porém os mecanismos básicos responsáveis por esta diversidade são a mutação e a recombinação. A seleção, a deriva genética e o fluxo de genes agem sobre a frequência dos alelos presentes nas populações alterando o nível de diversidade.

O estudo da diversidade genética é o processo pelo qual a variação entre indivíduos ou grupos de indivíduos ou populações é analisada por um método específico ou uma combinação de métodos. Os dados frequentemente envolvem medidas numéricas e em muitos casos, combinações de diferentes tipos de variáveis. Podem ser usados dados de pedigree, morfológicos ou de marcadores de DNA dependendo do objetivo do experimento, do nível de resolução requerido, dos recursos e infra-estrutura tecnológica disponível, etc.

Entre os métodos utilizados para estudar a diversidade genética está o de estimação de medidas de distância genética e similaridade muito úteis em estudos evolutivos e que podem ser calculadas por diferentes estatísticas dependendo do conjunto de dados. Discussões a respeito de várias medidas de distância genética estão disponíveis na literatura (Nei, 1973; Felsenstein, 1984; Almeida et al., 2000).

Outras abordagens têm sido propostas para quantificar o grau de diferenciação entre populações ou entre subdivisões de uma população. Testes de  $X^2$  usando estatísticas baseadas na frequência alélica tem grande poder para detectar diferenças entre populações ou subdivisões de populações quando a taxa de mutação é baixa. É possível também quantificar a extensão

da diferenciação pela estatística F de Wright (1951) ou a medida análoga de Cockerham (1969, 1973). Uma vez quantificadas as diferenças genéticas entre diferentes amostras/populações, é possível realizar seu agrupamento em diferentes categorias. Métodos estatísticos multivariados se constituem em importante estratégia para lidar com classificação de muitas amostras de acordo com sua variabilidade. Entre estes métodos destaca-se a análise de agrupamento, análise de componentes principais, análise de coordenadas principais e escalonamento multidimensional (Thompson et al., 1998; Almeida et al., 2000; Brown-Guedira et al., 2000).

A informação do pedigree pode ser útil para descrever a variabilidade genética em populações de animais. Os métodos propostos por MacCluer et al. (1986), Lacy (1989) e Boichard et al. (1997) tem sido aplicado com sucesso na análise de variabilidade genética de diversas raças bovinas.

A aplicação das diferentes metodologias de avaliação da diversidade genética é ampla, englobando estudos de taxonomia, evolução e ecologia, envolvendo diferentes organismos vivos. Na área da zootecnia, a crescente preocupação com a perda de diversidade genética dentro e entre as diferentes raças de animais domésticos tem levado a inúmeros trabalhos de descrição e análise genética de populações (Almeida et al., 2000; Vasconcellos et al., 2003; Carneiro et al, 2007; Martinez et al., 2008). Com o resultado destas análises espera-se obter parâmetros para o direcionamento da seleção não só para características de interesse econômico, mas também para o controle da perda de diversidade genética.

### 2.1.1. A origem genética dos bovinos domesticados

As espécies animais importantes para a alimentação e para a produção agrícola foram originadas através de um processo de domesticação de espécies selvagens iniciado há aproximadamente 12.000 anos atrás. Achados arqueológicos indicam que o bovino moderno foi originado do Auroque (*Bos primigenius*), um bovino selvagem de grandes dimensões que vivia na Europa, Ásia e norte da África (Isaac, 1962) e que está extinto há 400 anos, sendo que divergiu de seus ancestrais bovídeos há cerca de 1 milhão de ano atrás.

As raças bovinas atuais se distribuem em dois tipos geográficos: os taurinos encontrados principalmente na Europa, África e Ásia, e os índicos ocorrendo no Sul da Ásia (subcontinente Indiano) e Leste da África. A classificação original coloca estes grupos em espécies diferentes, porém, devido à sua inter-fertilidade, muitos autores consideram como subespécies *Bos taurus taurus* e *Bos taurus indicus*, respectivamente.

A visão, baseada somente em registros arqueológicos, de que todas as raças bovinas modernas têm suas raízes nos centros de domesticação da Ásia Ocidental, tem sido contestada com base em análises de dados de seqüência de DNA mitocondrial, marcadores microssatélites e, mais recentemente, SNPs. Estudos revelam que a divergência entre taurinos e zebuínos ocorreu há mais ou menos 250 mil anos fornecendo fortes evidências de que a domesticação ocorreu em centros independentes (Bradley et al, 1996; Loftus et al., 1994; MacHugh et al., 1997).

As raças bovinas atuais são o resultado do processo de

domesticação que envolveu adaptação, inter cruzamento e seleção e sua variabilidade é expressa tanto em características morfológicas como em produtivas tais como produção de carne, leite, adaptação às doenças e resistência a parasitas, etc. Inicialmente o tamanho efetivo da população de *Bos taurus* era muito grande (>50.000), porém na época da domesticação foi reduzido para cerca de 2.000 e subseqüentemente para menos de 100 indivíduos em algumas raças modernas (Goddard et al., 2008).

No Brasil, com exceção da região Sul, predominam as raças índicas. Os zebuínos foram introduzidos no Brasil no período compreendido entre 1920 e 1962 sendo estimado em 7.000 o número de animais que entraram no país importados da Índia (Santiago, 1985). A excelente adaptação dos animais Nelore às condições ambientais locais ocasionou um grande crescimento da população. Atualmente o rebanho de gado de corte do Brasil é de cerca de 175 milhões de cabeças (FNP, 2010) e estima-se que 80% seja de raças zebuínas e de mestiços (ABIEC, 2009).

### **2.1.2. Tamanho efetivo da população**

O tamanho efetivo ( $N_e$ ) é um importante parâmetro populacional que ajuda a explicar como as populações evoluíram e expandiram ou declinaram e ajuda a melhorar o entendimento e a modelagem da arquitetura genética das características de herança complexa (Reich and Lander, 2001).

O conceito de  $N_e$  foi introduzido por Wright (1931, 1938) e é definido como o tamanho de uma população idealizada (infinitamente grande, sem ação da mutação, migração e seleção) que daria origem à mudança na freqüência

gênica ou na taxa de endogamia observada na população real, isto é,  $N_e = q(1-q)/2\sigma_{\Delta q}^2$  ou  $1/2\Delta F$  sendo  $q$  = frequência gênica,  $\sigma_{\Delta q}^2$  = variância da mudança da frequência gênica e  $\Delta F$  = variação da endogamia. O  $N_e$  mede, então, a taxa de oscilação genética e a endogamia na população. Portanto, se a variância da mudança na frequência gênica ou a taxa de incremento da endogamia são conhecidas, seja através do cálculo das frequências gênicas (se o genótipo ou marcadores genéticos são conhecidos) ou do conhecimento do pedigree, o  $N_e$  pode ser calculado diretamente pelas fórmulas citadas (Caballero, 1994). Uma série de circunstâncias que envolvem a estimação do tamanho efetivo da população e suas implicações são discutidas por Caballero (1994).

Estudando o DNA de 19 raças que diferem geográfica e biologicamente, pesquisadores que compõem o Bovine Hapmap Consortium (2009) observaram que o  $N_e$  recente tem diminuído para todas as raças, refletindo fenômenos como o do “gargalo” associado à domesticação, formação das raças e, em algumas delas, intensa seleção recente para características produtivas. Um menor  $N_e$  sugere baixa diversidade genética, o que representa uma preocupação para a viabilidade das espécies. Neste mesmo trabalho a diversidade foi estudada entre e dentro de subespécies, sendo que entre os subgrupos taurinos e zebuínos a diversidade não foi considerada baixa, apesar do declínio do  $N_e$ . No entanto, a menor diversidade dentro das raças taurinas modernas poderia ser reflexo de uma baixa diversidade dentro da população ancestral pré-domesticação e/ou de efeitos de forte gargalo na formação das raças e forte seleção para docilidade e produtividade pós-domesticação. Os resultados para raças zebuínas sugeriram que o  $N_e$  da fase da pré-

domesticação da população índica, ocorrida no sul da Ásia – um centro de diversidade de espécies - foi muito maior do que do gado taurino.

Para uma característica de herdabilidade 0,30 (peso ao desmame, por exemplo) é necessário um valor de  $N_e=250$  para manter a variância aditiva no seu valor inicial (Hill, 2000). Goddard e Smith (1990) recomendam um tamanho efetivo mínimo de 40 enquanto que Meuwissen e Woollians (1994) recomendam valores entre 30 e 250 para prevenir decréscimos no valor adaptativo.

A estrutura da população determina o desenvolvimento da endogamia e, por sua vez, o  $N_e$ . O uso intensivo da inseminação artificial aliado às técnicas modernas de reprodução tem acelerado o processo de seleção e facilitado a migração de genes. Desta forma, raças muito especializadas e sujeitas a programas intensivos de melhoramento genético tem se baseado na utilização de um número pequeno de reprodutores (Gutierrez et al., 2003; Sorensen, 2005; Hammami et al, 2007; Vozzi et al, 2007; MacEachern et al. 2009).

Faria et al. (2009) calcularam o tamanho efetivo das principais raças zebuínas brasileiras e verificaram que houve uma significativa redução em consequência da baixa quantidade de reprodutores utilizados. Para a raça Nelore os valores foram de 85, 98, 71 e 68 para os períodos de 1979-1983, 1984-1988, 1989-1993 e 1994-1998, respectivamente. Segundo Magnabosco et al (1997), seis touros importados tem contribuído significativamente para a formação do rebanho Nelore brasileiro, com aproximadamente 20% dos genes presentes no atual rebanho sendo proveniente destes touros.

### 2.1.3. Endogamia

Populações com tamanho efetivo pequeno perdem diversidade genética em consequência da deriva genética. Mesmo na ausência de seleção, os alelos presentes numa geração podem, aleatoriamente, se tornar mais ou menos freqüentes, ou até mesmo extinguir-se nas gerações subseqüentes. Esta flutuação das freqüências alélicas resulta na diminuição da diversidade genética, refletida tanto pela perda de heterozigosidade como pela perda de variantes alélicas (Lacy, 1989). Níveis mais baixos de heterozigosidade frequentemente resultam na depressão de valores adaptativos, ou depressão endogâmica, enquanto que a falta de variantes alélicas restringe a resposta à seleção à longo prazo (Falconer & Mackay, 1996). Por isso a endogamia é considerado o principal fator genético que ameaça, à curto prazo, a sobrevivência das espécies, enquanto que a deriva genética é considerada a principal causa de perda de variação genética a longo prazo, ameaçando, portanto, a adaptabilidade das populações.

As estratégias avançadas de melhoramento genético empregadas atualmente nas populações de animais domésticos são efetivas em gerar progresso genético, contudo, focalizam em poucos animais considerados superiores, especialmente machos, o que aumenta a probabilidade de gerar animais endogâmicos (Verrier et al., 1993) ameaçando os programas de melhoramento genético em relação ao ganho genético. Portanto, a endogamia é um importante parâmetro para ser monitorado e controlado num programa de melhoramento genético (Sorensen et al., 2005).

Wright (1922) apresentou a fórmula básica para calcular o

coeficiente de endogamia ( $F$ ) que indica o grau de parentesco entre indivíduos com ascendência comum, ou, de outra forma, é a probabilidade de que, em um locus, dois alelos amostrados aleatoriamente de dois indivíduos sejam idênticos por descendência. O cálculo é baseado no número de gerações até o antepassado comum e pode ser obtido com a informação de pelo menos um dos pais. Porém, mesmo que ambos os pais sejam conhecidos, a endogamia irá ser subestimada se alguns de seus ancestrais não forem identificados (Lutaaya et al., 1999). Boichard et al. (1997) estudaram a robustez do cálculo de endogamia em populações com pedigree incompleto e concluíram que, nestas situações, o  $F$  é subestimado e o tamanho efetivo da população estimado a partir destes valores é superestimado. Este fenômeno foi particularmente mais claro a longo prazo, já que a estimação da endogamia baseia-se no maior caminho de parentesco, o qual é mais provável de ser afetado pela falta de informação.

Em populações que utilizam a inseminação artificial e realizam seleção sistemática com base em efeitos genéticos estimados, os pais não conhecidos, ou por motivo de perda de informação (menos freqüente), ou por motivo de manejo (mais freqüente), podem não diferir geneticamente, em média, daqueles outros pais identificados. Portanto a endogamia média real em animais com pais não identificados pode ser similar à de seus contemporâneos que tem os pais conhecidos. VanRaden (1992) propôs um algoritmo onde a endogamia dos animais cujos pais são desconhecidos é igual à média de seus contemporâneos que têm os pais conhecidos. Lutaaya et al. (1999) testaram este procedimento e concluíram que a endogamia calculada da forma regular



foi fortemente subestimada na geração atual se o pedigree contivesse um grande número de animais não identificados e que o cálculo proposto por VanRaden (1992) forneceu estimativas mais realistas da endogamia especialmente até 20% de perda de um dos pais.

#### **2.1.4. Parâmetros com base na probabilidade da origem do gene**

Conforme discutido anteriormente, uma importante limitação do coeficiente de endogamia para o estudo da diversidade genética de uma população é que ele é muito sensível à qualidade da informação de pedigree disponível (Lutaaya et al., 1999). Em muitas situações algumas informações são perdidas levando a um viés na estimação da taxa de endogamia. Além do mais, as populações domésticas são sujeitas à seleção o que, neste caso, torna menos direta a associação entre a endogamia e a variabilidade genética (Wray et al., 1990; Verrier et al., 1991).

Outra abordagem para estudos de diversidade genética através das relações de pedigree foi inicialmente proposta por Dickinson e Lush (1933), posteriormente elaborada na forma de probabilidades da origem do gene por James (1972) e finalmente modificado por Boichard et al. (1997) para levar em conta efeitos de gargalo no pedigree. Esta metodologia é baseada na estimação de qual proporção dos genes em uma população é derivada de qualquer ancestral específico e pode ser computada traçando o pedigree de todos os indivíduos e, para cada indivíduo, computar a contribuição genética proporcional de cada ancestral como a fração de linhas de ascendência que remonta a esse ancestral (James, 1972). Neste método, as contribuições

genéticas dos fundadores, isto é, ancestrais com pais desconhecidos, da população atual são medidas. Apesar da definição de um fundador ser também muito dependente da informação do pedigree, este método consegue avaliar como um pool genético original tem sido mantido através das gerações (Boichard et al., 1997).

Lacy (1989) propôs uma medida para sintetizar a distribuição das contribuições dos fundadores que é o número de fundadores que contribuem igualmente para produzir o mesmo nível de diversidade da população em estudo, ou, número equivalente de fundadores, ou ainda, número efetivo de fundadores. Foi também proposto estimar o número equivalente de genomas fundadores, isto é, o número de fundadores que contribuem igualmente com nenhuma perda aleatória de alelos fundadores na progênie para produzir o mesmo nível de diversidade da população em estudo.

Em programas intensos de seleção onde um número limitado de animais melhoradores é amplamente disseminado através da inseminação artificial, a computação do número de fundadores é superestimada, já que os vários fundadores contabilizados poderiam ser parentes. Para contornar este problema, Boichard et al. (1997) propuseram a medida de número de ancestrais (fundadores ou não) necessários para explicar a completa diversidade genética da população em estudo.

Estudando a robustez destas estatísticas, Boichard et al. (1997) concluíram que os parâmetros derivados da probabilidade da origem do gene também são afetados pela qualidade do pedigree, porém em menor grau. O número efetivo de ancestrais, por exemplo, foi bem menos afetado, uma vez

que se baseia em menores caminhos de parentesco.

Em suma, os parâmetros da probabilidade da origem do gene representam um simples, porém preciso procedimento para descrever a variabilidade genética após um pequeno número de gerações (Boichard et al., 1997).

## **2.2. O uso da genômica no melhoramento genético animal**

### **2.2.1. Marcadores Moleculares**

A maioria das características de interesse econômico na produção animal é controlada por muitos genes recebendo a denominação de características poligênicas, de herança complexa ou ainda de características quantitativas. Thoday (1961) sugeriu que se um gene de um caráter de herança complexa estivesse ligado a um gene de herança simples, os efeitos fenotípicos do primeiro poderiam ser indiretamente estudados com base nos efeitos do gene vizinho.

Com o advento das tecnologias modernas da Genética e da Biologia Molecular surgiram diversos tipos de marcadores moleculares que detectam polimorfismos genéticos diretamente no DNA. Recentemente, uma nova classe de marcadores tem revolucionado as análises genéticas: marcadores do tipo SNP (*Single Nucleotide Polymorphism*). O recente seqüenciamento do genoma bovino tem permitido detectar um grande número destes marcadores, pois eles são extremamente abundantes e estão distribuídos de forma homogênea pelo genoma. Estudos com humanos e com espécies de interesse zootécnico mostram que pode haver milhões de SNPs no genoma de um indivíduo (Bovine

Genome Sequencing and Analysis Consortium, 2009). Além disso, apresentam outras vantagens como automatização de análise, erros laboratoriais mínimos e baixos índices de erro de genotipagem (índices  $<0,01\%$ ) (Caetano, 2009).

Três tipos de marcadores podem ser definidos: 1) marcadores diretos: loci que codificam para a mutação funcional; 2) marcadores em Desequilíbrio de Ligação (DL): loci que estão em desequilíbrio de ligação com a mutação funcional em toda população e 3) marcadores em Equilíbrio de Ligação (EL): loci que estão em equilíbrio de ligação com a mutação funcional (Dekkers, 2004). Os três tipos de marcadores diferem nos métodos de detecção e na sua aplicação em programas de seleção (Anderson, 2001). Enquanto que os marcadores diretos e, num menor grau, os marcadores DL, são difíceis de detectar, porém são de fácil utilização na seleção, os marcadores EL se caracterizam por uma maior facilidade de detecção, porém são de uso limitado, pois a distância entre o QTL e o marcador não é suficientemente pequena para garantir que a ligação persista em toda a população.

Os marcadores moleculares foram e ainda têm sido utilizados para construção e saturação de mapas genéticos, que, por sua vez, têm possibilitado a cobertura e análise completa de genomas e, portanto, a localização de QTL. O mapeamento pelo desequilíbrio de ligação tem recebido, nos últimos anos, considerável atenção, pois apresenta uma melhor resolução e maior poder do que os demais métodos, porém requer uma densidade de marcadores compatível com a extensão do desequilíbrio na população de interesse. Assim, com o advento dos marcadores moleculares e dos mapas

genéticos cada vez mais densos, o número de trabalhos envolvendo associações entre esses marcadores e os caracteres de herança poligênica tem sido ampliado significativamente.

### **2.2.2. Desequilíbrio de Ligação**

O DL é uma associação não-aleatória entre alelos de diferentes loci, ou seja, quando a frequência de um determinado haplótipo (dois alelos de diferentes loci no mesmo cromossomo) é maior do que o esperado, os loci estão em desequilíbrio de ligação (Ardlie et al. 2002).

O DL pode ser o resultado da migração, mutação, seleção, tamanho (pequeno) da população ou outros eventos genéticos (Lander e Schork, 1994). Nas populações de animais domésticos, o tamanho da população é, geralmente, o que causa o DL, já que o tamanho efetivo de muitas destas populações é pequeno. O DL devido ao cruzamento (migração) é grande quando linhas endogâmicas são cruzadas, porém pequeno quando raças ou animais que não diferem marcadamente nas frequências gênicas são cruzados, sendo que ele desaparece depois de algumas gerações (Goddard, 1991). As mutações são eventos mais raros e que provavelmente ocorreram ao longo de muitas gerações atrás, sendo sua importância muito pequena no DL que observamos atualmente. A seleção, mesmo sendo, provavelmente, um evento causador de DL muito importante, tem o seu efeito localizado ao redor de genes específicos e, portanto, tem um efeito relativamente pequeno na quantidade total de DL do genoma.

Espera-se que a extensão do DL em populações decresça tanto em

função do tempo ( $t$ ) como da taxa de recombinação ( $r$ ) entre os marcadores. Porém, apesar desta tendência ser comumente observada em dados empíricos, marcadores próximos não estão sempre em DL (Ardlie et al, 2001) e, ao contrário, DL foi observado entre marcadores localizados a grandes distâncias um do outro (Collins et al., 1999; Stephens et al., 2001). Isto decorre do fato de que os fatores que governam o DL são numerosos, complexos e algumas vezes vagamente compreendidos. Ardlie et al. (2002) apresentam e descrevem alguns destes fatores relacionados à forças demográficas, moleculares e evolucionárias.

Uma vez que o DL é resultado, predominantemente, do tamanho da população, a extensão do DL em humanos, cujo tamanho efetivo é em torno de 10.000 (Kruglyak, 1999), é menor do que em animais domesticados onde o tamanho efetivo pode ser de 100 ou menos (Riquet et al., 1999). Níveis significantes de DL em humanos se estendem por uma distância menor do que 5Kb ( $\sim 0,005$ cM) (Dunning et al., 2000; Reich et al., 2001), enquanto que em bovinos, níveis consideráveis de DL podem se estender até mais do 5-10 cM (Fanir et al., 2000, Hayes et al., 2003; Sargolzaei et al., 2008; Kim e Kirkpatrick, 2008; de Roos et al., 2008). O nível de DL em gado de leite é, em geral, mais alto do que em gado de corte devido, principalmente, ao menor tamanho efetivo de população nas raças leiteiras (Goddard et al. 2006).

As medidas de DL mais comumente utilizadas são o  $D'$  (Lewontin, 1964) e o  $r^2$  (Hill e Robertson, 1968), ambas assumindo valores entre 0 e 1.  $D'$  foi utilizado em estudos envolvendo gado holandês que detectaram fortes DL por distâncias longas, contudo foi constatado que esta medida é superestimada

quando a amostra é pequena e pela presença de alelos raros na população (Ardlie et al., 2002; Du et al., 2007; Bohmanova et al., 2010). O  $r^2$  representa a correlação entre dois loci e foi demonstrado ser a medida mais adequada para estimar o DL entre marcadores bialélicos, como os SNPs, por ser menos sensível à frequência alélica e ao tamanho da amostra. (Zhao et al., 2007b; Bohmanova et al., 2010).

Trabalhos recentes apresentam valores de DL calculados pelo  $r^2$  com base em marcadores SNPs menores do que aqueles estimados pelo  $D'$  com marcadores microssatélites (Fanir et al., 2000; Tenesa et al., 2003) considerando as mesmas distâncias, ou, de outra forma, os níveis altos de DL se estendem por distâncias mais curtas. Valores de  $r^2$  ao redor de 0,2 foram encontrados a distâncias  $\leq 100$  kb (Goddard e Meuwissen, 2005; Gautier et al., 2007; Khatar et al., 2007; McKay et al. 2007; Marques et al., 2008; Sargolzaei et al., 2008; Kim e Kirkpatrick, 2008; Sargolzaei et al., 2009, Bohmanova et al. 2010).

Os estudos realizados para descrever a extensão e o padrão do DL em bovinos se tornaram mais detalhados à medida que os painéis disponíveis para genotipagem foram se tornando mais densos. Desde 2008 está disponível, tanto para trabalhos científicos como comercialmente, o Illumina BovineSNP50 BeadChip que genotipa para 54.609 SNP uniformemente espaçados (Illumina, 2010a) e, em 2010, foi lançado o BovineHD BeadChip que genotipa para mais que 777.000 SNP (Illumina, 2010b). Utilizando o SNP50 BeadChip para bovinos da raça Holandês, Sargolzaei et al. (2009) encontraram altos níveis de DL ( $r^2 > 0,3$ ) que foram capturados pelos SNPs e

que se estenderam por aproximadamente 40 kb. No entanto, a média de DL entre SNPs localizados a distâncias muito próximas (<10kb) foi menor do que as previamente reportadas usando um painel menos denso (10k) de marcadores (Sargolzaei et al., 2008), possivelmente porque foram observados alguns agrupamentos de SNPs muito próximos no painel de 10k, enquanto que no painel de 50k estes agrupamentos não foram observados, apesar da densidade ser maior. Uma descrição mais detalhada do padrão de DL pode ser usada para determinar a densidade de SNP requerida em diferentes regiões do genoma para mapeamento fino, seleção genômica e para uma melhor interpretação de resultados de mapeamento de QTLs (Sargolzaei et al., 2009).

Os estudos de DL têm sido realizados predominantemente em raças sujeitas à intensa seleção e criadas em países desenvolvidos. Com isso, os padrões de DL de populações de raças puras criadas extensivamente e de animais cruzados permanece pouco investigado. É esperado que, devido ao maior tamanho efetivo da população, a extensão do DL seja menor nestas populações. Estudando uma população de animais *Bos indicus* x *Bos taurus* da África ocidental, Thévenon et al. (2007) relataram que o DL se estendeu por distâncias mais curtas do que aquelas observadas em gado de países desenvolvidos. Lu et al. (2008) estudaram a extensão do DL em raças puras (Angus e Piemontês) e cruzadas de gado de corte e encontraram que a quantidade de DL decresceu rapidamente de 0,31 para 0,22 e 0,15 quando as distâncias entre os marcadores mudaram de 0-30kb para 30-60kb e para 60-100kb, respectivamente. Utilizando uma população mista de animais das raças Holandês, Jersey, e Angus, De Roos et al. (2008) encontrou  $r^2$  médio de 0,35,



0,25, 0,22, 0,14 e 0,06 para distâncias de 10, 20, 40, 100 e 1000 kb, respectivamente. Kelly et al. (2008) calculou o  $r^2$  para uma população de gado de corte multi-racial (progênie de vacas cruzadas acasaladas com touros de raça pura) com predominância das raças Angus, Simental, Limousin, Charolês e Piemontês. O valor do  $r^2$  médio foi de 0,21 para marcadores com distância de 30-35 kb entre eles, o que é um pouco mais baixo do que os encontrados em populações de raças puras.

No Brasil, os trabalhos envolvendo genotipagens densas são ainda muito escassos. Resultados da genotipagem de 25 touros de inseminação artificial da raça Gyr (*Bos taurus* x *Bos indicus*) com o painel de 50.000 marcadores da Illumina, sendo que somente 18098 foram úteis para a análise de acordo com a frequência gênica, mostraram que a média do DL, medido pelo  $r^2$ , entre marcadores adjacentes foi de 0,21 e que houve uma tendência exponencial de decaimento do DL em função da distância genética. Nos intervalos de 0 a 0,1Mb, 0 a 0,2Mb, 0 a 0,5Mb e 0 a 1Mb, a média do  $r^2$  foi de 0,20, 0,18, 0,14 e 0,11, respectivamente. O DL útil para estudos de associação ( $>0,3$ ) (Ardlie et al., 2002) não se estendeu por mais do que 0,5Mb. Porém, o pequeno tamanho da amostra sugere que possa haver viés nestas estimativas de DL e novos estudos com amostras maiores são recomendados pelos autores (Silva et al., 2010).

### **2.2.3. Seleção Assistida por Marcadores**

Os primeiros trabalhos envolvendo marcadores moleculares para caracterização genética de animais e plantas e como ferramenta para o

melhoramento animal datam do final dos anos 70 e final dos anos 80, quando as técnicas de análise do DNA mostraram avanços significativos. Métodos para estimar os efeitos de marcadores genéticos sobre características quantitativas, por exemplo, têm sido propostos desde a década de 70 (Geldermann, 1975; Fernando e Grossman, 1989).

A Seleção Assistida por Marcadores (SAM) tem sido proposta e implementada pela indústria de produção animal (Dekkers, 2004) e, segundo Spelman et al (1999), a sua vantagem sobre a seleção tradicional é proporcional à percentagem da variância genética explicada pelo marcador. A questão chave, portanto, é quantos QTLs são responsáveis pela variação das características quantitativas e quantos destes são necessários para explicar a maior parte da variância genética para uma característica quantitativa (Hayes et al., 2006). Apesar de várias publicações terem demonstrado a existência de QTLs para as principais características de produção, a aplicação destes resultados em populações comerciais tem sido vista com bastante cautela. Somente um número limitado de genes que afetam a característica de interesse foi identificado. Ainda que estes genes representem uma boa porção da variação, o restante permanece obscuro. No que se refere ao tipo de marcador utilizado, quando se trata de marcadores indiretos a seleção não ocorre diretamente no gene e, sim, no marcador que está a uma determinada distância e em desequilíbrio de ligação. Como, através da recombinação, este DL pode se desfazer, a eficiência da seleção para o QTL é reduzida. Além disso, a fase de ligação pode diferir de uma população para outra. Já os marcadores diretos estão localizados no próprio gene, porém podem não ser

causais, ou seja, não estão na região do gene que causa a mudança na característica, portanto, funcionam como os indiretos, com a vantagem de estarem mais próximos da região causal. Finalmente, mesmo considerando marcadores diretos causais, os efeitos do QTL são estimados com base em associações estatísticas entre marcadores e fenótipos. Então, considerando a necessidade de trabalhos de validação, o uso de informação molecular não remove a necessidade da informação fenotípica e, portanto, sofre as mesmas limitações da genética quantitativa, as quais são apresentadas por Dekkers e Hospital (2001) em conjunto com as oportunidades do uso de dados moleculares.

Dekkers (2004) avaliou a utilização dos marcadores moleculares em programas comerciais de melhoramento genético, seus sucessos e limitações, e discutiu estratégias para transpor as limitações. Spelman e van Arendonk (1997) demonstraram que a implementação da SAM com um QTL que foi falsamente identificado causa perda genética comparado com a seleção sem levar em conta o QTL. Estes autores concluem que o QTL deve ser verificado em um estudo de validação antes da implementação de um programa de SAM.

A pesquisa relacionada à QTL foi impulsionada pela saturação dos mapas genéticos aumentando a sua potência de detecção e a acurácia da seleção assistida por marcadores (Grapes et al., 2004; Zhao et al., 2005; Zhao et al., 2007a; Hayes et al. 2007). Exemplos de testes genéticos para QTL disponíveis comercialmente são dados por Dekkers (2004).

Porém, apesar dos esforços dos pesquisadores, relativamente poucos genes foram identificados como responsáveis pelas variações nas

características de importância econômica, resultado que apóia a tese de que as variações neste tipo de característica são consequência da soma de pequenos efeitos de muitos genes e também da ação do meio ambiente (modelo infinitesimal). Além disto, ainda que tenham sido identificados marcadores para genes com efeito significativo, a porção que explicam da variância genética é relativamente modesta em termos totais (10-15%) (Casas et al., 2000; Barendse, 2001; Ciobanu et al., 2004; Casas et al., 2005; Schenkel et al., 2006). Então, as vantagens do uso destes tipos de testes de marcadores ficam por conta do seu uso para características de difícil ou custosa medição, herdabilidade baixa ou de expressão tardia no animal, sendo que o processo de validação é necessário para o uso em uma população diferente daquela onde foi descoberto o marcador.

#### **2.2.4. Seleção Genômica**

A seleção genômica é uma forma de seleção assistida por marcadores do tipo DL que explora potencialmente todos os QTL do genoma (Meuwissen et al., 2001). Estes autores propuseram métodos de predição do valor genético total usando um mapa denso de marcadores cobrindo totalmente o genoma através de simulação. O valor genético foi estimado pela soma dos efeitos de pequenos segmentos do cromossomo (haplótipos) definidos pelos alelos marcadores contidos nestes segmentos. A correlação com o valor genético verdadeiro (acurácia) foi de 0,85. Desta forma foi demonstrado que é possível calcular valores genéticos com alta acurácia a partir da genotipagem para painéis densos de marcadores e os autores denominaram este processo

de seleção genômica. Note-se que, na seleção genômica, não é necessária a busca de marcadores individuais próximos a QTL, pois todos os marcadores e QTL são considerados na análise, sendo o mapa de marcadores assumido suficientemente denso. Além disso, não é necessário o teste para genes específicos se o propósito é o melhoramento genético geral para características quantitativas. Além da proposta destes autores para estimação dos efeitos dos marcadores, outros métodos têm sido propostos (Gianola et al., 2003; Xu, 2003; VanRaden, 2007).

O valor genético obtido da informação advinda dos efeitos estimados dos marcadores tem sido chamado de valor genômico direto (VGD), enquanto que a combinação do VGD e do VGE tem sido chamada de Valor Genético Genômico (VGG). A acurácia de ambas as informações depende: 1) do nível de DL entre os marcadores e os QTL; 2) do número de animais com fenótipo e genótipo na população referência (também chamada de população de treinamento ou de estimação) da qual os efeitos dos SNPs serão estimados; 3) da herdabilidade da característica ou da acurácia da DEP ou do VGE, se este for utilizado como resposta na equação de predição dos efeitos dos marcadores e 4) da distribuição dos efeitos dos QTL (Hayes et al, 2009). Os três últimos itens indicam o quão acurada é a predição dos efeitos dos marcadores (Goddard, 2009).

Para que os marcadores possam predizer os efeitos dos QTL em toda a população e através das gerações, eles devem estar em suficiente DL com os QTL. Este nível pode ser quantificado pelo  $r^2$  que é a proporção da variação causada pelos alelos num QTL que é explicada pelos marcadores.

Meuwissen et al (2001) simularam uma população com  $r^2=0,20$  entre marcadores adjacentes ( $r^2$  calculado a partir do  $N_e$  e da distância genética), portanto, com base neste trabalho, supôs-se que para atingir níveis de acurácia da VGD ao redor de 0,85, o  $r^2$  deveria ter pelo menos este valor. Calus et al (2008) estudaram o efeito do  $r^2$  médio entre marcadores adjacentes sobre a acurácia do VGD e encontraram que a acurácia aumentou drasticamente em função do aumento do  $r^2$ . Para  $r^2=0,1$ , a acurácia foi de 0,68, enquanto que para  $r^2=0,2$  a acurácia foi igual a 0,82.

Os efeitos dos marcadores são estimados em uma população de treinamento, ou seja, uma população com animais genotipados e que possuam fenótipos medidos, que são utilizados na predição genômica de outros animais que pertençam à mesma população tal que o DL entre marcadores e QTL persista da subpopulação de treinamento para a subpopulação de predição. Porém há o interesse na utilização dos efeitos estimados numa determinada população de treinamento na predição genômica de animais não relacionados a esta. Neste caso a confiabilidade das predições genômicas para estes indivíduos pode ser bem menor, tanto quanto forem geneticamente mais distantes. A tendência é de que os conjuntos de dados para estimação dos efeitos dos SNPs sejam acrescidos de animais de diferentes populações e raças, porém, até o momento, poucos trabalhos têm sido feitos neste sentido. Os resultados de Ibáñez-Escriche (2009), utilizando simulação de genotipagem densa, demonstraram que as acurácias baseadas nas equações de predição obtidas de animais cruzados para aplicação em animais puros não foram muito menores do que as acurácias baseadas nas equações estimadas na própria

população de animais puros e foram quase idênticas quando as raças cruzadas eram geneticamente próximas (origem comum). Já quando os animais da população de predição não têm parentesco com os animais da população de estimação ou de treinamento, mesmo sendo da mesma raça, as predições dos valores genéticos genômicos requerem uma densidade de marcadores e número de registros de treinamento substancialmente maiores (Meuwissen, 2009).

A densidade dos marcadores tem um importante efeito na confiabilidade dos VGG porque um maior número de marcadores, quando igualmente distribuídos pelo genoma, irá aumentar a probabilidade de que cada QTL esteja em alto DL com pelo menos um marcador. (Calus et al., 2008; Goddard, 2009). Apesar do DL se estender por mais de algumas centenas de Kb, em função do  $N_e$  atual em populações de animais domésticos, a persistência da fase de ligação entre diferentes raças ou populações somente pode ocorrer se o marcador estiver muito perto do QTL. Em raças taurinas, por exemplo, a persistência só ocorre a distâncias <10 kb (Gautier et al., 2007; De Roos et al., 2008). Estes autores encontraram correlações entre os valores de  $r^2$  de 0,77 em média (Gautier et al., 2007) e próximo a 1 para distâncias <10 kb (De Roos et al., 2008).

A herdabilidade da característica aliada ao número de fenótipos medidos no conjunto de dados de treinamento também afetam a confiabilidade do VGG. Para características com baixa herdabilidade, muitos fenótipos (e genótipos) são necessários para estimar acuradamente os efeitos dos marcadores (Meuwissen et al., 2001; Calus et al., 2008; Goddard, 2009). De

Roos et al. (2009) testaram a hipótese de que à medida que aumente o número de animais na população de treinamento provenientes de diferentes populações, a confiabilidade das predições genômicas poderiam aumentar, já que a confiabilidade depende do número de fenótipos medidos. Porém, quando os efeitos dos marcadores são muito diferentes entre as populações, esta estratégia pode diminuir a confiabilidade, dependendo da densidade dos marcadores.

Schaeffer (2006) propôs e discutiu uma estratégia para aplicar a seleção genômica no melhoramento de gado de leite no Canadá, assumindo que: 1) animais podem ser genotipados para milhares de SNPs de uma só vez onde os SNPs estão localizados a intervalos aproximados de 1cM através do genoma; 2) os efeitos para cada intervalo em uma característica podem ser estimados simultaneamente em um modelo onde estes efeitos são aleatórios; 3) um valor genético genômico pode ser obtido para cada animal somando os efeitos estimados de seu genótipo; 4) a acurácia deste valor genético genômico é na ordem de 80%; 5) o valor genético genômico pode ser estimado quando o animal nasce. O autor concluiu que não é possível ignorar as vantagens potenciais de esquemas de seleção genômica, pois as mudanças genéticas podem ser duas vezes maiores do que as obtidas no esquema de teste de progênie, sendo que a economia nos custos logísticos pode chegar a 92%.

Desde janeiro de 2009, nos Estados Unidos, e de agosto de 2009 no Canadá, provas nacionais de touros da raça Holandesa estão sendo realizadas com base em informações genômicas associadas com os VGE tradicionais. Um amplo trabalho avaliando a confiabilidade das predições genômicas foi



realizado para 27 características produtivas envolvendo 3.576 touros na população de treinamento genotipados para 38.416 marcadores selecionados do painel de 50K (Illumina BovineSNP50 BeadChip) e 1.759 touros na população de predição. As predições genômicas foram combinadas com o valor genético médio dos pais dos touros (VGP) através de um índice de seleção e comparadas com o VGP. Os valores genéticos combinados (VGG) foram mais acurados para todas as 27 características. Em média as confiabilidades dos VGG foram 23% maiores do que dos VGP, aumentando em média a confiabilidade de 27% para 50%, atingindo 78% para percentagem de gordura do leite (VanRaden et al., 2009).

Como parte da colaboração entre pesquisadores da University of Guelph-Canadá e do *United States Department of Agriculture* (USDA), Schenkel et al. (2009) apresentaram resultados de validação das predições genômicas no Canadá, utilizando provas domésticas e provas de touros estrangeiros do *Multi-trait Across-Country Evaluation* (MACE), envolvendo touros Holandeses provados genotipados com o painel de 50k, 6403 no grupo de estimação (treinamento) e 929 no grupo de validação (predição). Um total de 44 características foi analisado. O quadrado da correlação ( $\text{corr}^2$ ) entre os valores genéticos genômicos dos touros obtidos quando eles não tinham filhas avaliadas e os valores genéticos obtidos no último ano para estes mesmos touros (com muitas filhas avaliadas) foi calculado como forma de testar a habilidade de predizer o mérito de touros jovens usando informação genômica. As predições genômicas mostraram incremento do  $\text{corr}^2$  e de confiabilidade observada ( $\text{corr}^2$  ajustado para o fato de que a acurácia dos VGE não é perfeita

e que os touros jovens não são uma mostra aleatória) para 42 das 44 características avaliadas. Para características de produção o aumento médio do  $\text{corr}^2$  foi de 0,27 pontos e da confiabilidade foi de 0,29 pontos. O uso das provas do MACE de touros estrangeiros para predição em adição as provas domésticas aumentou o  $\text{corr}^2$  e a confiabilidade quando comparados com o uso exclusivo de provas domésticas, o que foi provocado, provavelmente, pelo fato do número de touros no grupo de predição ter quadruplicado ao incluírem-se as provas do MACE. Este resultado ilustra a importância de grandes bancos de dados de genótipos para a estimação dos efeitos dos marcadores.

Em gado de corte ainda predominam os trabalhos de simulação, porém bancos de dados importantes estão sendo avaliados e aumentados para que os benefícios da utilização da informação genômica sejam claramente entendidos. Nos EUA, os estudos atuais tem tido um forte enfoque na raça Angus através da genotipagem de cerca de 2000 touros de inseminação artificial cujo sêmen é amplamente utilizado por criadores desta raça, possuindo DEP acuradas para várias características produtivas (Garrick, 2010). O *U.S. Meat Animal Research Center* (MARC) tem trabalhado com cerca de 2.000 touros de 16 raças de corte (Angus, Beefmaster, Brahman, Brangus, Braunvieh, Charolês, Chiangus, Gelbvieh, Hereford, Maine-Anjou, Red Angus, Salers, Santa Gertrudes, Shorthorn, Simental) com o propósito de gerar informação a respeito da predição dos efeitos dos marcadores sobre populações cruzadas (Thallman, 2009). O *U.S. carcass merit Project* (CMP) possui um amplo banco de dados com mais de 8.000 animais que possuem medidas de carcaça, maciez da carne e outros atributos sensoriais. Alguns

destes animais, representando 13 raças, foram genotipados para marcadores específicos para estas características que foram validados pelo *National Beef Cattle Evaluation Consortium* (NBCEC). Atualmente, genotipagens para painéis densos estão sendo feitas objetivando a descoberta de genes e predição genômica. Outros bancos de dados fomentados por Universidades americanas e companhias privadas estão sendo avaliados (ver Garrick, 2010).

O Canadá e Austrália, além dos Estados Unidos, se constituem em importantes centros de pesquisa e investimento na área da genômica aplicada. As Universidades de Alberta e de Guelph, no Canadá, e o *Australian Cooperative Research Center for Beef Cattle* (Beef CRC), na Austrália, tem importantes projetos em colaboração com os EUA envolvendo fenótipos de difícil medição como eficiência alimentar e qualidade de carcaça.

No Brasil, projetos financiados por empresas governamentais de fomento à pesquisa e com a colaboração de companhias privadas de melhoramento genético animal que executam medições relacionadas à qualidade de carcaça, resistência a parasitas e outras medidas de importância econômica, tem sido aprovados.

Em gado de corte, três questões básicas estão sendo preferencialmente estudadas: 1) a proporção da variação genética que pode ser predita pelo conhecimento dos genótipos para painéis de 50K; 2) a perda da habilidade preditiva quando os efeitos estimados na população de treinamento são aplicados em animais de raças diferentes; 3) a habilidade de um painel reduzido predizer confiavelmente o desempenho (Garrick, 2010).

Um estudo envolvendo genotipagem densa (50k) de touros da raça

Angus e validação cruzada resultou em valores de correlação entre as predições genômicas e o desempenho dado pelas DEP na ordem de 0,5 a 0,7, ou seja, 25% a 50% da variância das DEP foi explicada pelos efeitos dos marcadores. Isto é equivalente a 6-16 produtos num teste de progênie convencional para uma característica com herdabilidade 25% (Garrick, 2009). O autor comenta que estas correlações são tão boas quanto àquelas obtidas em populações de gado de leite antes das predições genômicas serem incorporadas à informação da média dos pais.

Avaliando a possibilidade de utilização de painéis menos densos e, portanto, de menor custo, Garrick (2009) testou, utilizando os mesmos dados citados anteriormente, subconjuntos de 600 SNPs mais informativos selecionados do painel de 50K através de técnicas bayesianas e encontrou que houve uma pequena perda da habilidade preditiva em reduzir o painel desta maneira. Numa abordagem subsequente, este mesmo autor relatou resultados onde até 200 marcadores específicos para marmoreio foram escolhidos entre os 600 marcadores e as correlações entre as DEP genômicas e o fenótipo observado obtidas através de uma análise bivariada, semelhante à análise utilizada para estimar correlações genéticas entre um par de características, foram de 0,28, 0,29, 0,39 e 0,43 para 50, 100, 150 e 200 marcadores respectivamente. Porém, uma vez que esta abordagem disponibiliza painéis específicos para determinada característica, foram reunidos marcadores para mais de uma característica em um painel de 384 SNPs e testada a correlação entre a DEP genômica ou score do marcador e a DEP dada pelo teste de progênie em uma nova amostra de 275 touros Angus que não tinham sido

utilizados em nenhuma análise de treinamento. Os resultados foram de 0,59 para marmoreio, 0,32 para gordura subcutânea, 0,58 para área de olho de lombo, 0,44 para peso de carcaça, 0,39 para prenhez de novilha e 0,35 para peso ao ano. Então, tal painel poderia explicar de 10% a 35% da variância da DEP para estas características.

Outra abordagem para a genotipagem de animais para painéis menos densos (3k) envolve a técnica de imputação que consiste na predição de marcadores SNP não genotipados sendo possível transferir informações a partir de animais genotipados para painéis de alta densidade (Druet et al., 2010; Zhang e Druet, 2010) de tal forma que painéis de baixa densidade poderiam explicar similar magnitude da variação genética aditiva do que os painéis de alta densidade. A eficiência da técnica depende da taxa de erro da imputação, que, por sua vez, depende: a) do número de SNP no painel de baixa densidade; b) do parentesco entre os animais genotipados para os painéis de alta e baixa densidade; c) do número de indivíduos na população de referência (Zhang e Druet, 2010; Druet et al., 2010). A eficiência da imputação também depende de qual método é utilizado (Zhang e Druet, 2010; Sargolzaei et al., 2010c). Zhang e Druet (2010) relataram que os alelos dos marcadores não genotipados pelo painel menos denso são preditos com uma taxa média de erro de imputação entre 3% a 4% quando aproximadamente 3.000 SNP igualmente espaçados são genotipados. Sargolzaei et al. (2010a) estudaram o processo de imputação com base em três diferentes cenários onde as mães de animais jovens da raça Holandesa, genotipados para painéis de 3K, foram genotipadas para painéis de 50K, 3K e não genotipadas. Os autores

verificaram que 99,6%, 92,23% e 42,47% dos animais jovens apresentaram acurácia de imputação maior ou igual a 0,95, respectivamente. Sargolzaei et al. (2010b) investigaram o efeito da imputação do painel de 3K SNP para o painel de 50K SNP na acurácia dos Valores Genômicos Preditos e verificaram que a redução na confiabilidade foi, em média, ao redor de dois pontos (0,02) para várias características.

Em um estudo de simulação (Kizilkaya et al., 2010), foi avaliada a perspectiva de utilização dos resultados da população de treinamento de uma raça para prever o desempenho de animais de outras raças. Num cenário onde o número de genes causais variou de 50 a 500 para uma característica de  $h^2=0,50$  e na ausência de dominância e epistasia, os resultados mostraram que a habilidade preditiva do painel de 50K diminuiu consideravelmente quando o número de genes causais aumentou. Na melhor situação, a correlação entre as predições genômicas e o desempenho realizado caiu de 0,4, para 50 genes, para 0,2-0,3 para 500 genes. Além disso, foi testada a aplicação da estimação dos efeitos dos marcadores da população de raça pura na população de múltiplas raças e vice-versa, sendo que os melhores resultados foram obtidos quando os efeitos foram estimados na raça pura para prever as outras raças. Isto pode ser explicado pelo fato de que o DL é, na média, maior nas raças puras do que nas populações envolvendo várias raças.

## **CAPÍTULO II<sup>1</sup>**

---

<sup>1</sup> Conforme normas para publicação na Genetics and Molecular Biology

## **Pedigree analysis in a large Brazilian Nellore herd**

**F. V. Brito<sup>1</sup>, M. Sargolzaei<sup>2</sup>, J. Braccini<sup>1</sup>, J. A. Cobuci<sup>1</sup>, C.M. Pimentel, J. Barcellos and F. S. Schenkel<sup>2</sup>**

<sup>1</sup>Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul, Caixa Postal 15100, Porto Alegre, RS, Brasil, 91540-000.

<sup>2</sup>Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada, N1G2W1.

Key words: pedigree analysis, Nellore, effective population size, inbreeding, probability of gene origin

Corresponding author: Fernanda Varnieri Brito

Rua Guilherme Alves 170/304, Porto Alegre, RS, Brasil, 90480-000

[fernandabrito@gensys.com.br](mailto:fernandabrito@gensys.com.br)



## ABSTRACT

A large herd of Nellore cattle was studied, using a depth pedigree analyses. The inbreeding coefficient calculated taking into account the incomplete pedigree due to the use of multiple young sires mating was 1.73% for the last generation, which was higher than the regular inbreeding coefficient (0.25%). The effective population size was estimated to be 114, 245 and 101 for the time periods 1995-1999, 1999-2003 and 2003-2007, respectively. Parameters based on the probability of gene origin were used to describe the genetic diversity over time in the herd. The effective number of founders, ancestors and founder genomes decreased over time, showing loss of genetic diversity. In the last period (2003-2007), based on available pedigree information, one prominent ancestor contributed 10.6% to the gene pool of the herd and 30% of this pool was explained by 31 ancestors. The analysis of inbreeding under random mating indicated that the mating strategies used in the herd are lessening the inbreeding rates. However, it is important to continue monitoring the genetic diversity of this herd.

## INTRODUCTION

In recent decades, due to advances in animal breeding methods and in technification of farms, more Nellore cattle herds have been under intensive genetic improvement programs. These programs, based on mixed model methodology and the use of artificial insemination, brought more accurate

genetic evaluations and decreased generation intervals, resulting in short term genetic gains (Henderson, 1975; Fernando and Gianola, 1986; Kennedy et al. 1988), but increased inbreeding rates and reduces genetic diversity, since related animals are more likely to be selected (Quinton et al. 1992; Verrier et al., 1993). Therefore, inbreeding and genetic diversity parameters should be monitored in animal breeding programs. One way to analyze the genetic diversity of a population is based on information provided by the pedigree records. Besides the coefficient of inbreeding and effective population size, statistics based on the probability of gene origin (MacCluer et al., 1986, Lacy, 1989; Boichard et al., 1997) provide valuable information for studying a population undergoing several years of selection. In Brazil, the beef cattle population is mainly formed by Nellore cattle. Artificial insemination (AI) is used as a tool for genetic improvement, but because of large herds are managed in large paddocks, where the use of single sire mating is not feasible, the use of multiple selected young sires (MS, groups of bulls mated to groups of cows) is common. Due to the average high genetic value of the young sires, their use results in genetic gain, but also in loss of pedigree information resulting in an underestimation of the inbreeding coefficient (Lutaaya et al., 1999). However, missing pedigree information can be taken into account while computing inbreeding coefficients (VanRaden, 1992; Colleau and Sargolzaei, 2011). Therefore, this study was carried out to: 1) estimate the inbreeding coefficients and the effective population size of a large Nellore herd, under selection and that uses MS mating; 2) analyze the genetic variability of this herd using the statistics derived from the probability of gene origin.

## MATERIAL AND METHODS

### Data

A large Nellore cattle herd belonging to Agropecuaria Jacarezinho, which participates in a large breeding program in Brazil with others breeders, consisting of 146,291 animals born from 1984 to 2007, was investigated. In addition, the database of the Brazilian Association of Zebu Breeders was used to trace back the pedigree of the herd to a 1938 base, resulting in 162,244 genealogical records. The data set was divided into three periods of five years (1995-1999, 1999-2003 and 2003-2007) and the parameters of genetic diversity were calculated for each period. The genetic evaluation in this population is based on BLUP estimates of breeding values, using an individual animal model. Every year the best two-year old young bulls (20%) are used for MS mating and the top 1% bulls are used for AI. In addition, semen of bulls from other farms participating in the breeding program and from insemination companies are used.

### Methodology

**Inbreeding coefficient:** The regular inbreeding coefficients (Wright ,1922) were calculated using the algorithm described by Sargolzaei et al. (2005), which is based on the indirect method proposed by Colleau (2002).The inbreeding coefficients were also calculated using the method proposed by VanRaden

(1992;VRM), which accounts for missing pedigree information. This method assigns to the animals with unknown parent(s) the average inbreeding coefficient of the animals of the same generation. The inbreeding coefficients expected under random mating ( $F_{ST}$ ) proposed by Wright (1969) were calculated and compared to the regular inbreeding coefficients, resulting in a statistics that describes the deviation from random mating ( $f_{is}$ ), which allows to assess the mating strategies used in the breeding program.

**Effective population size ( $N_e$ ):** The  $N_e$  was calculated as:

$$N_e = 1 / (2\Delta F_t L)$$

where  $\Delta F_t$  is the annual rate of inbreeding.  $F$  was calculated as proposed by VanRaden (1992) and  $L$  was the average generation interval.

**Effective number of founders ( $E_f$ ):**  $E_f$  is defined as the number of equally contributing founders which would produce the same amount of genetic variability as in the current population. This contribution was calculated as:

$$E_f = 1 / \left( \frac{1}{n_g} \sum_{i \in G} \sum_{j \in F_0} t_{ij}^2 \right)$$

where  $F_0$  is the set of founders,  $G$  is the set of individuals in the reference population,  $n_g$  is the number of individuals, and  $t_{ij}$  is an element of matrix  $T$  that represents the fraction of the genes that individual  $i$  inherited from founder  $j$  (Lacy, 1989).

**Effective number of founder genomes ( $E_{fg}$ ):**  $E_{fg}$  is defined as the number of

equally contributing founders with no loss of founder alleles that would give the same amount of genetic diversity as in the current population. It was calculated as Lacy (1989):

$$\frac{1}{\bar{f}}$$

where  $\bar{f}$  is the average coancestry in the population.

**Effective number of non-founders ( $E_{nf}$ ):**  $E_{nf}$  accounts for loss of genetic diversity due to drift accumulated over non-founder generations. It was calculated as Caballero and Toro (2000):

$$\frac{1}{\sum_{i=1}^{n_a} p_i^2}$$

**Effective number of ancestors ( $E_a$ ):**  $E_a$  was proposed by Boichard et al. (1997) and determines the minimum number of ancestors (founders or not) required to explain the complete genetic diversity in the current population. It was calculated as:

where  $p_i$  is the marginal genetic contribution of ancestor  $i$  to the population, and  $n_a$  is the number of ancestors. In this study the 1000 most contributing ancestors were used.

**Software:** The CFC software (Sargolzaei et al., 2006b) was used to calculate

pedigree statistics, regular inbreeding coefficient,  $E_f$ ,  $E_{fg}$ , and  $E_{nf}$ . Inbreeding expected under random mating was calculated using the EVA software (Berg et al., 2007). The PEDIG package (Boichard, 2002) was used to calculate generation intervals,  $E_a$  and to identify ancestors with the highest genetic contributions.

## RESULTS AND DISCUSSION

Out of the total of 162,244 animals, 9.2% were considered founders (both sire and dam were unknown), 34.4% had both parents known, and 56% had only the dam known. The regular average inbreeding for animals born in 2007 was 0.25%, but was 1.73% when calculated by VRM, which was nearly 7 times higher (Figure 1a). This sizeable difference can be attributed to the large number of unknown sires and it indicates that VRM recovered some of the inbreeding missed by the regular method. Lutaya et al. (1999) tested the efficacy of this method by simulating the loss of dam information, ranging from 10% to 50%. The deletion of dam's information was done randomly and the dams could have come from any of the previous generations and, therefore, the authors chose to assign the average inbreeding coefficient of the animals born in the 4 years preceding the birth of their progeny. They concluded that, under these conditions, with VRM, the average inbreeding decreased only 30% when 80% of dams were identified and by 78% when 50% of dams were identified. VRM was able to recover most of the inbreeding with up to 20% of missing pedigree information, and was superior to the regular inbreeding for all fractions

of missing dams. As in the current data set the bulls used in MS matings were all selected from the same generation that preceded the birth of their offspring, VRM seemed to be more efficient than in Lutaya and colleagues' study, mainly with the highest levels of loss of pedigree records.

The average inbreeding increased across the years indicating a loss of diversity, possibly caused by selection and the consequent use of animals with some degree of relationship. Regular Inbreeding levels in Nelore have been reported by Faria et al. (2009) and were slightly higher (around 2.0%) than those found using VRM in the current study. For other Brazilian Zebu breeds, Faria et al. (2002) reported values of  $F$  ranging from 0.98 in the Polled Nelore to 10.13% in the Sindi breed. With a decreasing trend over the years, Faria et al. (2010) reported a  $F$  of 0.42% for the Brahman breed that, differently from the Nelore breed, would be expanding.

The rates of inbreeding fluctuated across the periods considered in the current study, showing an overall increasing trend in inbreeding (Figure 1a). However, Figure 1b shows that the deviations from random mating were mostly negative since 1995, suggesting that the average observed inbreeding of the herd was lower than the level expected under random mating, indicating that mating is occurring among animals less closely related. The inbreeding under random mating was on average 1.3 times higher than the inbreeding coefficient calculated using the actual matings, reaching levels 4 times higher in 2005 (Figure 1b). This difference indicates an appropriate control of matings by using a software called PAD, which was adopted in the population since the late '90s (Roso, 1998). Despite the fact that the inbreeding coefficients in Figure 1b have

been calculated by the regular method, the difference in inbreeding levels between the random mating and actual matings is still relevant. The increase in inbreeding rate and the reduction of the difference between inbreeding under random and actual matings after 2005 (Figures 1a and 1b) should be monitored.

As for the rates of inbreeding, the  $N_e$  showed fluctuations since 1995 (Table 1). In the first period, with the intensification of use of AI and the use of an important ancestor (Kulal AJ) that was born in 1993 (Table 2), the estimated  $N_e$  was only 114. In the subsequent period  $N_e$  was increased, which could be speculated to be a consequence of the beginning of intense use of the PAD mating software. More recently, there was a decrease in  $N_e$ , which could be explained by an increasing use of descendants from an important ancestor (Kulal AJ, Table 2) who contributed to 10.6% of gene pool of the herd in the 2003-2007 period and, despite of the control of matings between closely related animals.

Studying the DNA of 19 breeds that differ geographically and biologically, the Bovine Hapmap Consortium (2009) observed that the recent  $N_e$  has decreased for all breeds, reflecting phenomena like bottleneck associated with domestication, breed formation, and, in some breeds, recent intense selection for production traits. Faria et al. (2009) estimated the  $N_e$  of the main Brazilian zebu breeds and found a significant decrease due to the use of few sires. The values for Nelore were 85, 98, 71 and 68 across the studied periods (79-83, 84-88, 89-93 and 94-98, respectively). According Magnabosco et al. (1997), six bulls have significantly contributed to the formation of Nelore cattle in Brazil,



with approximately 20% of genes of the current herds coming from these bulls. For the period of 2003-2007 in the current study, eight ancestors contributed to 20% of the gene pool (Table 2), while 31 ancestors contributed to 30%. These contributions might have been underestimated due to high missing pedigree information. However, studies involving other herds of Brazilian Nelore reported similar contributions from the first 30 main ancestors (Faria et al., 2009; Vozzi et al., 2006; Vozzi et al., 2007).

Despite the fact that a mating system is being used to control inbreeding rates, the statistics based on the probability of gene origin show that loss of genetic diversity is occurring (Table 1). The  $E_f$  values are high because of the high number of unknown parents. The  $E_{fg}$  is decreasing and its value for animals born in recent years was around 70. It shows that, besides the unequal founder contributions, bottleneck and genetic drift were important causes of genetic diversity loss in this herd.  $E_{nf}$  is only slightly higher than  $E_{fg}$ , what indicates that most of the genetic drift that occurred in the herd accumulated over non-founder generations. The ratio between  $E_f$  and  $E_a$ , which assess possible bottlenecks, increased with the time, indicating that the herd probably is not going through a recent bottleneck. Faria et al. (2010) reported ratios of 17.5 and 3.7 for two periods for Brahman breed, showing the existence of a bottleneck process in this breed.

The ratio between  $E_{fg}$  and  $E_f$  indicates the magnitude of genetic drift and the lower the value, the greater the impact of genetic drift on loss of diversity. In this study the ratios were 0.25, 0.17 and 0.14 for the three periods, indicating that the importance of genetic drift is increasing. Nevertheless recent values of  $E_f$ ,

$E_a$  and  $E_{fq}$  are larger than those presented by Faria et al. (2009) for the entire registered Nelore population in Brazil, which indicates a greater genetic diversity in this particular evaluated herd.

## **CONCLUSIONS**

Even though the pedigree records contained a large number of unknown sires in the analyzed Nelore herd, the use of VanRaden's method for calculating inbreeding coefficients taking into account the missing pedigree records as well as the use of the statistics based on the probability of gene origin seemed suitable for assessing the genetic diversity in the herd. Although the analyses showed loss of diversity, the mating strategies used in the herd seem to be lessening the rate of loss. However, it is important to continue monitoring the diversity of this population.

## **ACKNOWLEDGEMENTS**

The authors thank Agropecuaria Jacarezinho and the Brazilian Association of Zebu breeders (ABCZ) for providing the data for this study. The first author was sponsored by a scholarship from the Brazilian Federal Agency for Support and Evaluation of Graduate Education – Capes.

## REFERENCES

- Berg P, Sørensen MK and Nielsen J(2007) EVA interface user manual. <http://eva.agrsci.dk/>. Accessed on Fev. 2009.
- Boichard D (2002) Pedig: a Fortran package for pedigree analysis suited for large populations. In: *Proc 7th WCGALP*, Montpellier, France. CD-ROM, communication 28-13.
- Boichard D, Maignel L and Verrier E (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genet Sel Evol* 29:5–23.
- Bovine HapmapConsortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324:528-532.
- Caballero A and Toro MA (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet Res Camb* 75: 331-343.
- Colleau JJ (2002) An indirect approach to the extensive calculation of relationship coefficients. *Genet Sel Evol* 34:409-421.
- FariaFJC, FilhoAEV, Madalena FE and Josahkian LA (2002) Pedigree analysis in the Brazilian zebu breeds. In: *Proc 7th World Congress on Genetic Applied to Livestock Production*, Montpellier, France. CD-ROM, communication 26-29.
- Faria FJC, FilhoAEV, Madalena FE and Josahkian LA (2009) Pedigree analysis in the Brazilian zebu breeds. *J Anim Breed Genet* 126:148–153.

- Faria LC, Queiroz SA, Vozzi PA, Lôbo RB, Magnabosco CU and Oliveira JA (2010) Variabilidade genética em uma população de bovinos da raça Brahman no Brasil. Análise de pedigree. (2010) In: *Anais XV Congresso Mundial da raça Brahman*, Uberaba, Brasil.
- Fernando RL and Gianola D (1986) Optimal properties of the conditional mean as a selection criterion. *Theor Appl Genet* 72:822-825.
- Henderson CR, (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
- Kennedy BW, Schaeffer LR and Sorensen DA (1988) Genetic properties of animal models. In: *ProcAnim Model Workshop*, Edmonton, Alberta. *J Dairy Sci* 71(suppl 2):17-26.
- Lacy RC (1989) Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. *Zoo Biol* 8:111-123.
- LutaayaE, Misztal I, Bertrand JK and Mabry JW (1999) Inbreeding in populations with incomplete pedigrees. *J Anim Breed Genet* 116:475–480.
- MacCluer JW, VandeBerg JL, Read B and Ryder OA (1986) Pedigree analysis by computer simulation. *Zool Biol* 5:147-160.
- Magnabosco CU, Cordeiro CMT, Trovo JBF, Mariante AS, Lôbo RB (1997) Catálogo de linhagens do germoplasma zebuino: raça Nelore. Embrapa-Cenargen, Documentos, Brasília, 23.
- Quinton M, Smith C and Goddard ME (1992) Comparison of selection methods at the same level of inbreeding. *J Anim Sci* 70: 1060-1067.

- Roso VM and Fries LA (1998) PAD – Um programa para planejar acasalamentos em bovinos de corte. In: *Proc II Simposio Nacional da SBMA*, Uberaba, Brazil. p. 359-360.
- Sargolzaei M, Iwaisaki H and Colleau JJ (2005) A fast algorithm for computing inbreeding coefficients in large populations. *J Anim Breed Genet* 122:325-331.
- Sargolzaei M and Colleau JJ (2011) MIM: an indirect method to assess inbreeding and coancestry in large incomplete pedigrees of selected dairy cattle. *J Anim Breed Genet* 1-11.
- Sargolzaei M and Colleau JJ (2006a) Decomposing inbreeding and coancestry into ancestral components. In: *Proc 8th WCGALP*, Belo Horizonte, Brazil, CD-ROM, communication 30-03.
- Sargolzaei M, Iwaisaki H and Colleau JJ (2006b) CFC: a tool for monitoring genetic diversity. In: *Proc 8th WCGALP*, Belo Horizonte, Brazil, CD-ROM, communication 27-28.
- VanRaden PM (1992) Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J Dairy Sci* 75:3136-3144.
- Verrier E, Colleau JJ, Foulley JL (1993) Long-term effects of selection based on the animal model BLUP in a finite population. *Theor Appl Genet* 87:446-454.
- Vozzi PA, Marcondes CR, Bezerra LAF and Lôbo RB (2007) Pedigree analysis in the breeding program for Nelore cattle. *Genet Mol Res* 6(4):1044-1050.

- Vozzi PA, Marcondes CR, Magnabosco CU, Bezerra LAF and Lôbo RB (2006) Structure and genetic variability in Nelore (*Bosindicus*) cattle by pedigree analysis. *Genet Mol Biol* 29(3):482–485.
- Vozzi, PA, Marcondes CR, Bezerra LAF and Lôbo RB (2007). Pedigree analysis in the breeding program for Nelore cattle. *Genet and Mol Res* 6(4):1044-1050.
- Wright S (1922) Coefficients of Inbreeding and relationship. *Amer Nat* 56:330-338.
- Wright S (1969) *Evolution and the genetics of populations* The University of Chicago Press, Chicago, USA.

Table 1: Generation interval (Gen Int), inbreeding calculated by VanRaden's method ( $\Delta F$ ) and effective numbers in the Nellore herd

Period	Gen Int	$\Delta F$	$N_e^1$	$E_f^2$	$E_a^2$	$E_{fg}^2$	$E_{nf}^2$
1995-1999	6,85	0.00064	114	607	239	152	204
1999-2003	6,47	0.00032	245	436	87	74	89
2003-2007	6,83	0.00073	101	479	77	67	78

<sup>1</sup> $N_e$ =effective population size; <sup>2</sup> $E_f$ ,  $E_a$ ,  $E_{fg}$ ,  $E_{nf}$ = effective numbers of founders, ancestors, founder genomes and non-founders.

Table 2: The eight ancestors with highest genetic contributions to the gene pool of animals born between 2003 and 2007.

Ancestor	Sex	Birth Year	Number of offspring	Number of descendants	Genetic contributions (%)		
					Total	Marginal	Cumulated
Kulal AJ	M	1993	7,593	20,745	10.6	10.6	10.6
Gim de Garça	M	1976	333	29,154	2.1	2.1	12.7
Tiete	M	1987	1,351	7,409	1.6	1.6	14.3
Cabral AJ	M	1997	892	1,722	1.5	1.5	15.7
Solimões AJ	M	1995	777	2,700	1.3	1.3	17.1
Karvardi	M	1955	89	50,127	1.8	1.2	18.3
Di Cavalcanti AJ	M	1998	646	1,408	1.2	1.0	19.3
Rondon AJ	M	1997	533	956	0.9	0.9	20.1



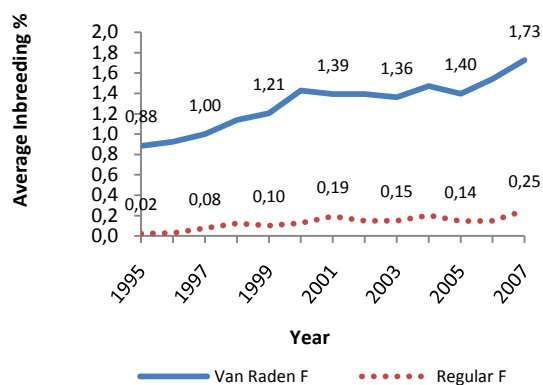


Figure 1a

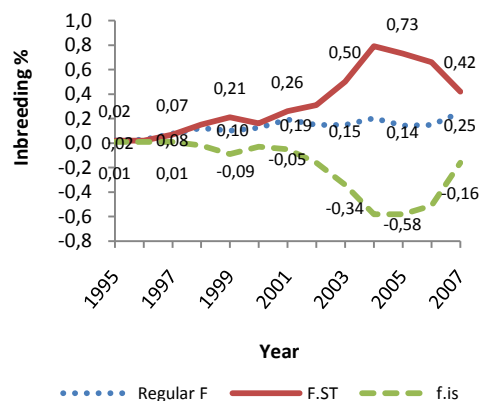


Figure 1b

Figure 1a .Average regular inbreeding coefficient and average inbreeding coefficient calculated by VanRaden's method; Figure 1b. Average regular inbreeding coefficient, inbreeding expected under random mating (F.ST) and deviation from random mating (f.is)

## **CAPITULO III<sup>1</sup>**

---

<sup>1</sup> Conforme normas para publicação na BMC Genetics

**Accuracy of genomic selection in a simulated population mimicking the  
extent of linkage disequilibrium in beef cattle**

**Fernanda V Brito<sup>1,2\*</sup>, José Braccini Neto<sup>2</sup>, Mehdi Sargolzaei<sup>1,3</sup>, Jaime A  
Cobuci<sup>2</sup>, Flavio S Schenkel<sup>1</sup>**

<sup>1</sup>Centre for Genetic Improvement of Livestock, University of Guelph, Guelph,  
ON, Canada, N1G2W1.

<sup>2</sup>Departamento de Zootecnia, Universidade Federal do Rio Grande do Sul,  
Caixa Postal 15100, Porto Alegre, RS, Brasil, 91540-000.

<sup>3</sup>L'Alliance Boviteq, Saint-Hyacinthe, QC, Canada, J2T5H1.

\*Corresponding author

Email addresses:

FVB: [fbrito@uoguelph.ca](mailto:fbrito@uoguelph.ca)

JBN: [jose.braccini@ufrgs.br](mailto:jose.braccini@ufrgs.br)

MS: [msargol@uoguelph.ca](mailto:msargol@uoguelph.ca)

JAC: [jaime.cobuci@ufrgs.br](mailto:jaime.cobuci@ufrgs.br)

FSS: [schenkel@uoguelph.ca](mailto:schenkel@uoguelph.ca)

## **ABSTRACT**

### **Background**

The success of genomic selection depends mainly on the extent of linkage disequilibrium (LD) between markers and quantitative trait loci (QTL), the number of animals in the training set (TS) and the heritability ( $h^2$ ) of the trait. The extent of LD depends on the genetic structure of the population and the density of markers. The aim of this study was to calculate accuracy of Direct Genomic Estimated Breeding Values (DGEBV) using GBLUP method for different levels of marker densities, heritabilities and sizes of the TS of simulated populations that mimicked the extent and pattern of reported LD in beef cattle.

### **Results**

The accuracy of DGEBV increased significantly ( $p < 0.05$ ) with the increase in the number of bulls in the TS (480, 960 or 1920), trait  $h^2$  (0.10, 0.25 or 0.40) and marker densities (40k or 800k). Increasing the number of animals in the TS by 4-fold and using phenotypes to calculate marker effects was not sufficient to maintain or increase the accuracy of DGEBV obtained using estimated breeding values (EBVs) when the trait  $h^2$  was lower than 0.40 for both marker densities. Comparing to accuracies of parentage average (PA), the gains by using DGEBV would be of 27%, 13% and 10% for trait  $h^2$  equal to 0.10, 0.25 and 0.40, respectively, considering the scenario with 40k markers and 1920 bulls in TS.

## **Conclusions**

As reported in dairy cattle, the size of the TS and the extent of LD have major impact on the accuracy of DGEBV. Based on the findings of this simulation study, bigger TS, as well as denser marker panels, aiming to increase the level of LD between markers and QTL, will likely be needed in beef cattle.

It is expected that, in the absence of a sufficient number of bulls with accurate EBVs for estimating the marker effects, the required number of bulls in the TS to produce accurate DGEBV will be even higher. A denser panel is expected to increase accuracy of DGEBV, but its use should be accompanied by an increase in number of animals in the TS.

## **BACKGROUND**

Genomic selection is a LD marker-assisted selection that potentially explores all QTL in the genome [1]. The breeding value (BV) is estimated by the sum of the effects of marker alleles or haplotypes covering the entire genome and its accuracy could be as high as 0.85 [1].

The BV calculated from the estimated effects of markers is often called DGEBV and the blended value between DGEBV and traditional EBV is often called Genomic Estimated Breeding Value (GEBV). The accuracy of DGEBV and GEBV depends on: 1) the level of LD between markers and QTL; 2) the number of animals in the TS; 3) the heritability of the trait and 4) the distribution of QTL effects [2].

Since 2009, national evaluations of Holstein bulls are being performed in the USA and Canada based on GEBV. Before that, a comprehensive study evaluating the reliability of GEBV of many economic traits in thousands of Holstein bulls genotyped with the Illumina BovineSNP50 Genotyping BeadChip (Illumina, Inc., San Diego, California, USA) demonstrated that genomic PA of young animals were substantially higher than the reliability of traditional PA [3,4].

In beef cattle, the situation is quite different with respect to feasibility of genomic selection. There are several reasons for this, which include the genetic structure of herds, which involve a greater number of breeds with different origin and history of selection, resulting in significant differences in genetic parameters, such as LD and effective population size ( $N_e$ ), which influence the accuracy of genomic selection [5,6,7]. Simulation studies are still predominant, despite of the availability of databases, mainly in *bos taurus* breeds [e.g., 8]. However, there is a gap of information concerning several beef cattle breeds, especially with respect to *bos indicus* (zebu) breeds.

The aim of this study was to evaluate the potential accuracy of DGEBV for two levels of marker densities, different heritability levels and different sizes of TS in simulated populations that mimicked the extent and pattern of the reported LD in beef cattle under selection.

## METHODS

### Simulation

#### Population structure

Using the QMSim software [9], populations were simulated based on forward-in-time process [10] with either 40k or 800k single nucleotide polymorphism (SNP) markers, and 750 QTL across the 29 *bos taurus* autosomes (BTAs). In the first simulation step, 1000 generations with a constant size of 1000 were simulated, followed by 1020 generations with a gradual decrease in population size from 1000 to 200 in order to create initial LD and establish mutation-drift equilibrium in historical generations. The number of individuals of each sex remained the same in this step and the mating system was based on the random union of gametes, randomly sampled from both the male and female gametic pools. In the second step, an expansion of the population was created by initially randomly selecting 100 founder males and 100 founder females from the last generation of the historical population. In order to obtain a large number of individuals with more than 50 offspring, eight generations were simulated with five offspring per dam and an exponential growth of the number of dams. The mating was again based on the random union of gametes. In the case of simulating 800k SNP markers, the expansion of the population was carried over six generations, instead of eight, due to computer time and memory requirements which resulted in a smaller population size.

In the next simulation step, two recent generation sets were simulated by selecting 640 males and 32000 females (POP1) or 160 males and 8000

females (POP2) from the last generation of the two expand populations with 40k or 800k SNP markers, respectively. Then ten generations were simulated. Generations 3 to 8 was used as TS and EBVs for this group were estimated by excluding information for generations 9 and 10. The prediction set for both POP1 and POP2 included only animals born in generation ten.

The parameters used in the recent generations mimicked a real production system as close as possible with one progeny per dam, 50% of male progeny, random mating system, replacement ratio of 60% for sires and 20% for dams, selection for high values of EBV, and culling for low values of EBV. The breeding values were estimated by BLUP, using Henderson's mixed linear equations [11] for an individual animal model, considering the true additive genetic variance. The rate of missing sire and dam information was 5%. A single trait with heritability of 0.10, 0.25 or 0.40 and phenotypic variance of 1.0 was simulated. The true breeding value of an individual was equal to the sum of the QTL additive effects and the phenotypes were generated by adding random residuals to the true breeding values. The whole simulation process was repeated 10 times. The parameters of simulation process are summarized in Table 1, while Figure 1 depicts the simulation steps.

### **Genome**

The simulated genome consisted of 29 pairs of autosomes with length identical to the real bovine genome based on Btau\_3.1 assembling [12] totalizing 2333 cM. In most reported simulation studies, just one chromosome was simulated, due to the computing time and memory requirements. The advantage of simulating the real number of autosomes with length identical to the bovine



genome is to create a more realistic scenario with respect to the number of physically unlinked marker and QTL loci. The SNP markers were evenly distributed and the initial number of markers was chosen such that it would generate two densities of segregating bi-allelic loci with minor allele frequency (MAF) $>0.1$ : 40k or 800k. The markers were neutral in their effect on the trait. A number of QTL was simulated to generate 750 segregating loci with two, three or four alleles and MAF $>0.1$ , whose positions were randomly distributed. Additive allelic effects were randomly sampled from gamma distribution with shape parameter equal to 0.4. The rate of missing marker genotypes was 0.01 and the rate of marker genotyping error was 0.005. A recurrent mutation rate of  $1e-4$  for both markers and QTLs was considered to establish mutation-drift equilibrium in historical generations. The parameters used for simulating the genome are given in Table 1 and a summary of simulated SNP markers for each autosome is given in Table 2. The levels of LD in the recent generations of POP1 and POP2 are given in Table 2 and the extent of LD in Figure 3 and Table 3. Due to the computing time required to calculate all possible pair-wise LD, statistics on Figure 3 and Table 3 are for BTA 1 and for one replicate only. However, these statistics should illustrate well the extent and decay of simulated LD.

### **Training and prediction sets**

The training set was composed of bulls from generations three to eight for both POP1 and POP2. The reason for excluding the first two generations was to use data from a population undergoing selection, which would be mostly the case in

beef cattle. In order to estimate the effects of 40k segregating markers, all the bulls ( $n=1920$ ) that had more than 50 offspring (average progeny size=73) were selected from POP1 for the three levels of heritability (0.10, 0.25 or 0.40). Of these, 960 and 480 bulls were randomly sampled so that the average accuracy of EBVs was kept the same across the TS sizes within each level of heritability. Therefore, the average progeny size was approximately the same (73) for the three TSs with an average accuracy of EBVs equal to 0.79, 0.90 or 0.94 for heritability of 0.10, 0.25 or 0.40, respectively. For the purpose of estimating the effects of 800k markers in POP2, only 480 bulls had more than 50 offspring (average progeny size=73) because of the smaller population size and, therefore, these were selected from POP2 for the three levels of heritabilities (0.10, 0.25 or 0.40). However, considering that this number of animals with highly accurate EBV ( $n=480$ ) is realistic for most beef cattle breeding programs, this scenario would be useful for assessing the accuracy of genomic selection using a high density marker panel.

Two other additional TSs were created by randomly selecting 7680 and 1920 animals from POP1 and POP2, respectively. In this case, the phenotypes were used to estimate the marker effects, what represented a 4-fold increase in the TS. Figure 2 depicts these simulated scenarios with respect to the TS sizes.

The prediction sets were composed of 8000 and 2000 individuals randomly chosen from the 10<sup>th</sup> generation of POP1 and POP2, respectively. A gap of one generation was left between the training and prediction sets, so that animals in both sets were not too closely related.

## Genetic evaluations

The EBVs of animals were calculated by the standard animal model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of trait phenotypes,  $\mu$  is the overall mean,  $\mathbf{Z}$  is an incidence matrix and  $\mathbf{a}$  is the vector of animals' additive genetic effects and  $\mathbf{e}$  is a vector of random errors. The variance of  $\mathbf{y}$  was assumed to be  $\mathbf{ZAZ}'\sigma_a^2 + \mathbf{I}\sigma_e^2$  where  $\mathbf{A}$  is the additive genetic relationship matrix and  $\sigma_e^2$  is the residual variance.

A ridge regression model was used to estimate SNP effects for using in computing DGEBV. The model can be written as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of either trait phenotypes or EBVs calculated from trait phenotypes and pedigree data by BLUP,  $\mu$  is the overall mean,  $\mathbf{X}$  is the matrix of marker genotypes for each animal and  $\boldsymbol{\beta}$  is the vector of marker effects, and  $\mathbf{e}$  is a vector of random errors.

The marker genotypes of an individual are re-coded as the number of copies of one of the SNP alleles, *i.e.*, 0, 1 or 2. A common ridge regression factor was used, assuming a common variance for the marker effects, *i.e.*  $\sigma_a^2/n$  where  $n$  is the number of markers. This is equivalent to GBLUP method proposed by VanRaden [13].

The DGEBV was computing for the animals in the prediction set as

$$\hat{\mathbf{a}} =$$

where  $\mathbf{X}$  is the matrix of marker genotypes for each animal in the prediction set and  $\hat{\mathbf{a}}$  is the vector of estimated marker effects.

The gebv software [14] was used to estimate marker effects and DGEBV.

The accuracy of DGEBV was calculated as the correlation between the DGEBV and the true breeding value and standard errors were computed as the standard deviation of the accuracies across the 10 replicates, divided by  $\sqrt{10}$ .

The differences between the average accuracies of DGEBV across scenarios were tested by t-test and the accuracies of DGEBV were regressed on number of animals in the TS, using a quadratic regression.

### Linkage Disequilibrium

Linkage Disequilibrium was measured by  $r^2$ , which is the squared correlation of the alleles at 2 loci [15]:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

where  $D=f(AB)-f(A)f(B)$ , and  $f(AB)$ ,  $f(A)$ ,  $f(a)$ ,  $f(B)$ ,  $f(b)$  are observed frequencies of haplotypes AB and of alleles A, a, B, b, respectively.

It was demonstrated that  $r^2$  is a more suitable measure to estimate LD for biallelic markers, such as SNPs, because  $r^2$  is less sensitive to allelic frequency and sample size than  $D'$  [16,17,18].

### Effective population size

The effective population size was calculated based on  $r^2$ , assuming a model without mutation [19]:

$$r^2 = \frac{1}{4N_e c + 1}$$

where  $c$  is linkage map distance (in morgans) between two SNPs.

The effective population size in earlier generations was estimated at different time points, back to 500 generations, by taking the average LD for a range of short distances to minimize the chance of an event of recombination, increasing the accuracy of estimation of  $N_e$  in the remote past (Figure 4). The age of  $N_e$  for any distance was estimated by  $1/(2c)$ . These estimates are an approximation of true values [20].

## RESULTS AND DISCUSSION

### Linkage disequilibrium

The overall  $r^2$  between adjacent SNP across all autosomes for moderate heritability (0.25) were 0.25 and 0.32 for POP1 (40k) and POP2 (800k), respectively (Table 2). This average  $r^2$  between adjacent SNPs for POP1 is similar to the value reported by [21] for Gyr cattle in Brazil ( $r^2 = 0.21$ ), using bulls genotyped with the Illumina BovineSNP50 chip and similar to the  $r^2$  reported by [22] for a multi-breed herd genotyped with the Illumina BovineSNP50 chip in Canada ( $r^2 = 0.21$  between markers 30-35kb apart). The trend of decay of LD with the increase in physical distance was, as expected, exponential for both POP1 and POP2 (Figure 3) in agreement with other published studies [21,23,24].

Table 3 shows the average  $r^2$  for different distance ranges up to 1Mb. The highest average estimated  $r^2$  was 0.24 for a distance range of 0-0.10 Mb. The  $r^2$

values decreased with increasing distance between SNP, but not as sharply as the decrease reported for Holstein cattle [23]. In the current simulation study, for moderate heritability (0.25) and POP1 (40k), the % of  $r^2 > 0.30$  for ranges 0-0.1Mb, 0-0.2Mb, 0-0.5Mb and 0-1Mb was 28%, 25%, 15% and 9%, respectively. These results are similar to findings of [21] for Gyr cattle, who reported a % of  $r^2 > 0.30$  of 23%, 20%, 14% and 10% for the same ranges, respectively.

Using the Illumina BovineSNP50 chip, [25] reported  $r^2$  of 0.31, 0.22 and 0.15 for Angus and crossbred cattle for distance ranges between 0-0.03Mb, 0.03-0.06Mb and 0.06-0.10Mb, respectively. The simulated results in the present study for the distance range 0-0.10Mb seem in line with those results reported by [25].

Mckay et al. [26] reported the  $r^2$  for eight breeds of cattle varying from 0.15 to 0.20 for a distance range from 0-0.1Mb, which is similar to the  $r^2$  observed in the simulated data for the same distance range (Table 3).

Table 4 shows the average LD ( $r^2$ ) between adjacent SNPs and distribution of SNP pairs across different LD ranges on chromosome 1 in the recent generations for three heritability levels and POP1 and POP2. The % of SNP pairs with  $r^2 > 0.30$  did not differ across the heritabilities, probably because the number of generations under selection was small (10). When comparing the marker density, 43% of SNP pairs showed high LD ( $r^2 > 0.30$ ) in POP2 (800k), while 33% of the SNP pairs showed high LD in POP1 (40k) for moderate heritability (0.25). This difference of ten points seems not very high, but when it is translated to number of SNP pairs in high LD in each marker density, it

becomes quite significant and has an impact on the accuracy of genomic selection, as shown later the in results.

### **Effective population size**

Figure 4 shows the estimated historical effective size [20] over time for POP1 (40k) and POP2 (800k). In the simulation process the initial  $N_e$  was 1000, which was properly estimated and depicted in the Figure. With the creation of LD generated by mutations and genetic drift and, in more recent generations, by selection,  $N_e$  dropped to ~250 and ~120, for POP1 (40k) and POP2 (800k), respectively. The smaller  $N_e$  for POP2 than POP1 is due the smaller population size (88,160 vs. 352,640 individuals for POP2 and POP1, respectively). These values of  $N_e$  are similar to the estimates calculated based on inbreeding rates for a Brazilian Nellore cattle herd reported by Brito et al. [27], but are larger than those presented by Faria et al. [28] for the entire registered Nellore population in Brazil.

### **Accuracy of DGEVB**

Overall, the accuracy of DGEVB increased significantly ( $p < 0.05$ ) with the increase in the number of bulls in the TS, heritability of the trait and density of markers (Table 5). When the EBV was used to estimate the marker effects in POP1 (40k), the number of bulls in the TS, which varied from 480 to 1920, increased the accuracy of DGEVB for all three heritability levels. The DGEVB accuracies were regressed on the number of bulls in the TS, using a quadratic regression. The estimated regressions were  $y = 0.256 + 0.00026x - 5.2e^{-8}x^2$

( $p < 0.0001$ ;  $R^2 = 0.91$ ),  $y = 0.220 + 0.00036x - 8.2e^{-8}x^2$  ( $p < 0.0001$ ;  $R^2 = 0.96$ ) and  $y = 0.228 + 0.00037x - 8.3e^{-8}x^2$  ( $p < 0.0001$ ;  $R^2 = 0.97$ ) for heritability = 0.10, 0.25 and 0.40, respectively. Figure 5 depicts the estimated regressions for each heritability level. It is expected that the accuracy of DGE BV would increase 0.15, 0.20 and 0.21 points for an increase from 480 to 1480 bulls in the TS, and 0.11, 0.12 and 0.14 points from 920 to 1920 bulls in the TS, for heritability of 0.10, 0.25 and 0.40 respectively. These results do not fully agree with those from [3], who used Holstein bulls genotyped for the Illumina BovineSNP50 chip and reported that gains from genomic data increased almost linearly with number of bulls in the TS. Although it not possible to extrapolate out of the range of number of bulls in the TS, the current results suggest that it will be necessary genotyping a large number of bulls to obtain large gains with genomic selection, and a plateau might be expected, limiting the possible gain in accuracy of DGE BV for the density of markers considered (40k).

The highest accuracy (0.64) was observed for heritability of 0.40 and 1920 bulls in the TS. This value implies that 41% of genetic variance of the trait was explained by marker effects. These results are similar to those reported by [8], using data from 2,000 Angus bulls. The author reported correlations from 0.5 to 0.7 between conventional EBV and genomic predictions. The author, however, did not report the accuracy of conventional EBV used to calculate the correlations.

The effect of heritability was not as evident for category of 480 bulls in the TS, as it was for 960 and 1920 bulls (Figure 6), reflecting the fact that the sample size became a more limiting factor than the heritability level for the category of



480 bulls.

The response variable used for calculating the marker effects also affected the accuracy of DGEVB. Figure 6 shows that when the phenotypes were used as response variable to estimate marker effects in POP1 (40k), but with a larger number of animals in the TS ( $n=7680$ ), the accuracy of DGEVB was higher than that from using EBVs from 480 bulls for all heritability levels ( $p<0.05$ ). However, when EBVs from 1920 bulls instead of own phenotypes from 7680 bulls were used to estimate the marker effects, the accuracy of DGEVB was higher for heritabilities of 0.10 and 0.25 ( $p<0.05$ ). Therefore, in this study, genotyping 4-fold more animals (7680 vs. 1920) and using own phenotypes instead of EBVs to estimate the marker effects was not sufficient to increase the accuracy of DGEVB for moderate to low heritability traits.

Schenkel et al. [4] reported that substantial increase in the size of TS by the use of foreign bulls with MACE proofs in addition to the bulls with domestic proofs increased the accuracy of DGEVB for all economically important traits in Canadian dairy cattle. This shows the importance of large TS size for estimating the marker effects.

Figure 7 depicts the DGEVB accuracy for the 800k markers scenario. Because the computing time and memory requirements, it was possible to simulate only 480 bulls with accurate EBV in the TS. The estimated accuracies of DGEVB were 0.43, 0.46 and 0.48 for heritability 0.10, 0.25 and 0.40, respectively. The accuracies of DGEVB in this scenario were higher ( $p<0.05$ ) than those from 40k SNP markers and 480 bulls in the TS and similar to those with 40k SNP markers and 960 bulls in the TS. These results indicate that the higher level of

LD between the 800k SNP markers would require 2 times less animals in the TS. Muir [29], using simulated data, reported that increasing number of markers had contradictory effects on accuracy of GEBV, because of co-linearity between the effects of markers. Then it might be expected that the 800k SNP scenario would benefit more from an increase in the TS size and differences with respect to the accuracy of DGEBV from the 40k SNP scenario would increase.

When the number of bulls genotyped with 800k SNP markers was 1920 and own phenotypes were used for estimating marker effects, the accuracy of DGEBV increased only for heritability 0.40 ( $p < 0.10$ ), while for the other two levels of heritability the accuracy decreased ( $p < 0.05$ ) compared to accuracies from 480 bulls in the TS and using EBV for estimating marker effects.

One of the main advantages of using genomic information is to obtain accurate EBV early in the animal's life, when the own phenotypic information has not been measured yet. Then, if we consider that the average EBV of the parents is the only available information in this stage, we can compare the accuracies provided by PA and the accuracies of DGEBV. Since the reliability of PA is a quarter of the sum of reliabilities of the EBV of parents, considering the average reliability of EBV of bulls in the TS reported in this study and the average reliability of the EBVs of the dams in the simulated data, the expected accuracies for PA would be 0.44, 0.53 and 0.58 for heritability = 0.10, 0.25 and 0.40, respectively. These accuracies are smaller than those accuracies of DGEBV considering 1920 bulls in the TS for POP1 (40k) for all heritability levels. Therefore, the gains in accuracy by using DGEBV would be of 27%, 13% and 10% compared to accuracy of PA for these conditions.

Schenkel et al. [4] reported results from a large genomic prediction validation study in Canada for 44 traits of dairy cattle, using 6403 bulls in the TS genotyped for the Illumina BovineSNP50 chip. The genomic predictions showed an increase of reliability for 42 of the 44 traits. For production traits the average gain in reliability was of 41%, using domestic proofs and 74%, using MACE proofs. Therefore, the gains reported in the current study that tried to mimic the LD structure that one might expect in a beef cattle population were lower than those reported in dairy cattle.

The main issue that has been discussed in most studies involving genomic selection is whether it will bring additional benefits, considering the cost of genotyping. An alternative to reduce the cost of genotyping and, thus, increase the number of genotyped animals, is the imputation of genotypes, which would allow genotyped animals with a less dense SNP panels (e.g., 3k) to be imputed to a denser SNP panels (e.g., 50k), using a reference population genotyped with the denser SNP panel [30]. This alternative should be considered primordial in the projects that involve genomic selection, mainly in countries where the number of available bulls with accurate EBV for genotyping is much smaller than those considered in this study. In this way, with the use of imputation, more bulls, even those with less accurate EBV, could be incorporated to the TS.

## CONCLUSIONS

As for dairy cattle, gains that will be achieved by incorporating genomic information into beef cattle evaluation are mainly dependent on the number of animals in the TS and the level of LD between markers and QTL, but in beef cattle, bigger TSs will be likely needed, as well as denser panels to increase levels of LD between markers and QTL. It is expected that, in the absence of a sufficient number of bulls with accurate EBV for estimating the marker effects, the required number of bulls in the TS to produce accurate DGEBV will be even higher. A denser panel is expected to increase DGEBV accuracy, but to fully exploit the increased level of LD between markers and QTL, its use should be accompanied by an increase in number of bulls in the TS. Therefore, as the number of animals in the TS is crucial for the success of genomic selection, it is important that the institutions involved in genomic selection research and application in beef cattle are aware that a joint collaborative work, sharing genotypes, can facilitate to achieve the goal of incorporating genomic information in breeding programs with substantial gains for producers and the beef industry.

Both in dairy cattle and in beef cattle, the proofs for many traits are already quite accurate and one of the main benefits of genomic selection is to predict more accurate breeding values early in the life of the animal compared to parent average. The gains in accuracy by using DGEBV and, therefore, before the animal having an accurate EBV, showed more advantage for traits with a moderate to low heritability.

Additional simulation studies are recommended, aiming to evaluate the impact of specific levels and extent of LD in beef cattle in different populations and production systems, the effectiveness of other genomic selection methods, such as Bayesian methods, as well as the number of animals in the TS that would result in successful incorporation of genomic information in the prediction of genetic values, considering also possible strategies for imputation. The results of the current investigation provide preliminary insights on the issues related to the use of genomic selection in beef cattle.

#### **AUTHORS' CONTRIBUTIONS**

FVB carried out the simulations, performed the simulated data analysis and drafted the manuscript. MS participated in the design of the simulations process and wrote the code of the main software used in the study. JBN and JAC critically revised the manuscript. FSS contributed to the design of the simulations and analysis of simulated data. In addition, FSS revised the manuscript. All authors read and approved the final manuscript.

#### **ACKNOWLEDGEMENTS**

The first author was sponsored by a scholarship from the Brazilian Federal Agency for Support and Evaluation of Graduate Education – Capes.

## REFERENCES

1. Meuwissen THE, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819-1829.
2. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: **Invited review: Genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**:433-443.
3. VanRaden PM, Van Tassel CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS: **Invited review: reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92**:16-24.
4. Schenkel FS, Sargolzaei M, Kistemaker G, Jansen GB, Sullivan P, Van Doormaal BJ, VanRaden PM, Wiggans GR: **Reliability of genomic evaluation of Holstein cattle in Canada.** In *Proceedings of Interbull International Workshop – genomic information in genetic evaluations. 2009; Uppsala, Sweden.* [<http://www-interbull.slu.se/bulletins/framesida-pub.htm>].
5. Taylor JF, Decker JE, Kim J, McClure MC, Mckay SD, Rolf MM, Taxis T, Vasco D, Schnabel RD: **Prospects for genomic selection in beef cattle.** In *Proceedings of the 41<sup>th</sup> Beef Improvement Federation Annual Research Symposium: 2009; Sacramento, USA.*

6. Garrick DJ: **The nature, scope and impact of some whole-genome analysis in beef cattle.** In *Proceedings of the 9<sup>th</sup> World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig*. Edited by German Society for Animal Science; 2010: 0023.
7. Goddard ME, Hayes BJ, Meuwissen THE: **Genomic selection in farm animal species – lessons learnt and future perspectives.** In *Proceedings of the 9<sup>th</sup> World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig*. Edited by German Society for Animal Science; 2010: 701.
8. Garrick DJ: **The nature and scope of some whole-genome analysis in US beef cattle.** In *Proceedings of 41<sup>th</sup> Beef Improvement Federation Annual Research Symposium: 2009; Sacramento, USA*.  
[[http://www.bifconference.com/bif2009/proceedings/G3\\_pro\\_Garrick.pdf](http://www.bifconference.com/bif2009/proceedings/G3_pro_Garrick.pdf)]
9. Sargolzaei M, Schenkel FS: **QMSim: a large-scale genome simulator for livestock.** *Bioinformatics* 2009, **25**(5):680-681.
10. Carvajal-Rodríguez A: **Simulation of genomes: a review.** *Curr Genomics* 2008, **9**:155-159.
11. Henderson CR: **Best linear unbiased estimation and prediction under a selection model.** *Biometrics* 1975, **31**:423-447.

12. Snelling WM, Chiu R, Schein JE, Hobbs M, Abbey CA, Adelson DL, Aerts J, Bennett GL, Bosdet IE, Boussaha M, Brauning R, Caetano AR, Costa MM, Crawford AM, Dalrymple BP, Eggen A, Van der Wind AE, Floriot S, Gautier M, Gill CA, Green RD, Holt R, Jann O, Jones SJM, Kappes SM, Keele JW, Jong PJ, Larkin M, Lewin HA, McEwan JC, McKay S, Marra MA, Mathewson CA, Matukumalli LK, Moore SS, Murdoch B, Nicholas FW, Osoegawa K, Roy A, Salih H, Schibler L, Schnabel RD, Silveri L, Skow LC, Smith TPL, Sonstegard TS, Taylor JF, Tellam R, Van Tassell CP, Williams JL, Womack JE, Wye NH, Yang G, Zhao S for the International Bovine BAC Mapping Consortium: **A physical map of the bovine genome.** *Genome Biology* 2007, **8**:R165.
13. VanRaden, PM: **Efficient Methods to Compute Genomic Predictions.** *J Dairy Sci* 2008, **91**:4414–4423
14. Sargolzaei M, Schenkel FS, VanRaden PM: **gebv: Genomic breeding value estimator for livestock.** *Technical report to the Dairy Cattle Breeding and Genetics Committee. University of Guelph, October 07, 2009.*
15. Hill WG, Robertson A: **Linkage disequilibrium in finite populations.** *Theor Appl Genet* 1968, **38**:226-231.
16. Du FX, Clutter AC, Lohuis MM: **Characterizing linkage disequilibrium in pig populations.** *Int J Biol Sci* 2007, **3**:166-178.



17. Zhao H, Nettleton D, Dekkers JCM: **Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms.** *Genet Res Camb* 2007, **89**:1-6.
18. Bohmanova J, Sargolzaei M, Schenkel FS: **Characteristics of linkage disequilibrium in North American Holsteins.** *BMC Genomics* 2010, **11**(421).
19. Sved JA: **Linkage Disequilibrium and homozygosity of Chromosome segments in finite populations.** *Theor Pop Biol* 1971, **2**:125-141.
20. Hayes BJ, Visscher PM, McPartlan C, Goddard ME: **Novel multilocus measure of linkage disequilibrium to estimate past effective population size.** *Genome Res* 2003, **13**:635-643.
21. Silva CR, Neves HHR, Queiroz SA, Sena JAD, Pimentel ECG: **Extent of linkage disequilibrium in Brazilian Gyr dairy cattle based on genotypes of AI sires for dense SNP markers.** In *Proceedings of the 9<sup>th</sup> World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig*. Edited by German Society for Animal Science; 2010: PP1-29.

22. Kelly MJ, Sargolzaei M, Wang Z, Kolbehdari D, Stothard P, Schenkel FS, Moore SS, Miller SP: **Estimated linkage disequilibrium in a multi-breed beef herd based on the Illumina BovineSNP50 BeadChip.** *J Anim Sci* 2008, **86**(E-Suppl 2):410.
23. Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR: **Extent of linkage disequilibrium in Holstein Cattle in North America.** *J Dairy Sci* 2008, **91**:2106-2117.
24. Kim ES, Kirkpatrick BW: **Linkage disequilibrium in the North American Holstein population.** *Anim Genet* 2009, **40**:279-288.
25. Lu D, Sargolzaei M, Kelly M, Vander Voort G, Wang Z, Mah J, Plastow G, Moore SS, Miller SP: **Extent of linkage disequilibrium in purebred and crossbred beef cattle.** *J Dairy Sci* 2008, **92**(E-Suppl 1):43.
26. McKay SD, Schnabel D, Murdoch M, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassel CP, Williams JL, Taylor JF, Moore SS: **Whole genome linkage disequilibrium maps in cattle.** *BMC Genetics* 2007, **8**(74).

27. Brito FV, Sargolzaei M, Braccini Neto J, Cobuci JA, Schenkel FS: **Pedigree analysis in a large Brazilian Nelore herd.** In *Proceedings of the 9<sup>th</sup> World Congress on Genetics Applied to Livestock Production: 1-6 August 2010; Leipzig.* Edited by German Society for Animal Science; 2010: PP1-57.
28. Faria FJC, Filho AEV, Madalena FE, Josahkian LA: **Pedigree analysis in the brazilian zebu breeds.** *J Anim Breed Genet* 2009, **126**:148-153.
29. Muir WM: **Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters.** *J Anim Breed Genet* 2007, **124**:342-355.
30. Sargolzaei M, Schenkel FS, Chesnais J: **Comparison between two methods of imputation (AIPL vsBoviteq/CGIL) for 6,246 genotypes provided by AIPL.** In *Dairy Cattle Breeding and Genetics Committee meeting, October 5, 2010, Guelph, Canada.*

**Table 1. Parameters of the simulation process**

<b>Population structure</b>	<b>POP1/POP2</b>
<b>Step 1: Historical generations (HG)</b>	
Number of generations(size) – phase 1	1000(1000)
Number of generations(size) – phase 2	1020(200)
<b>Step 2: Expanded generations (EG)</b>	
Number of founders males from HG	100
Number of founders females from HG	100
Number of generations	8 / 6
Number of offspring per dam	5
<b>Step 3: Recent generations</b>	
Number of founders males from EG	640 / 160
Number of founders females from EG	32000 / 8000
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 model
Ratio of missing sire and dam	5%
Heritability of the trait	10%, 25%, or 40%
Phenotypic variance	1.0
<b>Genome</b>	
Number of chromosomes	29
Total length	2335cM
Number of markers	40,000 / 800,000
Marker distribution	Evenly spaced
Number of QTL	750
QTL distribution	Random
MAF for markers	0.1
MAF for QTL	0.1
Additive allelic effects for markers	Neutral
Additive allelic effects for QTL	Gamma distribution (shape=0.40)
Rate of missing marker genotypes	0.01
Rate of marker genotyping error	0.005
Rate of recurrent mutation	0.0001

POP1: population 1; POP2: population 2; EBV: estimated breeding value; BV: breeding value; QTL: quantitative trait loci; MAF: minimum allele frequency. The whole simulation process was repeated 10 times.

**Table 2. Average Linkage Disequilibrium ( $r^2$ ) between adjacent SNP markers.**

BTA	Length (Mb)	POP1 (40k)		POP2 (800k)	
		Number of SNP	Average $r^2$ (SD)	Number of SNP	Average $r^2$ (SD)
1	146	2712	0.25(0.25)	47539	0.33(0.30)
2	126	2205	0.25(0.25)	38629	0.33(0.30)
3	116	2060	0.24(0.25)	38824	0.31(0.29)
4	111	2011	0.23(0.24)	37889	0.30(0.28)
5	119	1749	0.24(0.24)	32230	0.32(0.30)
6	112	1962	0.26(0.26)	38405	0.32(0.29)
7	101	1687	0.25(0.26)	33209	0.31(0.29)
8	104	1748	0.24(0.25)	33833	0.32(0.30)
9	95	1369	0.26(0.26)	30472	0.31(0.29)
10	96	1413	0.27(0.26)	30513	0.32(0.29)
11	102	1551	0.25(0.25)	33513	0.30(0.28)
12	78	1070	0.24(0.26)	23437	0.31(0.29)
13	83	1249	0.25(0.25)	27385	0.32(0.30)
14	82	1290	0.25(0.25)	27452	0.33(0.30)
15	75	1081	0.27(0.27)	22531	0.33(0.30)
16	73	1150	0.24(0.25)	22891	0.33(0.30)
17	70	1195	0.27(0.26)	23281	0.31(0.29)
18	63	987	0.23(0.23)	19391	0.32(0.29)
19	63	966	0.25(0.26)	20719	0.31(0.29)
20	68	1133	0.24(0.24)	21775	0.35(0.31)
21	63	1032	0.25(0.26)	19720	0.32(0.30)
22	60	946	0.26(0.26)	19212	0.31(0.29)
23	49	772	0.25(0.26)	15958	0.31(0.29)
24	60	901	0.24(0.24)	18087	0.32(0.30)
25	42	749	0.25(0.25)	14861	0.31(0.29)
26	48	840	0.25(0.26)	15999	0.30(0.30)
27	43	670	0.26(0.26)	13563	0.32(0.30)
28	40	704	0.25(0.26)	13407	0.32(0.30)
29	45	806	0.26(0.26)	14420	0.33(0.30)
Overall	2333	38008	0.25(0.26)	749145	0.32(0.30)

The results are for each of the 29 autosomes in the recent generations for moderate heritability (0.25) and POP1 (40k) and POP2 (800k) over 10 replicates. SD: Standard Deviation.

**Table 3. Average Linkage Disequilibrium ( $r^2$ ) for different distances between closely located SNP pairs.**

Distance range (Mb)	POP1 (40k)				POP2 (800k)			
	Pairs (n)	$r^2$ (SD)	Frequency $r^2 > 0.30$ (%)		Pairs (n)	$r^2$ (SD)	Frequency $r^2 > 0.30$ (%)	
$h^2=0.10$	0.00-0.10	5049	0.23 (0.24)	1512 (29.95)	4275	0.24 (0.25)	1272 (29.75)	
	0.10-0.20	5074	0.17 (0.19)	1063 (20.95)	4474	0.18 (0.20)	919 (20.54)	
	0.20-0.30	6265	0.14 (0.16)	885 (14.13)	4290	0.14 (0.16)	644 (15.01)	
	0.30-0.40	4938	0.12 (0.14)	510 (10.33)	4358	0.12 (0.14)	493 (11.31)	
	0.40-0.50	5042	0.10 (0.13)	425 (8.43)	4280	0.10 (0.13)	359 (8.39)	
	0.50-0.60	6199	0.09 (0.11)	370 (5.97)	4196	0.09 (0.12)	279 (6.65)	
	0.60-0.70	4994	0.08 (0.10)	228 (4.57)	4208	0.09 (0.11)	242 (5.75)	
	0.70-0.80	5975	0.07 (0.09)	187 (3.69)	4268	0.08 (0.10)	190 (4.45)	
	0.80-0.90	6183	0.07 (0.09)	167 (2.70)	4334	0.07 (0.10)	170 (3.92)	
	0.90-1.00	4917	0.06 (0.08)	128 (2.60)	4163	0.07 (0.09)	149 (3.58)	
Overall	3932610	0.008 (0.03)	6151 (0.16)	3126250	0.01 (0.03)	6062 (0.19)		
$h^2=0.25$	0.00-0.10	5248	0.22 (0.24)	1495 (28.49)	4494	0.24 (0.25)	1375 (30.60)	
	0.10-0.20	5224	0.17 (0.19)	1077 (20.62)	4740	0.18 (0.20)	1006 (21.22)	
	0.20-0.30	6528	0.13 (0.16)	855 (13.10)	4630	0.14 (0.17)	724 (15.64)	
	0.30-0.40	5147	0.11 (0.14)	461 (8.96)	4560	0.12 (0.14)	488 (10.70)	
	0.40-0.50	5112	0.09 (0.12)	315 (6.16)	4596	0.11 (0.13)	406 (8.83)	
	0.50-0.60	6466	0.08 (0.11)	343 (5.31)	4469	0.10 (0.12)	333 (7.45)	
	0.60-0.70	5146	0.07 (0.09)	186 (3.61)	4418	0.09 (0.11)	250 (5.66)	
	0.70-0.80	5227	0.07 (0.09)	139 (2.66)	4502	0.08 (0.10)	226 (5.02)	
	0.80-0.90	6431	0.06 (0.08)	178 (2.77)	4359	0.08 (0.10)	207 (4.75)	
	0.90-1.00	5129	0.06 (0.08)	95 (1.85)	4416	0.07 (0.09)	154 (3.49)	
Overall	4096952	0.006 (0.03)	5689 (0.14)	3283438	0.011 (0.03)	6543 (0.20)		
$h^2=0.40$	0.00-0.10	5439	0.22 (0.24)	1517 (27.89)	4191	0.24 (0.24)	1265 (30.18)	
	0.10-0.20	5468	0.17 (0.19)	1070 (19.57)	4225	0.17 (0.20)	865 (20.47)	
	0.20-0.30	6801	0.13 (0.16)	886 (13.03)	4263	0.14 (0.16)	621 (14.57)	
	0.30-0.40	5435	0.11 (0.14)	543 (9.99)	4251	0.12 (0.14)	457 (10.75)	
	0.40-0.50	5444	0.09 (0.12)	361 (6.63)	4286	0.10 (0.12)	333 (7.77)	
	0.50-0.60	6782	0.08 (0.10)	293 (4.32)	4284	0.09 (0.11)	270 (6.30)	
	0.60-0.70	5377	0.07 (0.09)	176 (3.27)	4103	0.09 (0.11)	232 (5.65)	
	0.70-0.80	5374	0.07 (0.09)	150 (2.79)	4260	0.08 (0.10)	205 (4.81)	
	0.80-0.90	6733	0.06 (0.08)	155 (2.30)	4285	0.07 (0.09)	136 (3.17)	
	0.90-1.00	5342	0.06 (0.07)	75 (1.40)	4199	0.06 (0.08)	104 (2.48)	
Overall	4290985	0.006 (0.03)	5710 (0.13)	3072154	0.010 (0.03)	5305 (0.17)		

The results are for chromosome 1 in the recent generations for three heritability levels and POP1 (40k) and POP2 (800k) for one replicate. For POP2 (800k), 40k markers were randomly sampled to represent the 800k panel.

**Table 4. Average Linkage Disequilibrium ( $r^2$ ) between adjacent SNPs pairs and distribution across different  $r^2$  ranges.**

		Number of SNP Pairs and (%)					
		0.10		0.25		0.40	
LD ( $r^2$ ) range							
POP1 (40K)	0.00-0.10	975	(36.97)	1040	(38.35)	1060	(37.80)
	0.10-0.20	463	(17.56)	468	(17.26)	501	(17.87)
	0.20-0.30	305	(11.57)	316	(11.65)	331	(11.80)
	0.30-0.40	232	(8.80)	232	(8.55)	248	(8.84)
	0.40-0.50	192	(7.28)	195	(7.19)	182	(6.49)
	0.50-0.60	126	(4.78)	142	(5.24)	136	(4.85)
	0.60-0.70	110	(4.17)	98	(3.61)	107	(3.82)
	0.70-0.80	83	(3.15)	78	(2.88)	81	(2.89)
	0.80-0.90	62	(2.35)	66	(2.43)	79	(2.82)
	0.90-1.00	89	(3.38)	77	(2.84)	79	(2.82)
Average LD ( $r^2$ )		0.26		0.25		0.25	
POP2 (800K)	0.00-0.10	13980	(29.54)	14124	(29.71)	14939	(31.03)
	0.10-0.20	8034	(16.98)	7443	(15.66)	8072	(16.77)
	0.20-0.30	5388	(11.38)	5442	(11.45)	5571	(11.57)
	0.30-0.40	4149	(8.77)	4240	(8.92)	4156	(8.63)
	0.40-0.50	3250	(6.87)	3418	(7.19)	3321	(6.90)
	0.50-0.60	2859	(6.04)	2837	(5.97)	2665	(5.54)
	0.60-0.70	2359	(4.98)	2479	(5.21)	2274	(4.72)
	0.70-0.80	1928	(4.07)	2162	(4.55)	2121	(4.41)
	0.80-0.90	1929	(4.08)	2032	(4.27)	1924	(4.00)
	0.90-1.00	3451	(7.29)	3362	(7.07)	3095	(6.43)
Average LD ( $r^2$ )		0.33		0.33		0.32	

LD: Linkage Disequilibrium; POP1: population 1; POP2: population 2. Results for average over 10 replicates.

**Table 5. Accuracy of Direct Genomic Estimated Breeding Value.**

		POP1(40k)				POP2(800k)	
$h^2$		480	960	1920	7680	480	1920
PHE	0,10	-	-	-	<b>0,44a,a</b>	-	<b>0,30b,a</b>
	0,25	-	-	-	<b>0,56a,b</b>	-	<b>0,41b,b</b>
	0,40	-	-	-	<b>0,65a,c</b>	-	<b>0,50b,c</b>
EBV	0,10	<b>0,37a,a</b>	<b>0,45b,a</b>	<b>0,56c,a</b>	-	<b>0,43b,a</b>	-
	0,25	<b>0,37a,a</b>	<b>0,49b,b</b>	<b>0,60c,b</b>	-	<b>0,46d,ab</b>	-
	0,40	<b>0,39a,b</b>	<b>0,51b,c</b>	<b>0,64c,c</b>	-	<b>0,48d,b</b>	-

The accuracies of Direct Genomic Estimated Breeding Value (DGE BV) are for animals in the prediction set, considering two levels of marker densities, different numbers of bulls in the training set (TS), three heritability levels, and two alternate response variables used for calculating the marker effects – phenotype (PHE) or Estimated Breeding Value (EBV). The accuracy of EBV used for calculating the marker effects was 0.79, 0.90 or 0.94 for the heritability 0.10, 0.25 or 0.40, respectively, regardless the population and number of bulls in the TS. The average progeny size was 73 for all TS sizes, except for those where the phenotypic record was used to estimate the marker effects. The results are presented as the average over 10 replicates. Different letters indicate significant differences ( $p < 0.05$ ) by t-test (the first letter indicates differences within rows, while the second letter indicates differences within columns).



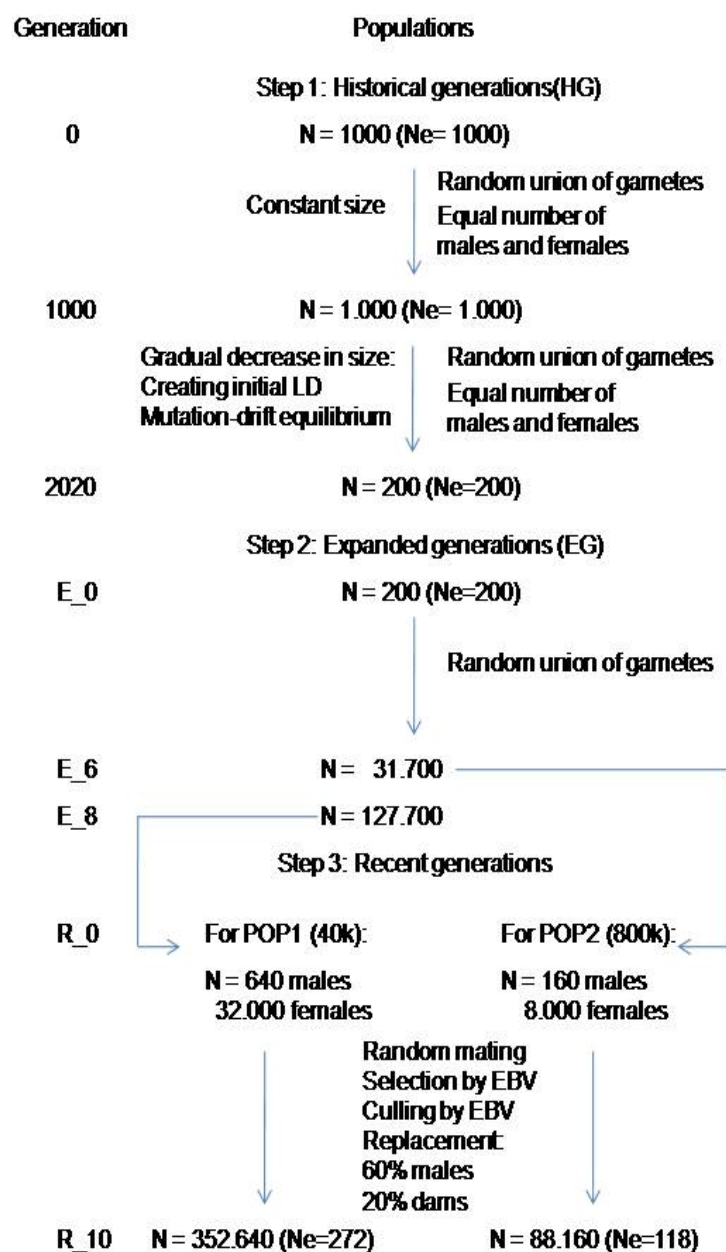
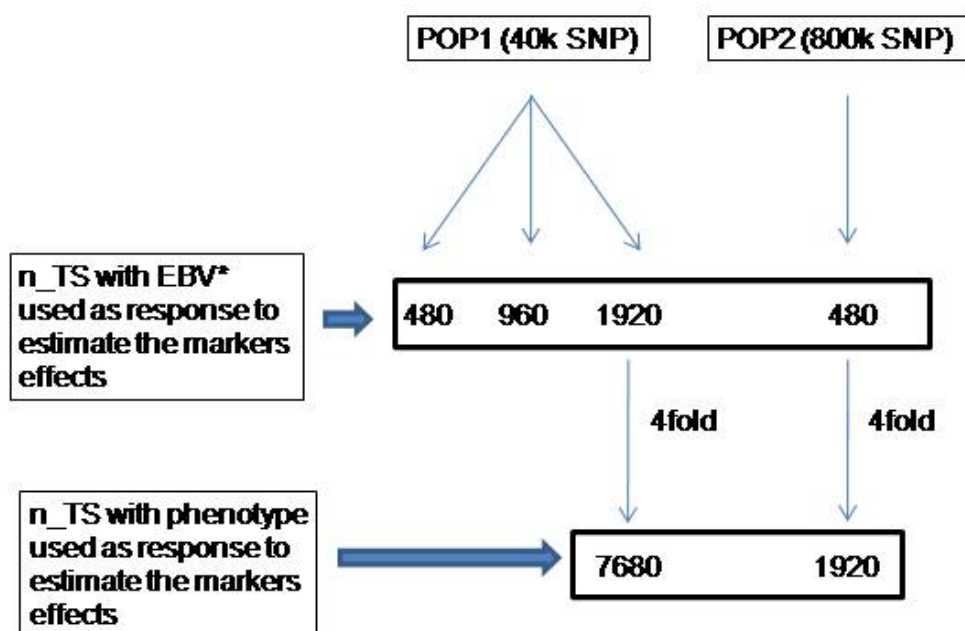
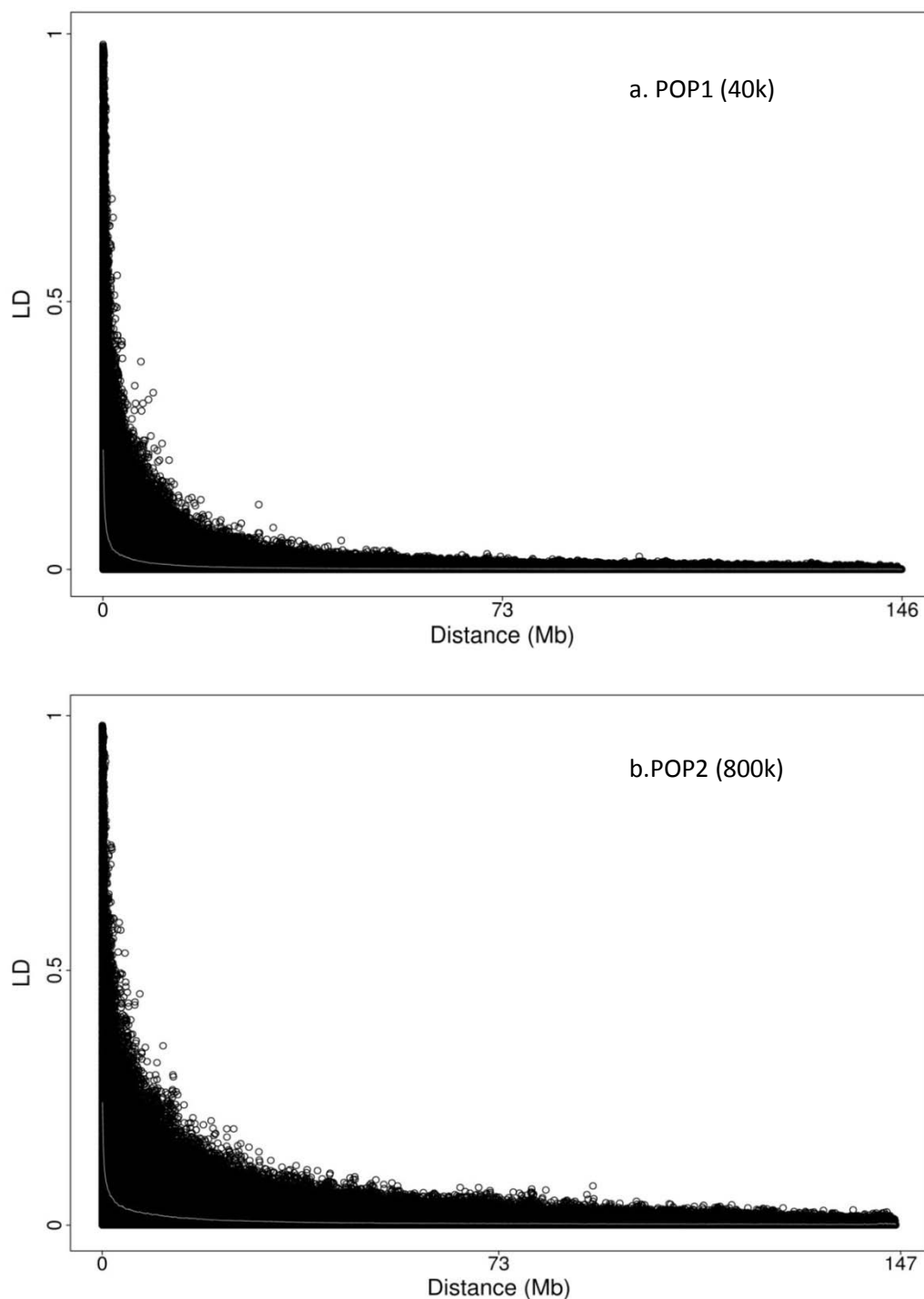


Figure 1. Schematic representation of the simulation steps.



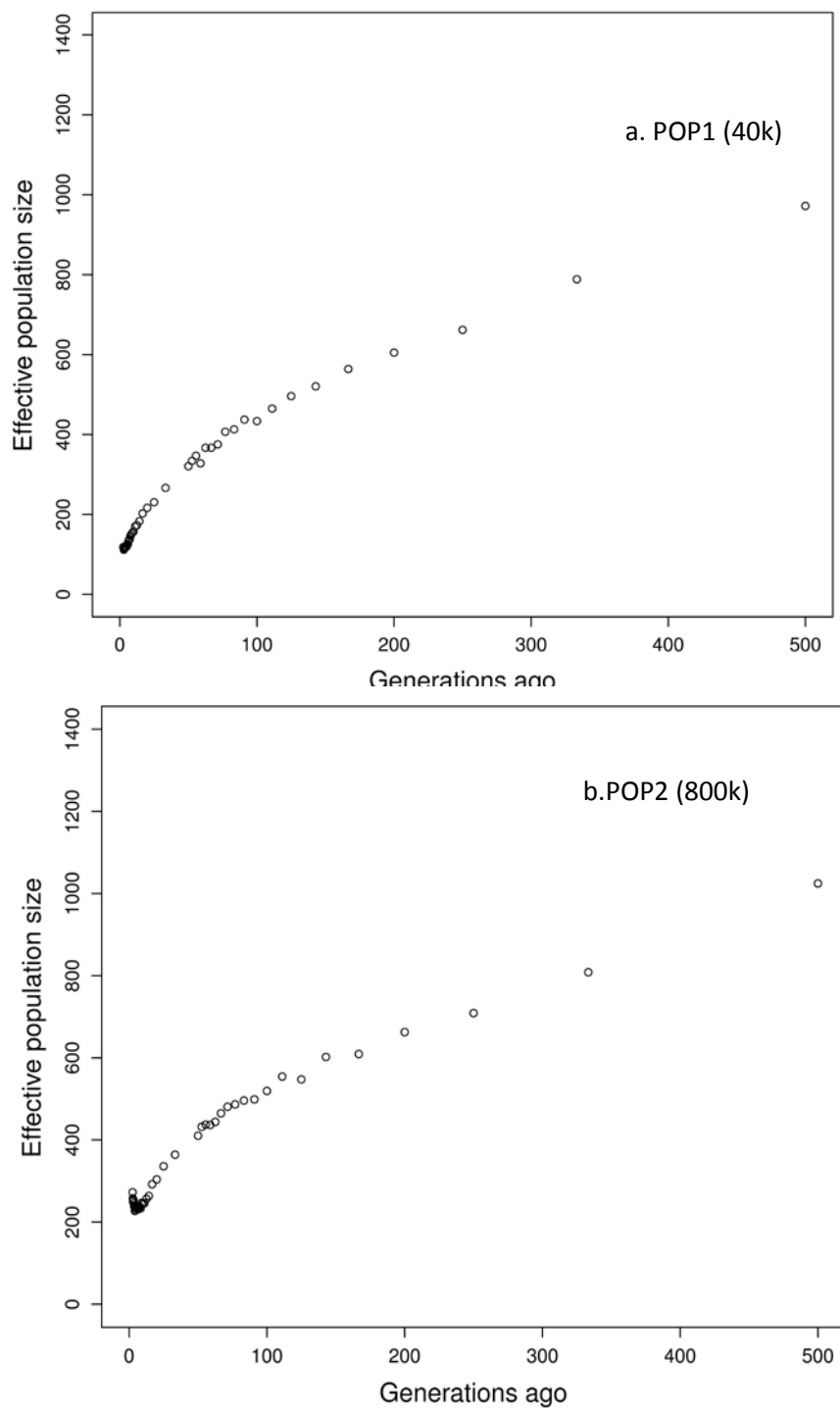
**Figure 2. Schematic representation of the simulated scenarios.**

n\_TS: number of bulls in the training set. Accuracies of EBV for all n\_TS were 0.79, 0.90 or 0.94 for heritability =0.10, 0.25 or 0.40, respectively. The average progeny size was 73 for all n\_TS and heritability levels.



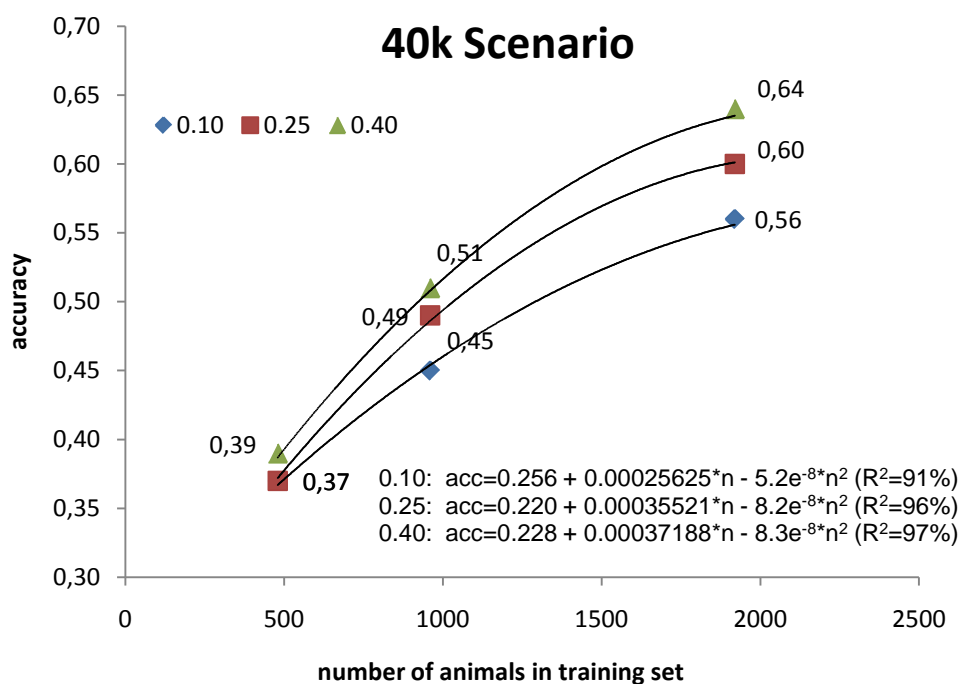
**Figure 3. Extent of Linkage Disequilibrium (LD).**

The extent of LD is shown on chromosome 1 in the recent generations for the scenario of moderate heritability (0.25) and POP1 (a) and POP2 (b). The mean  $r^2$  over successive intervals of 0.1 Mb is plotted in white. For the graph for POP2 (800k), 40,000 markers were picked randomly to represent the 800k panel. The results are for one replicate.



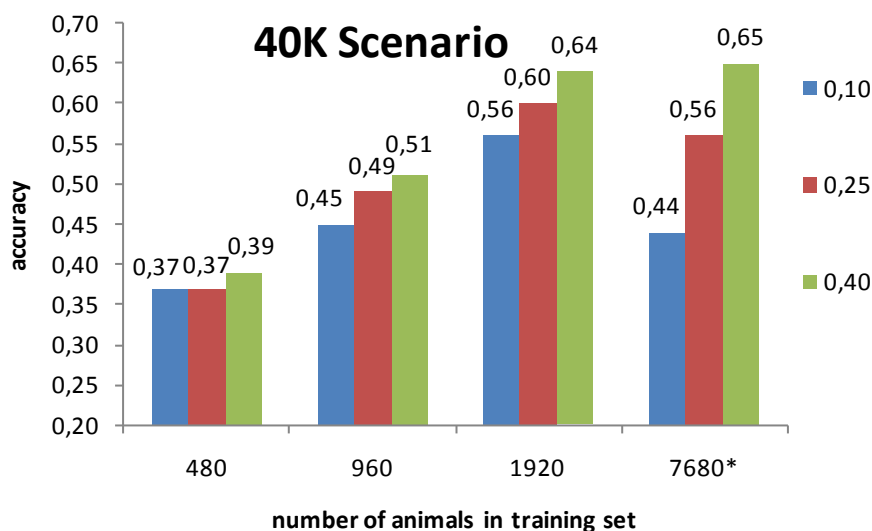
**Figure 4. Effective Population Size.**

Estimated historical effective population size over time for POP1 (a) and POP2 (b). The results are for one replicate.



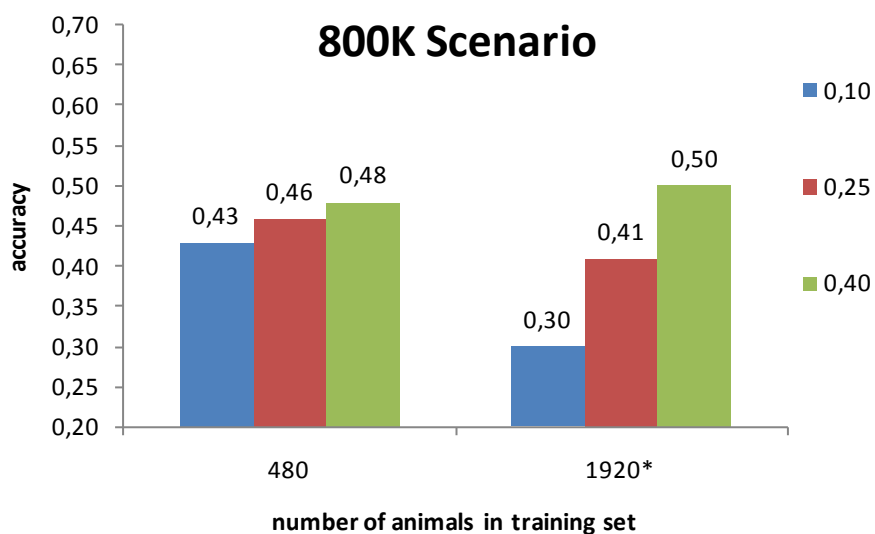
**Figure 5. Accuracy of Direct Genomic Estimated Breeding Value (DGEBV) for 40k Scenario.**

DGEBV as a function of number of animals in the training set, considering three levels of heritability and 40k markers. The results are presented as the average of 10 replicates.



**Figure 6. Accuracy of Direct Genomic Estimated Breeding Value (DGEBV) for 40k Scenario.**

DGEBV for different number of animals in the training set, considering three heritability levels and 40k markers. The marker effects were estimated using the estimated breeding values (EBVs) of the animals, with the exception of scenario 7680\*, for which phenotypes were used. The results are presented as the average of 10 replicates.



**Figure 7. Accuracy of Direct Genomic Estimated Breeding Value (DGEBV) for 800k Scenario.**

DGEBV for different number of animals in the training set, considering three heritability levels and 800k markers. The marker effects were estimated using the estimated breeding values (EBVs) of the animals, with the exception of scenario 1920\*, for which phenotypes were used. The results are presented as the average of 10 replicates.

### 3. CONSIDERAÇÕES FINAIS

A análise de pedigree realizada em um rebanho grande de gado de corte submetido há vários anos de seleção pela DEP mostrou que houve perda de diversidade genética, porém, possivelmente através de estratégias de acasalamentos, houve controle desta perda. O uso continuado de seleção acurada baseada na informação própria e de parentes e o acasalamento de animais mais aparentados do que a média da população levam a uma diminuição do  $N_e$  da população, o que por sua vez, aumenta a deriva genética, resultando na perda de diversidade. Ainda que, neste trabalho, tenha sido detectada perda de diversidade genética, os valores de  $N_e$  são maiores do que em raças leiteiras sujeitas a intensa seleção, e raças de corte cujo o uso intensivo de poucos reprodutores resultou em  $N_e$  menores que 100.

Parâmetros de DL para gado de corte encontrados na literatura foram utilizados para simular populações que resultaram em DL similares aos da literatura e de  $N_e$  similares aos estimados nesta investigação . O processo de simulação atingiu o objetivo de gerar uma estrutura de DL similar às de populações de gado de corte sujeitas a seleção.

A acurácia da DGEV aumentou significativamente ( $p < 0,05$ ) com o aumento do número de touros no grupo de treinamento, herdabilidade da característica e densidade dos marcadores. O número de touros no GT teve um efeito quadrático sobre as acurácias do VGD para todas as herdabilidades



( $p < 0,0001$ ). Os resultados mostraram que, assim como para gado de leite, em gado de corte os ganhos ao incorporar a informação genômica são dependentes, principalmente, do número de animais no grupo de treinamento, mas com a diferença de que, em gado de corte provavelmente um número ainda maior de animais no grupo de treinamento será necessário devido às diferenças na estrutura do DL assim como painéis mais densos também serão necessários, uma vez que, para 40k marcadores, as acurácias do VGD parecem atingir um limite. Como a variável utilizada para a estimação dos efeitos dos marcadores afeta a acurácia da DGEVB, é esperado que, na ausência de um número suficiente de touros com VGE acurado, o número requerido de touros no grupo de treinamento para obter-se uma DGEVB acurada irá ser ainda maior. É esperado que um painel mais denso de marcadores diminua este requerimento, porém deverá ser acompanhado de um aumento no número de animais no grupo de treinamento para reduzir o efeito da super-parametrização. Portanto, é clara a importância da união de esforços de instituições envolvidas na pesquisa e produtores no sentido de, através de um trabalho conjunto e do compartilhamento de genótipos, conseguir atingir a meta de incorporar informação genômica nos programas de melhoramento genético com ganhos substanciais para produtores e indústria de gado de corte.

Os resultados da simulação mostraram que o maior nível de acurácia da DGEVB (0,64), considerando um painel de 40K, foi alcançado com 1920 touros com VGE altamente acurados (0,94) e herdabilidade média a alta (0,40), sendo similar à acurácia (0,65) obtida quando se utilizou um número 4

vezes maior de animais e o fenótipo ao invés do VGE para estimação dos efeitos dos marcadores. Se considerarmos que a acurácia do VGE de animais jovens baseado na informação dos pais, para esta herdabilidade, é de aproximadamente 0,58, o ganho ao utilizar-se a informação genômica ao invés do VGE é de 10 pontos percentuais. Além disso, observou-se que, à medida que o nível de herdabilidade decresceu, a acurácia da DGE BV também decresceu, porém em menor proporção, mostrando uma maior vantagem do uso da seleção genômica para características de média a baixa herdabilidade. Considerando  $h^2=0,25$  e  $h^2=0,10$ , as acurácias esperadas para o VGE baseado na média dos pais são  $\sim 0,53$  e  $\sim 0,44$ , respectivamente. Observando-se o resultado para  $n=1920$ , as acurácias são 0,60 e 0,56 para estes níveis de herdabilidade, mostrando um ganho proporcionalmente maior (13 e 27 pontos percentuais, respectivamente) do que para a herdabilidade mais alta (0,40) evidenciando a vantagem do uso da informação genômica para características de baixa herdabilidade.

Como os testes de progênie produzem EBV bastante acuradas, a vantagem da seleção genômica é mais evidente para animais jovens. Se considerarmos que, em gado de leite, onde a seleção genômica já está sendo realizada, as características de interesse são medidas bem tarde na vida do animal e em somente um sexo, ao contrário do que ocorre em gado de corte, onde muitas características são medidas em ambos os sexos e podem ser tomadas mais cedo na vida do animal, espera-se que, para gado de corte, o ganho adicional advindo da seleção genômica, seja para aquelas características de difícil medição ou que são medidas mais tarde como as de

carcaça e qualidade de carne, por exemplo, que, frequentemente, são medidas após o abate do animal. Há também as medidas em um dos sexos, como as de fertilidade, que, aliada à baixa herdabilidade, tornam-se um dos focos principais da seleção genômica em gado de corte.

Outra questão que envolve a viabilidade da incorporação da informação genômica nos programas de melhoramento genético é o custo da genotipagem. Uma alternativa para reduzir este custo é a imputação de genótipos, o que permitiria genotipar um número maior de animais para painéis menos densos (3k), portanto de menor custo, e imputar painéis mais densos (50k, por exemplo) usando uma população referência genotipada para o painel mais denso. Esta alternativa deverá ser considerada como primordial nos projetos envolvendo seleção genômica, principalmente no Brasil, onde o número de touros disponíveis para genotipagem com VGE altamente acuradas é muito inferior ao relatado neste trabalho onde se atingiu a maior acurácia ( $n=1920$  para  $h^2=0,40$ ). Assim, poderiam ser agregados ao grupo de treinamento touros com VGE menos acuradas, porém em um número maior.

Considerando os resultados desta pesquisa, recomendam-se estudos adicionais com o objetivo de avaliar as conseqüências de extensões de DL específicos em gado de corte em diferentes populações e sistemas de produção, assim como o número de animais no grupo de treinamento que resultaria no sucesso da incorporação da informação genômica na predição dos valores genéticos, considerando, também, possíveis estratégias de imputação genotípica.

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, S.E.M.; MACHADO, M.S.N.; STEIGLEDER, C.S. et al. Genetic diversity in a Brazilian bovine herd based on four microsatellite loci. **Genet. Mol. Biol.**, Ribeirão Preto, v.23, n.2, p.347-350, 2000.

ANDERSON, L. Genetic dissection of phenotypic diversity in farm animals. **Genetics**, Pittsburg, v.2, p.130-138, 2001.

ANUÁRIO da pecuária brasileira – ANUALPEC 2010. São Paulo: Informa Economics FNP, 2010. 368 p.

ARDLIE, K.G.; LIU-CORDERO, S.N.; EBERLE, M.A. et al. Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. **Am. J. Hum. Genet.**, Amsterdam, v.69, p.582-589, 2001.

ARDLIE, K.G.; KRUGLYAK, L.; SEIELSTAD, M. 2002. Patterns of linkage disequilibrium in the human genome. **Nature Reviews: Genetics**, London, v.3, p. 299-309, 2002.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNE. **Rebanho Bovino Brasileiro**. Disponível em <[http://www.abiec.com.br/3\\_rebanho.asp](http://www.abiec.com.br/3_rebanho.asp)>. Acesso em: jul. 2010.

BARENDSE, W.; BUNCH, R.; THOMAS, S. et al. The TG5 DNA marker test for marbling capacity in Australian feedlot cattle. In: BEEF QUALITY CRC MARBLING SYMPOSIUM, 2001, Coffs Harbour, Australia. **Proceedings...** Coffs Harbour: CRC Beef, 2001. p. 52-57.

BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F.S. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, London, v.11, n.421, 2010.

BOICHARD, D.; Maignel, L.; VERRIER, E. The value of using probabilities of gene origin to measure genetic variability in a population. **Genet. Sel. Evol.**, Paris, v.29, p.5-23, 1997.

BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The genome sequence of taurine cattle: a window to ruminant biology and evolution **Science**, Washington, v.324, p.522-528, 2009.

BOVINE HAPMAP CONSORTIUM. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds **Science.**, Washington, v.324, p.528-532, 2009.

BRADLEY, D.G.; MACHUGH, D.E.; CUNNINGHAM, P. et al. Mitochondrial diversity and the origins of African and European cattle. **Proc. Natl. Acad. Sci.**, Washington, v.93, p.5131-5135, 1996.

BROWN, W.L. Genetic diversity and genetic vulnerability – an appraisal. **Econ. Bot.**, New York, v.37, n.1, p.4-12, 1983.

BROWN-GUEDIRA, G.L.; THOMPSON, J.A.; NELSON, R.L. et al. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. **Crop Sci.**, Madison, v.40, p.815-823, 2000.

CABALLERO, A. Developments in the prediction of effective population size. **Heredity**, Midlothian, v.73, p. 657-679, 1994.

CAETANO, A.R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **R. Bras. Zootec.**, Viçosa, v.38(supl especial), p. 64-71, 2009.

CALUS, M.P.L.; MEUWISSEN, T.H.E.; de ROOS, A.P.W. et al. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, Pittsburg, v.178, p.553-561, 2008.

CARNEIRO, T.X.; GONÇALVES, E.C; SCHNEIDER, M.P.C et al. Diversidade genética e eficiência de DNA microssatélites para o controle genealógico da raça Nelore. **Arq. Bras. Med. Vet. Zootec.**, Belo Horizonte, v.59, n.5, p.1257-1262, 2007.

CASAS, E.; SHACKELFORD, S.D.; KEELE, J.W et al. Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. **J. Anim. Sci.**, Champaign, v.78, p.560-569, 2000.

CASAS, E.; WHITE, S.N.; RILEY, D.G. et al. Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle. **J. Anim. Sci.**, Champaign, v.83, p.13-19, 2005.

CIOBANU, D.C.; BASTIAANSEN, W.M.; LONERGAN, S.M. et al. New alleles in calpastatin gene are associated with meat quality traits in pigs. **J. Anim. Sci.**, Champaign, v.82, p.2829-2839, 2004.

COCKERHAM, C.C. Variance of gene frequencies. **Evolution**, Lawrence, v.23, p.72-84, 1969.

COCKERHAM, C.C. Analysis of gene frequencies. **Genetics**, Pittsburg, v.74, p.679-700, 1973.

COLLINS, A.; LONJOU, C.; MONTON, N.E. Genetic epidemiology of single-nucleotide polymorphisms **Proc. Natl. Acad. Sci. USA**, Washington, v.96, p.15173-15177, 1999.

DE ROOS, A.P.W.; HAYES, B.J.; SPELMAN, R.J. et al. Linkage disequilibrium and persistence of phase in Holstein-Friesian Jersey and Angus cattle. **Genetics**, Pittsburg, v.179, p.1503-1512, 2008.

DE ROOS, A.P.W.; HAYES, B.J.; GODDARD, M.E. Reliability of genomic predictions across multiple populations. **Genetics**, Pittsburg, v.183, p.1545-1553, 2009.

DEKKERS, J.C.M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews: Genetics**, London, v.3, p. 22-32, 2001.

DEKKERS, J.C.M. Commercial application of marker -and gene- assisted selection in livestock: strategies and lessons. **J. Anim. Sci.**, Champaign, v.82(e-suppl.), p. E313-E328, 2004.

DICKINSON, W.F.; LUSH, J.L. Inbreeding and the genetic history of the Rambouillet sheep in America. **J. Heredity**, Newport, v.24, p.19-33, 1933.

DRUET, T.; SCHROOTEN, C.; ROOS, A.P.W. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. **J. Dairy Sci.**, Champaign, v.93 p.5443–5454, 2010.

DU, F.X.; CLUTTER, A.C.; LOHUIS, M.M. Characterizing linkage disequilibrium in pig populations **Int. J. Biol. Sci.**, New South Wales, v.3, p.166-178, 2007.

DUNNING, A.M.; DUROCHER, F.; HEALEY, S. et al. The extent of linkage disequilibrium in four populations with distinct demographic histories. **Am. J. Hum. Genet.**, Amsterdam, v.67, p.1544-1554, 2000.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4<sup>th</sup> ed. Harlow, UK: Longman Scientific and Technical, 1996.

FARIA, F.J.C.; FILHO, A.E.V., MADALENA, F.E. et al. Pedigree analysis in the Brazilian Zebu breeds **J. Anim. Breed. Genet.**, Malden, v.126, p.148-153, 2009.

FARNIR, F.; COPPIETERS, J.; ARRANZ, J.J. et al. Extensive genome-wide linkage disequilibrium in cattle. **Genome Res.**, Woodbury, v.10, p.220-227, 2000.

FELSENSTEIN, J. Distance methods for inferring phylogenies: a justification. **Evolution**, Lawrence, v.38, p.16-24, 1984.

FERNANDO R.L.; GROSSMAN M. Marker assisted selection using best linear unbiased prediction. **Genet. Sel. Evol.**, Paris, v.21, p.467-477, 1989.

GARRICK, D.J. The nature and scope of some whole-genome analysis in US beef cattle In: BEEF IMPROVEMENT FEDERATION ANNUAL RESEARCH SYMPOSIUM , 41, 2009, Sacramento. **Proceedings...**Sacramento: Beef Improvement Federation, 2009. Disponível em: <[http://www.bifconference.com/bif2009/proceedings/G3\\_pro\\_Garrick.pdf](http://www.bifconference.com/bif2009/proceedings/G3_pro_Garrick.pdf)>. Acesso em: set. 2010.

GARRICK, D.J. The nature, scope and impact of some whole-genome analysis in beef cattle In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 9., 2010, Leipzig. **Proceedings...**Leipzig: German Society for Animal Science, 2010. 0023. 1 CD-ROM.

GAUTIER, M.; FARAUT, T.; MOAZAMI-GOUDARZI, K. et al. Genetic and haplotypic structure in 14 European and African cattle breeds. **Genetics**, Pittsburg, v.177, p.1059-1070, 2007.

GELDERMANN, H. Investigation on inheritance of quantitative characters in animals by gene markers: 1. Methods. **Theor. Appl. Genet.**, Stuttgart. v.46, p.319-330, 1975.

GIANOLA, D.; PEREZ-ENCISO, M.; TORO, M.A. On marker-assisted prediction of genetic value: beyond the ridge. **Genetics**, Pittsburg, v.163, p.347-365, 2003.

GODDARD, M.G.; SMITH, C. Optimum number of bull sires in dairy cattle breeding. **J. Dairy Sci.**, Champaign, v.73, p.1113-1122, 1990.

GODDARD, M.E. Mapping genes for quantitative traits using linkage disequilibrium. **Genet. Sel. Evol.**, Paris, v.23(Suppl 1), p.131s-134s, 1991.

GODDARD, M.E.; MEUWISSEN, T.H.E. The use of linkage disequilibrium to map quantitative trait loci. **Aust. J. Exp. Agr.**, Collingwood, v.45, p.837-845, 2005.

GODDARD, M.E.; HAYES, B.J.; MCPARTLAN, H. et al. Can the same genetic markers be used in multiple breeds? In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Proceedings...**Belo Horizonte, 2006. 1 CD-ROM.

GODDARD, M.E.; HAYES, B.J.; MACEACHERN, S. et al. Evolutionary history of cattle based on genomics. In: CANADIAN SOCIETY OF ANIMAL SCIENCE ANNUAL MEETING, 2008, Guelph. **Abstracts...** Guelph: CSAS, 2008. p. 2.

GODDARD, M. Genomic selection: prediction of accuracy and maximization of long term response. **Genetica**, Gif-sur-Yvette, v.136, p. 245-257, 2009.

GRAPES, L.; DEKKERS, J.C.M.; ROTHSCILD, M.F. et al. Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. **Genetics.**, Pittsburg, v.166, p.1561-1570, 2004.

GUTIERREZ, J.P.; ALTARRIBA, J.; DIAZ, C. et al. 2003. Pedigree analysis of eight Spanish beef cattle breeds. **Genet. Sel. Evol.**, Paris, v.35, p.43-63, 2003.

HAMMAMI, H.; CROQUET, C.; STOLL, J. et al. Genetic diversity and joint-pedigree analysis of two importing Holstein populations. **J. Dairy Sci.**, Champaign, v.90, p.3530-3541, 2007.

HAYES, B.J.; VISSCHER, P.M.; MCPARTLAN, C. et al. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. **Genome Res.**, Woodbury, v.13, p.635-643, 2003.

HAYES, B.J.; CHAMBERLAIN A.J.; GODDARD, M.E. Use of markers in linkage disequilibrium with QTL in breeding programs. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Proceedings...**Belo Horizonte, 2006. 1 CD-ROM.

HAYES, B.J.; CHAMBERLAIN, A.J.; MCPARTLAN, H. et al. Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. **Genet. Res.**, Mackay, v.89, p.215-220, 2007.

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

HILL, W.G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theor. Appl. Genet.**, Stuttgart, v.38, p.226-231, 1968.

HILL, W.G. Maintenance of quantitative genetic variation in animal breeding programmes. **Livest. Prod. Sci.**, Amsterdam, v.63, p.99-109, 2000.

IBANEZ-ESCRICHE, N. Genomic selection of purebreds for crossbred performance. **Genet. Sel. Evol.**, Paris, v.41, p.12, 2009.

ILLUMINA. **BovineSNP50 Genotyping BeadChip**. San Diego, Califórnia, 2010a. Disponível em:  
<[http://www.illumina.com/Documents/products/datasheets/datasheet\\_bovine\\_snp50.pdf](http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf)>. Acesso em: set. 2010.



- ILLUMINA. **BovineHD Genotyping BeadChip**. San Diego, Califórina, 2010b. Disponível em: <[http://www.illumina.com/Documents/products/datasheets/datasheet\\_bovineHD.pdf](http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf)>. Acesso em: set. 2010.
- ISAAC, E. On the domestication of cattle. **Science**, Washington, v.137, p.195-204, 1962.
- JAMES, J.W. Computation of genetic contributions from pedigrees **Theor. Appl. Genet.**, Stuttgart, v.42, p.272-273, 1972.
- KELLY, M.J.; SARGOLZAEI, M.; WANG, Z. et al. Estimated linkage disequilibrium in a multi-breed beef herd based on the Illumina BovineSNP50 BeadChip. **J. Anim. Sci.**, Champaign, v.86(E-Suppl 2), n.410, 2008.
- KHATKAR, M.S.; ZENGER, K.R.; HOBBS, M. et al. A primary assembly of a bovine haplotype block map based on a 15k SNP panel genotyped in Holstein-Friesian cattle. **Genetics**, Pittsburg, v.176, p.763-772, 2007.
- KIM, E.S.; KIRKPATRICK, B.W. Linkage disequilibrium in the North American Holstein population. **Anim. Genet.**, Oxford, v.40, p.279-288, 2009.
- KIZILKAYA, K.; FERNANDO, R.L.; GARRICK, D.J. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. **J. Anim. Sci.**, Champaign, v.88, p. 544-551, 2010.
- KRUGLYAK, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. **Nature Genetics**, London, v.22, p.139-144, 1999.
- LACY, R.C. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. **Zoo. Biol.**, Malden, v.8, p.111-123, 1989.
- LANDER, E.S.; SCHORK, N.J. Genetic dissection of complex traits. **Science**, Washington, v.265, p.2037-2048, 1994.
- LEWONTIN, R.C. The interactions of selection and linkage. I. General considerations; heterotic models. **Genetics**, Pittsburg, v.49, p.49-67, 1964.
- LOFTUS, R.T.; MACHUGH, E.; BRADLEY, D.G. et al. Evidence for two independent domestications of cattle. **Proc. Natl. Acad. Sci USA**, Washington, v.91, p.2757-2761, 1994.
- LU, D.; SARGOLZAEI, M.; KELLY, M. et al. Extent of linkage disequilibrium in purebred and crossbred beef cattle. **J. Dairy Sci.**, Champaign, v.92(E-Suppl 1): n.43, 2008.

LUTAAYA, E.; MISZTAL, I.; BERTRAND, J.K. et al. Inbreeding in populations with incomplete pedigrees. **J. Anim. Breed. Genet.**, Malden, v.116, p.475-480, 1999.

MACCLUER, J.W.; VANDEBERG, J.; READ, B. et al. Pedigree analysis by computer simulation. **Zoo. Biol.**, Malden, v.5, p.147-160, 1986.

MACEACHERN, S.; HAYES, B.; MCEWAN, J. et al. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos Taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. **BMC Genomics**, London, v.10, n.181, 2009.

MACHUGH, D.E.; SHRIVER, M.D.; LOFTUS, R.T. et al. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos Taurus* and *Bos indicus*) **Genetics**, Pittsburg, v.146, n.3, p.1071-1086, 1997.

MAGNABOSCO, C.U.; CORDEIRO, C.M.T.; TROVO, J.B.F. et al. Catálogo de linhagens do germoplasma zebuino: raça Nelore. **Embrapa-Cenargen, Documentos**, Brasília, n. 23, 1997.

MARQUES, E.; SCHNABEL, D.; STOTHARD, P. et al. High density linkage disequilibrium maps of chromosome 14 in Holstein and Angus cattle. **BMC Genetics**, London, v.9, n.45, 2008.

MARTINEZ, R.A.; GARCIA, D.; GALLEGO, J.L. et al. Genetic variability in Colombian Creole cattle populations estimated by pedigree information. **J. Anim. Sci.**, Champaign, v.86, p.545-552, 2007.

MCKAY, S.D.; SCHNABEL, D.; MURDOCH, M. et al. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, London, v.8, n.74, 2007.

MEUWISSEN, T.H.E.; WOOLLIAMS, J.A. Effective sizes of livestock populations to prevent a decline in fitness. **Theor. and Appl. Genet.**, Stuttgart, v.89, p.1019-1026, 1994.

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

MEUWISSEN, T.H.E. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. **Genet. Sel. Evol.**, Paris, v.41, n.35, 2009.

NEI, M. Analysis of gene diversity in subdivided populations. **Proc. Natl. Acad. Sci. USA**, Washington, v.70, p.3321-3323, 1973.

RAO, V.R.; HODGKIN, T. Genetic diversity and conservation and utilization of plant genetic resources. **Plant Cell Tiss. Organ Cult.**, Urbana, v.68, p.1-19, 2002.

REICH, D.E.; LANDER, E.S. On the allelic spectrum of human disease. **Trends Genet.**, Amsterdam, v.17, p.502-510, 2001.

RIQUET, J.; COPPIETERS, W.; CAMBISANO, N. et al. Identity-by-descent fine-mapping of QTL in outbred populations: application to milk production in dairy cattle. **Proc. Natl. Acad. Sci.**, Washington, v.96, p.9252-9257, 1999.

SANTIAGO, A.A. **O Zebu na Índia, no Brasil e no Mundo**. Instituto Campineiro de Ensino Agrícola, Campinas. 1985.

SARGOLZAEI, M., SCHENKEL, F.S., JANSEN, G.B. et al. Extent of Linkage Disequilibrium in Holstein Cattle in North America, **J. Dairy Sci.**, Champaign, v.91, p.2106-2117, 2008.

SARGOLZAEI, M.; SCHENKEL, F.S. Extent and pattern of linkage disequilibrium in North American Holstein cattle evaluated using a 50k SNP panel In: KENNEDY CONFERENCE ON QUANTITATIVE GENETICS AND ANIMAL BREEDING, 2008. **Abstracts...** Guelph. 2009. pp. 125-126.

SARGOLZAEI, M.; SCHENKEL, F.S.; CHESNAIS, J. Comparison between two methods of imputation (AIPL vs Boviteq/CGIL) for 6,246 genotypes provided by AIPL. In: DAIRY CATTLE BREEDING AND GENETICS COMMITTEE MEETING, 2010. **Report...** Canada:University of Guelph, 2010.

SCHENKEL, F.S.; MILLER, S.P.; JIANG, Z. et al. Association of a single nucleotide polymorphism in the calpastatin gene with carcass and meat quality traits of beef cattle. **J. Anim. Sci.**, Champaign, v.84, p.291-299, 2006.

SCHENKEL, F.S.; SARGOLZAEI, M.; KISTEMAKER, G. et al. Reliability of genomic evaluation of Holstein cattle in Canada In: INTERBULL INTERNATIONAL WORKSHOP – GENOMIC INFORMATION IN GENETIC EVALUATIONS, 39, 2009. **Proceedings...** Uppsala, Sweden Disponível em: <<http://www-interbull.slu.se/bulletins/framesida-pub.htm>>. Acesso em: nov. 2010.

SHAEFFER, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. **J. Anim. Breed. Genet.**, Malden, v.123, p.218-223, 2006.

SILVA, C.R.; NEVES, H.H.R.; QUEIROZ, S.A. et al. Extent of linkage disequilibrium in Brazilian Gyr dairy cattle based on genotypes of AI sires for dense SNP markers In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 9., 2010, Leipzig. **Proceedings...** Leipzig: German Society for Animal Science, 2010. PP1-29. 1 CD-ROM.

SMITH, L.A.; CASSEL, B.G.; PEARSON, R.E. 1998. The effects of inbreeding on the lifetime performance of dairy cattle. **J. Dairy Sci.**, Champaign, v.81, p.2729-2737, 1998.

SORENSEN, A.C.; SORENSEN M.K.; BERG, P. Inbreeding in Danish dairy cattle breeds. **J. Dairy Sci.**, Champaign, v.88, p.1865-1872, 2005.

SPELMAN, R.J.; VAN ARENDONK, J.A.M. Effect of inaccurate parameter estimates on genetic response to marker assisted selection in an outbred population. **J. Dairy Sci.**, Champaign, v.80, p.3399-3410, 1997.

SPELMAN, R.J.; GARRICK, D.J.; VAN ARENDONK, J.M. Utilization of genetic variation by marker assisted selection in commercial dairy cattle populations. **Livest. Prod. Sci.**, Amsterdam, v.59, p.51-60, 1999.

STEPHENS, J.C.; Schneider, J.A.; Tanguay, D.A. et al. Haplotype variation and linkage disequilibrium in 313 human genes. **Science**, Washington, v.293, p.489-493, 2001.

TENESA, A.; KNOTT, S.A.; WARD, D. et al. Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. **J. Anim. Sci.**, Champaign, v.81, p.617-623, 2003.

THALLMAN, R.M. Whole-genome selection project involving 2,000 industry sires. In: ANNUAL RESEARCH SYMPOSIUM AND ANNUAL MEETING, 41., 2009, Grass Valley, California. **Proceedings...** Beef Improvement Federation, 2010. C4.1. Disponível em: <<http://beefimprovement.org/PDFs/BIF%202009%20PROGRAM.pdf>>. Acesso em: ago. 2010.

THÉVENON, S.; DAYO, G.K.; SYLLA, S. et al. 2007. The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. **Anim. Genet.**, Oxford, v.38, p.277-286, 2007.

THODAY, J.M. Location of polygenes. **Nature**, London, v.191, p.368-370, 1961.

THOMPSON, J.A.; NELSON, R.L.; VODKIN, L.O. Identification of diverse soybean germplasm using RAPD markers. **Crop Sci.**, Madison, v.38, p.1348-1355, 1998.

UNITED NATIONS ENVIRONMENT PROGRAMME 2010. Disponível em: <<http://www.unep.org>>. Acesso em: mar. 2010.

VANRADEN, P.M. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. **J. Dairy Sci.**, Champaign, v.75, p.3136-3144, 1992.

VANRADEN, P.M. Efficient estimation of breeding values from dense genomic data. **J. Dairy Sci.**, Champaign, v.90(suppl), p.374-375, 2007.

VANRADEN, P.M.; VAN TASSEL, C.P.; WIGGANS, G.R. et al. Invited review: reliability of genomic predictions for North American Holstein bulls **J. Dairy Sci.**, Champaign, v.92, p.16-24, 2009.

VASCONCELLOS, L.P.M.K.; TAMBASCO-TALHARI, D.; PEREIRA, A.P. et al. Genetic characterization of Aberdeen Angus cattle using molecular markers. **Genet. Mol. Biol.**, Ribeirão Preto, v.26, n.2, p.133-137, 2003.

VERRIER, E.; COLLEAU, J.J.; FOULLEY, J.L. Methods for predicting genetic response to selection in small populations under additive genetic models: a review. **Livest. Prod. Sci.**, Amsterdam, v.29, p.93-114, 1991.

VERRIER, E.; COLLEAU, J.J.; FOULLEY, J.L. Long-term effects of selection based on the animal model BLUP in a finite population. **Theor. Appl. Genet.**, Stuttgart, v.87, p.446-454, 1993.

VOZZI, P.A.; MARCONDES, C.R.; BEZERRA, L.A.F. et al. Pedigree analysis in the breeding program for Nellore cattle **Genet. Mol. Res.**, Ribeirão Preto, v.6, n.4, p.1044-1050, 2007.

WRAY, N.R.; WOOLLIAMS, J.A.; THOMPSON, R. Methods for predicting rates of inbreeding in selected populations. **Theor. Appl. Genet.**, Stuttgart, v.80, p.503-512, 1990.

WRIGHT, S. Coefficients of inbreeding and relationship. **Amer. Soc. Nat.**, Chicago, v.56, n.645, p.330-338, 1922.

WRIGHT, S. Evolution in mendelian populations. **Genetics**, Pittsburg, v.16, p.97-159, 1931.

WRIGHT, S. Size of population and breeding structure in relation to evolution. **Science**, Washington, v.87, p.430-431, 1938.

WRIGHT, S. The genetic structure of populations. **Ann. Eugen.**, New York, v.15, p.323-354, 1951.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Pittsburg, v.163, p.789-801, 2003.

ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **J. Dairy Sci.**, Champaign, v.93 p.5487-5494, 2010.

ZHAO, H.; NETTLETON, D.; SOLLER, M. et al. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. **Genet. Res.**, Mackay, v.86, p.77-87, 2005.

ZHAO, H.; FERNANDO, L.; DEKKERS, J.C.M. Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. **Genetics**, Pittsburg, v.175, p.1975-1986, 2007a.

ZHAO, H.; NETTLETON, D.; DEKKERS, J.C.M. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms **Genet. Res. Camb.**, Cambridge, v.89, p.1-6, 2007b.

## 5. APÊNDICES

### APÊNDICE 1: Normas para publicação - Genetics and Molecular Biology

Conteúdo disponível em <http://www.scielo.br/revistas/gmb/iinstruc.htm>

**Genetics and Molecular Biology** (formerly named Revista Brasileira de Genética/Brazilian Journal of Genetics - ISSN 0100-8455) is published by the Sociedade Brasileira de Genética ([Brazilian Society of Genetics](#)).

The Journal considers contributions that present the results of original research in genetics, evolution and related scientific disciplines.

Although **Genetics and Molecular Biology** is an official publication of the Brazilian Society of Genetics, contributors are not required to be members of the Society.

It is a fundamental condition that submitted manuscripts have not been and will not be published elsewhere. With the acceptance of a manuscript for publication, the publishers acquire full and exclusive copyright for all languages and countries.

Manuscripts considered in conformity with the scope of the journal, judged by the Editor in conjunction with the Editorial Board, are reviewed by the Associate Editor and two or more external reviewers. Acceptance by the Editor is based on the quality of the work as substantial contribution to the field and on the overall presentation of the manuscript.

#### 1. Manuscripts should be submitted to:

Angela M. Vianna – Morgante, Editor-in-Chief  
Genetics and Molecular Biology  
E-mail: [editor@gmb.org.br](mailto:editor@gmb.org.br)

#### 2. A submission package sent to the Editorial Office must contain:

**a)** The manuscript that must be submitted by the Corresponding Author. This is the person who will also check the page proofs, and arranges for the payment that may incur during the editorial process.

**b)** An accompanying cover letter, signed by the corresponding author, stating that the data have not been published and are not under consideration elsewhere, and that all authors have approved the submission of the manuscript. It must also inform the e-mail addresses of all other authors so that they can be contacted by the Editorial Office for confirmation of the submission. Possible conflicts of interest (e.g. due to funding, consultancies) must be

disclosed.

**c)** An electronic copy of the text, tables and figures, including supplementary material to be published online only. Formats for text are Word or RTF in Windows platform. Images in TIFF or JPEG formats should be sent in separate files (For Figures, see detailed instructions in 3.1.h).

**d)** Manuscripts including photos or any other identifiable data of human subjects must be accompanied by a copy of the signed consent by the individual or his/her guardian.

Failure to adhere to these guidelines can delay the handling of your contribution and manuscripts may be returned before being reviewed

### 3. Categories of Contribution

#### 3.1. Research Articles

Manuscripts must be written in English in double-spaced, 12-point type throughout; formatted to A4 paper with 2.5 cm margins; marked with consecutive line and page numbers, beginning with the cover page.

The following elements must start on a new page and be ordered as they are listed below:

a) **The title page** must contain: a concise and informative title; the authors' names (first name at full length); the authors' institutional affiliation, including department, institution, city, state or province and country; different affiliations indicated with superscript Arabic numbers; a short running title of about 35 characters, including spaces; up to five key words; the corresponding author's name, postal address, phone and fax numbers and email address.

b) **The Abstract** must be a single paragraph that does not exceed 200 words and summarizes the main results and conclusions of the study. It should not contain references.

c) **The text** must be as succinct as possible. Text citations: articles should be referred to by authors' surnames and date of publication; citations with two authors must include both names; in citations with three or more authors, name the first author and use et al. List two or more references in the same citation in chronological order, separated by semi-colons. When two or more works in a citation were published in the same year, list them alphabetically by the first author surname. For two or more works by the same author(s) in a citation, list them chronologically, with the years separated by commas. (Example: Freire-Maia et al., 1966a, 1966b, 2000). Only articles that are published or in press should be cited. In the case of personal communications or unpublished results, all contributors must be listed by initials and last name (et al. should not be used). Numbers: In the text, numbers nine or less must be written out except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numerals for numbers larger than nine. Binomial Names: Latin names of genera, species and infraspecific taxa must be printed in italics; names of orders and families should appear in the Title and also when first mentioned in the text. URLs for programs, data or other sources should be listed in the Internet Resources Section, immediately after the References



Section, not in the text. The text includes the following elements:

*Introduction* - Description of the background that led to the study.

*Material (or Subjects) and Methods* - Details relevant to the conduct of the study. Statistical methods should be explained at the end of this section.

*Results* - Undue repetition in text and tables should be avoided. Comment on significance of results is appropriate but broader discussion should be part of the Discussion section.

*Discussion* - The findings of the study should be placed in context of relevant published data. Ideas presented in other publications should not be discussed solely to make an exhaustive presentation.

Some manuscripts may require different formats appropriate to their content.

d) **The Acknowledgments** must be a single paragraph that immediately follows the discussion and includes references to grant support.

e) **The References Section:** references must be ordered alphabetically by the first author surname; references with the same first author should be ordered as follows: first, as single author in chronological order; next, with only one more co-author in alphabetical order by the second author; and finally followed by references with more than two co-authors, in chronological order, independent of the second author surnames. In references with more than 10 authors only the first ten should be listed, followed by et al. Use standard abbreviations for journal titles as suggested byNCBI (<http://www.ncbi.nlm.nih.gov/journals/>).

Only articles that are published or in press should be included in this section. Works submitted for publication but not yet accepted, personal communications and unpublished data must be cited within the text. "Personal communication" refers to individuals other than the authors of the manuscript being submitted; "unpublished data" refers to data produced by at least one of the authors of the manuscript being submitted. Works of restricted circulation (e.g., theses not available in public databases, congress abstracts not published in regular journals or public databases) should not be listed in this section.

*Sample journal article citation:*

Breuer ME and Pavan C (1955) Behaviour of polytene chromosomes of *Rhynchosciara angela* at different stages of larval development. *Chromosoma* 7:371-386.

Yonenaga-Yassuda Y, Rodrigues MT and Pellegrino KCM (2005) Chromosomal banding patterns in the eyelid-less microteiid lizard radiation: The X1X1X2X2:X1X2Y sex chromosome system in *Calyptommatus* and the karyotypes of *Psilophtalmus* and *Tretioscincus* (Squamata, Gymnophthalmidae). *Genet Mol Biol* 28:700-709.

*Sample book citation:*

Dobzhansky T (1951) Genetics and Origin of Species. 3rd edition. Columbia University Press, New York, 364 pp.

*Sample chapter-in-book citation:*

Crawford DC and Howard-Peebles PN (2005) Fragile X: From cytogenetics to molecular genetics. In: Gersen SL and Keagle MB (eds) The Principles of Clinical Cytogenetics. 2nd edition. Humana Press, New Jersey, pp 495-513.

*Sample electronic article citation:*

Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD and Van Dyke T (2004) pRb inactivation in mammary cells reveals common mechanisms for tumor initiation and progression in divergent epithelia. Plos Biol 2:194-205. <http://www.plosbiology.org>.

f) **Internet Resources Section:** this section should contain a list of URLs referring to data presented in the text, software programs and other Internet resources used during data processing. Date of consultation must be stated.

*Sample Internet resource citation:*

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/OMIM> (September 4, 2005)  
LEM Software, [http://dir.niehs.nih.gov/dirbb/weinbergfiles/hybrid\\_design.htm](http://dir.niehs.nih.gov/dirbb/weinbergfiles/hybrid_design.htm) (September 4, 2005)

g) **Tables** must be in Word format prepared with table tool, inserted at the end of the main text file, each table starting on a new page. A concise title should be provided above the table. Tables must be numbered consecutively in Arabic numerals. Each column must have a title in the box head. Footnotes typed directly below the table should be indicated in lowercase superscript numbers.

h) **Figures** must be numbered consecutively in Arabic numerals. Images should be in TIFF or JPEG format and provided in separate files. Identify each illustration by the first author name and the number of the respective figure. Figures in Word, PowerPoint or Excel format cannot be published. Only sequence data can be presented in Word format. Journal quality reproduction will require grayscale resolution yielding 300 dpi, color figures should be at 600 dpi. These resolutions refer to the output size of the file; if it is anticipated that images will be enlarged or reduced, the resolutions should be adjusted accordingly. Figures composed of several elements should be sent as a single panel, obeying the print size definitions of the journal (single or two columns width). Scanned figures should not be submitted. Color illustrations can be accepted. Figure legends must be included in the main text file and should be typed on a new page that immediately follows the tables.

i) **Nomenclature** should adhere to current international standards.

j) **Sequences** may appear in text or in figure. DNA, RNA and protein sequences equal to or greater than 50 units must be entered into public databases. The accession number must be provided and released to the general public together with publication of the article.

k) **Data access:** reference should be made to availability of detailed data and materials used for reported studies.

l) **Ethical issues:** reports of experiments on live vertebrates must include a statement that the institutional review board approved the work and the protocol number must be provided. For experiments involving human subjects, authors must also include a statement that informed consent was obtained from all subjects. If photos or any other identifiable data are included, a copy of the signed consent must accompany the manuscript.

m) **Supplementary Material:** Data that the authors consider of importance for completeness of a study, but which are too extensive to be included in the published version, can be submitted as Supplementary Material. This material will be made available together with the electronic version. In case a manuscript contains such material, it should be appropriately identified within the text file. Supplementary material in tables should be identified as Table S1, Table S2, etc., in case of figures, they should be named accordingly, Figure S1, Figure S2. In addition, a list of this material should be presented at the end of the manuscript text file, containing the following statement: Supplementary material - the following online material is available for this article:

- Table S1 < short title >

- Figure S1 - < short title >

This material is available as part of the online article from <http://www.scielo.br/gmb>

### 3.2 Short Communications

Present brief observations that do not warrant full-length articles. They should not be considered preliminary communications. They should be 15 or fewer typed pages in double spaced 12-point type, including literature cited. They should include an Abstract no longer than five percent of the paper's length and no further subdivision with introduction, material and methods, results and discussion in a single section. Up to to four items (tables and/or figures) may be submitted. The title page and reference section format is that of full-length article.

### 3.3 Letters to the Editor

Relate or respond to recent published items in the journal. Discussions of political, social and ethical issues of interest to geneticists are also welcome in this form.

### **3.4 Review Articles**

Review Articles are welcome. The Editor should be contacted prior to submission.

### **3.5 Book Reviews**

Publishers are invited to submit books on Genetics, Evolution and related disciplines, for review in the journal. Aspiring reviewers may propose writing a review.

### **3.6 History, Story and Memories**

Accounts on historical aspects of Genetics relating to Brazil.

## **4. Proofs and Copyright Transfer**

Page proofs will be sent to the corresponding author. Changes made to page proofs, apart from printer's errors, will be charged to the authors. Notes added in proof require Editorial approval. A form of consent to publish and transfer of copyright will have to be signed by the corresponding author, also on behalf of any co-authors.

## **5. Reprints**

Reprints are free of charge and provided as a pdf-file.

## APÊNDICE 2: Normas para publicação – BMC Genetics

Conteúdo disponível em:

<http://www.biomedcentral.com/bmcgenet/ifora/>

### Instructions for *BMC Genetics* authors

#### General information

You are advised also to read [About this journal](#), which includes other relevant information.

#### Submission process

Manuscripts must be submitted by one of the authors of the manuscript, and should not be submitted by anyone on their behalf. The submitting author takes responsibility for the article during submission and peer review.

To facilitate rapid publication and to minimize administrative costs, *BMC Genetics* accepts only [online submission](#).

Files can be submitted as a batch, or one by one. The submission process can be interrupted at any time - when users return to the site, they can carry on where they left off.

See below for examples of acceptable word processor and graphics file formats. Additional files of any type, such as movies, animations, or original data files, can also be submitted as part of the publication.

During submission you will be asked to provide a cover letter. Please use this to explain why your manuscript should be published in the journal, to elaborate on any issues relating to our editorial policies detailed in the instructions for authors, and to declare any potential competing interests.

Assistance with the process of manuscript preparation and submission is available from the customer support team ([info@biomedcentral.com](mailto:info@biomedcentral.com)).

We also provide a collection of links to useful tools and resources for scientific authors, on our [Tools for Authors](#) page.

#### Publication and peer review processes

Submitted manuscripts will be sent to peer reviewers, unless they are either out of scope or below threshold for the journal, or the presentation or written English is of an unacceptably low standard. They will generally be reviewed by two experts with the aim of reaching a first decision as soon as possible.

Reviewers do not have to sign their reports but are welcome to do so. They are asked to declare any competing interests.

It is journal policy to publish work deemed by peer reviewers to be a coherent and sound addition to scientific knowledge and to put less emphasis on interest levels, provided that the research constitutes a useful contribution to the field. In addition to their comments for the authors, reviewers are asked whether the writing is of acceptable quality. Where possible, the final decision is made on the basis that the peer reviewers are in accordance with one another, or that at least there is no strong dissenting view. In cases where there is strong disagreement either among peer reviewers or between the authors and peer reviewers, advice is sought from a member of the journal's Editorial Board. The journal allows a maximum of two revisions of any manuscript. All appeals should be directed to the Series Editor, Biology who holds ultimate responsibility for editorial decisions

Reviewers are also asked to indicate which articles they consider to be especially interesting or significant. These articles may be given greater prominence and greater external publicity, and the authors may be asked if they would prefer to have the manuscript published in *BMC Biology*

Once an article is accepted, it is published in *BMC Genetics* immediately as a provisional PDF file. The paper will subsequently be published in both fully browseable web form, and as a formatted PDF. The article will then be available through *BMC Genetics*, BioMed Central and PubMed Central, and will also be included in PubMed.

Authors will be able to check the progress of their paper through the submission system at any time by logging into [My BioMed Central](#), their personalized section of the site.

### **Article-processing charges**

*BMC Genetics* levies an article-processing charge for every accepted article, to cover the costs incurred by open access publication. In 2011 the article-processing charge is £1175/US\$1900/€1345. Generally, if the submitting author's institution is a [BioMed Central member](#) the cost of the article processing charge is covered by the membership, and no further charge is payable. In the case of authors whose institutions are [supporter members](#) of BioMed Central, however, a discounted article processing charge is payable by the author. Please [click here](#) to check if your institution is a BioMed Central member. We routinely waive charges for authors from low-income countries. For further details, see [more information about article-processing charges](#).

### **Editorial policies**

Any manuscripts, or substantial parts of it, submitted to the journal must not be under consideration by any other journal. In general, the manuscript should not

have already been published in any journal or other citable form, although it may have been deposited on a preprint server. Information on duplicate/overlapping publications can be found [here](#). Authors are required to ensure that no material submitted as part of a manuscript infringes existing copyrights, or the rights of a third party. Authors who publish in *BMC Genetics* retain copyright to their work ([more information](#)). Correspondence concerning articles published in *BMC Genetics* is encouraged through the [online comment system](#).

Submission of a manuscript to *BMC Genetics* implies that all authors have read and agreed to its content, and that any experimental research that is reported in the manuscript has been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the [Helsinki Declaration](#), and any experimental research on animals must follow internationally recognized guidelines. A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate. Informed consent must also be documented. Manuscripts may be rejected if the editorial office considers that the research has not been carried out within an ethical framework, e.g. if the severity of the experimental procedure is not justified by the value of the knowledge gained.

*BMC Genetics's* publisher, BioMed Central, has a legal responsibility to ensure that its journals do not publish material that infringes copyright, or that includes libellous or defamatory content. If, on review, your manuscript is perceived to contain potentially libellous content the journal Editors, with assistance from the publisher if required, will work with authors to ensure an appropriate outcome is reached.

Generic drug names should generally be used. When proprietary brands are used in research, include the brand names in parentheses in the Methods section.

*BMC Genetics* requires authors to declare any competing financial or other interest in relation to their work. If any author has a competing interest, it should be declared in the covering letter.

Any 'in press' articles cited within the references and necessary for the reviewers' assessment of the manuscript should be made available if requested by the editorial office.

Submission of a manuscript to *BMC Genetics* implies that readily reproducible materials described in the manuscript, including all relevant raw data, will be freely available to any scientist wishing to use them for non-commercial purposes. Nucleic acid sequences, protein sequences, and atomic coordinates should be deposited in an appropriate database in time for the accession number to be included in the published article. In computational studies where the sequence information is unacceptable for inclusion in databases because of

lack of experimental validation, the sequences must be published as an additional file with the article.

*BMC Genetics* also supports initiatives aimed at improving the reporting of biomedical research. We recommend authors refer to the [MIBBI Portal](#) for prescriptive checklists for reporting biological and biomedical research where applicable.

### **Nucleotide sequences**

Nucleotide sequences can be deposited with the [DNA Data Bank of Japan \(DDBJ\)](#), [European Molecular Biology Laboratory \(EMBL/EBI\) Nucleotide Sequence Database](#), or [GenBank](#) (National Center for Biotechnology Information).

### **Protein sequences**

Protein sequences can be deposited with [SwissProt](#) or the [Protein Information Resource \(PIR\)](#).

### **Structures**

Protein structures can be deposited with one of the members of the [Worldwide Protein Data Bank](#). Nucleic Acids structures can be deposited with the [Nucleic Acid Database](#) at Rutgers. Crystal structures of organic compounds can be deposited with the [Cambridge Crystallographic Data Centre](#).

### **Chemical structures and assays**

Structures of chemical substances can be deposited with [PubChem Substance](#). Bioactivity screens of chemical substances can be deposited with [PubChem BioAssay](#).

### **Microarray data**

Where appropriate, authors should adhere to the standards proposed by the [Microarray Gene Expression Data Society](#) and must deposit microarray data in one of the public repositories, such as [ArrayExpress](#), [Gene Expression Omnibus \(GEO\)](#) or the [Center for Information Biology Gene Expression Database \(CIBEX\)](#).

### **Computational modeling**

We encourage authors to prepare models of biochemical reaction networks using the [Systems Biology Markup Language](#) and to deposit the model with the [BioModels database](#), as well as submitting it as an additional file with the manuscript.



## Plasmids

We encourage authors to deposit copies of their plasmids as DNA or bacterial stocks with [Addgene](#), a non-profit repository, or [PlasmID](#), the Plasmid Information Database at Harvard.

BioMed Central is a member of the Committee on Publication Ethics (COPE). Authors who have appealed against a rejection but remain concerned about the editorial process can refer their case to COPE. For more information, visit [www.publicationethics.org](http://www.publicationethics.org).

BioMed Central endorses the World Association of Medical Editors (WAME) [Policy Statement on Geopolitical Intrusion on Editorial Decisions](#)

## Preparing main manuscript text

### File formats

The following word processor file formats are acceptable for the main manuscript document:

- Microsoft Word (version 2 and above)
- Rich text format (RTF)
- Portable document format (PDF)
- TeX/LaTeX (use [BioMed Central's TeX template](#))
- DeVice Independent format (DVI)
- Publicon Document (NB)

Users of other word processing packages should save or convert their files to RTF before uploading. Many free tools are available which ease this process.

TeX/LaTeX users: We recommend using [BioMed Central's TeX template and BibTeX stylefile](#). If you use this standard format, you can submit your manuscript in TeX format (after you submit your TEX file, you will be prompted to submit your BBL file). If you have used another template for your manuscript, or if you do not wish to use BibTeX, then please submit your manuscript as a DVI file. We do not recommend converting to RTF.

Note that [figures](#) must be submitted as separate image files, not as part of the submitted DOC/ PDF/TEX/DVI file.

### Article types

When submitting your manuscript, you will be asked to assign one of the following types to your article:

Research article

Database

Methodology article

Software

Please read the descriptions of each of the article types, choose which is appropriate for your article and structure it accordingly. If in doubt, your manuscript should be classified as a Research article, the structure for which is described below.

### **Manuscript sections for Research articles**

Manuscripts for Research articles submitted to *BMC Genetics* should be divided into the following sections:

- Title page
- Abstract
- Background
- Results
- Discussion
- Conclusions
- Methods (can also be placed after Background)
- List of abbreviations used(if any)
- Authors' contributions
- Authors' information (if any)
- Acknowledgements and Funding
- References
- Figure legends (if any)
- Tables and captions (if any)
- Description of additional data files (if any)

You can download a template (Mac and Windows compatible; Microsoft Word 98/2000) for your article. For instructions on use, see below.

The **Accession Numbers** of any nucleic acid sequences, protein sequences or atomic coordinates cited in the manuscript should be provided, in square brackets and include the corresponding database name; for example, [EMBL:AB026295, EMBL:AC137000, DDBJ:AE000812, GenBank:U49845, PDB:1BFM, Swiss-Prot:Q96KQ7, PIR:S66116].

The databases for which we can provide direct links are: EMBL Nucleotide Sequence Database (EMBL), DNA Data Bank of Japan (DDBJ), GenBank at the NCBI (GenBank), Protein Data Bank (PDB), Protein Information Resource (PIR) and the Swiss-Prot Protein Database (Swiss-Prot).

## **Title page**

This should list: the title of the article, which should include an accurate, clear and concise description of the reported work, avoiding abbreviations; and the full names, institutional addresses, and e-mail addresses for all authors. The corresponding author should also be indicated.

## **Abstract**

The abstract of the manuscript should not exceed 350 words and must be structured into separate sections: **Background**, the context and purpose of the study; **Results**, the main findings; **Conclusions**, brief summary and potential implications. Please minimize the use of abbreviations and do not cite references in the abstract.

## **Background**

The background section should be written from the standpoint of researchers without specialist knowledge in that area and must clearly state - and, if helpful, illustrate - the background to the research and its aims. The section should end with a very brief statement of what is being reported in the article.

## **Results and Discussion**

The Results and Discussion may be combined into a single section or presented separately. They may also be broken into subsections with short, informative headings.

## **Conclusions**

This should state clearly the main conclusions of the research and give a clear explanation of their importance and relevance. Summary illustrations may be included.

## **Methods**

This should be divided into subsections if several methods are described.

## **List of abbreviations**

If abbreviations are used in the text, either they should be defined in the text where first used, or a list of abbreviations can be provided, which should precede the authors' contributions and acknowledgements.

## **Authors' contributions**

In order to give appropriate credit to each author of a paper, the individual contributions of authors to the manuscript should be specified in this section.

An "author" is generally considered to be someone who has made substantive intellectual contributions to a published study. To qualify as an author one should 1) have made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) have been involved in drafting the manuscript or revising it critically for important intellectual content; and 3) have given final approval of the version to be published. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content. Acquisition of funding, collection of data, or general supervision of the research group, alone, does not justify authorship.

We suggest the following kind of format (please use initials to refer to each author's contribution): AB carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. JY carried out the immunoassays. MT participated in the sequence alignment. ES participated in the design of the study and performed the statistical analysis. FG conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

All contributors who do not meet the criteria for authorship should be listed in an acknowledgements section. Examples of those who might be acknowledged include a person who provided purely technical help, writing assistance, or a department chair who provided only general support.

### **Authors' information**

You may choose to use this section to include any relevant information about the author(s) that may aid the reader's interpretation of the article, and understand the standpoint of the author(s). This may include details about the authors' qualifications, current positions they hold at institutions or societies, or any other relevant background information. Please refer to authors using their initials. Note this section should not be used to describe any competing interests.

### **Acknowledgements and Funding**

Please acknowledge anyone who contributed towards the study by making substantial contributions to conception, design, acquisition of data, or analysis and interpretation of data, or who was involved in drafting the manuscript or revising it critically for important intellectual content, but who does not meet the criteria for authorship. Please also include their source(s) of funding. Please also acknowledge anyone who contributed materials essential for the study.

Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements.

Please list the source(s) of funding for the study, for each author, and for the manuscript preparation in the acknowledgements section. Authors must

describe the role of the funding body, if any, in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

## References

All references must be numbered consecutively, in square brackets, in the order in which they are cited in the text, followed by any in tables or legends. Reference citations should not appear in titles or headings. Each reference must have an individual reference number. Please avoid excessive referencing. If automatic numbering systems are used, the reference numbers must be finalized and the bibliography must be fully formatted before submission.

Only articles and abstracts that have been published or are in press, or are available through public e-print/preprint servers, may be cited; unpublished abstracts, unpublished data and personal communications should not be included in the reference list, but may be included in the text and referred to as "unpublished data", "unpublished observations", or "personal communications" giving the names of the involved researchers. Notes/footnotes are not allowed. Obtaining permission to quote personal communications and unpublished data from the cited author(s) is the responsibility of the author. Journal abbreviations follow Index Medicus/MEDLINE. Citations in the reference list should contain all named authors, regardless of how many there are.

Examples of the *BMC Genetics* reference style are shown below. Please take care to follow the reference style precisely; references not in the correct style may be retyped, necessitating tedious proofreading.

## Links

Web links and URLs should be included in the reference list. They should be provided in full, including both the title of the site and the URL, in the following format: **The Mouse Tumor Biology Database** [http://tumor.informatics.jax.org/mtbwi/index.do]. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

## ***BMC Genetics* reference style**

Style files are available for use with popular bibliographic management software:

- [BibTeX](#)
- [EndNote style file](#)
- [Reference Manager](#)
- [Zotero](#)

*Article within a journal*

1. Koonin EV, Altschul SF, Bork P: **BRCA1 protein products: functional motifs.** *Nat Genet* 1996, **13**:266-267.

*Article within a journal supplement*

2. Orengo CA, Bray JE, Hubbard T, LoConte L, Sillitoe I: **Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction.** *Proteins* 1999, **43**(Suppl 3):149-170.

*In press article*

3. Kharitonov SA, Barnes PJ: **Clinical aspects of exhaled nitric oxide.** *Eur Respir J*, in press.

*Published abstract*

4. Zvaifler NJ, Burger JA, Marinova-Mutafchieva L, Taylor P, Maini RN: **Mesenchymal cells, stromal derived factor-1 and rheumatoid arthritis [abstract].** *Arthritis Rheum* 1999, **42**:s250.

*Article within conference proceedings*

5. Jones X: **Zeolites and synthetic mechanisms.** In *Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore*. Edited by Smith Y. Stoneham: Butterworth-Heinemann; 1996:16-27.

*Book chapter, or article within a book*

6. Schnepf E: **From prey via endosymbiont to plastids: comparative studies in dinoflagellates.** In *Origins of Plastids. Volume 2*. 2nd edition. Edited by Lewin RA. New York: Chapman and Hall; 1993:53-76.

*Whole issue of journal*

7. Ponder B, Johnston S, Chodosh L (Eds): **Innovative oncology.** In *Breast Cancer Res* 1998, **10**:1-72.

*Whole conference proceedings*

8. Smith Y (Ed): *Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore*. Stoneham: Butterworth-Heinemann; 1996.

*Complete book*

9. Margulis L: *Origin of Eukaryotic Cells*. New Haven: Yale University Press; 1970.

*Monograph or book in a series*

10. Hunninghake GW, Gadek JE: **The alveolar macrophage**. In *Cultured Human Cells and Tissues*. Edited by Harris TJR. New York: Academic Press; 1995:54-56. [Stoner G (Series Editor): *Methods and Perspectives in Cell Biology*, vol 1.]

*Book with institutional author*

11. Advisory Committee on Genetic Modification: *Annual Report*. London; 1999.

*PhD thesis*

12. Kohavi R: **Wrappers for performance enhancement and oblivious decision graphs**. *PhD thesis*. Stanford University, Computer Science Department; 1995.

*Link / URL*

13. **The Mouse Tumor Biology Database**

[<http://tumor.informatics.jax.org/mtbwi/index.do>]

*Link / URL with author(s)*

14. Neylon, C: **Open Research Computation: an ordinary journal with extraordinary aims**.

[[http://blogs.openaccesscentral.com/blogs/bmcblog/entry/open\\_research\\_computation\\_an\\_ordinary](http://blogs.openaccesscentral.com/blogs/bmcblog/entry/open_research_computation_an_ordinary)]

### **Microsoft Word template**

Although we can accept manuscripts prepared as Microsoft Word, RTF or PDF files, we have designed a Microsoft Word template that can be used to generate a standard style and format for your article. It can be used if you have not yet started to write your paper, or if it is already written and needs to be put into *BMC Genetics* style.

[Download the template](#) (compatible with Mac and Windows Word 97/98/2000/2003/2007) from our site, and save it to your hard drive. Double click the template to open it.

### **How to use the *BMC Genetics* template**

The template consists of a standard set of headings that make up a *BMC Genetics* Research article manuscript, along with dummy fragments of body text. Follow these steps to create your manuscript in the standard format:

- Replace the dummy text for Title, Author details, Institutional affiliations, and

the other sections of the manuscript with your own text (either by entering the text directly or by cutting and pasting from your own manuscript document).

- If there are sections which you do not need, delete them (but check the rest of the Instructions for Authors to see which sections are compulsory).
- If you need an additional copy of a heading (e.g. for additional figure legends) just copy and paste.
- For the references, you may either manually enter the references using the reference style given, or use bibliographic software to insert them automatically. We provide style files for EndNote, Reference Manager and Zotero.

For extra convenience, you can use the template as one of your standard Word templates. To do this, put a copy of the template file in Word's 'Templates' folder, normally C:\Program Files\Microsoft Office\Templates on a PC. The next time you create a new document in Word using the File menu, the template will appear as one of the available choices for a new document.

### **Preparing illustrations and figures**

Figures should be provided as separate files. Each figure should comprise only a single file. There is no charge for the use of color.

Please read our figure preparation guidelines for detailed instructions on maximising the quality of your figures,

### **Formats**

The following file formats can be accepted:

- **EPS** (preferred format for diagrams)
- **PDF** (also especially suitable for diagrams)
- **PNG** (preferred format for photos or images)
- Microsoft Word (figures must be a single page)
- PowerPoint (figures must be a single page)
- TIFF
- JPEG
- BMP
- CDX (ChemDraw)
- TGF (ISIS/Draw)

### **Figure legends**

The legends should be included in the main manuscript text file rather than being a part of the figure file. For each figure, the following information should be provided: Figure number (in sequence, using Arabic numerals - i.e. Figure 1, 2, 3 etc); short title of figure (maximum 15 words); detailed legend, up to 300 words.

**Please note that it is the responsibility of the author(s) to obtain permission from the copyright holder to reproduce figures or tables that**



have previously been published elsewhere.

### **Preparing a personal cover page**

If you wish to do so, you may submit an image which, in the event of publication, will be used to create a cover page for the PDF version of your article. The cover page will also display the journal logo, article title and citation details. The image may either be a figure from your manuscript or another relevant image. You must have permission from the copyright holder to reproduce the image. Images that do not meet our requirements will not be used.

Images must be 300dpi and 155mm square (1831 x 1831 pixels for a raster image).

Allowable formats - EPS, PDF (for line drawings), PNG, TIFF (for photographs and screen dumps), JPEG, BMP, DOC, PPT, CDX, TGF (ISIS/Draw).

### **Preparing tables**

Each table should be numbered in sequence using Arabic numerals (i.e. Table 1, 2, 3 etc.). Tables should also have a title that summarizes the whole table, maximum 15 words. Detailed legends may then follow, but should be concise.

Smaller tables considered to be integral to the manuscript can be pasted into the document text file. These will be typeset and displayed in the final published form of the article. Such tables should be formatted using the 'Table object' in a word processing program to ensure that columns of data are kept aligned when the file is sent electronically for review; this will not always be the case if columns are generated by simply using tabs to separate text. Commas should not be used to indicate numerical values. Color and shading should not be used.

Larger datasets can be uploaded separately as additional files. Additional files will not be displayed in the final, published form of the article, but a link will be provided to the files as supplied by the author.

Tabular data provided as additional files can be uploaded as an Excel spreadsheet (.xls) or comma separated values (.csv). As with all files, please use the standard file extensions.

### **Preparing additional files**

Although *BMC Genetics* does not restrict the length and quantity of data in a paper, there may still be occasions where an author wishes to provide data sets, tables, movie files, or other information as additional information. These files can be uploaded using the 'Additional Material files' button in the manuscript submission process.

The maximum file size for additional files is 20 MB each, and files will be virus-scanned on submission.

Any additional files will be linked into the final published article in the form supplied by the author, but will not be displayed within the paper. They will be made available in exactly the same form as originally provided.

If additional material is provided, please list the following information in a separate section of the manuscript text, at the end of the document text file:

- File name
- File format (including name and a URL of an appropriate viewer if format is unusual)
- Title of data
- Description of data

Additional datafiles should be referenced explicitly by file name within the body of the article, e.g. 'See additional file 1: Movie1 for the original data used to perform this analysis'.

### **Formats and uploading**

Ideally, file formats for additional files should not be platform-specific, and should be viewable using free or widely available tools. The following are examples of suitable formats.

- Additional documentation
  - PDF (Adobe Acrobat)
- Animations
  - SWF (Shockwave Flash)
- Movies
  - MOV (QuickTime)
  - MPG (MPEG)
- Tabular data
  - XLS (Excel spreadsheet)
  - CSV (Comma separated values)

As with figure files, files should be given the standard file extensions. This is especially important for Macintosh users, since the Mac OS does not enforce the use of standard extensions. Please also make sure that each additional file is a single table, figure or movie (please do not upload linked worksheets or PDF files larger than one sheet).

### **Mini-websites**

Small self-contained websites can be submitted as additional files, in such a way that they will be browsable from within the full text HTML version of the article. In order to do this, please follow these instructions:

1. Create a folder containing a starting file called index.html (or index.htm) in the root
2. Put all files necessary for viewing the mini-website within the folder, or sub-folders
3. Ensure that all links are relative (ie "images/picture.jpg" rather than "/images/picture.jpg" or "http://yourdomain.net/images/picture.jpg" or "C:\Documents and Settings\username\My Documents\mini-website\images\picture.jpg") and no link is longer than 255 characters
4. Access the index.html file and browse around the mini-website, to ensure that the most commonly used browsers (Internet Explorer and Firefox) are able to view all parts of the mini-website without problems, it is ideal to check this on a different machine

Compress the folder into a ZIP, check the file size is under 20 MB, ensure that index.html is in the root of the ZIP, and that the file has .zip extension, then submit as an additional file with your article

## Style and language

### General

Currently, *BMC Genetics* can only accept manuscripts written in English. Spelling should be US English or British English, but not a mixture.

Gene names should be in italic, but protein products should be in plain type.

There is no explicit limit on the length of articles submitted, but authors are encouraged to be concise. There is no restriction on the number of figures, tables or additional files that can be included with each article online. Figures and tables should be sequentially referenced. Authors should include all relevant supporting data with each article.

*BMC Genetics* will not edit submitted manuscripts for style or language; reviewers may advise rejection of a manuscript if it is compromised by grammatical errors. Authors are advised to write clearly and simply, and to have their article checked by colleagues before submission. In-house copyediting will be minimal. Non-native speakers of English may choose to make use of a copyediting service.

### Help and advice on scientific writing

The abstract is one of the most important parts of a manuscript. For guidance, please visit our page on ["Writing titles and abstracts for scientific articles"](#)

Tim Albert has produced for BioMed Central a [list of tips](#) for writing a scientific manuscript. [MedBioWorld](#) also provides a list of resources for science writing.

### Abbreviations

Abbreviations should be used as sparingly as possible. They can be defined when first used or a list of abbreviations can be provided preceding the acknowledgements and references.

### Typography

- Please use double line spacing.
- Type the text unjustified, without hyphenating words at line breaks.
- Use hard returns only to end headings and paragraphs, not to rearrange lines.
- Capitalize only the first word, and proper nouns, in the title.
- All pages should be numbered.
- Use the *BMC Genetics* reference format.
- Footnotes to text should not be used.
- Greek and other special characters may be included. If you are unable to reproduce a particular special character, please type out the name of the symbol in full.

**Please ensure that all special characters used are embedded in the text, otherwise they will be lost during conversion to PDF.**

- Genes, mutations, genotypes, and alleles should be indicated in italics, and authors are required to use approved gene symbols, names, and formatting. Protein products should be in plain type.

### Units

SI Units should be used throughout (liter and molar are permitted, however).

Last revised: 9 August 2010

### APÊNDICE 3 – Resultados originais

**Apêndice 3a** – Resultados para a População 1 (40k) com 480 animais no grupo de treinamento e 8000 animais no grupo de predição para  $h^2$  0,10, 0,25 e 0,40 utilizando o VGE como resposta.

rep	0,10			0,25			0,40		
	y_conf	AC	conf	y_conf	ac	conf	y_conf	ac	conf
1	0,63	0,41	0,16	0,81	0,36	0,13	0,88	0,37	0,13
2	0,63	0,32	0,10	0,81	0,33	0,11	0,88	0,38	0,15
3	0,63	0,33	0,11	0,81	0,36	0,13	0,88	0,38	0,14
4	0,63	0,36	0,13	0,81	0,38	0,14	0,88	0,44	0,19
5	0,63	0,40	0,16	0,81	0,38	0,14	0,88	0,38	0,15
6	0,63	0,38	0,14	0,81	0,35	0,12	0,88	0,39	0,15
7	0,63	0,34	0,12	0,81	0,37	0,14	0,88	0,37	0,14
8	0,63	0,39	0,16	0,81	0,42	0,18	0,88	0,40	0,16
9	0,63	0,38	0,14	0,81	0,38	0,14	0,88	0,36	0,13
10	0,63	0,36	0,13	0,81	0,39	0,15	0,88	0,40	0,16
Média		<b>0,37</b>	<b>0,14</b>		<b>0,37</b>	<b>0,14</b>		<b>0,39</b>	<b>0,15</b>
DP		<b>0,030</b>	<b>0,021</b>		<b>0,024</b>	<b>0,019</b>		<b>0,023</b>	<b>0,018</b>
EP		<b>0,010</b>	<b>0,007</b>		<b>0,008</b>	<b>0,006</b>		<b>0,007</b>	<b>0,006</b>

rep=replicação; y\_conf: confiabilidade média da resposta; ac: acurácia do VGG; conf: confiabilidade do VGG; DP: desvio padrão; EP: erro padrão.

**Apêndice 3b** – Resultados para a População 1 (40k) com 960 animais no grupo de treinamento e 8000 animais no grupo de predição para  $h^2$  0,10, 0,25 e 0,40 utilizando o VGE como resposta.

rep	0,10			0,25			0,40		
	y_conf	ac	conf	y_conf	ac	conf	y_conf	ac	conf
1	0,63	0,49	0,24	0,81	0,50	0,25	0,88	0,51	0,26
2	0,63	0,42	0,17	0,81	0,46	0,21	0,88	0,52	0,27
3	0,63	0,42	0,18	0,81	0,48	0,23	0,88	0,53	0,28
4	0,63	0,44	0,19	0,81	0,49	0,24	0,88	0,50	0,25
5	0,63	0,49	0,24	0,81	0,48	0,23	0,88	0,50	0,25
6	0,63	0,45	0,21	0,81	0,48	0,23	0,88	0,48	0,23
7	0,63	0,46	0,21	0,81	0,49	0,24	0,88	0,51	0,26
8	0,63	0,47	0,22	0,81	0,51	0,26	0,88	0,51	0,26
9	0,63	0,43	0,18	0,81	0,48	0,23	0,88	0,49	0,24
10	0,63	0,47	0,22	0,81	0,49	0,24	0,88	0,53	0,28
Média		<b>0,45</b>	<b>0,21</b>		<b>0,49</b>	<b>0,24</b>		<b>0,51</b>	<b>0,26</b>
DP		<b>0,026</b>	<b>0,025</b>		<b>0,013</b>	<b>0,013</b>		<b>0,016</b>	<b>0,016</b>
EP		<b>0,008</b>	<b>0,008</b>		<b>0,004</b>	<b>0,004</b>		<b>0,005</b>	<b>0,005</b>

rep=replicação; y\_conf: confiabilidade média da resposta; ac: acurácia do VGG; conf: confiabilidade do VGG; DP: desvio padrão; EP: erro padrão.

**Apêndice 3c** – Resultados para a População 1 (40k) com 1920 animais no grupo de treinamento e 8000 animais no grupo de predição para  $h^2$  0,10, 0,25 e 0,40 utilizando o VGE como resposta.

rep	0,10			0,25			0,40		
	y_conf	ac	conf	y_conf	ac	conf	y_conf	ac	conf
1	0,63	0,58	0,34	0,81	0,62	0,38	0,88	0,63	0,39
2	0,63	0,55	0,30	0,81	0,57	0,33	0,88	0,62	0,39
3	0,63	0,53	0,28	0,81	0,60	0,36	0,88	0,65	0,42
4	0,63	0,54	0,29	0,81	0,59	0,34	0,88	0,63	0,40
5	0,63	0,58	0,34	0,81	0,59	0,34	0,88	0,63	0,40
6	0,63	0,54	0,29	0,81	0,58	0,34	0,88	0,63	0,39
7	0,63	0,55	0,31	0,81	0,61	0,37	0,88	0,63	0,40
8	0,63	0,57	0,32	0,81	0,62	0,39	0,88	0,64	0,41
9	0,63	0,55	0,30	0,81	0,61	0,37	0,88	0,62	0,39
10	0,63	0,57	0,33	0,81	0,62	0,38	0,88	0,67	0,44
Média		<b>0,56</b>	<b>0,31</b>		<b>0,60</b>	<b>0,36</b>		<b>0,64</b>	<b>0,40</b>
DP		<b>0,018</b>	<b>0,022</b>		<b>0,018</b>	<b>0,021</b>		<b>0,015</b>	<b>0,016</b>
EP		<b>0,006</b>	<b>0,007</b>		<b>0,006</b>	<b>0,007</b>		<b>0,005</b>	<b>0,005</b>

rep=replicação; y\_conf: confiabilidade média da resposta; ac: acurácia do VGG; conf: confiabilidade do VGG; DP: desvio padrão; EP: erro padrão.

**Apêndice 3d** – Resultados para a População 1 (40k) com 7680 animais no grupo de treinamento e 8000 animais no grupo de predição para  $h^2$  0,10, 0,25 e 0,40 utilizando o fenótipo como resposta.

rep	0,10			0,25			0,40		
	y_conf	ac	conf	y_conf	ac	Conf	y_conf	ac	conf
1	0,10	0,40	0,16	0,25	0,60	0,36	0,40	0,64	0,41
2	0,10	0,45	0,20	0,25	0,54	0,29	0,40	0,63	0,40
3	0,10	0,43	0,19	0,25	0,54	0,29	0,40	0,66	0,43
4	0,10	0,44	0,19	0,25	0,57	0,32	0,40	0,65	0,43
5	0,10	0,46	0,21	0,25	0,53	0,28	0,40	0,66	0,44
6	0,10	0,40	0,16	0,25	0,53	0,28	0,40	0,64	0,41
7	0,10	0,45	0,20	0,25	0,55	0,30	0,40	0,65	0,42
8	0,10	0,45	0,20	0,25	0,58	0,33	0,40	0,66	0,43
9	0,10	0,47	0,22	0,25	0,57	0,32	0,40	0,64	0,41
10	0,10	0,40	0,16	0,25	0,58	0,33	0,40	0,66	0,44
Média		<b>0,44</b>	<b>0,19</b>		<b>0,56</b>	<b>0,31</b>		<b>0,65</b>	<b>0,42</b>
DP		<b>0,026</b>	<b>0,022</b>		<b>0,024</b>	<b>0,026</b>		<b>0,011</b>	<b>0,014</b>
EP		<b>0,008</b>	<b>0,007</b>		<b>0,008</b>	<b>0,008</b>		<b>0,003</b>	<b>0,004</b>

rep=replicação; y\_conf: confiabilidade média da resposta; ac: acurácia do VGG; conf: confiabilidade do VGG; DP: desvio padrão; EP: erro padrão.

**Apêndice 3e** – Resultados para a População 2 (800k) com 480 animais no grupo de treinamento e 2000 animais no grupo de predição para  $h^2$  0,10, 0,25 e 0,40 utilizando o VGE como resposta.

rep	0,10			0,25			0,40		
	y_conf	ac	conf	y_conf	ac	conf	y_conf	ac	conf
1	0,63	0,42	0,18	0,81	0,46	0,22	0,88	0,55	0,30
2	0,63	0,45	0,20	0,81	0,41	0,16	0,88	0,42	0,18
3	0,63	0,47	0,22	0,81	0,47	0,22	0,88	0,44	0,19
4	0,63	0,45	0,20	0,81	0,48	0,23	0,88	0,48	0,23
5	0,63	0,35	0,12	0,81	0,50	0,25	0,88	0,51	0,26
6	0,63	0,49	0,24	0,81	0,41	0,17	0,88	0,48	0,23
7	0,63	0,47	0,23	0,81	0,41	0,17	0,88	0,50	0,25
8	0,63	0,37	0,14	0,81	0,48	0,23	0,88	0,48	0,23
9	0,63	0,41	0,17	0,81	0,50	0,25	0,88	0,47	0,22
10	0,63	0,41	0,17	0,81	0,43	0,19	0,88	0,44	0,19
Média		<b>0,43</b>	<b>0,19</b>		<b>0,46</b>	<b>0,21</b>		<b>0,48</b>	<b>0,23</b>
DP		<b>0,045</b>	<b>0,039</b>		<b>0,037</b>	<b>0,034</b>		<b>0,038</b>	<b>0,036</b>
EP		<b>0,014</b>	<b>0,012</b>		<b>0,012</b>	<b>0,011</b>		<b>0,012</b>	<b>0,012</b>

rep=replicação; y\_conf: confiabilidade média da resposta; ac: acurácia do VGG; conf: confiabilidade do VGG; DP: desvio padrão; EP: erro padrão.

**Apêndice 3f** – Resultados para a População 2 (800k) com 1920 animais no grupo de treinamento e 2000 animais no grupo de predição para  $h^2$  0,10, 0,25 e 0,40 utilizando o fenótipo como resposta.

rep	0,10			0,25			0,40		
	y_conf	ac	conf	y_conf	ac	conf	y_conf	ac	conf
1	0,10	0,34	0,12	0,25	0,43	0,18	0,40	0,50	0,25
2	0,10	0,34	0,11	0,25	0,38	0,14	0,40	0,48	0,23
3	0,10	0,30	0,09	0,25	0,39	0,15	0,40	0,52	0,27
4	0,10	0,35	0,12	0,25	0,41	0,17	0,40	0,53	0,28
5	0,10	0,22	0,05	0,25	0,43	0,19	0,40	0,49	0,24
6	0,10	0,37	0,14	0,25	0,34	0,12	0,40	0,48	0,23
7	0,10	0,29	0,09	0,25	0,40	0,16	0,40	0,51	0,26
8	0,10	0,25	0,06	0,25	0,41	0,17	0,40	0,51	0,26
9	0,10	0,27	0,07	0,25	0,46	0,21	0,40	0,55	0,30
10	0,10	0,31	0,10	0,25	0,41	0,17	0,40	0,47	0,22
Média		<b>0,30</b>	<b>0,10</b>		<b>0,41</b>	<b>0,17</b>		<b>0,50</b>	<b>0,25</b>
DP		<b>0,048</b>	<b>0,029</b>		<b>0,032</b>	<b>0,025</b>		<b>0,025</b>	<b>0,025</b>
EP		<b>0,015</b>	<b>0,009</b>		<b>0,010</b>	<b>0,008</b>		<b>0,008</b>	<b>0,008</b>

rep=replicação; y\_conf: confiabilidade média da resposta; ac: acurácia do VGG; conf: confiabilidade do VGG; DP: desvio padrão; EP: erro padrão.

## 6. VITA

Fernanda Varnieri Brito, filha de Luiz Fernando Brito e Vera Regina Varnieri Brito, nasceu aos 16 dias do mês de abril de 1964, em Porto Alegre, Rio Grande do Sul, Brasil.

Freqüentou da 1ª à 5ª série do ensino primário na escola Nossa Senhora do Bom Conselho, em Porto Alegre e da 6ª a 8ª série do ensino primário e o ensino médio no Colégio de Aplicação da UFRGS, em Porto Alegre. Ingressou no curso de Agronomia da Universidade Federal do Rio Grande do Sul no ano de 1983. Cumpriu estágio curricular na estação experimental de São Gabriel do Instituto de Pesquisas Zootécnicas do RS formando-se Engenheira Agrônoma em janeiro de 1988. Ingressou no curso de mestrado em zootecnia da Faculdade de Agronomia da UFRGS, área de produção animal, em março de 1989, sob a orientação do Prof. Luiz Alberto Fries, obtendo o grau de Mestre em Zootecnia em março de 1992. Fundou a empresa GenSys Consultores Associados S/S Ltda em 1991, onde exerce atividades de consultoria e assessoria na área de melhoramento genético animal, estando afastada desde o início do curso de doutorado em zootecnia cujo ingresso se deu em março de 2007 sob a orientação do Prof. José Braccini Neto.