

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**ESTUDO GENÔMICO DA PRODUÇÃO DE LEITE E SEUS
CONSTITUINTES EM BOVINOS DA RAÇA GUZERÁ**

Daniel Jordan de Abreu Santos
Médico Veterinário

2015

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL

**ESTUDO GENÔMICO DA PRODUÇÃO DE LEITE E SEUS
CONSTITUINTES EM BOVINOS DA RAÇA GUZERÁ**

Daniel Jordan de Abreu Santos

Orientador: **Prof. Dr. Humberto Tonhati**

Co-orientadores: **Dra. Maria Gabriela Campolina Diniz Peixoto**
Dr. Marcos Vinícius Gualberto Barbosa da Silva

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

Santos, Daniel Jordan de Abreu
S237e Estudo genômico da produção de leite e seus constituintes em
 bovinos da raça Guzerá. / Daniel Jordan de Abreu Santos. – –
 Jaboticabal, 2015
 viii, 129 p. ; 28 cm

 Tese (doutorado) - Universidade Estadual Paulista, Faculdade de
 Ciências Agrárias e Veterinárias, 2015
 Orientador: Humberto Tonhati
 Banca examinadora: Humberto Tonhati, Guilherme Venturini,
 Henrique Nunes, Rusbel Raul Aspilcueta Borquis e Francisco Ribeiro
 Araújo Neto

 Bibliografia

 1. Desequilíbrio de ligação. 2. "GWAS". 3. Modelos Bayesianos.
 4. Seleção genômica. 5. Zebu. I. Título. II. Jaboticabal-Faculdade de
 Ciências Agrárias e Veterinárias.

 CDU 637.1:636.2

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

DADOS CURRICULARES DO AUTOR

DANIEL JORDAN DE ABREU SANTOS – filho de Francisco Antônio Ismério dos Santos e de Liliane Jordan de Abreu Santos, nasceu em 20 de janeiro de 1984, na cidade de Cantagalo-RJ. Em 18 de dezembro de 2008 obteve o título de Médico Veterinário pelo Centro Universitário Serra dos Órgãos – Teresópolis-RJ. Em março de 2009 ingressou no Programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias – UNESP – Campus de Jaboticabal-SP, obtendo o grau de Mestre em 29 de Abril de 2011 sob orientação do Prof. Dr. Humberto Tonhati. Em agosto do mesmo ano iniciou o curso de doutorado no Programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias – UNESP – Campus de Jaboticabal-SP sob a orientação do Prof. Dr. Humberto Tonhati. Este também realizou o estágio de Doutorado no exterior, sendo recebido na “Universität für Bodenkultur” – Viena, Áustria, de Janeiro de 2014 à 2015, sob orientação do professor Dr. Johann Sölkner.

AGRADECIMENTOS

A Deus por tudo.

Aos meus pais por todo suporte e apóio.

À minha irmã pelas orações e aos meus outros familiares pela torcida.

Ao professor Dr. Humberto Tonhati pela orientação e amizade.

Aos amigos e também componentes da banca examinadora, Dr. Raul Borquis, Dr. Francisco, Dr. Guilherme e Dr. Henrique pelas valiosas sugestões e companherismo ao longo do período de doutoramento.

Ao meu coorientador Dr. Marcos Vinícius e a Embrapa Gado de Leite pelo aprendizado ao longo do período junto ao Laboratório de Bioinformática e Genômica Animal.

À minha co-orientadora Dra. Maria Gabriela Campolina Diniz Peixoto e o CBMG pelas valorosas contribuições e concessão das informações utilizadas neste trabalho.

À UNESP e ao Programa de Pós-Graduação em Genética e Melhoramento Animal da Faculdade de Ciências Agrárias e Veterinárias, pela oportunidade de realizar mais uma etapa da minha vida.

À Fundação de Amparo a Pesquisa (FAPESP), pela concessão da bolsa de estudo no país e de estágio no exterior e por todo recurso disponibilizado na minha formação.

À Professora Lúcia Galvão de Albuquerque e aos Professores do Departamento de ciências Exatas pelos conselhos e ensinamentos transmitidos.

À RKV: Krusty, Thiguinho, Marcolino, Murilo, André, João, Mineiro, Fábio, Dona Teresa, Cristina, Nega e a todos os outros ex-moradores ...Valeu!!!

Aos amigos de pós-graduação em melhoramento genético da salinha do Tonhati, Filhote, Gregório, Naudim, Diércles, Camila e Daniele e todos os outros amigos do departamento de melhoramento genético da FCAV.

À Universidade de Boku, ao professor Dr. Johann Sölkner, Solomon, Adam e todos os demais amigos de Viena.

A todos meus sinceros agradecimentos

SUMÁRIO

	Página
CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	1
Introdução.....	1
Revisão de Literatura.....	2
Objetivos.....	19
Referências Bibliográficas.....	20
CAPÍTULO 2 – ESTIMAÇÃO DO DESEQUILÍBRIO DE LIGAÇÃO E TAMANHO EFETIVO POPULACIONAL EM BOVINOS DA RAÇA GUZERÁ	35
Resumo.....	35
Introdução.....	36
Material e Métodos.....	37
Resultados e Discussão.....	39
Conclusões.....	44
Referências Bibliográficas.....	44
CAPÍTULO 3 – SELEÇÃO GENÔMICA PARA PRODUÇÃO DE LEITE, GORDURA E PROTEÍNA EM BOVINOS DA RAÇA GUZERÁ	49
Resumo.....	49
Introdução.....	50
Material e Métodos.....	51
Resultados e Discussão.....	59
Conclusões.....	73
Referências Bibliográficas.....	73
CAPÍTULO 4 – MAPEAMENTO AMPLO DO GENÔMA PARA A PROSPECÇÃO DE REGIÕES RELACIONADOS À PRODUÇÃO DE LEITE, GORDURA E PROTEÍNA EM BOVINOS DA RAÇA GUZERÁ.....	81
Resumo.....	81
Introdução.....	82
Material e Métodos.....	83
Resultados.....	85
Conclusões.....	93
Referências Bibliográficas.....	93

CAPÍTULO 5 – COMPARAÇÃO DE PAINÉIS DE SNPs DE ALTA DENSIDADE EM BOVINOS DA RAÇA GUZERÁ.....	101
Resumo.....	101
Introdução.....	102
Material e Métodos.....	104
Resultados e discussão.....	106
Conclusões.....	121
Referências Bibliográficas.....	122
CAPÍTULO 6 – CONSIDERAÇÕES FINAIS.....	127

ESTUDO GENÔMICO DA PRODUÇÃO DE LEITE E SEUS CONSTITUINTES EM BOVINOS DA RAÇA GUZERÁ

RESUMO – Painéis comerciais contendo milhares de SNPs a custo acessível revolucionaram os estudos genéticos na pecuária, principalmente por meio das análises de associação ampla e seleção genômica. A seleção genômica tem um aspecto prático, por ser diretamente aplicado aos programas de melhoramento, possibilitando aumento de acurácia das avaliações para as características quantitativas, como a produção de leite e seus constituintes. Como base nisso objetivou-se com esta tese verificar a distribuição das frequências dos polimorfismos e calcular o desequilíbrio de ligação (DL) dos segmentos cromossômicos no genoma de bovinos da raça Guzerá; estudar a associação dos marcadores com a produção de leite e seus constituintes; e comparar diferentes modelos para avaliação genômica com diferentes distribuições *a priori* para o efeito dos marcadores. Dessa forma, foi avaliado o DL entre marcadores de um painel de 50 k da Illumina® e estimado o tamanho efetivo populacional. Para isto foram utilizados 50 touros e 853 vacas Guzerá que também participaram dos estudos de associação e seleção genômica. A média de r^2 foi de 0,16 para a distância 100 kb no genoma destes animais. A densidade do painel de 50 k foi considerada suficiente para proporcionar DL entre os segmentos cromossômicos para a predição de valores genéticos genômicos. Já as estimativas do tamanho efetivo populacional foram reduzidas com o decorrer das gerações, indicando aumento da intensidade de seleção para a raça ao longo das gerações. O baixo tamanho efetivo observado para as gerações recentes (137) indicaram a importância de se considerar a endogamia nas decisões de acasalamentos para manter a diversidade genética da raça. As avaliações genômicas foram realizadas pelos métodos GBLUP, BayesC, BayesC π , Lasso e por meio de dois modelos multicaracterística para a produção de leite (PL), gordura (PG) e proteína (PP). As acurácias das predições variaram de 0,65 a 0,80. As metodologias de predição bayesianas, GBLUP e os modelos de avaliação genômica multicaracterística foram equivalentes para a predição dos valores genômicos dos animais desta população. No entanto, o GBLUP é a melhor opção considerando um ajuste global, além de exigir menor demanda computacional e facilidade em convergência em relação aos outros modelos. As predições genômicas foram adequadas para a orientação de seleção de animais da raça, quando estas foram baseadas nos efeitos dos marcadores e poligênico. Também foram calculadas as porcentagens de variâncias explicadas para cada SNP. Janelas composta por sete SNPs adjacentes que desviaram substancialmente das demais regiões genômicas foram investigadas. Foram observados dois sinais fortes relacionados à explicação da variância fenotípica no cromossomo 8 e 28 para PL, PG e PP. Nestas regiões foram encontrados 30 genes, sendo que alguns deles tinham alta expressão no tecido mamário, ou estavam relacionado à fisiologia de excreção e ao metabolismo glicídico. Dentro destas duas regiões, haviam cinco candidatos à QTLs que foram reportadas para PL e PP na literatura. Um quarto estudo adicional objetivando avaliar o desempenho do “Axiom® Genome-Wide BOS 1 Array” da Affymetrix e o “BovineHD BeadChip” da Illumina® foi realizado, considerando quarenta e oito touros com representividade genética da população. Ambos os painéis apresentaram

cobertura genômica adequada, possuindo alta densidade de marcadores espalhados pelo genoma, sendo que parte destes, apresentou cobertura genética direta (direta nos genes). Apesar das estimativas do DL terem sido maiores com o painel da Illumina, foram observados níveis altos e moderados de DL para pequenas distâncias com os dois painéis. Estes níveis podem ser adequados para auxiliar na obtenção de valores genéticos genômicos para animais da raça Guzerá. No entanto, a combinação destes dois painéis comerciais não indicou viabilidade.

Palavras-chaves: Desequilíbrio de ligação, “GWAS”, Modelos bayesianos, Seleção genômica, Zebu.

GENOMIC STUDY FOR MILK YIELD AND ITS CONSTITUENTS IN GUZERÁ CATTLE

ABSTRACT – Commercial panels containing thousands of SNPs revolutionized genetic studies in livestock, especially with the genome-wide association study and genomic selection. The genomic selection is also directly applied to breeding programs enabling increasing on accuracy of the genetic evaluation for quantitative traits such as milk yield and its constituents. Thus, the aim of this thesis was to study the distribution of frequencies of polymorphisms and the linkage disequilibrium (LD) in the Guzerá cattle; was also to study the association of the markers with the dairy traits and to compare different models for genomic evaluation with different prior distributions for the effect of markers. The LD and the effective size were estimated using a 50 k Illumina® panel for 50 sires and 853 cows. For 100 kb the average of r^2 was 0.16. The density of this 50 k panel was considered sufficient to provide LD between chromosomal segments for estimation of genomic breeding values. The effective size was reduced over the course of generations, indicating increase intensity of selection. The low estimate for effective size for recent generations indicated the importance of considering inbreeding for mating to maintain the genetic diversity of this breed. Genomic evaluations were performed by GBLUP, BayesC, BayesC π , Lasso and two multi-trait models for milk yield (MY), fat (MF) and protein (MP). The accuracies for predictions ranged from 0.65 to 0.80. Bayesian prediction, GBLUP and multi-traits models were equivalent for predict the genomic breeding values. However, the best option was GBLUP considering overall fit, lower computational requirement and facility of convergence. The results indicated that genomic predictions were adequate to assist animal selection. In other study, the percentages of explained variance for each SNP were calculated. Windows, composed of seven adjacent SNPs deviated substantially from the other genomic regions were investigated. Two strong signals were observed and related to the explanation of phenotypic variance in BTA8 and BTA28 for MY, MF and MP. In these regions were found 30 genes and some of them had high expression in mammary tissue, or were related to the metabolism of glucose and excretion physiology. Within these regions, were reported five candidates to QTLs for MY and MP in the literature. An additional study for evaluate the performance of “Axiom® Genome-Wide BOS 1 Array” from Affymetrix and the BovineHD BeadChip from Illumina® was performed using forty eight sires. These sires had genetic representation for this population. Both panels had adequate genomic coverage, high density of markers distributed in the genome. Some markers showed also direct genetic coverage (directly inside the genes). The LD estimates were higher with the Illumina panel. But we observed high and moderate levels of LD for short distances with both panels. It may be suitable to obtaining genomic breeding values for these animals. However, the combination of both panels indicated no viability.

Keywords: Linkage disequilibrium, GWAS, Bayesian models, Genomic Selection, Zebu.

CAPÍTULO 1 - CONSIDERAÇÕES GERAIS

1. INTRODUÇÃO

Dentre as raças zebuínas consideradas de dupla-aptidão a raça Guzerá apresenta, em conjunto, maior potencial lácteo e frigorífico. Tal aptidão produtiva ainda é muito baixa quando comparada às raças taurinas especializadas que apresentam níveis maiores de produção. A especialização destas raças europeias deve-se à maior intensidade de seleção para características produtivas, bem como, maior período de exposição a este tipo de seleção. A raça Guzerá, como os zebuínos em geral, teve menor intensidade de seleção artificial e maior de seleção natural, o que conferiu alta adaptabilidade às condições tropicais.

Introduzida no Brasil no século XIX com a finalidade de servir de força motriz (trabalho), a raça Guzerá passou a ser selecionada para produção de leite e carne por iniciativa dos criadores (MELO; PENNA 2000). Nas últimas décadas, programas de melhoramento atuando em conjunto com ABCZ (Associação Brasileira dos criadores de Zebu) auxiliam no progresso genético da raça para características de interesse econômico (LÔBO et al., 2014; PEIXOTO et al., 2014).

O programa de Melhoramento do Guzerá para leite, executado pela Embrapa Gado de Leite e pelo Centro Brasileiro de Melhoramento do Guzerá, é responsável pelas avaliações moleculares e genético-quantitativas das características ligadas à produção de leite, seja por teste progênie ou pelo núcleo MOET (múltipla ovulação e transferência de embriões) (PEIXOTO et al., 2014). Entretanto, a acurácia dos valores genéticos pela avaliação dos animais em esquema de famílias MOET não é muito alta quando comparada ao teste de progênie (BABIUK; PHILLIPS; MOO-YOUNG, 2013). Neste sentido, a tecnologia de seleção genômica pode ser de grande utilidade para as avaliações genéticas da raça, já que as acurácia dos valores genéticos obtidas por meio das avaliações genômicas para touros jovens podem ser mais interessantes que os valores genéticos obtidos pela avaliação tradicional (PSZCZOLA; MULDER; CALUS, 2011). Além disso, as avaliações genômicas tendem à discriminar os melhores animais (geneticamente superiores), ao

invés das melhores famílias e predizem valores genéticos diferenciados para irmãos completos que ainda não possuem informações de desempenho próprio ou de suas filhas (DAETWYLER et al., 2007).

As avaliações genômicas já são parte importante de muitos programas de melhoramento genético para gado leiteiro (HAYES et al., 2009; VANRADEN et al., 2009; HARRIS; JOHNSON 2010). Em alguns destes programas, os touros antigos que foram provados pelo teste de progênie foram genotipados e essas informações têm sido compartilhadas entre programas de outros países, como a cooperação norte-americana (MUIR; DOORMAAL; KISTEMAKER 2010) e o projeto EuroGenomics (LUND et al., 2011).

A seleção genômica por empregar simultaneamente milhares de marcadores distribuídos ao longo do genoma, possibilitará o aumento de acurácia por incrementar informações moleculares as avaliações para características quantitativas, como a produção de leite e seus constituintes (MEUWISSEN; HAYES; GODDARD, 2001). Quando usada como critério para a pré-seleção para o teste progênie, as avaliações genômicas poderão reduzir gastos para provar animais que teriam baixo desempenho no teste (HAYES et al., 2009). Sendo assim, diante das vantagens apresentadas é justificado o estudo de seleção genômica para produção de leite e seus constituintes na raça.

2. REVISÃO DE LITERATURA

2.1 Base da análise genômica: Desequilíbrio de Ligação.

As características quantitativas leiteiras passaram a ser estudadas e até mesmo selecionadas em muitos programas de melhoramento genético, com auxílio de marcadores moleculares os quais indicam as variantes de determinada região do DNA responsável pela expressão das características. Estes marcadores podem ser diretos, que indicam exatamente a mutação causal de um gene, ou indiretos, que marcam regiões que estão próximas à mutação causal ou à regiões relacionadas com estas em apenas algumas famílias (DEKKERS, 2004).

Quando se trabalha com painéis de muitos marcadores em características quantitativas, estes marcadores são praticamente indiretos com

proximidade das mutações causais (DEKKERS, 2004). Desse modo a distância entre o *locus* do marcador e a mutação implica na probabilidade de ocorrer o processo de recombinação. A recombinação é um fenômeno que ocorre durante a formação dos gametas (espermatozoide e óvulo) e envolve a troca aleatória de material genético entre os cromossomos homólogos (GRIFFT et al., 2007). A taxa de ocorrência de recombinação entre dois *loci* é proporcional à distância física entre eles em um mesmo cromossomo. Desta forma, quanto menor a distância entre dois *loci*, mais lento será para se atingir o equilíbrio entre as frequências genótípicas observadas e esperadas desses *loci* sob gerações de acasamentos ao acaso (FALCONER; MACKAY, 1997).

A medida de desequilíbrio de ligação (DL) indica uma associação não aleatória entre dois *loci* considerados, baseados em suas frequências genótípicas e alélicas (FALCONER; MACKAY, 1997). A principal causa do DL, embora não exclusiva, é o “linkage” (ligação) entre os *loci* (proximidade física). A técnica de seleção genômica explora o DL entre os marcadores, pois assume que os efeitos dos segmentos cromossômicos analisados representam também o DL entre o marcador e um possível QTL (“Quantitative Trait Locus”) (DE ROOS et al., 2008). Neste sentido as primeiras abordagens a serem empregadas anteriormente à implementação da seleção genômica em um programa de melhoramento genético ou até mesmo ao estudo genômico, são os estudos visando caracterizar a extensão, a distribuição e o decaimento do DL em uma população.

Existem muitos índices empregados para mensurar o DL, sendo que para marcadores bialélicos os principais são $|D'|$ (LEWONTIN, 1988) e r^2 (HILL; ROBERTSON, 1968) em que ambos variam de 0 a 1 (BOHAMANOVA; SARGOLZAEI; SCHENKEL, 2010). A medida $|D'|$ é uma tentativa de parametrizar o DL (GUPTA; RUSTGI; KULWAL, 2005), e se tornou útil em comparações entre *loci* quando diferentes frequências são analisadas. Entretanto, o $|D'|$ tende a ser fortemente inflacionado em pequenas amostras e na presença de alelos raros (TEARE et al., 2002). Já a medida r^2 não possui estas limitações e pode ser interpretada como um coeficiente de correlação de frequências entre os alelos. O r^2 também é preferido para estudos de associação também porque existe uma relação inversa entre este e o tamanho da amostra necessária para detectar a associação entre um QTL e os SNPs

(WALL; PRITCHARD, 2003).

Inicialmente Farnir et al. (2000) e Tenesa et al. (2003) estudando o DL em bovinos de leite com microssatélites relataram alto níveis de DL por longas distâncias. Entretanto estudos baseados em SNPs mostraram DL alto em curtas distâncias. Para a raça Holandesa, McKay et al., (2007) (2.670 SNPs) e Bohamano, Salgozaei e Schenkel (2010) (38.590 SNPs) encontraram $r^2 \geq 0,2$ em distâncias de até 100 kb. Outros autores como Khatkar et al. (2008) (9.195 SNPs), Qanbari et al. (2010) (40.854 SNPs) e Sargolzaei et al. (2008) (5.564 SNPs) observaram $r^2 \geq 0,2$ em populações de bovinos da raça Holandês em distâncias inferiores a 100 kb. Já De Roos et al. (2008) (3.072 SNPs) encontraram valores de $r^2 \leq 0,15$ para “Holstein”, Jersey e Angus em distâncias de até 100 kb. Silva et al. (2010) analisando o DL em 25 touros da raça Gir com um painel de 54.000 SNPs obteve o valor de 0,21 para média de r^2 em 100 kb. Entretanto Santos et al. (2013a) encontraram valores de r^2 iguais a 0,17 para as raças Gir (1959 animais) e Sindi (116 animais) trabalhando com um painel de 54.000 SNPs.

A variação observada na extensão das estimativas de DL publicadas depende de vários fatores tais como: história e estrutura populacional; tamanho efetivo populacional (HAYES et al., 2003); o próprio tamanho da amostra, pois quando pequenas pode levar à superestimação (YAN et al., 2009); densidade e distribuição de marcadores, o tipo de método utilizado para a reconstrução de haplótipos, rigor de filtragem dos SNPs (limiar de MAF e equilíbrio de Hardy-Weinberg); e o uso de apenas haplótipos maternos, ou de ambos os haplótipos maternos e paternos (BOHAMANOVA; SARGOLZAEI; SCHENKEL 2010).

Todos os resultados observados na literatura consultada indicam que a densidade dos marcadores utilizados foi suficiente para proporcionar desequilíbrio de ligação entre os segmentos cromossômicos, e consequentemente entre os SNPs e os QTL, para uma eficiente estimação dos valores genéticos genômicos, principalmente para distâncias menores que 100 Kb (r^2 moderados a alto). Meuwissen, Hayes e Goddard (2001) ao propor a ideia de seleção genômica utilizaram marcadores adjacentes com $r^2 > 0,20$ indicando que tais marcadores, com esse DL com QTL, podem obter uma explicação efetiva da variação fenotípica resultante do efeito do QTL. Com base nisto, de Roos et al. (2008) estimaram que seria necessário um painel de

no mínimo 50.000 SNPs para se alcançar $r^2 \geq 0,20$ entre os marcadores adjacentes entre as raças por ele estudada. Entretanto, a tendência atual para a seleção genômica de bovinos leiteiros é utilizar painel de alta densidade (>600.000 SNPs). Apesar do aumento do número de informações a serem processadas, estas são mais acuradas em relação à captura do efeito do QTL pelos marcadores.

O DL também é importante na construção de blocos haplotípicos e das fases haplotípicas. Estes blocos por vezes podem ser utilizados como unidades para as análises genômicas ao invés do próprio SNP (CALUS et al. 2008), nos algoritmos de imputação (BROWNING; BROWNING, 2010), e na detecção genômica de alelos letais (VANRADEN et al. 2011). Segundo Khatkar et al. (2007), os haplótipos são regiões cromossômicas de alto DL e que normalmente apresentam baixa diversidade, sendo representando tipicamente por regiões de baixa recombinação flanqueadas por “hotspots” de recombinação. Já quando o DL é estimado em duas populações com os mesmos SNPs é possível estudar a persistência de fase de ligação entre elas. A persistência de fase de ligação refere-se quanto um segmento cromossômico permanece inalterado ao longo de uma determinada distância física, em diferentes subpopulações, raças ou espécies. Esta medida é baseada nas correlações dos r^2 entre duas populações ao longo das distâncias físicas (DE ROOS et al., 2008). Segundo Badke et al., (2012) a persistência de fase de ligação é uma medida do grau de concordância de fase de DL para os pares de SNPs entre duas populações. Através desta medida também é possível especular o tempo de divergência em número de gerações passadas entre duas populações (DE ROOS et al., 2008).

2.2 Evolução das avaliações: tradicional, assistida por marcadores e genômica.

O melhoramento genético animal é alcançado por meio da seleção de indivíduos que são considerados superiores para reprodução. A classificação desta superioridade (avaliação), a partir das informações dos indivíduos participantes de um programa de melhoramento genético, é o principal recurso para orientação dos acasalamentos. Quanto maior é a semelhança entre a classificação sugerida com a classificação genética real, maior será a garantia

que a superioridade será transmitida para a geração seguinte. A precisão dos métodos de avaliação, conhecida como acurácia, é um dos principais componentes que determinam o ganho genético em uma população, sendo de grande importância dentro de um programa de melhoramento a utilização de métodos de avaliação genética adequados. Inicialmente, estas avaliações eram baseadas no fenótipo, ou seja, os animais que tinham melhores desempenhos eram escolhidos, ou no caso da produção de leite, os filhos das vacas mais produtivas. O valor genético era obtido pela simples multiplicação do desvio fenotípico da média do rebanho pela herdabilidade. Devido à complexidade das várias fontes de informação, havia certa dificuldade quando se desejava utilizar simultaneamente informações de registro único, registros repetidos, informações das progênes, do pedigree e de outras características correlacionadas para predição do valor genético de um indivíduo.

No final da metade do século XX, passou-se a ser utilizada a metodologia dos índices de seleção (HAZEL; LUSH 1943), que considerava as relações entre as medidas fenotípicas, bem como as relações (covariâncias) entre animais com fenótipos (critérios de seleção - presentes na mão esquerda) e os animais avaliados (objetivo de seleção - presentes na mão direita). Com esta metodologia era possível processar todos os tipos de fontes de informação, de forma que o índice era o valor genético dos animais. Os índices são considerados como BLP (*Best Linear Prediction*) e apresentam como principais propriedades a diminuição do erro de predição, maximização da correlação entre o valor genético estimado e verdadeiro (acurácia) e a maximização da probabilidade de classificação correta pelo valor genético predito. Desta forma, houve um incremento de acurácia no índice por agregar informações da colateralidade de outros animais. A partir desta metodologia passou-se a dar maior importância ao pedigree dos animais para aplicação em avaliações, além da convencional relação progenitor-progênie.

Posteriormente, com o advento dos modelos mistos (HENDERSON, 1949) as avaliações passaram a fornecer estimativas mais acuradas do valor genético. No início, por meio do modelo touro que considerava relações moderadas entre os animais e, em seguida utilizando-se o modelo animal, que incluíam a matriz de parentesco na análise e considerava, portanto, relações completas entre todos os animais bem como com seus ancestrais. Nesta

metodologia foi possível estimar simultaneamente efeitos fixos (BLUE - Best Linear Unbiased Estimator) e aleatórios (BLUP – Best Linear Unbiased Prediction). Nesses modelos a inversa da matriz de parentesco entre os animais, multiplicada pela variância genética aditiva é adicionada à sub-matriz resultante da multiplicação de duas matrizes de incidência dos efeitos genético aditivo na mão esquerda da equação de mínimos quadrados generalizado. Desse modo a solução BLUP é obtida para todos os animais que estão presentes na matriz formada pelo pedigree. Esta metodologia tem propriedades semelhantes ao do índice de seleção, porém resulta no valor genético predito diretamente, diferente dos índices em que primeiro é obtido os ponderadores e depois os valores genéticos. Com isso o valor genético estimado passou a ser utilizado mais intensamente como ferramenta de seleção nos programas de melhoramento genético, sempre divulgados em sumários na forma de PTA ou DEP (metade do valor genético predito).

A produção de leite, gordura e proteína são características constantemente trabalhadas nos rebanhos participantes de controles leiteiros. Como essas características se repetem ao longo de um período (curva de lactação), é possível empregar além dos modelos unicaracterística e de repetibilidade os modelos que consideram o dia de controle como o multicaracterística de dimensão finita ou de regressão aleatória (JAMROZIK; SCHAEFFER; DEKKERS, 2007; SANTOS et al. 2013b). Tradicionalmente, a produção de leite tinha como critério de seleção a avaliação da produção acumulada até os 305 dias (P305), podendo ser utilizados dados de uma só lactação (unicaracterística) ou de várias lactações como repetições de uma mesma característica (repetibilidade) (EL FARO; ALBUQUERQUE, 2003). As desvantagens destes métodos de avaliação não se encontram somente nos modelos de avaliação, mas também da própria utilização da P305 como critério. Para a avaliação genética da estimativa da P305, é necessário que as vacas tenham as lactações encerradas para que sejam avaliadas. Caso contrário, para que se incluam essas informações nas avaliações, as lactações devem ser projetadas para 305 dias. Entretanto esse procedimento não é realizado em alguns programas de melhoramento para leite, como de animais zebuínos, já que poderia acarretar problemas de superestimação da P305 para animais de lactações curtas.

Alternativamente a P305, existem os próprios registros de controles que podem ser utilizados como critério de seleção. Os modelos que consideram os dias do controle (*test-day models*) também são utilizados em larga escala por alguns programas de melhoramento de taurinos leiteiros do mundo inteiro, destacando o Canadá, Alemanha, Holanda, Áustria, Brasil entre outros (ALBUQUERQUE, 2004). Dentre os *test-day models*, o modelo de dimensão finita consideram os registros de produção, como características distintas, e os modelos de regressão aleatória abordam esses controles como dados longitudinais, de modo que consideram as produções em função do tempo, ou seja, consideram a curva de lactação dos animais. A vantagem dos modelos de regressão aleatória está justamente na estimação da curva genética ao invés de estimativas pontuais dos valores genéticos, o que permite avaliar a persistência genética da produção leite como consequência (JAMROZIK; SCHAEFFER, 1997). Apesar da maior acurácia obtida com estes modelos em relação ao unicaracterística ou de repetibilidade, a demanda computacional para o processamento aumentou pelo aumento do número de parâmetros a serem estimados.

Com o intuito de incrementar a acurácia e também reduzir o intervalo de geração, foram estudadas, e até mesmo aplicadas em alguns programas, as informações moleculares além das informações fenotípicas e de pedigree. Primeiramente, foi usada a seleção assistida por marcadores (MAS). Esta metodologia foi aplicada em gado de leite principalmente na pré-seleção de animais, como para entrada de touros jovens em programas de teste de progênie (KASHI; HALLERMAN; SOLLER, 1990; MACKINNON; GEORGES, 1998). A MAS utiliza simultaneamente dados fenotípicos e dados de marcadores moleculares em DL com alguns QTLs e foi recomendada com o objetivo de aumentar o ganho genético anual para as características de importância econômica em várias espécies animais (DEKKERS, 2004). Na MAS, a predição do valor genético total por meio do BLUP é obtida incluindo juntamente com a informação poligênica, os dados dos marcadores moleculares como efeito fixo ou aleatório (DEKKERS, 2004), ou utilizando-se índices que combinem essas duas informações, em que os ponderadores podem ser mudados de acordo com o objetivo (DEKKERS; VAN ARENDONK, 1998).

Meuwissen e Goddard (1996) mostraram, a partir de dados simulados, que a MAS é especialmente útil para aquelas características de baixa herdabilidade, com alto custo de mensuração, medidas tardiamente na vida do animal ou após a seleção. Entretanto, para implementação da MAS, algumas problemáticas foram observadas. Pelo fato das características quantitativas envolverem grande quantidade de genes para sua expressão, somente uma proporção pequena da variância genética total era capturada pelo pequeno número de marcadores moleculares, (geralmente relacionados aos genes de efeito maior) limitando o progresso ou ganho genético (SOLBERG et al. 2008). Outro ponto é que quando a avaliação é para características com muitos *loci*, como as leiteiras, existe certa dificuldade de incorporação dessas informações moleculares nos modelos, como na determinação dos ponderadores a serem usados, principalmente quando se objetiva selecionar múltiplas características. Outro limitante é que para a aplicação da MAS para pré-seleção de touros, é necessário que as famílias sejam grandes o suficiente (principalmente com muitos irmãos completos) para que seja possível a seleção dentro de família. Além disso, a MAS requer estabelecimento do DL entre marcadores e QTLs para cada família em avaliação, o que pode não representar todas as famílias da população em questão. Em função desses aspectos, a implementação da MAS tem sido limitada e os ganhos em eficiência são muito reduzidos (DEKKERS, 2004).

Visando contornar as problemática da MAS, alternativas moleculares estão sendo amplamente estudadas. A partir do trabalho de Meuwissen, Hayes e Goddard (2001) concretizou-se a metodologia de seleção genômica (SG) e iniciou-se a comercialização, à custos acessíveis, de painéis com milhares de marcadores de polimorfismos de nucleotídeo único (SNP) espalhados pelo genoma bovino (VAN TASSELL et al., 2008). Os SNPs são formas de variação abundantes de polimorfismos de DNA presentes em todo o genoma e têm se tornado preferenciais em relação aos outros tipos de marcadores moleculares, uma vez que apresentam menores taxas relativas de mutação e são de fácil genotipagem (ROMUALDI et al. 2002). Esse tipo de marcador é comumente utilizado para a detecção e localização de QTL para características complexas (HAYES et al. 2008). Na SG, a idéia central não é utilizar marcadores específicos para os QTLs, mas sim utilizar uma quantidade expressiva deles,

espalhados por todo o genoma, de forma a promover uma cobertura ampla de todas as regiões do DNA presumindo-se que sempre existem marcadores próximos e em DL com um possível QTL (DE ROOS et al., 2008). Por meio desta propriedade é possível envolver todo o genoma para a análise de uma determinada característica, o que implica em uma eficiente explicação dos efeitos genéticos totais, permitindo também a utilização destas informações em modelos estatísticos para estimação de valores genéticos, aqui denominados de valores genéticos genômicos (MEUWISSEN; HAYES; GODDARD, 2001). Desta forma, a SG é a seleção de indivíduos considerados superiores a partir desses valores genéticos genômicos (GEBV).

A SG não possui os mesmos limitantes da MAS e de acordo com Dekkers (2007), Meuwissen (2007) e Muir (2007), e quando comparada à seleção tradicional quantitativa, baseada no BLUP, fornece: 1) predições dos valores genéticos com maior acurácia, principalmente para as características que se expressam em um único sexo e/ou sejam de baixa herdabilidade, 2) menor incremento na endogamia (menor tendência a seleção familiar), 3) antecipação do processo de seleção no caso de características mensuradas tardiamente na vida do animal, além de 4) viabilizar a avaliação de características de difícil mensuração ou de alto custo.

A SG é a mais promissora ferramenta que surgiu para aumentar a taxa de ganho genético na pecuária (WEIGEL et al., 2010). O incremento de acurácia das predições genômicas é observado principalmente em animais jovens não havendo mudanças significativas nos touros já provados (SCHAEFFER, 2006). As provas são usadas para pré-seleção para o teste progênie e também para seleção de animais quando estes são selecionados pelo GEBV na avaliação genômica total. As provas genômicas para gado de leite possuem um fluxo que envolve uma população de referência e outra população a ser selecionada. Desta forma, a população de referência é constituída por animais que necessariamente possuem informações acuradas da característica. Esta população servirá de base genética para a predição dos efeitos dos marcadores. Os efeitos dos marcadores são multiplicados pelos genótipos dos animais que se desejam avaliar, e os valores genéticos genômicos destes animais é predito pela somatória dos resultados destas multiplicações, sem a utilização de informações fenotípicas desse animal ou de

suas filhas (HAYES et al., 2009).

A determinação da população de referência, como o seu tamanho e constituição, tem grande influência na precisão das predições genômicas (HAYES et al., 2009; VANRADEN et al. 2009). No caso de rebanhos leiteiros a constituição é dependente principalmente da composição (touro e vaca) e a genotipagem seletiva de acordo com a estrutura da variável resposta. Na maioria dos países, apenas os touros, principalmente de alta acurácia foram genotipados e incluídos na população de referência (LOBERG; DÜRR, 2009). Entretanto, em muitos países, existe um número limitado de touros que são testados e isso dificulta a obtenção acurada de predições dos coeficientes dos marcadores na população de referência de modo que diferentes estratégias podem ser utilizadas. Uma das estratégias é a inclusão de genótipos de fêmeas na população de referência. Em muitos programas de melhoramento genético as matrizes, candidatas a mãe de touros, são genotipadas (SPANGLER et al., 2008). Por outro lado, a utilização das informações mais acuradas, implicam na utilização dos melhores animais como referência. Em estudo de simulação de gado leiteiro, Jiménez-Monteiro, González-Recio e Alenda (2011) concluíram que a amostragem bi-caudal fêmeas juntamente com a amostragem usual para machos deve ser considerada, muito embora estes autores não tenham avaliados esta combinação com desvios de rendimento das filhas (DYD) para machos ou qualquer outro pseudo-fenótipo.

O primeiro passo para a seleção genômica é gerar a variável resposta para a análise genômica que depende das fontes disponíveis (desempenho próprio, das filhas ou dos pais). Diferentes informações podem ser usadas desde o próprio fenótipo, como registros únicos, repetidos ou a média da progênie, até os pseudo-fenótipo como DYD (VANRADEN; WIGGANS, 1991), EBV (estimated breeding value) e EBV-desregressado (GARRICK; TAYLOR; FERNANDO, 2009). Para as características leiteiras os pseudo-fenótipos são preferidos por atribuir informações dos machos para análise genômica. Dentre estes, o DYD é o mais usado porque além dos touros apresentarem impacto maior sobre programas de melhoramento genético do que as vacas, seus DYDs são mais acurados do que os fenótipos das vacas (CALUS, 2009). Os EBV-desregressados pode ser considerado um tipo de “deregressed-proff” quando são utilizados touros acurados na população de referência, pois

combinam diferentes fontes de informação além das informações de suas filhas. De um modo geral, os “deregressed-proff” são considerados equivalentes ao DYD (SIGURDSSON; BANOS, 1995). Os EBV-desregressados são utilizados em avaliações genômicas de características leiteiras em caso em que existam vacas e touros na população de referência. Entretanto quando se utiliza qualquer pseudo-fenótipo desregressado há um incremento individual desproporcional à variável resposta que leva a necessidade de se considerar heterogeneidade do resíduo pelos modelos estatísticos com ponderadores que variam de acordo com a desregressão e efeitos considerados (GARRICK; TAYLOR; FERNANDO, 2009).

Os resultados das avaliações genômicas quando são usados como critérios para a pré-seleção de animais para o teste progênie possibilita a redução de gastos para provar touros que teriam baixo desempenho no teste (HAYES et al., 2009). Schaeffer (2006) reportou que as mudanças genéticas podem ser de três à quatro vezes maiores que os atuais programas de teste progênie e economizar com custos logísticos em até 97%. Além disso, os custos de genotipagem também são susceptíveis de diminuir ao longo do tempo, o que tornaria a SG mais fácil de administrar. Schrooten et al., (2005) reportaram que o progresso genético aumentou de 19 para 31% comparado com o teste progênie, quando os marcadores explicaram 50% da variância genética. VanRaden et al., (2009) reportaram que a habilidade de predição para características leiteiras usando as predições genômicas foi 50% contra 27% pelo o PTA tradicional. Eles também reportaram que os ganhos para os touros provados foram altamente significativos, embora menores do que os touros jovens devido à elevada confiabilidade inicial dos touros antigos. Vale ressaltar que a SG já é aplicada e mostraram resultados promissores em vários países como USA, Canada, Nova Zelândia e Holanda, onde os valores genéticos dos touros leiteiros têm sido preditos com uso de informações genômicas (HAYES et al., 2009).

Apesar destas últimas inovações na área genômica trazerem as vantagens descritas acima, as informações moleculares exigiram o aumento de recursos estatísticos e computacionais, limitando o uso das informações em muitas análises, como as realizadas por meio de modelos multicaracterísticas e pelos modelos que consideram o dia do controle. Embora estes modelos sejam

aplicados facilmente substituindo a matriz de parentesco obtida a partir do pedigree (A) pela matriz de parentesco genômica (G) (KOIVULA et al, 2012; TSURUTA et al, 2011), ainda não existe uma análise robusta de natureza multivariada com modelos que consideram diferentes variâncias para os diferentes marcadores, sendo esta, talvez, uma das maiores perspectivas para otimização de análises genômicas das características leiteiras.

2.3 Metodologias de avaliação genômica

A acurácia das predições das avaliações genômicas é afetada por muitos fatores, entre eles estão: o número de animais em conjunto de dados de treinamento; a densidade do painel de marcadores, que determina o nível de DL entre marcadores e QTL; a relação entre os animais do conjunto de referência e de validação; a relação entre os animais dentro do conjunto de referência; a herdabilidade da característica; a acurácia das observações ou pseudo-fenótipo usado; e o modelo e metodologia estatística aplicada (GODDARD, 2009; HAYES et al., 2009; GODDARD; HAYES, 2009). Neste sentido, as principais metodologias e modelagens consideradas para as predições genômicas serão detalhadas a seguir:

2.3.1 Principais efeitos ajustados nos modelos de predição dos valores genômicos.

Os diferentes modelos de avaliação genômica podem considerar inclusão ou não da informação dos poligenes juntamente com o efeito dos marcadores.

2.3.1.1 Modelo com efeito dos marcadores:

Para um determinado *locus* de um SNP existem dois alelos possíveis (1, 2), podendo existir três possíveis genótipos: “11”, “12” e “22”. Arbitariamente será designado com o valor $+1/2 a_j$ para o alelo 1 e o valor de $-1/2 a_j$ para o alelo 2, de acordo com a parametrização clássica na qual a_j é a metade da diferença entre dois homozigotos. O modelo pode ser representado na seguinte na notação matricial:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wa} + \mathbf{e},$$

em que $\mathbf{y}$ é o vetor das observações; $\mathbf{b}$ é o vetor dos efeitos ambientais (incluindo a média); $\mathbf{a}$ é o vetor dos efeitos dos marcadores; $\mathbf{X}$ e $\mathbf{W}$ são as correspondentes matrizes de incidência. É assumido que os resíduos são independentes e apresentam distribuição $\mathbf{e} \sim N(0, I\sigma_e^2)$. Para os efeitos dos marcadores ($\mathbf{a}$), diferentes distribuições *a priori* podem ser usadas.

2.3.1.2 Modelo com efeito dos marcadores e o componente poligênico:

Uma extensão do modelo anterior, incluindo o efeito dos poligenes, pode ser representado da seguinte forma:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wa} + \mathbf{Zu} + \mathbf{e},$$

em que $\mathbf{u}$ é o vetor dos efeitos poligênicos aditivos que representa aqueles genes que não são considerados pelos marcadores. Esse efeito aleatório possui uma distribuição *a priori* $\mathbf{u} \sim N(0, A\sigma_u^2)$, com A sendo a matriz de parentesco animal obtida pelo pedigree.

2.3.2 Metodologias de predição dos efeitos dos marcadores e dos valores genômicos:

Para predição dos efeitos dos marcadores bem como os valores genômicos, diferentes metodologias são descritas na literatura sendo as principais:

2.3.2.1 BLUP Genômico:

As abordagens para o BLUP genômico são as mais utilizadas devido ao baixo número de parâmetros a serem estimados e a facilidade computacional no processo de estimação (SU et al., 2012; VANRADEN, 2008). Entretanto, em geral apresentam a desvantagem de estimar a mesma variância para todos os marcadores, e não realizar a seleção de covariáveis (marcadores), salvo em algumas implementações, como a utilização da matriz genômica (G) ponderada e a os modelos de regressão de cumeieira ("Ridge Regression") com efeitos heteroscedásticos (SHEN; ALAM; RONNEGARD, 2013). Existem duas abordagens principais na literatura, que, no entanto, são equivalentes. A primeira foi descrita por Meuwissen, Hayes e Goddard (2001) como RR-BLUP:

$$y = \mu + \sum_{i=1}^n X_i a_i + e$$

em que y é a variável resposta, μ é a média geral da população, X_i é uma matriz com os alelos de cada animal para cada marcador i , n é o número de marcadores ao longo do genoma e a_i é o vetor dos efeitos dos marcadores i , onde é assumido que $a_i | \sigma_a^2 \sim N(0, \sigma_a^2)$, e e é o efeito do erro aleatório onde é assumido que $e | \sigma_e^2 \sim N(0, I\sigma_e^2)$.

A outra abordagem foi descrita por VanRaden (2008) como G-BLUP:

$$y = \mu + Zg + e$$

Onde g é o vetor com os valores genéticos genômicos obtidos diretamente. É assumido que $g | \sigma_g^2 \sim N(0, \sigma_g^2 G)$, onde G é a matriz de parentesco genômico dada por $G = \frac{(XX')}{2 \sum_{i=1}^m p_i(1-p_i)}$, sendo p_i igual à frequência

alélica do marcador i . Z é matriz de incidência para g .

2.3.2.2 Procedimentos bayesianos:

Os procedimentos bayesianos apresentam vantagens em relação ao método BLUP, pois consideram variâncias diferentes para cada marcador e realizam seleção de covariáveis (com a exceção do BayesA). Seguindo o mesmo modelo geral apresentado para o RR-BLUP, os métodos bayesianos assumem distribuição para as variâncias residuais e distribuição específica para os efeitos dos marcadores de acordo com cada tipo de método usado.

a) BayesA (MEUWISSEN; HAYES; GODDARD, 2001)

No BayesA é assumindo uma distribuição *à priori* distribuição qui-quadrado escalonada invertida para variâncias dos efeitos dos marcadores (σ_i^2). Sendo a_i o efeito do marcador i assume-se que:

$$a_i | \sigma_i^2 \sim N(0, \sigma_i^2)$$

$$\sigma_i^2 \sim \chi^{-2}(v, S)$$

em que S é o parâmetro de escala e v é o número de graus de liberdade.

b) *BayesB* (MEUWISSEN; HAYES; GODDARD, 2001)

O método BayesB apresenta as mesmas suposições que o BayesA para uma fração fixa com probabilidade $1-\pi$ dos marcadores, e uma outra parte (π) apresenta efeitos nulos. Sendo assim:

$$a_i | \pi, \sigma_i^2 \sim (1-\gamma_i)N(0, \sigma_i^2 = 0) + \gamma_i N(0, \sigma_i^2)$$

$$P(\gamma_i = 0) = \pi$$

$$P(\gamma_i = 1) = (1 - \pi)$$

$$\sigma_i^2 \sim \chi^{-2}(v, S) \text{ com probabilidade } (1 - \pi),$$

em que os valores $v=4,234$ e $S=0,0429$ são a média e variância de σ_{gi}^2 , dado que $\sigma_{gi}^2 > 0$ (MEUWISSEN; HAYES; GODDARD, 2001).

c) *BayesC π* (KIZILKAYA; FERNANDO; GARRICK, 2010)

Assim como no BayesB, o BayesC π também possui uma fração π dos marcadores com variância igual à zero, porém para este é assumido uma distribuição *à priori* beta com parâmetros α e β .

$$a_i | \pi, \sigma_i^2 \sim \pi N(0, \sigma_i^2 = 0) + (1 - \pi) N(0, \sigma_i^2)$$

$$\sigma_i^2 \sim \chi^{-2}(v, S)$$

$$S \sim \text{Gama}(\alpha, \beta)$$

No método BayesC uma variância comum é especificada para todos os marcadores (σ^2) e π tem distribuição uniforme $U(0,1)$, que equivale ao método RR-BLUP com seleção de covariáveis.

2.3.2.3 *Lasso Bayesiano* (DE LOS CAMPOS et al. 2009)

O método Lasso Bayesiano também realiza seleção de covariáveis porém não por meio de uma fração π de marcadores, mas via regressão penalizada, onde uma distribuição controlada pelo parâmetro λ é declarada para toda coleção de variâncias dos marcadores. Segundo Yi e Xu (2008) este método minimiza a soma de quadrados dos resíduos limitando a soma dos valores absolutos dos efeitos dos marcadores, $t \geq \sum |\beta_j|$ para $t \geq 0$, se a resposta dos preditores estão padronizados. Esse tipo especial de restrição permite que

alguns coeficientes de regressão sejam exatamente iguais a zero. Cabe ressaltar que, com a utilização do modelo LASSO, o grau de restrição efeitos dos marcadores é estimado a partir de informações do conjunto de dados, sem necessidade de predeterminação do grau de restrição nos parâmetros. Desta forma tem-se que:

$$\begin{aligned} a_i | \tau_i \sigma_e^2 &\sim N(0, \tau_i^2 \sigma_e^2) \\ \tau_i^2 | \lambda^2 &\sim \text{Exp}(\lambda^2) \\ \lambda^2 &\sim \text{Gamma}(\varphi_1, \varphi_2) \end{aligned}$$

em que a variância dos efeitos dos marcadores é dada como $\sigma_i^2 = \tau^2 \sigma_e^2$. Os efeitos dos marcadores serão dados condicionalmente ao parâmetro λ :

$$p(a | \lambda) = \prod_{j=1}^p \frac{\lambda}{2} \exp(-\lambda |a_i|)$$

Legarra et al., (2011) propôs uma melhoria desta metodologia para predição dos efeitos dos SNPs, conhecida como LASSO bayesiano melhorado. A principal diferença desta implementação em relação ao LASSO descrito anteriormente está nas variâncias dos efeitos dos marcadores, que não é mais dado em função variância residual, mas sim em função variância genética aditiva.

$$a_i | \tau_i \sigma_g^2 \sim N(0, \tau_i^2 \sigma_g^2)$$

Então a variância dos efeitos dos marcadores é dada como $\sigma_i^2 = \tau^2 \sigma_g^2$. Em que σ_g^2 é a variância genética aditiva.

2.4 Análises de associação ampla do genoma (GWAS):

As análises de associação ampla do genoma (GWAS) é uma das principais estratégias para investigação das mutações subjacentes aos marcadores em características complexas. Esta técnica utiliza uma série de SNPs espalhados em todo o genoma para investigar a relação entre a variação genética explicada por esses marcadores e variação fenotípica da característica. Alguns destes SNPs podem ser a própria variante causal ou podem estar em alto DL com a variante causal, tendendo a serem herdados

juntos ao longo das gerações (PEARSON; MANOLIO, 2008; BUSH; MOORE, 2012). No entanto, a GWAS é apenas a primeira parte no processo de descoberta de um gene (CANTOR; LANGE; SINSHEIMER, 2010; HARDY; SINGLETON, 2009). Após a detecção da região genômica, os genes são anotados de acordo com as coordenadas genômicas para espécie. As conclusões dos GWAS necessitam de maior validação, como a utilização de recursos de sequenciamento e avaliação funcional dos genes. O GWAS foi empregado com sucesso na identificação de genes em humanos (SAXENA et al., 2007; SUN et al., 2011) e em diferentes espécies (COLE et al., 2011; FLINT; ESKIN, 2012; XIE et al., 2012). Para as características leiteiras a maioria dos estudos indicam o gene DGTA1, no cromossomo 14, como a região mais significativa (JIANG et al., 2010; WANG et al., 2013). Isto demonstra a consistência e a viabilidade desta metodologia para gado leiteiro, já que este gene é comprovadamente o de efeito maior nestas raças (GRISART et al., 2004).

Várias abordagens foram propostas para a GWAS: utilizando-se análises de associação por meio de marcadores únicos seja por testes de correlação ou de regressão genética aditiva (CLEVELAND; DEEB, 2009); modelos lineares mistos (LM) (YU et al., 2006; SAHANA et al., 2010); modelos de haplótipos (HM) (DASHAB et al., 2012) e genealogia baseada no modelo misto (GenMix) (DASHAB et al., 2012). Também estão sendo empregadas as mesmas modelagens utilizada para SG, principalmente as que consideram encolhimento dos efeitos dos marcadores com seleção de covariáveis como o Lasso (TIBSHIRANI, 1994), os modelos Bayesianos, como BayesC (PETERS et al., 2013), e o “single-step” GWAS (ssGWAS) que é uma adaptação do GBLUP (WANG et al., 2012).

Assim como na SG, a densidade do painel de marcadores, modelo estatístico empregado, a herdabilidade da característica, tamanho da amostra também têm impactos sobre GWAS. Considerando o método com testes de marcadores únicos existem dois outros problemas comuns: os testes múltiplos, a estratificação da população e o grau de parentesco entre os indivíduos da população. O enorme número de testes estatísticos realizados apresenta um potencial para resultados falso-positivos, levando a um novo rigor no nível aceitação da significância estatística e requisitando replicação dos resultados

encontrados (PEARSON; MANOLIO, 2008). Para resolver os múltiplos problemas de teste, correção de Bonferroni e teste de permutação podem ser usados. Já para estrutura, as análises são realizadas dentro de cada estrutura ou pode-se realizar correção do fenótipo para subestrutura populacional por meio dos componentes principais da matriz de relacionamento genômico (PRICE et al., 2006). Considerando que exista uma leve subestruturação e os indivíduos são pouco aparentados, a técnica “Genomic Control” para desinflacionar o valor de associação pode ser adotado (DEVLIN; ROEDER, 1999). Já quando os indivíduos da população são aparentados, a abordagem por meio dos modelos mistos é preferida por discriminar o efeito poligênico do efeito do marcador.

3. OBJETIVOS

O objetivo geral deste tese foi a detecção de polimorfismos presentes no DNA que estejam associados às características de produção de leite, gordura e proteína da raça Guzerá, bem como a avaliação de modelos para predição de valores genéticos genômicos, visando a futura seleção genômica de animais para estas características. Especificamente, os objetivos da tese foram:

- Verificar a distribuição das frequências dos polimorfismos e calcular o desequilíbrio de ligação no genoma de bovinos da raça Guzerá.
- Comparar diferentes modelos para avaliação genética genômica em relação à diferentes distribuições *a priori* para a predição dos efeitos do marcadores, considerando a inclusão ou não da informação dos poligenes (modelo infinitesimal).
- Estudar a associação dos polimorfismos com produção de leite e seus constituintes nas vacas da raça Guzerá.
- Comparar painéis de comerciais de alta densidade que serão utilizados em uma futura genotipagem de animais desta população.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBUQUERQUE, L. G. Regressão aleatória: nova tecnologia pode melhorar a qualidade das avaliações genéticas. In: SIMPÓSIO DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL, 5., 2004, Pirassununga. **Anais...** Pirassununga: Sociedade Brasileira de Melhoramento Animal, 2004. 1 CD-ROM.
- BABIUK, L. A.; PHILLIPS, J. P.; MOO-YOUNG, M. **Animal biotechnology: comprehensive biotechnology, first supplement**. Oxford: Pergamon Press 1989. 213 p.
- BADKE, Y. M.; BATES, R. O.; ERNST, C. E.; SCHWAB, C.; STEIBEL, J. P. Estimation of linkage disequilibrium in four US pig breeds. **BMC Genomics**, Londres, v. 13, p. 24, 2012. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-13-24>>.
- BOHAMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, Londres, v. 11, p. 421-432, 2010. Disponível em: Disponível em: <<http://dx.doi.org/10.1186/1471-2164-11-421>>.
- BROWNING, S. R.; BROWNING, B. L. High-resolution detection of identity by descent in unrelated individuals. **The American Journal of Human Genetics**, Chicago, v. 86, n. 4, p. 526-539, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.ajhg.2010.02.021>>.
- BUSH, W. S.; MOORE, J. H. Chapter 11: Genome-Wide Association Studies. **PLoS Computational Biology**, Cambridge, v. 8, n. 12, p. e1002822, 2012. Disponível em: <<http://dx.doi.org/10.1371/journal.pcbi.1002822>>.
- CALUS, M. P. L. Genomic breeding value prediction: methods and procedures. **Animal**, Cambridge, v. 4, n. 2, p. 157-164, 2009. Disponível em: <<http://dx.doi.org/10.1017/S1751731109991352>>.

- CALUS, M. P. L.; MEUWISSEN, T. H. E.; DE ROOS, A. P. W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, Bethesda, v. 178, n. 1, p. 553-561, 2008. Disponível em: <<http://dx.doi.org/10.1534/genetics.107.080838>>.
- CANTOR, R. M.; LANGE, K.; SINSHEIMER, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. **The American Journal of Human Genetics**, Chicago, v. 86, n. 1, p. 6–22, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.ajhg.2009.11.017>>.
- CLEVELAND, M.; DEEB, N. Evaluation of a genome-wide approach to multiple marker association considering different marker densities. **BMC Proceedings**, Londres, v. 3, suppl. 1, p. s5, 2009.
- COLE, J. B.; WIGGANS, G. R.; MA, L.; SONSTEGARD, T. S.; LAWLOR, T. J.; CROOKER, B. A.; VAN TASSELL, C. P.; YANG, J.; WANG, S.; MATUKUMALLI, L. K.; DA, Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics**, Londres, v. 12, 2011. Disponível em: <<http://www.biomedcentral.com/1471-2164/12/408>>. Acesso em: 20 ago. 2015.
- DAETWYLER H. D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS J. A. Inbreeding in genome-wide selection. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 369-376, 2007. Disponível em: <<http://dx.doi.org/10.1111/j.1439-0388.2007.00693.x>>.
- DASHAB, G. R.; KADRI, N. K.; SHARIATI, M. M.; SAHANA, G. Comparison of linear mixed model analysis and genealogy-based haplotype clustering with a Bayesian approach for association mapping in a pedigreed population. **BMC Proceedings**, Londres, v. 6, suppl 2, p. s4, 2012. Disponível em: <<http://dx.doi.org/10.1186/1753-6561-6-S2-S4>>.

- DE LOS CAMPOS, G.; NAYA, H.; GIANOLA, D.; CROSSA, J.; LEGARRA, A.; MANFREDI, E.; WEIGEL, K.; COTES, J. M. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, Bethesda, v. 182, n. 1, p. 375-385, 2009. Disponível em: <<http://dx.doi.org/10.1534/genetics.109.101501>>.
- DE ROOS, A. P. W.; HAYES, B. J.; SPELMAN, R. J.; GODDARD, M. E. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. **Genetics**, Bethesda, v. 179, n. 3, p. 1503-1512, 2008. Disponível em: <<http://dx.doi.org/10.1534/genetics.107.084301>>.
- DEKKERS, J. C. M. Prediction of response to marker assisted and genomic selection using selection index theory. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 331-341, 2007.
- DEKKERS, J. C. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, Champaign, v. 82, suppl., p. s313–s328, 2004.
- DEKKERS, J. C. M.; VAN ARENDONK, J. A. M. Optimum selection for quantitative traits with information on an identified locus in outbred populations. **Genetical Research**, Cambridge, v. 71, p. 257-275, 1998.
- DEVLIN, B.; ROEDER, K. Genomic control for association studies. **Biometrics**, Arlington, v. 55, n. 4, p. 997-1004, 1999.
- EL FARO, L.; ALBUQUERQUE, L. G. Estimação de parâmetros genéticos para produção de leite no dia do controle e produção acumulada até 305 dias, para as primeiras lactações de vacas da raça Caracu. **Revista Brasileira de Zootecnia**, Viçosa, v. 32, p. 284-264, 2003. Disponível em: <<http://dx.doi.org/10.1590/S1516-35982003000200006>>.
- FALCONER, D. S.; MACKAY, T. C. **Introduction to quantitative genetics**. 4th. ed. Edinburgh Gate: Benjamin Cummings, 1997. 450 p.

- FARNIR, F.; COPPIETERS, W.; ARRANZ, J. J.; BERZI, P.; CAMBISANO, N.; GRISART, B.; KARIM, L.; MARCQ, F.; MOREAU, L.; MNI, M.; NEZER, C.; SIMON, P.; VANMANSHOVEN, P.; WAGENAAR, D.; GEORGES, M. Extensive genome-wide linkage disequilibrium in cattle. **Genome Research**, Nova York, v. 10, n. 2, p. 220-227, 2000.
- FLINT, J.; ESKIN, E. Genome-wide association studies in mice. **Nature Reviews Genetics**, New York, v. 13, p. 807-817, 2012. Disponível em: <<http://dx.doi.org/10.1038/nrg3335>>.
- GARRICK, D. J.; TAYLOR, J. F.; FERNANDO R. L. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetics Selection Evolution**, Londres, v. 41, 2009. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-41-55>>.
- GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, Dordrecht, v. 136, n. 2, p. 245–257, 2009. Disponível em: <<http://dx.doi.org/10.1007/s10709-008-9308-0>>.
- GODDARD, M. E.; HAYES, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, New York, v. 10, p. 381–391, 2009. Disponível em: <<http://dx.doi.org/10.1038/nrg2575>>.
- GRIFFITH, A. J. F.; WESSLER, S. R.; LEWONTIN, R. C.; CARROLL, S. B. **Introduction to genetic analysis**. 9th. ed. Nova York: W. H Freeman, 2007. 800 p.
- Grisart, B.; Farnir, F.; Karim, L.; Cambisano, N.; Kim, J. J.; KvasZ, A.; MNI, M.; SIMON, P.; FRÈRE, J. M.; COPPIETERS, W.; GEORGES, M. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. **Proceedings of the National Academy of Sciences of the United State of America**, Washington, v. 101, n. 8, p. 2398–2403, 2004. Disponível em: <<http://dx.doi.org/10.1073/pnas.0308518100>>.

- GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. **Plant Molecular Biology**, Glasgow, v. 57, p. 461–485, 2005. Disponível em: <<http://dx.doi.org/10.1007/s11103-005-0257-z>>.
- HARDY, J.; SINGLETON, A. Genomewide association studies and human disease. **The New England Journal of Medicine**, Boston, v. 360, p. 1759–1768, 2009. Disponível em: <<http://dx.doi.org/10.1056/NEJMra0808700>>.
- HARRIS, B. L.; JOHNSON, D. L. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. **Journal of Dairy Science**, Champaign, v. 93, n. 3, p. 1243–1252, 2010. Disponível em: <<http://dx.doi.org/10.3168/jds.2009-2619>>.
- HAYES, B. J.; VISSCHER, P. M.; MCPARTLAN, H. C.; GODDARD, M. E. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. **Genome Research**, Nova York, v. 13, n. 4, p. 635–643, 2003.
- HAYES, B. J.; LIEN, S.; NILSEN, H.; OLSEN, H. G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISWSEN, T. E. The origin of selection signatures on bovine chromosome 6. **Animal Genetics**, Oxford, v. 39, n. 2, p. 105-116, 2008. Disponível em: <<http://dx.doi.org/10.1111/j.1365-2052.2007.01683.x>>.
- HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, p. 433-443, 2009. Disponível em: <<http://dx.doi.org/10.3168/jds.2008-1646>>.
- HAZEL, L. N.; LUSH J. L. The efficiency of three methods of selection. **Journal of Heredity**, Oxford, v. 33, p. 393-399, 1943.
- HENDERSON, C. R. Estimation of changes in herd environment. **Journal of Dairy Science**, Champaign, v. 32, p. 709, 1949.

- HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Berlin, v. 38, p. 226-231, 1968.
- JAMROZIK, J.; SCHAEFFER, L. R. Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. **Journal of Dairy Science**, Champaign, v. 80, p. 762-770, 1997.
- JAMROZIK, J.; SCHAEFFER, L. R.; DEKKERS, J. C. M. Genetic evaluation of dairy cattle using test day yields and random regression model. **Journal of Dairy Science**, Champaign, v. 80, p. 1217-1226, 1997.
- JIANG, L.; LIU, J.; SUN, D.; MA, P.; DING, X.; YU, Y.; ZHANG, Q. Genome wide association studies for milk production traits in chinese holstein population. **Plos One**, São Francisco, v. 5, 2010. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0013661>>.
- JIMÉNEZ-MONTERO, J. A.; GONZÁLEZ-RECIO, O.; ALENDIA, R. Genotyping strategies for genomic selection in small dairy cattle populations. **Animal**, Cambridge, v. 6, n. 8, p. 1216-1224, 2011. Disponível em: <<http://dx.doi.org/10.1017/S1751731112000341>>.
- KASHI, Y.; HALLERMAN, E.; SOLLER, M. Marker assisted selection of candidate bulls for progeny testing programmes. **Animal Production**, Cambridge, v. 51, n. 1, p. 63-74, 1990. Disponível em: <<http://dx.doi.org/10.1017/S0003356100005158>>.
- KHATKAR, M. S.; ZENGER, K. R.; HOBBS, M.; HAWKEN, R. J.; CAVANAGH, J. A. L.; BARRIS, W.; MCCLINTOCK, A. E.; MCCLINTOCK, S.; THOMSON, P.C.; TIER, B.; NICHOLAS, F. W.; RAADSMA, H. W. A primary assembly of a bovine haplotype block map based on a 15,036 single nucleotide polymorphism panel genotyped in Holstein Friesian cattle. **Genetics**, Bethesda, v. 176, p. 763-772, 2007. Disponível em: <<http://dx.doi.org/10.1534/genetics.106.069369>>.

- KHATKAR, M. S.; NICHOLAS, F. W.; COLLINS, A. R.; ZENGER, K. R.; CAVANAGH, J. A.; BARRIS, W.; SCHNABEL, R. D.; TAYLOR, J. F.; RAADSMA, H. W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics**, Londres, v. 9, 2008. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-9-187>>.
- KIZILKAYA, K.; FERNANDO, R. L.; GARRICK, D. J. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. **Journal of Animal Science**, Champaign, v. 88, n. 2, p. 544–551, 2010. Disponível em: <<http://dx.doi.org/10.2527/jas.2009-2064>>.
- KOIVULA, M.; STRANDÉN, I.; PÖSÖ, J.; AAMAND, G. P.; MÄNTYSAARI, E. A. Single step genomic evaluations for the nordic red dairy cattle test day data. **Interbull Bulletin**, Upsala, v. 46, p. 28-31, 2012.
- LEGARRA, A.; ROBERT-GRANIÉ, C.; CROISEAU, P.; GUILLAUME, F.; FRITZ S. Improved lasso for genomic selection. **Genetics Research**, Cambridge, v. 93, n. 1, p. 77–87, 2011. Disponível em: <<http://dx.doi.org/10.1017/S0016672310000534>>.
- LEWONTIN, R. C. On measures of gametic disequilibrium. **Genetics**, Bethesda, v. 120, n. 3, p. 849-852, 1988.
- LOBERG, A.; DÜRR, J. W. Interbull survey on the use of genomic information. In: THE INTERBULL TECHNICAL WORKSHOP, 39., 2009, Uppsala. **Proceedings...** p. 3-14.
- LÔBO, R. B.; BEZERRA, L. A. F.; OLIVEIRA, H. N.; PENNA, V. M.; BERGMANN, J. A. G.; LEITE, J. F.; MAGNABOSCO, C. U. **Sumário de touros das raças Nelore, Guzerá, Brahman e Tabapuã**. Ribeirão Preto: ANCP, 2014.
- LUND, M. S.; DE ROOS, S. P. W.; DE VRIES, A. G.; DRUET, T.; DUCROCQ, V.; FRITZ, S.; GUILLAUME, F.; GULDBRANDTSEN, B.; LIU, Z.; REENTS, R. A common reference population from four European Holstein

populations increases reliability of genomic predictions. **Genetics Selection Evolution**, Londres, v. 43, n. 43, 2011. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-43-43>>.

MACKINNON, M. J.; GEORGES, M. A. J. Marker-assisted preselection of young dairy sires prior to progeny-testing. **Livestock Production Science**, Amsterdam, v. 54, n. 3, p. 229-250, 1998. Disponível em: <[http://dx.doi.org/10.1016/S0301-6226\(97\)00169-3](http://dx.doi.org/10.1016/S0301-6226(97)00169-3)>.

MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; WOUTER, C. W.; CREWS, D.; DIAS NETO, E.; GILL, C. A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J. F.; STEPHEN, S.; MOORE, S. S. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, Londres, v. 8, p. 74, 2007. Disponível em: <<http://dx.doi.org/10.1186/1471-2156-8-74>>.

MELO, V. J. M; PENNA, V. M. P. Programas de melhoramento da raça Guzerá. In: SIMPÓSIO DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL, 3., 2000, Belo Horizonte. **Anais...** Belo Horizonte: Sociedade Brasileira de Melhoramento Animal, 2000. 1 CD-ROM.

MEUWISSEN, T. H. E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 321-322, 2007. Disponível em: <<http://dx.doi.org/10.1111/j.1439-0388.2007.00708.x>>.

MEUWISSEN, T. H. E.; GODDARD, M. E. The use of marker haplotypes in animal breeding schemes. **Genetics Selection Evolution**, Londres, v. 28, n. 2, p. 161-176, 1996. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-28-2-161>>.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819–1829, 2001.

- MUIR, W. M. Comparison of genomic and traditional BLUP – estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 342-355, 2007.
- MUIR, B.; DOORMAAL, B. V.; KISTEMAKER, G. International genomic cooperation –North American perspective. **Interbull Bull**, Uppsala, v. 41, p. 71–76, 2010.
- PEARSON, T.; MANOLIO T. How to interpret a genome-wide association study. **JAMA**, Chicago, v. 299, n. 11, p. 1335–1344, 2008. Disponível em: <<http://dx.doi.org/10.1001/jama.299.11.1335>>.
- PEIXOTO, M. G. C. D.; SANTOS, G. G. dos; BRUNELI, F. A. T.; PENNA, V. M.; VERNEQUE, R. S.; PANETTO, J. C. C.; MACHADO, C. H. C.; MACHADO, M. A.; LOBO, R. B.; CARVALHO, M. R. S. **Programa Nacional do Guzerá para leite: resultados do teste de progênie, do programa de melhoramento genético de zebuínos da ABCZ e do núcleo MOET**. Juiz de Fora: Embrapa Gado de Leite, 2014. 72 p. (Documentos, 168).
- PETERS, S. O.; KIZILKAYA, K.; GARRICK, D. J.; FERNANDO, R. L.; REECY, J. M.; WEABER, R. L.; SILVER, G. A.; THOMAS, M. G. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. **Journal of Dairy Science**, Champaign, v. 91, n. 2, p. 605-612, 2013. Disponível em: <<http://dx.doi.org/10.2527/jas.2012-5580>>.
- PRICE, A. L.; PATTERSON, N. J.; PLENGE, R. M.; WEINBLATT, M. E.; SHADICK, N. A.; REICH, D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, New York, v. 38, p. 904-909, 2006. Disponível em: <<http://dx.doi.org/10.1038/ng1847>>.

- PSZCZOLA, M.; MULDER, H. A.; CALUS, M. P. L. Effect of enlarging the reference population with (un)genotyped animals on the accuracy of genomic selection in dairy cattle. **Journal of Dairy Science**, Champaign, v. 94, n. 1, p. 431-441, 2011.
- QANBARI, S.; PIMENTEL, E. C. G.; TETENS J.; THALLER, G.; LICHTNER, P.; SHARIFI, A. R.; SIMIANER, H. The pattern of linkage disequilibrium in German Holstein cattle. **Animal Genetics**, Oxford, v. 41, n. 4, p. 346–356, 2010. Disponível em: <<http://dx.doi.org/10.1111/j.1365-2052.2009.02011.x>>.
- ROMUALDI, C.; BALDING, D.; NASIDZE, I.S.; RISCH, G.; ROBICHAUX, M.; SHERRY, S. T.; STONEKING, M.; BATZER, M. A.; BARBUJANI, G. Patterns of human diversity, within and amongc, inferred from biallelic DNA polymorphisms. **Genome Research**, Nova York, v. 12, n. 4, p. 602-612, 2002.
- SAHANA, G.; GULDBRANDTSEN, B.; JANSS, L.; LUND, M. S. Comparison of association mapping methods in a complex pedigreed population. **Genetic Epidemiology**, New York, v. 34, n. 5, p. 455–462, 2010. Disponível em: <<http://dx.doi.org/10.1002/gepi.20499>>.
- SANTOS, D. J. A.; UTSUNOMIYA, A. T. H.; TONHATI, H.; PEIXOTO, M. G. C. D.; PANETTO, J. C. C.; SILVA, M. V. G. B. Desequilíbrio de ligação em diferentes raças zebuínas e em uma população F2 Gir X Holandês. In: REUNIÓN DE LA ALPA, 23., 2013, Havana, Cuba. **Anais...Havana: Associação Latina Americana de Produção Animal**, 2013a. 1 CD-ROM.
- SANTOS, D. J. A.; PEIXOTO, M. G. C. D.; BORQUIS, R. R. A.; VERNEQUE, R. S.; PANETTO, J. C. C.; TONHATI H. Genetic parameters for test-day milk yield, 305-day milk yield, and lactation length in Guzarat cows. **Livestock Science**, Amsterdam, v. 152, n. 2-3, p. 114-119, 2013b. Disponível em: <<http://dx.doi.org/10.1016/j.livsci.2012.12.012>>.

SARGOLZAEI, M.; SCHENKEL, F. S.; JANSEN, G. B.; SCHAEFFER, L. R.
Extent of linkage disequilibrium in Holstein cattle in North America.
Journal of Dairy Science, Champaign, v. 91, n. 5, p. 2106-2117, 2008.
Disponível em: <<http://dx.doi.org/10.3168/jds.2007-0553>>.

SAXENA, R.; VOIGHT, B. F.; LYSSENKO, V.; BURTT, N. P.; BAKKER, P. I.;
CHEN, H.; ROIX, J. J.; KATHIRESAN, S.; HIRSCHHORN, J. N.; DALY,
M. J.; HUGHES, T. E.; GROOP, L.; ALTSHULER, D.; ALMGREN,
P.; FLOREZ, J. C.; MEYER, J.; ARDLIE, K.; BENGTSSON BOSTRÖM,
K.; ISOMAA, B.; LETTRE, G.; LINDBLAD, U.; LYON, H. N.; MELANDER,
O.; NEWTON-CHEH, C.; NILSSON, P.; ORHO-MELANDER,
M.; RÅSTAM, L.; SPELIOTES, E. K.; TASKINEN, M. R.; TUOMI,
T.; GUIDUCCI, C.; BERGLUND, A.; CARLSON, J.; GIANNINY,
L.; HACKETT, R.; HALL, L.; HOLMKVIST, J.; LAURILA, E.; SJÖGREN,
M.; STERNER, M.; SURTI, A.; SVENSSON, M.; SVENSSON,
M.; TEWHEY, R.; BLUMENSTIEL, B.; PARKIN, M.; DEFELICE,
M.; BARRY, R.; BRODEUR, W.; CAMARATA, J.; CHIA, N.; FAVA,
M.; GIBBONS, J.; HANDSAKER, B.; HEALY, C.; NGUYEN, K.; GATES,
C.; SOUGNEZ, C.; GAGE, D.; NIZZARI, M.; GABRIEL, S. B.; CHIRN, G.
W.; MA, Q.; PARIKH, H.; RICHARDSON, D.; RICKE, D.; PURCELL, S.
Genome-wide association analysis identifies loci for type 2 diabetes and
triglyceride levels. **Science**, New York, v. 316, n. 5829, p. 1331–1336,
2007.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle.
Journal of Animal Breeding and Genetics, Malden, v. 123, n. 4, p. 218-
223, 2006.

SCHROOTEN, C.; BOVENHUIS, H.; VAN ARENDONK, J. A. M.; BIJMA, P.
Genetic progress in multistage dairy cattle breeding schemes using
genetic markers. **Journal of Dairy Science**, Champaign, v. 88, n. 4, p.
1569-1581, 2005. Disponível em: <[http://dx.doi.org/10.3168/jds.s0022-0302\(05\)72826-5](http://dx.doi.org/10.3168/jds.s0022-0302(05)72826-5)>.

- SHEN, X.; ALAM, M.; FIKSE, F.; RONNEGARD, L. A novel generalized ridge regression method for quantitative genetics. **Genetics**, Bethesda, v. 193, p. 1255-1268, 2013. Disponível em: <<http://dx.doi.org/10.1534/genetics.112.146720>>.
- SIGURDSSON, A.; BANOS, G. Dependent variables in international sire evaluations. **Acta Agriculturae Scandinavica**, Uppsala, v. 45, n. 4, p. 209-217, 1995. Disponível em: <<http://dx.doi.org/10.1080/09064709509413079>>.
- SILVA, C. R.; NEVES, H. H. R.; QUEIROZ, S. A.; SENA, J. A. D.; PIMENTEL, E. C. G. 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, Alemanha. **Anais...** 1 CD-ROM.
- SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Genomic selection using different marker types and densities. **Journal of Animal Science**, Champaign, v. 86, n. 10, p. 2447-2454, 2008. Disponível em: <<http://dx.doi.org/10.2527/jas.2007-0010>>.
- SPANGLER, M. L.; SAPP, R. L.; BERTRAND, J. K.; MAC NEIL, M. D.; REKAYA, R. Different methods of selecting animals for genotyping to maximize the amount of genetic information known in the population. **Journal of Animal Science**, Champaign, v. 86, n. 10, p. 2471-2479, 2008. Disponível em: <<http://dx.doi.org/10.2527/jas.2007-0492>>.
- SU, G.; BRØNDUM, R. F.; MA, P.; GULDBRANDTSEN, B.; AAMAND, G. P.; LUND, M. S. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in nordic holstein and red dairy cattle populations. **Journal of Dairy Science**, Champaign, v. 95, n. 8, p. 4657-4665, 2012. Disponível em: <<http://dx.doi.org/10.3168/jds.2012-5379>>.

- SUN, L. D.; XIAO, F. L.; LI, Y.; ZHOU, W. M.; TANG, H. Y.; TANG, X. F.; ZHANG, H.; SCHAARSCHMIDT, H.; ZUO, X. B.; FOELSTER-HOLST, R.; HE, S. M.; SHI, M.; LIU, Q.; LV, Y. M.; CHEN, X. L.; ZHU, K. J.; GUO, Y. F.; HU, D. Y.; LI, M.; LI, M.; ZHANG, Y. H.; ZHANG, X.; TANG, J. P.; GUO, B. R.; WANG, H.; LIU, Y.; ZOU, X. Y.; ZHOU, F. S.; LIU, X. Y.; CHEN, G.; MA, L.; ZHANG, S. M.; JIANG, A. P.; ZHENG, X. D.; GAO, X. H.; LI, P.; TU, C. X.; YIN, X. Y.; HAN, X. P.; REN, Y. Q.; SONG, S. P.; LU, Z. Y.; ZHANG, X. L.; CUI, Y.; CHANG, J.; GAO, M.; LUO, X. Y.; WANG, P. G.; DAI, X.; SU, W.; LI, H.; SHEN, C. P.; LIU, S. X.; FENG, X. B.; YANG, C. J.; LIN, G. S.; WANG, Z. X.; HUANG, J. Q.; FAN, X.; WANG, Y.; BAO, Y. X.; YANG, S.; LIU, J. J.; FRANKE, A.; WEIDINGER, S.; YAO, Z. R.; ZHANG, X. J. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. **Nature Genetics**, New York, v. 43, n. 7, p. 690–694, 2011. Disponível em: <<http://dx.doi.org/10.1038/ng.851>>.
- TEARE, M. D.; DUNNING, A. M.; DUROCHER, F.; RENNART, G.; EASTON D. F. Sampling distribution of summary linkage disequilibrium measures. **Annals Human Genetics**, Cambridge, v. 66, pt. 3, p. 223–233, 2002.
- TENESA, A.; KNOTT, S. A.; WARD, D.; SMITH, D.; WILLIAMS, J. L.; VISSCHER P. M. Estimation of linkage disequilibrium in a sample of United Kingdom dairy cattle population using unphased genotypes. **Journal of Animal Science**, Champaign, v. 81, n. 3, p. 617–623, 2003.
- TIBSHIRANI, R. Regression selection and shrinkage via the lasso. **Journal of the Royal Staistical Society**, Londres, v. 58, n. 1, p. 267–288, 1994.
- TSURUTA, S.; AGUILAR, I.; MISZTAL, I.; LAWLOR, T. J. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. **Journal of Dairy Science**, Champaign, v. 94, n. 8, p. 4198–4204, 2011. Disponível em: <<http://dx.doi.org/10.3168/jds.2011-4256>>.
- VANRADEN, P. M. Efficient methods to Compute Genomic predictions. **Journal of Dairy Science**, Champaign, v. 91, n. 11, p. 4414–4423, 2008. Disponível em: <<http://dx.doi.org/10.3168/jds.2007-0980>>.

- VANRADEN, P. M.; WIGGANS, G. R. Derivation, calculation and use of national animal model information. **Journal Dairy Science**, Champaign, v. 74, n. 8, p. 2737-2746, 1991.
- VANRADEN, P. M.; VAN TASSEL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, Champaign, v. 92, n. 1, p. 16–24, 2009. Disponível em: <<http://dx.doi.org/10.3168/jds.2008-1514>>.
- VANRADEN, P. M.; OLSON, K. M.; NULL, D. J.; HUTCHISON, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. **Journal of Dairy Science**, Champaign, v. 94, n. 12, p. 6153–6161, 2011. Disponível em: <<http://dx.doi.org/10.3168/jds.2011-4624>>.
- VAN TASSELL, C. P.; SMITH, T. P. L.; MATUKUMALLI, L. K.; TAYLOR, J. F.; SCHNABEL, R. D.; LAWLEY, C. T.; HAUDENSCHILD, C. D.; MOORE, S. S.; WARREN, W. C.; SONSTEGARD, T. S. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nature Methods**, New York, v. 5, n. 3, p. 247-252, 2008. Disponível em: <<http://dx.doi.org/10.1038/nmeth.1185>>.
- WALL, J. D.; PRITCHARD, J. K. Haplotype blocks and linkage disequilibrium in the human genome. **Nature Reviews Genetics**, New York, v. 4, n. 8, p. 587-597, 2003.
- WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, Cambridge, v. 94, n. 2, p. 73-83, 2012. Disponível em: <<http://dx.doi.org/10.1017/S0016672312000274>>.
- WANG, H.; JIANG, L.; LIU, X.; YANG, J.; WEI, J.; XU, J.; ZHANG, Q.; LIU, J. F. A Post-GWAS replication study confirming the *PTK2* gene associated with milk production traits in Chinese Holstein. **PLoS ONE**, São Francisco, v. 8, n. 12, 2013. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0083625>>.

- WEIGEL, K. A.; VAN TASSELL, C. P.; O' CONNEL, J. R.; VANRADEN, P. M.; WIGGANS, G. R. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. **Journal of Dairy Science**, Champaign, v. 93, n. 5, p. 2229-2238, 2010. Disponível em: <<http://dx.doi.org/10.3168/jds.2009-2849>>.
- XIE, L.; LUO, C.; ZHANG, C.; ZHANG, R.; TANG, J.; NIE, Q.; MA, L.; HU, X.; LI, N.; DA, Y.; ZHANG, X. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. **PLoS One**, São Francisco, v. 7, 2012. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0030910>>.
- YAN, J.; SHAH, T.; WARBURTON, M. L.; BUCKLER, E. S.; MCMULLEN, M. D.; CROUCH, J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP Markers. **PLoS ONE**, São Francisco, v. 4, 2009. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0008451>>.
- YI, N.; XU, S. Bayesian LASSO for quantitative trait loci mapping. **Genetics**, Bethesda, v. 179, n. 2, p. 1045-1055, 2008. Disponível em: <<http://dx.doi.org/10.1534/genetics.107.085589>>.
- YU, J.; PRESSOIR, G.; BRIGGS, W. H.; BI, I. V.; YAMASAKI, M.; DOEBLEY, J. F.; MCMULLEN, M. D.; GAUT, B. S.; NIELSEN, D. M.; HOLLAND, J. B.; KRESOVICH, S.; BUCKLER, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, p. 203–208, 2006. Disponível em: <<http://dx.doi.org/10.1038/ng1702>>.

CAPÍTULO 2 - ESTIMAÇÃO DO DESEQUILÍBRIO DE LIGAÇÃO E TAMANHO EFETIVO POPULACIONAL EM BOVINOS DA RAÇA GUZERÁ

RESUMO - O desequilíbrio de ligação (DL) é a base para a maioria dos estudos genômicos, pois representam o DL entre um marcador e um possível QTL. Desta forma, objetivou-se com este estudo avaliar e caracterizar o desequilíbrio de ligação do painel de 50 k da Illumina® e o tamanho efetivo da população de Guzerá que será usada nos estudos de associação e seleção genômica. Para este estudo, foram usados 50 touros e 973 fêmeas da raça Guzerá pertencentes ao programa de melhoramento do Guzerá para Leite, sob coordenação da Embrapa gado de leite em parceria com o Centro Brasileiro de Melhoramento do Guzerá e a ABCZ. O DL foi avaliado utilizando a estatística r^2 para todos os pares de marcadores em cada cromossomo. Para estimar o tamanho efetivo, a esperança do r^2 ao longo da distância física foi obtida como a média do r^2 em intervalos regulares de 100 Kb, para distâncias acima de 100 Kb, e intervalos de 5Kb para distâncias abaixo de 100kb. De modo geral, a utilização de um painel com menor densidade para as raças Guzerá proporcionaria cobertura inadequada para estudos de associação e seleção genômica. Os resultados obtidos indicam que a densidade do painel utilizado foi suficiente para proporcionar DL entre os segmentos cromossômicos e são adequados para a estimação de valores genéticos para curtas distâncias (r^2). O tamanho efetivo atual foi de 137 e apresentou tendência de ser reduzido com o decorrer das gerações, indicando aumento da intensidade de seleção para a raça nas últimas gerações. O baixo tamanho efetivo observado para as gerações recentes traz implicações da consideração da endogamia nas decisões dos acasalamentos para manter a diversidade genética da raça.

Palavras chaves: Genômica, Seleção genômica, Zebu

INTRODUÇÃO

A recente disponibilidade de genotipagem em larga escala, com milhares de informações de marcadores SNP, determinou uma nova tendência nas análises genéticas modernas da relação entre genótipo e fenótipo e em suas aplicações no melhoramento genético animal. As principais implementações e inovações são as análises de associação global (COLE et al., 2011; JIANG et al., 2010) e a seleção genômica ampla (MEUWISSEN; HAYES; GODDARD, 2001). Estas duas técnicas exploram o desequilíbrio de ligação (DL) entre os marcadores, pois assumem que o DL entre os segmentos cromossômicos analisados representa também o DL entre um marcador e um possível QTL ("Quantitative Trait Locus") (DE ROOS et al., 2008). A medida de DL indica então uma associação não aleatória entre dois *loci* considerados, baseados em suas frequências genótípicas e alélicas (FALCONER; MACKAY, 1997). Desta forma as primeiras abordagens a serem empregadas anteriormente à implementação da seleção genômica (ou de qualquer estudo genômico envolvendo uma plataforma de marcadores) são os estudos visando caracterizar a extensão, a distribuição e o decaimento do DL na população.

A estimativa do DL é influenciada por fatores relacionados às técnicas de estimação e a fatores populacionais. Como fatores técnicos, encontram-se o tamanho e a representividade da amostra, pois quando pequenas podem levar a estimação imprecisa, principalmente à superestimação (YAN et al., 2009); densidade e distribuição de marcadores; o tipo de método utilizado para a reconstrução de haplótipos; rigor de filtragem dos SNPs (limiar das menores frequências alélicas e equilíbrio Hardy-Weinberg); e o uso de apenas haplótipos maternos ou ambos os haplótipos parentais (BOHAMANOVA; SARGOLZAEI; SCHENKEL, 2010). Os fatores populacionais estão relacionados à história e estrutura da população, dentre os quais, vale destacar o tamanho efetivo populacional (N_e), que influencia inversamente o DL (HAYES et al., 2003).

O N_e refere-se ao número de indivíduos aptos à reprodução e que mostrariam a mesma dispersão das frequências alélicas que uma população idealizada sob deriva genética aleatória ou também, a mesma quantidade de endogamia que essa população idealizada (WRIGHT, 1938). Segundo Hayes

et al. (2003), o DL em distâncias genômicas curtas é dependente do tamanho efetivo relacionado à história antiga da população, enquanto o desequilíbrio à longa distância depende do tamanho efetivo relacionado à história recente da população.

Embora o DL e o Ne sejam fatores que influenciam a acurácia das análises de associação global e seleção genômica (HAYES et al, 2009), o padrão de DL ainda não foi exaustivamente caracterizado na raça Guzerá, e as previsões de Ne são limitadas àquelas inferidas a partir dos dados de pedigree (PANETTO et al., 2010; PEIXOTO et al., 2010). Assim, objetivou-se com este estudo avaliar e caracterizar o desequilíbrio de ligação do painel BovineSNP50 da Illumina® e o tamanho efetivo da população de Guzerá para futuros estudos de associação e seleção genômica.

MATERIAL E MÉTODOS

Para este estudo, foram usados 50 touros e 973 fêmeas da raça Guzerá pertencentes ao programa de melhoramento do Guzerá para Leite, sob coordenação da Embrapa Gado de Leite em parceria com o Centro Brasileiro de Melhoramento do Guzerá e a ABCZ (Associação Brasileira dos Criados de Zebu). A genotipagem foi realizada utilizando um painel de 50K (“BovineSNP50 BeadChip”) para as fêmeas e um HD (“BovineHD BeadChip”) da Illumina® para os touros. Entretanto, para padronização das análises, foram utilizadas informações moleculares referentes ao conjunto de marcadores presentes no painel de 50K. O controle de qualidade dos dados genômicos foi realizado com o auxílio do “software” PLINK (Purcell et al. 2007). Para o controle de qualidade das amostras, foram adotados a “call rate” maior que 0,90 e heterozigosidade até mais ou menos 3 desvios-padrão da média. Já como critérios para o controle de qualidade do marcador foram adotados “call rate” maior que 0,98, MAF maior que 2%, “p-value” do teste estatístico do equilíbrio de Hardy-Weindberg com até 10^{-6} , correlação entre marcadores até 0,998, além da eliminação de SNPs coincidentes e com possíveis erros de posicionamento físico em relação ao mapa de referência (UMD v3.1). A quantidade de animais e de SNPs utilizados no final do processo foi 1.005 indivíduos (45 touros e 960 fêmeas) e 25.024 SNPs.

As fases dos haplótipos foram construídas com o “software” fastPHASE (SCHEET; STEPHENS, 2006). O desequilíbrio de ligação (DL) foi avaliado utilizando a estatística r^2 (HILL; ROBERTSON, 1968) para todos os pares de marcadores em cada cromossomo, utilizando o programa estatístico R (R CORE TEAM, 2013). O r^2 foi calculado da seguinte forma:

$$r^2 = \frac{(freq.AB * freq.ab - freq.Ab * freq.aB)^2}{(freq.A * freq.a * freq.B * freq.b)}$$

Em que $freq.A$, $freq.a$, $freq.B$ e $freq.b$ são as frequências dos alelos A, a, B e b, respectivamente, e $freq.AB$, $freq.ab$, $freq.aB$ e $freq.Ab$ são as frequências dos haplótipos Ab, ab, aB e Ab na população, respectivamente.

Foram identificados e removidos possíveis SNPs “misplaced” em relação ao genoma de referência (UMD3.1). Para isso foram analisados os dez maiores valores de r^2 de cada SNP. Níveis críticos de r^2 de acordo com a distância física foram utilizados: valores de r^2 acima de 0,81 até 10Mb; valores de r^2 de 0,40 a 0,80 até 14 Mb; valores de r^2 de 0,20 a 0,39 até 17 Mb; valores de r^2 de 0,10 a 0,19 até 21 Mb, abaixo de 0,20 até 27 Mb. A fim que não fosse eliminado SNPs com posicionamento correto que possui um alto r^2 com um “misplaced” foi considerado o SNP deveria possuir mais de 60% dos r^2 com outros 10 SNPs acima do nível crítico para ser considerado um possível “misplaced”. A retirada dos SNPs se inicia pela totalidade dos dez SNPs até o nível de tolerância (60 %).

A fim de se obter uma melhor verificação acerca da cobertura dos marcadores contidos no painel, foram construídos cenários com distintas densidades de marcadores pela remoção sequencial dos SNPs (100%, 50%, 25% e 10%) e analisados o DL entre os SNPs adjacentes. Também a partir de cenários menores é possível estudar a adequação de um painel customizado para a raça e que poderia ser utilizado em estudos de imputação. Quando 50 % dos SNPs foram mantidos, significa que o segundo SNP na sequência foi removido e mantido o terceiro, quando 25% foram mantidos, significa que foram removidos 2 SNPs a partir do primeiro e mantido o quarto, e assim por diante.

Para estimar o tamanho efetivo, a esperança do r^2 ao longo da distância física foi obtido como a média do r^2 em intervalos regulares de 100 Kb, para

distâncias acima de 100 Kb, e intervalos de 5Kb para distâncias abaixo de 100kb. Uma vez que a esperança do r^2 pode ser dado em função do tamanho efetivo populacional (N_e) e da distância em Morgan (c) como $E(r^2) \approx 1/[1 + 4N_e c]$, o tamanho efetivo populacional à T gerações passadas foi estimado da seguinte forma:

$$N_T = (1/4c)(1/\bar{r}^2 - 1)$$

em que N_T é o tamanho efetivo populacional à T gerações passadas, c é a distância entre os marcadores em morgans, assumindo 1Mb= 1cM, e $T=1/2c$ (Hayes et al., 2003). Pares de marcadores com distâncias $c < 10^{-6}$ (~100bp), isto é, $T > 500.000$, não foram usadas pois a aproximação é válida apenas para c muito maior que a taxa de mutação (~ 10^{-8} por *locus* por geração).

RESULTADOS E DISCUSSÃO

Todas as estimativas do r^2 em função da distância física foram analisadas para cada cromossomo. Foi observado que alguns cromossomos apresentaram alguns SNPs com níveis de r^2 maiores que o esperado em distâncias maiores que 10 Mb. Estes SNPs nesta situação são denominados como “misplaced” segundo Bohamanova, Sargolzaei e Schenkel (2010), pois são referentes a um possível erro de coordenadas pelo UMD v3.1 “assembly”. Estes autores propõem um algoritmo para correção das posições aparentemente equivocadas, porém os mesmos aconselham a não utilização destes SNPs em algumas análises, devido imprecisão do novo posicionamento. Foram identificados e eliminados um total de 201 SNPs em todo o genoma. Após a eliminação dos “misplaced”, as estatísticas de r^2 foram consideradas como adequadas para ser avaliada em função da distância física. Como esperado, foi observada a tendência de diminuição do r^2 quando se aumentou a distância física entre os marcadores. Foi observada a típica curva de declínio acentuado em distâncias até 1 Mb aproximadamente, e declínio suave em distâncias maiores. A maior média de r^2 (0,19) pode ser considerado um valor moderado, e foi observada em distâncias entre marcadores com até 40 Kb.

Tabela 1 - Desequilíbrio de ligação (r^2) entre pares (N) de SNPs separados por diferentes distâncias para todos os autossomos.

Distância Mb	N ¹	Média	Mediana	DP ²	Inf95 ³	Sup95 ³	% $r^2 > 0,15$ ⁴	% $r^2 > 0,20$ ⁴	% $r^2 > 0,3$ ⁴
0-0,04	7608	0,19	0,07	0,26	0,00	0,85	34,54	28,43	21,37
0,04-0,06	5766	0,16	0,05	0,23	0,00	0,75	30,13	24,99	18,33
0,06-0,1	11639	0,13	0,04	0,20	0,00	0,62	25,25	20,14	13,7
0,1-0,2	28032	0,1	0,03	0,17	0,00	0,48	18,93	14,52	9,36
0,2-0,5	82162	0,06	0,02	0,12	0,00	0,28	11,18	7,81	4,48
0,5-1	132931	0,05	0,02	0,09	0,00	0,20	7,55	4,95	2,54
0,5-1	259095	0,04	0,01	0,08	0,00	0,16	5,71	3,65	1,76
1-2	739498	0,03	0,01	0,06	0,00	0,13	3,73	2,28	1,08
2-5	1149128	0,02	0,01	0,05	0,00	0,08	1,75	1,06	0,46
5-10	2038494	0,01	0,00	0,03	0,00	0,05	0,13	0,33	0,08
10-20	4413934	0,01	0,00	0,01	0,00	0,02	0,05	0,03	0,00
20-50	2971037	0,00	0,00	0,01	0,00	0,01	0,00	0,00	0,00
50-100	397864	0,00	0,00	0,01	0,00	0,01	0,00	0,00	0,00

SNP: single-nucleotide polymorphism. ¹N: número de pares de SNPs. ²Desvio padrão.

³Percentil 5% e 95%. ⁴Porcentagem de pares de SNPs com $r^2 > 0,15$, $0,20$ e $0,3$.

Valores de r^2 considerados como útil para ser empregados em análises genômicas divergem na literatura, podendo ser maiores que 0,20 em estudos de simulação (MEUWISSEN; HAYES; GODDARD, 2001) e maiores que 0,15, 0,20, 0,25 e 0,30 em estudos com dados reais (MCKAY et al., 2007; BOHAMANOVA; SALGOZAEI; SCHENKEL, 2010; QANBARI et al., 2010). McKay et al., (2007) sugeriram que com um SNP a cada 100 kb seriam possível obter uma média de r^2 variando de 0,15 à 0,20 entre os marcadores, e isto representaria uma correlação suficiente entre os marcadores e um QTL ($r^2 \approx 0,3$). Desta forma, para distâncias de 0 a 40 kb e 40 kb à 60 kb, 35% e 30% dos marcadores apresentaram um limiar de r^2 maior que 0,15 (Tabela 1). Para estas mesmas distâncias 28% e 25% dos marcadores apresentaram limiar de r^2 maior que 0,20, e 21% e 18% dos marcadores apresentaram r^2 maior que 0,3, respectivamente. Sendo assim, marcadores com estas distâncias podem propiciar níveis de DL adequados para análises genômicas. No entanto, Bohamanova, Salgozaei e Schenkel (2010) trabalhando com uma população da raça Holandesa, encontraram proporções maiores que neste trabalho (30% e 23% dos SNPs com $r^2 > 0,3$, para as mesmas distâncias).

A distância de 100 kb foi adotada como padrão de comparação devido a maioria dos estudos utilizarem esta distância para a discussão dos níveis de r^2 encontrados. Desta forma, foi observada neste estudo uma média r^2 igual à

0,16. Este resultado está em concordância com as médias estimadas por McKay et al., (2007) para seis raças taurinas e duas zebuínas (0,15 a 0,20) e por de Roos et al. (2008) para a raça Holandesa, Jersey e Angus ($r^2 \leq 0,15$). Resultados maiores que o encontrado no presente estudo foram reportados por Khatkar et al. (2008) ($>0,3$), Sargolzaei et al. 2008 ($> 0,50$), Qanbari et al., (2010) (0,21) e Bohamanova, Salgozaei e Schenkel (2010) (0,22), para a raça Holandesa e por Silva et al. (2010), estudando 25 touros da raça Gir (0,21). Já valores menores foram encontrados por Espigolan et al. (2013) para a raça Nelore (0,11). Em geral os zebuínos apresentam um menor DL que os taurinos, possivelmente devido a menor intensidade de seleção para características de interesse econômico. A variação na extensão das estimativas de DL observada entre a população estudada da raça Guzerá e as demais citadas, pode ser atribuída, principalmente, às diferenças no tamanho efetivo das populações e as metodologias de estimação do r^2 .

A partir cenários com distintas densidades de marcadores pela remoção sequencial dos SNPs foi possível estudar a adequação de um painel customizado para a raça (Tabela 2). Pelos resultados, foi possível observar que os valores médios de r^2 tiveram uma considerável diminuição, quando apenas 50%, 25% e 10% marcadores adjacentes foram analisados ($r^2 \leq 0,10$). A porcentagem de r^2 útil também declinou principalmente quando foram considerados $r^2 > 0,2$ com 50% e 25 % dos SNPs e $r^2 > 0,3$ com 100% e 50% dos SNPs mantidos nas análises. Os resultados indicaram também que a utilização de um painel com menor densidade para a população estudada proporcionaria cobertura inadequada para estudos de associação e para seleção genômica. A média de todos os SNPs adjacentes foi r^2 de 0,14. Este resultado é satisfatório com um painel de 50k para populações zebuínas, considerando que Espigolan et al. (2013) encontraram 0,17 usando um painel de alta densidade (770 mil SNPs) em uma população da raça Nelore.

Tabela 2 - Média de r^2 entre pares de marcadores SNPs adjacentes para diferentes densidades de marcadores.

% SNP Mantido	Nº marcadores	Média	% $r^2 > 0,2$	r^2 % $> 0,3$
100	25.024	0,14	22	15
50	12.512	0,10	15	9
25	6.256	0,06	8	5
10	2.503	0,04	5	2

Por meio da Figura 3, é possível observar que o tamanho efetivo populacional da raça era próximo à 20.000 há 10.000 gerações passadas, ou seja, mais recente que a divergência de *Bos primigenius taurus* e *Bos primigenius indicus*, processo que ocorreu há mais de 100.000 gerações atrás (MACHUGH et al. 1997). Depois da domesticação, há aproximadamente 10.000 anos, o tamanho efetivo foi reduzido para alguns milhares (~15 mil), indicando a influência da seleção artificial na estrutura populacional e a maior intensidade de seleção com o avançar das gerações. Resultado semelhante foi também encontrado por de Roos et al., (2008) trabalhando com raças taurinas leiteiras. Há 1000 gerações, o tamanho efetivo populacional era maior que 2.000 e tendeu a ser reduzido para valores entre mil e dois mil. Desde o início da entrada maciça dos zebuínos no Brasil, aproximadamente 15 gerações atrás, o tamanho efetivo populacional era próximo à 300 e tendeu a acentuar o seu declínio. Segundo de Roos et al. (2008) as estimativas de tamanho efetivo não são muito acuradas quando se assume que o tamanho efetivo populacional é constante para todas as gerações, mas ela é aproximadamente verdadeira se o tamanho efetivo populacional é modificado linearmente com o tempo. Entretanto, vários fatores podem influenciar a extensão do DL, de modo que as estimativas devem ser consideradas como aproximadas.

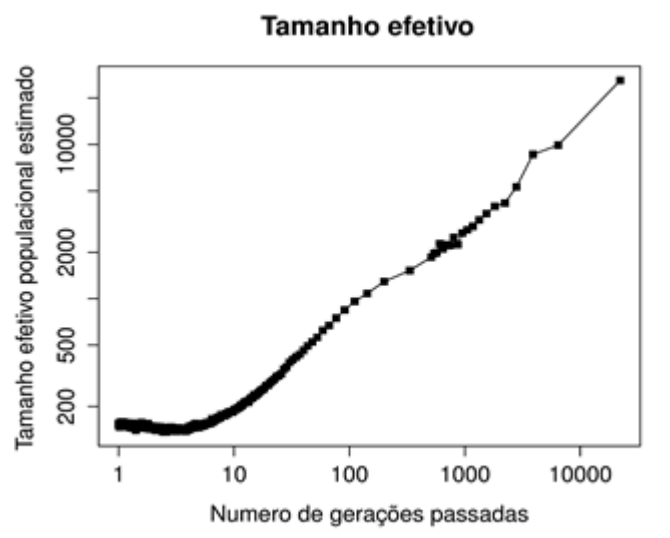


Figura 3 - Tamanho efetivo populacional ao longo da história da população, estimado da média de r^2 para diferentes marcadores para diferentes raças zebuínas.

Pelas estimativas, o tamanho efetivo populacional da raça na geração atual é de aproximadamente de 137. Este valor apesar de baixo, está de acordo com os valores referencias da FAO (1998), que recomenda um tamanho efetivo de no mínimo 50 para que a taxa de endogamia da população seja mantida abaixo do nível de um por cento por geração. Valores menores que este foram observados por De Roos et al., (2008) utilizando a mesma metodologia para estudar raças taurinas, que reportaram valores variando de 32 (Holandês) à 135 (Jersey). No entanto, valores semelhantes foram obtidos para esta mesma população por Faria et al. (2002) e Faria et al. (2009) utilizando as informações do pedigree, que encontraram valores de 117 e 104, respectivamente. Foi observado uma leve tendência de aumento do tamanho efetivo a partir da sétima geração até a primeira. Tendência semelhante foi observada por Peixoto et al. (2010), utilizando as informações de pedigree desta população que observaram uma oscilação para 5 gerações com leve tendência ao aumento nas gerações atuais. Estes autores encontram valores para o tamanho efetivo popuacional de 384, 49, 61, 131 e 98, para as cinco últimas gerações, resultando uma média de 144. Vale ressaltar que os dados de genotipagens são de animais oriundos desde a quarta geração até atual, o que pode ter contribuído para uma única estimativa para as últimas gerações desta população. Segundo Peixoto et al. (2010), o baixo tamanho efetivo para

esta mesma população pode ser consequência do uso intenso de poucos touros (geneticamente superiores) na reprodução.

O baixo tamanho efetivo observado para a raça Guzerá revela a importância da consideração da endogamia nas decisões dos acasalamentos nos programas de melhoramento genético para manter a diversidade genética desta população. Desta forma, valem destacar como alternativas, as avaliações genômicas, cujas informações para orientação de acasalamento possibilitariam um menor incremento da endogamia que as informações obtidas com a avaliação tradicional quantitativa, uma vez que as primeiras tendem a discriminar os melhores animais (geneticamente superiores), ao invés das melhores famílias (DAETWYLER et al., 2007). Os resultados obtidos indicam que a densidade do painel utilizado pode ser suficiente para capturar o DL entre os SNPs e os QTLs, para possível predição de valores genéticos genômicos, principalmente para distâncias menores que 100 kb (r^2 modelados à alto). No entanto, a utilização de um painel de maior densidade, como o “BovineHD BeadChip” da Illumina® obteria níveis melhores de r^2 ($>0,20$) para a realização de estudos genômicos. Neste sentido, estratégias de imputação com o painel utilizado para um de maior densidade devem ser estudadas.

CONCLUSÃO

A densidade do painel utilizado foi considerada suficiente para proporcionar desequilíbrio de ligação entre os segmentos cromossômicos para estudos de associação ampla e para predição de valores genéticos genômicos.

O tamanho efetivo foi reduzido com o decorrer das gerações. O baixo tamanho efetivo das gerações recentes traz implicações da consideração da endogamia nas decisões dos acasalamentos para manter a diversidade genética da raça.

REFERÊNCIAS BIBLIOGRÁFICAS

BOHAMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, Londres, v. 11, p. 421, 2010. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-11-421>>.

- COLE, J. B.; WIGGANS, G. R.; MA, L.; SONSTEGARD, T. S.; LAWLOR, T. J.; CROOKER, B. A.; VAN TASSELL, C. P.; YANG, J.; WANG, S.; MATUKUMALLI, L. K.; DA, Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics**, Londres, v. 12, p. 408, 2011. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-12-408>>.
- DAETWYLER, H. D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS, J. A. Inbreeding in genome-wide selection. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 369-376, 2007. Disponível em: <<http://dx.doi.org/10.1111/j.1439-0388.2007.00693.x>>.
- DE ROOS, A. P. W.; HAYES, B. J.; SPELMAN, R. J.; GODDARD, M. E. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. **Genetics**, Bethesda, v. 179, n. 3, p. 1503–1512, 2008. Disponível em: <<http://dx.doi.org/10.1534/genetics.107.084301>>.
- ESPIGOLAN, R.; BALDI, F.; BOLIGON, A. A.; SOUZA, F. R. P.; GORDO, D. G. M.; TONUSSI, R. L.; CARDOSO, D. F.; OLIVEIRA, H. N.; TONHATI, H.; SARGOLZAEI, M.; SCHENKEL, F. S.; CARVALHEIRO, R.; FERRO, J. A.; ALBUQUERQUE, G. L. Study of whole genome linkage disequilibrium in Nellore cattle. **BMC Genomics**, Londres, v. 14, p. 305, 2013. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-14-305>>.
- FALCONER, D. S.; MACKAY, T. C. **Introduction to quantitative genetics**. 4th. ed. Edinburgh Gate: Benjamin Cummings, 1997. 450 p.
- FAO. Food and Agriculture Organization. **Título**. 2011. Disponível em: <<http://faostat.fao.org>>. Acesso em: 03 abr. 2015.
- FARIA, F. J. C.; VERCESI FILHO, A. E.; MADALENA, F. E.; JOSAHKIAN, L. A. Pedigree analysis in the Brazilian Zebu Breeds. **Journal of Animal Breeding Genetics**, Malden, v. 126, n. 2, p. 148–153, 2009. Disponível em: <<http://dx.doi.org/10.1111/j.1439-0388.2008.00767.x>>.

- FARIA, F. J. C.; VERCESI FILHO, A. E.; MADALENA, F. E.; JOSAHKIAN, L. A. Pedigree analysis in the Brazilian Zebu Breeds. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6., 2002, Montpellier. **Anais...** 1 CD-ROM.
- HAYES, B. J.; VISSCHER, P. M.; MCPARTLAN, H. C.; GODDARD M. E. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. **Genome Research**, Nova York, v. 13, n. 4, p. 635–643, 2003.
- HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, p. 433-443, 2009. Disponível em: <<http://dx.doi.org/10.3168/jds.2008-1646>>.
- HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Berlin, v. 38, n. 6, p. 226-231, 1968. Disponível em: <<http://dx.doi.org/10.1007/BF01245622>>.
- JIANG, L.; LIU, J.; SUN, D.; MA, P.; DING, X.; YU, Y.; ZHANG, Q. Genome wide association studies for milk production traits in Chinese Holstein population. **Plos One**, São Francisco, v. 5, 2010. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0013661>>.
- KHATKAR, M. S.; NICHOLAS, F. W.; COLLINS, A. R.; ZENGER, K. R.; CAVANAGH, J. A.; BARRIS, W.; SCHNABEL, R. D.; TAYLOR, J. F.; RAADSMA, H. W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high density SNP panel. **BMC Genomics**, Londres, v. 9, p. 187, 2008. Disponível em: Disponível em: <<http://dx.doi.org/10.1186/1471-2164-9-187>>.
- MACHUGH, D. E.; SHRIVER, M. D.; LOFTUS, R. T.; CUNNINGHAM, P.; BRADLEY, D. G. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). **Genetics**, Bethesda, v. 146, n. 3, p. 1071–1086, 1997.

- MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; WOUTER, C. W.; CREWS, D.; DIAS NETO, E.; GILL, C. A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J. F.; STEPHEN, S.; MOORE, S. S. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, Londres, v. 8, p. 74, 2007. Disponível em: <<http://dx.doi.org/10.1186/1471-2156-8-74>>.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819–1829, 2001.
- PANETTO, J. C. C.; GUTIÉRREZ, J. P.; FERRAZ, J. B. S.; CUNHA, D. G.; GOLDEN, B. L. Assessment of inbreeding depression in a Guzerat dairy herd: Effects of individual increase in inbreeding coefficients on production and reproduction. **Journal of Dairy Science**, Champaign, v. 93, n. 10, p. 4902-4912, 2010. Disponível em: <<http://dx.doi.org/10.3168/jds.2010-3197>>.
- PEIXOTO, M. G. C. D.; POGGIAN, C. F.; VERNEQUE, R. S.; EGITO, A. A.; CARVALHO, M. R. S.; PENNA, V. M.; BERGMANN, J. A. G.; VICCINI, L. F.; MACHADO, M. A. Genetic basis and inbreeding in the Brazilian Guzerat (*Bos indicus*) subpopulation selected for milk production. **Livestock Science**, Amsterdam, v. 131, n. 2-3, p. 168–174, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.livsci.2010.03.015>>.
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: a toolset for whole-genome association and population-based linkage analysis. **American Journal of Human Genetics**, Chicago, v. 81, n. 3, p. 559-575, 2007.
- QANBARI, S.; PIMENTEL, E. C. G.; TTENS J.; THALLER, G.; LICHTNER, P.; SHARIFI, A. R.; SIMIANER, H. The pattern of linkage disequilibrium in German Holstein cattle. **Animal Genetics**, Oxford, v. 41, n. 4, p. 346–356, 2010. Disponível em: <<http://dx.doi.org/10.1111/j.1365-2052.2009.02011.x>>.

R CORE TEAM. R: A **Language and environment for statistical computing**.

Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>.

SARGOLZAEI, M.; SCHENKEL, F. S.; JANSEN, G. B.; SCHAEFFER, L. R.

Extent of linkage disequilibrium in Holstein cattle in North America.

Journal of Dairy Science, Champaign, v. 91, n. 5, p. 2106-2117, 2008.

Disponível em: <<http://dx.doi.org/10.3168/jds.2007-0553>>.

SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale

population genotype data: applications to inferring missing genotypes and

haplotypic phase. **American Journal of Human Genetics**, Chicago, v.

78, p. 629-644, 2006.

SILVA, C. R.; NEVES, H. H. R.; QUEIROZ, S. A.; SENA, J. A. D.; PIMENTEL,

E. C. G. 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, Alemanha. **Anais...** 1 CD-ROM.

WRIGHT, S. Size of population and breeding structure in relation to

evolution. **Science**, New York, v. 87, n. 2263, p. 430-2264, 1938.

YAN, J.; SHAH, T.; WARBURTON, M. L.; BUCKLER, E. S.; MCMULLEN, M.

D.; CROUCH, J. Genetic characterization and linkage disequilibrium

estimation of a global maize collection using SNP markers. **PLoS ONE**,

São Francisco, v. 4, n. 12, p. e5481, 2009. Disponível em

<<http://dx.doi.org/10.1371/journal.pone.0008451>>.

CAPÍTULO 3 – SELEÇÃO GENÔMICA PARA PRODUÇÃO DE LEITE, GORDURA E PROTEÍNA EM BOVINOS DA RAÇA GUZERÁ

RESUMO - A seleção genômica é uma promissora ferramenta para incrementar a taxa de ganho genético nos programas de melhoramento genético. Desta maneira, objetivou-se com este estudo avaliar a adequação das avaliações genômicas por meio do estudo das estimativas das acurácias das predições genômicas e da comparação com as predições obtidas pela avaliação tradicional para a produção de leite (PL), gordura (PG) e proteína (PP) dos animais pertencentes ao programa de melhoramento genético da raça Guzerá para leite. Também foi objetivo comparar diferentes metodologias de predição genômicas para estas características nesta população. 909 animais da raça Guzerá foram genotipadas utilizando painel de 50k (“BovineSNP50 BeadChip”) e 777k (“BovineHD BeadChip”) da Illumina®, no entanto, foram utilizadas apenas informações moleculares referentes ao conjunto de marcadores presentes no painel de 50 k. Para o controle de qualidade das amostras foram adotados “call rate” maior que 0,90 e heterozigosidade dentro de ± 3 desvios padrão da média. Já os critérios para qualidade do marcador foram adotados “call rate” superior à 0,98, MAF maior que 2%, *p-value* do equilíbrio de Hardy-Weinberg com até 10^{-6} , correlação entre marcadores até 0,998, além da eliminação de SNPs coincidentes e com possíveis erros de posicionamento físico em relação ao mapa de referência. O número de animais utilizado depois do controle de qualidade foi 903, sendo 45 touros e 858 vacas. O número de SNPs final foi de 25.024. As informações utilizadas referentes às características estudadas, foram os valores genéticos desregressados. Foram empregadas as metodologias GBLUP (uni, tricaracterística e associado a matriz H), BAYESC, BAYESC π , LASSO e dois modelos multicaracterística. Para validação da habilidade de predição dos modelos foi utilizada a metodologia de validação cruzada, considerando a técnica de “k-folder”. Os critérios de comparação entre os modelos foram acurácia empírica, viés das predições genômicas; quadrado médio do erro de predição; média das acurácias estimadas individualmente associadas à da avaliação genômica geral; e um índice de desempenho dos modelos associado ao ajuste global. O impacto das avaliações genômicas foi estudado com a acurácia esperada das predições genômicas e por meio da estimativa da correlação de Spearman entre os valores genéticos genômicos e, entre estes e o valores genéticos obtidos pela avaliação tradicional, bem como através da análise de coincidência das predições destes modelos. As predições genômicas foram adequadas para seleção de animais da raça Guzerá e as acurácias empíricas das predições foram em média de 0,77 para PL, 0,72 para PG e 0,79 para PP. Houve algumas divergências na classificação dos animais pelas avaliações tradicional e genômicas. As metodologias de predição bayesianas, GBLUP e os modelos genômicos multicaracterística são equivalentes para a predição dos valores genômicos dos animais desta população. No entanto, o GBLUP é a melhor opção, considerando um ajuste global.

Palavras chaves: Inferência Bayesiana, SNPs, Zebu.

INTRODUÇÃO

A seleção genômica (SG) é a mais recente e promissora ferramenta para incrementar a taxa de ganho genético na pecuária (WEIGEL et al., 2010). Uma das principais razões para esse incremento, é a diminuição do intervalo de gerações, por aumentar a relação acurácia/tempo das provas genéticas. Além disso, outro benefício pode ser alcançado quando as avaliações genômicas são usadas estrategicamente como critérios para a pré-seleção de animais para a participação no teste progênie, pois pode-se reduzir gastos com animais que teriam baixo desempenho no teste (HAYES et al., 2009; SCHAEFFER, 2006). Entretanto, a acurácia das predições oriundas das avaliações genômicas são afetadas por muitos fatores, entre eles estão: o número de animais no conjunto de dados de referência ; a densidade do painel de marcadores, que determina o nível de desequilíbrio de ligação entre marcadores e QTL (“Quantitative Trait Locus”); a relação entre os animais do conjunto de referência e validação; a relação entre os animais dentro do conjunto de referência; a herdabilidade da característica; a acurácia das observações ou dos pseudo-fenótipos utilizados; o modelo e metodologia estatística utilizada (GODDARD, 2009; GODDARD; HAYES, 2009; HAYES et al., 2009).

As avaliações genômicas são parte importante de muitos programas de melhoramento genético para gado leiteiro (HAYES et al., 2009; VANRADEN et al., 2009; HARRIS; JOHNSON 2010). Em alguns destes programas, os touros antigos que foram provados pelo teste de progênie foram genotipados e essas informações foram compartilhadas entre programas de outros países, como a cooperação norte-americana (MUIR; DOORMAAL; KISTEMAKER 2010) e o projeto EuroGenomics (LUND et al., 2011).

No Brasil, já são utilizadas informações moleculares objetivando a seleção, como a identificação dos animais para o alelo A e B da kappa-Caseína (PEIXOTO et al., 2014). O teste progênie tradicional e o núcleo MOET (múltipla ovulação e transferência de embriões), que acelera o progresso genético pela avaliação precoce dos animais, são considerados no programa de melhoramento da raça para leite (PEIXOTO et al., 2014). Entretanto, a acurácia dos valores genéticos pela avaliação das famílias MOET não é muito alto

quando comparado ao teste de progênie (BABIUK; PHILLIPS; MOO-YOUNG, 1989). A acurácia dos valores genéticos genômicos para touros jovens podem ser mais altos que os valores genéticos obtidos com os modelos mistos, cujo o ganho em acurácia é limitado ao aumento das informações provenientes das famílias. Além disso, as avaliações genômicas tendem a discriminar os melhores animais, ao invés das melhores famílias (DAETWYLER et al., 2007) e predizem valores genéticos diferenciados para irmãos completos que ainda não possuem informações de desempenho próprio ou de suas filhas.

Neste sentido, a SG, ao possibilitar o aumento de acurácia das predições dos valores genéticos por incrementar informações de marcadores moleculares nas avaliações para as características quantitativas, pode auxiliar no progresso genético da raça. Desta maneira, objetivou-se com este estudo avaliar a adequação das avaliações genômicas por meio do estudo das estimativas das acurácias das predições genômicas e da comparação com as predições obtidas pela avaliação tradicional para a produção de leite, gordura e proteína de animais pertencentes ao programa de melhoramento genético da raça Guzerá. Também foi objetivo comparar diferentes metodologias de predição genômicas para estas características nesta população.

MATERIAL E MÉTODOS

Avaliação genético-quantitativa

Foram analisadas informações referentes à produção de leite, proteína e gordura pertencentes a vacas distribuídas em 36 rebanhos participantes do programa de melhoramento genético do Guzerá para Leite, realizado pela Embrapa Gado de Leite em parceria com a ABCZ (Associação Brasileira dos Criadores de Zebu). A partir das produções do dia no controle foram calculadas as produções, acumuladas até 305 dias de leite (PL), gordura (PG) e proteína (PP). Para PL e PG foram consideradas informações até a quarta lactação, com a restrição de que a vaca tivesse informação da primeira lactação. Para PP, foram utilizadas apenas as informações referentes à primeira lactação. A descrição dos dados utilizados está apresentada na Tabela 1.

O grupo de contemporâneos foi definido como rebanho, ano e estação de parto. A estação de parto foi considerada como seca no período de abril a setembro, e como chuvosa no período de outubro a março. Foram

considerados grupos de contemporâneos com mais de quatro animais e registros da característica compreendidos entre mais ou menos três desvios padrões da média do grupo contemporâneo.

Tabela 1 - Número de animais (N), número de registros (NR), número de grupos contemporâneos (GC), média, desvio padrão (DP) e coeficiente de variação para a produção de leite (PL), gordura (PG) e proteína (PP).

Característica	N	NR	GC	Média	DP	CV%
PL	3773	5423	233	2142	1032	48,18
PG	2029	2474	100	112	57	50,89
PP	1627	1627	59	70	29	41,43

Foram utilizados 6039 animais na matriz de parentesco, sendo que 698 destes eram touros, e destes 366 possuíam filhas com registros. Para as análises genéticas para PL foi utilizado um modelo animal de repetibilidade, para PG um modelo bicaracterística de repetibilidade com a PL como âncora, e para PP um modelo animal bicaracterística com PL como âncora.

Os componentes de variância foram estimados com o método de Máxima Verossimilhança Restrita (REML) com o programa Wombat (MEYER et al., 2006). Foram considerados como efeitos fixos os grupos de contemporâneos, e a idade da vaca ao parto (linear e quadrático) como covariável. Como efeito aleatório foi considerado o genético direto e de ambiente permanente (no caso da PL e PG), além do residual. A descrição dos parâmetros genéticos estimados e as estatísticas dos valores genéticos preditos (EBVs) para os animais genotipados estão na Tabela 2.

Tabela 2 - Variância genética aditiva (σ_g^2), residual (σ_e^2), de ambiente permanente (σ_{ap}^2), fenotípica (σ_p^2), herdabilidade (h^2) e média e desvio padrão dos valores genéticos (EBVs) preditos para os animais genotipados para a produção de leite (PL), gordura (PG) e proteína (PP).

Característica	σ_g^2	σ_e^2	σ_{ap}^2	σ_p^2	h^2	EBV
PL	158690	381560	71147	611400	0,26±0,004	7,86±8,81
PG	439	616	508,19	1564	0,28±0,03	14,76±14,20
PP	197	572,79	-	769	0,27±0,02	286±260

Análise genômica

1- *Dados utilizados*

Foram genotipados 909 animais da raça Guzerá utilizando um painel de 50k (“BovineSNP50 BeadChip”) (50 touros) e 777k (“BovineHD BeadChip”) (859 vacas) da Illumina®. Entretanto para padronização da análise foram utilizadas apenas informações moleculares referentes ao conjunto de marcadores presentes no painel de 50 k. O controle de qualidade dos dados genômicos foram realizados com o auxílio do PLINK (Purcell et al., 2007). Para o controle de qualidade das amostras foram adotados “call rate” de 0,90 e heterozigosidade de ± 3 desvios-padrão da média. Já os critérios para qualidade do marcador foram adotados “call rate” superior à 0,98, MAF maior que 2%, *p-value* do equilíbrio de Hardy-Weinberg com até 10^{-6} , correlação entre marcadores até 0,998, além da eliminação de SNPs coincidentes e com possíveis erros de posicionamento físico em relação ao mapa de referência. O número de animais utilizado depois do controle de qualidade foi 903, sendo 45 touros e 858 vacas. O número de SNPs final foi de 25.024. Após isso, foram imputados os genótipos perdidos utilizando o software fastPHASE (SCHEET; STEPHENS, 2006). As informações utilizadas referentes às características estudadas, foram os valores genéticos desregressados (drEBVs) segundo Garrick, Taylor e Fernando (2009). A média e o desvio-padrão dos drEBVs dos animais genotipados para PP, PG e PL foram respectivamente 11 ± 25 , 20 ± 34 e 380 ± 662 .

2 – Análises estatísticas

Foram avaliadas as principais representantes das metodologias encontradas na literatura, desde modelos que consideram homogeneidade de variância para o efeito do marcador, os que consideram heterogeneidade de variância e seleção de covariáveis com fração fixa e móvel, e encolhimento dos efeitos dos marcadores via regressão penalizada. Desta forma foram empregados o GBLUP (SNP-BLUP), BAYESC, BAYESC π e LASSO bayesiano por meio do “software” GS3 (LEGARRA; RICARD; FILANGI, 2015), o que incluíram o efeito dos marcadores e efeito poligênicos, além do residual. Matricialmente o modelo G-BLUP pode ser representado da seguinte forma:

$$y = \mu + Ma + Zu + e \quad (3)$$

em que y é o vetor com os drEBVs; μ é a média populacional; a é o vetor dos efeitos dos marcadores, u é o vetor solução dos efeitos poligênicos; e o vetor de resíduos; e, X , M e Z são as correspondentes matrizes de incidência. Para os efeitos a , u e e foi assumido priores com $N(0, G\sigma_g^2)$, $N(0, A\sigma_u^2)$ e $N(0, D\sigma_e^2)$ respectivamente, em que σ_g^2 , σ_u^2 e σ_e^2 são as variâncias genética aditiva explicada pelos marcadores, poligênica e residual; G é a matriz de parentesco genômico baseado nas frequências alélicas dos marcadores, computada como $G = MM' / 2\sum p_i(1-p_i)$ (VANRADEN, 2008); A é a matriz de parentesco baseado no pedigree e D é uma matriz diagonal com elementos $d_{ij} = (1/w_i)$. A ponderação w_i contabiliza a heterogeneidade de variância devido à diferença de acurácia na avaliação genética. A ponderação foi definida como $w_i = r_i^2 / (1 - r_i^2)$, em que r_i^2 é a “reliability” da avaliação genética (quadrado da acurácia).

Nos métodos BayesC e BayesC π instalou-se uma função de mistura para efeitos dos SNPs de modo que, a partir de uma distribuição binomial, uma fração com probabilidade π dos marcadores não possuiu efeito. Desta forma y , μ , u , e e foram definidos como na equação (3), e a (efeito de marcadores) foi definido como $a_i | \pi, \sigma_i^2 \sim \pi N(0, \sigma_i^2 = 0) + (1 - \pi) N(0, \sigma_i^2)$, em que a variância dos marcadores (σ_i^2) seguiu uma distribuição à *priori* qui-quadrado invertida com v graus de liberdade e parâmetro de escala S . Um valor arbitrário de 4 foi assumido para v , enquanto que os parâmetros de escala foram obtidos a partir de uma distribuição gama ($S \sim \text{Gama}(\alpha, \beta)$). Para o BayesC π foi assumido uma distribuição à *priori* não informativa para π , ou seja, uma distribuição uniforme foi obtida a partir de uma distribuição beta com parâmetros $\alpha = 1$ e $\beta = 1$. Para o aqui denominado BayesC, foi assumido uma distribuição informativa beta ($\alpha = 10^8$, $\beta = 10^{10}$) para π , implicando que este parâmetro seja mantido fixo em torno de 0,99.

Assim como nos modelos de mistura descritos acima, no LASSO apenas as pressuposições para os efeitos dos marcadores se diferem da equação 3. Nesta metodologia é assumida uma distribuição exponencial à *priori* para variâncias dos efeitos dos SNPs. A implementação para melhorar as estimativas da metodologia LASSO, a qual considera as variâncias

diferenciadas para resíduos e para o efeito dos marcadores, denominada de “Improved Lasso” (LEGARRA et al. 2011), foi empregada. Sendo assim a *priori* para efeitos de SNP individuais (a_i) pode ser representada como:

$$\Pr(a_i | \tau_i^2) = N(0, \tau_i^2) \quad \text{e} \quad \Pr(\tau_i^2 | \lambda) = \lambda^2 \exp(-\lambda^2 | \tau_i^2 |)$$

Esta parametrização implica que as variâncias individuais para cada SNP, isto é τ_i^2 , são estimadas condicionalmente a uma regularização do parâmetro λ . Este parâmetro foi estimado considerando uma distribuição à *priori* gama com limites entre 0 e 10^7 .

Além dos modelos univariados, dois outros modelos multivariados foram considerados para avaliação genômica das três características simultaneamente, utilizando-se o software BLUPF90 (MIZTAL et al., 2014). Um destes foi um modelo GBLUP que considerou apenas o efeito dos marcadores (TRI-G) e o outro considerou o efeito de marcadores e o de pedigree conjuntamente na forma da matriz H (TRI-H) (MISZTAL; LEGARRA; AGUILAR, 2009). Para estes modelos não foi considerado a ponderação para os resíduos.

Os parâmetros obtidos com os modelos empregados bem como as razões de variâncias, referentes as proporções de explicação de alguns componentes, também foram estudados e usadas na comparação entre os modelos. Estes foram: a variância do efeito do marcador (σ_a^2); a variância do efeito poligênico (σ_u^2); variância do resíduo (σ_e^2); a variância genética aditiva explicada pelos marcadores (σ_{gm}^2) dado como $\sigma_a^2 \pi \sum p_i q_i$, em que p e q são as frequências alélicas; a variância do drRBV (σ_T^2); a proporção da variância genética aditiva não explicada pelos marcadores (c), dado como $(\sigma_u^2 / (\sigma_{gm}^2 + \sigma_u^2))$; $h^2_M = \sigma_{gm}^2 / (\sigma_T^2)$; e $h^2_G = (\sigma_{gm}^2 + \sigma_u^2) / (\sigma_T^2)$.

3 – Delineamento para validar as predições genômicas

Para validação da habilidade de predição dos modelos foi utilizada a metodologia de validação cruzada, considerando a técnica de “k-folder”. Isto foi feito visando explorar o número reduzido de animais nas populações de treinamento e validação, reflexo da pequena população da raça Guzerá no país. Sendo assim, foram particionados aleatoriamente sete diferentes conjuntos de treinamento e validação com 774 e 129 animais,

respectivamente. O método de validação entre gerações não foi utilizado devido à diferença de acurácia dos valores genéticos entre os grupos de validação (menor magnitude) e treinamento (medida de correlação pouco confiável). Desta forma foi possível observar que as acurácias dos valores genéticos das subpopulações foram de mesma magnitude, e consideradas suficientes (maior que 0,40) para avaliar a habilidade de predição dos modelos (Tabela 3).

Tabela 3 Descrição do número de animais (N), média e quantil 5% (Q5%) e 95% (Q95%) das acurácias dos valores genéticos da população de referência (REF) e validação (VAL), separadas em diferentes partes (k1-7).

Partes	N		Estatística	Proteína		Gordura		Leite	
	REF	VAL		REF	VAL	REF	VAL	REF	VAL
K1	774	129	Média	0,66	0,65	0,70	0,70	0,67	0,67
			Q5%	0,40	0,42	0,43	0,46	0,40	0,42
			Q95%	0,91	0,90	0,93	0,92	0,93	0,91
K2	774	129	Média	0,66	0,65	0,69	0,69	0,67	0,66
			Q5%	0,40	0,51	0,43	0,58	0,40	0,52
			Q95%	0,91	0,84	0,93	0,86	0,93	0,85
K3	774	129	Média	0,65	0,66	0,69	0,70	0,67	0,67
			Q5%	0,40	0,48	0,43	0,51	0,40	0,51
			Q95%	0,91	0,88	0,93	0,90	0,93	0,90
K4	774	129	Média	0,66	0,65	0,69	0,69	0,67	0,66
			Q5%	0,40	0,53	0,43	0,59	0,40	0,53
			Q95%	0,91	0,90	0,93	0,91	0,93	0,91
K5	774	129	Média	0,66	0,65	0,69	0,69	0,67	0,67
			Q5%	0,40	0,55	0,43	0,56	0,40	0,56
			Q95%	0,90	0,91	0,92	0,93	0,93	0,92
K6	774	129	Média	0,65	0,67	0,69	0,71	0,67	0,68
			Q5%	0,40	0,43	0,43	0,47	0,40	0,43
			Q95%	0,91	0,90	0,93	0,92	0,92	0,93
K7	774	129	Média	0,66	0,65	0,69	0,69	0,67	0,67
			Q5%	0,42	0,40	0,46	0,43	0,42	0,40
			Q95%	0,91	0,83	0,93	0,86	0,93	0,85
Total	903	903	Média	0,66		0,69		0,67	
			Q5%	0,40		0,43		0,40	
			Q95%	0,91		0,93		0,93	

Para reduzir o número de processamento para cada população foram inicialmente estimados os componentes de variância para todos os animais (903) utilizando 500.000 iterações, com amostragem a cada 50 e descarte inicial de 50.000. Para os dados à *posteriori* foram realizados os testes de convergência dos parâmetros de cada modelo utilizando o pacote BOA (Bayesian Output Analysis) do programa estatístico R (R CORE TEAM, 2013). Os componentes de variância assim estimados foram utilizados como valor inicial para as 30.000 iterações, com amostragem a cada 50 e descarte inicial

de 3.000 para as sete populações de treinamento geradas para cada característica.

As predições dos valores genômicos (GEBV) foram obtidas a partir dos efeitos dos SNPs utilizando a seguinte fórmula:

$$GEBV = \sum_i^n M_i \hat{a}_i$$

em que n é o número de segmentos de marcadores ao longo do genoma, M_i é uma matriz com os genótipos de cada animal para cada marcador i , e $\hat{a}$ é o vetor dos efeitos de cada marcador i . Os GEBV foram calculados a partir dos marcadores (GEBV_M) e a partir dos marcadores mais o efeito poligênico (GEBV_T). Para o caso dos modelos tri-características, as soluções do GEBV foram obtidas diretamente pelas matrizes de relacionamentos (G e H).

4- Critérios de comparação

Foram considerados cinco critérios relacionados ao delineamento de *cross-validação*, outro relacionado à avaliação genômica geral, e um índice entre estes foi utilizado para avaliar os modelos empregados para a predição do GEBV. São eles:

a-) uma medida de acurácia empírica (r_{IT} e r_{IM}) calculada pela divisão do coeficiente de correlação de Pearson entre o drEBV e o GEBV_M ou GEBV_T pela média das acurácia dos drEBV dos animais da população de validação.

b-) viés das predições genômicas, medido como o coeficiente de regressão linear do drEBV no GEBV_M (b_M) e no GEBV_T (b_T) (OLSON et al., 2011).

c-) quadrado médio do erro de predição entre o drEBV e GEBV_M (MSE_M) ou GEBV_T (MSE_T) dos animais do conjunto de validação.

d-) média das acurácias estimadas individualmente e associadas aos GEBV_M e GEBV_T da avaliação genômica geral (r_{cM} e r_{cT}). Estas acurácias são baseadas na variância do erro de predição obtidos como elementos da matriz inversa das equações do modelo misto (MME), como $\sqrt{1 - PEV_i / \sigma_g^2}$, em que σ_g^2 é a variância genética aditiva e a PEV é a variância do erro de predição. Uma

segunda estimativa de acurácia foi obtida pela média dos PEVs e utilizada na fórmula $\sqrt{1 - PEV_m / \sigma_g^2}$.

Para os modelos tri-características e o GBLUP, a PEV é facilmente obtida com a matriz de parentesco genômico. Entretanto, para modelos com heterogeneidade de variâncias e seleção de marcadores foi adotada uma aproximação via ponderação da matriz G. A ponderação é dada em função de maximizar a matriz G para prever os valores genéticos semelhantes aos estimados pelos modelos que consideram variância heterogênea para os efeitos dos marcadores. Desta forma, existe uma matriz diagonal D tal que $MDa = Ma^*$, em que **a** e **a*** são as soluções para os efeitos dos marcadores obtidas pelo GBLUP e por outro modelo qualquer. A variância de cada parte desta igualdade será $MDD'M' \sigma_a^2 = MIM'\sigma_{a^*}^2$. Desenvolvendo a expressão, chega-se que $DD'=I (\sigma_{ai}^2 / \sigma_a^2)$. Sendo assim a matriz diagonal usada para a nova G ($MDD'M'$) terá seus elementos iguais à $\sigma_{ai}^2 / (\sigma_{gm}^2 / 2\sum p_i q_i)$ que prediz, via matriz G, valores genéticos de outros modelos. Posteriormente, as PEVs destes valores genéticos genômicos foram obtidos pela inversa da MME com esta nova matriz G.

e-) Índice de desempenho dos modelos composto pelos seguintes critérios com seus ponderadores:

$$IND = r_i^{*2} + r_c^{*2} - [MSE^* / 2(\sigma_T^2) + ((1-b^*)^2 / (b^*)^2)]$$

Sendo que r_i^* , r_c^* , MSE^* e b^* estão descritos nos tópicos acima e σ_T^2 é a variância do drEBV. * Refere-se aos $GEBV_M$ ou $GEBV_T$.

5-) Impacto das avaliações genômicas

O impacto das avaliações genômicas foi analisado sob duas perspectivas: das predições dos valores genéticos genômicos e sua acurácia. Sendo assim, as acurácias empíricas (r_i) e estimadas (r_c) foram comparadas com a acurácia esperada (r_e) das predições genômicas, que foi calculada de com a adaptação da fórmula proposta para prever a acurácia do GBLUP (DAETWYLER et al., 2010):

$$r_{ggG} = \sqrt{\frac{N_p h^2}{N_p h^2 + M_e}}$$

em que N_p é o número de indivíduos da população e h^2 é a herdabilidade da característica. O M_e é o número eficaz de segmentos de cromossomos e foi calculado como $M_e = 2N_e * L / \ln(4N_e * L)$ e (M_e), em que foi assumido o valor de 30 Morgan como a estimativa de tamanho do genoma (L) e 137 como a estimativa do tamanho efetivo desta população (N_e). Como o fenótipo utilizado foi o drEBV a herdabilidade foi considerada como a média do quadrado das acurácias drEBV de todos os animais, como sugerido por Neves et al., (2014). No entanto, este valor foi ponderado pela média das acurácias dos drEBV de todos os animais, de modo a obter uma base de similar magnitude numérica à acurácia empírica e estimada ($r_e = \frac{r_{ggG}}{h}$).

O impacto relacionado às predições dos valores genéticos genômicos foi avaliado por meio da estimativa da correlação de Spearman entre os GEBVs obtidos com os modelos genômicos e, entre estes e o valores genéticos obtidos pela avaliação tradicional (BLUP) bem como por meio da análise de coincidência das predições destes modelos.

RESULTADOS E DISCUSSÃO

Parâmetros estimados pelos modelos de avaliação genômica

Diferente das estimativas dos parâmetros obtidas diretamente a partir do fenótipo real, as estimativas dos parâmetros obtidos a partir do drEBVs não são de interpretação biológica fácil, já que esta medida é relativa à uma avaliação genética anterior, ou seja, é dependente da base genética avaliada e da metodologia de avaliação utilizada. No entanto, a comparação dos parâmetros entre os modelos estudados, pode fornecer subsídios sobre o desempenho destes modelos e sobre a viabilidade das avaliações genômicas.

As variâncias de marcador (σ_a^2) estimadas para PP (Tabela 4), apresentaram médias à *posteriori* diferentes entre os modelos, principalmente as obtidas com BayesC, porém todos os modelos apresentaram estimativas semelhantes quando o intervalo de alta densidade (IAD) de 95% foi

considerado. No entanto para PG (Tabela 5) e PL (Tabela 6) as estimativas foram diferentes pelo IAD de 95 %. A tendência do BayesC estimar valores maiores comparado aos outros modelos, pode ser entendida porque apenas 1% dos marcadores foram utilizados para explicar a variância genética das características.

Tabela 4 - Média, limite inferior (Inf5%) e superior (Sup95%) do intervalo de alta densidade da distribuição *a posteriori* das estimativas dos parâmetros para produção de proteína de acordo com cada modelo estudado e para o parâmetro λ do LASSO.

Parâmetro	Estatísticas	GBLUP	BayesC	BayesC π	LASSO	TRI-G	TRI-H
σ^2_a	Média	0,037	2,131	0,053	0,028	0,05	-
	Inf5%	0,024	0,034	0,025	0,016	0,04	-
	Sup95%	0,053	4,055	0,103	0,040	0,07	-
σ^2_u	Média	185	303	207	223	-	-
	Inf5%	65	144	85	140	-	-
	Sup95%	327	497	353	321	-	-
σ^2_e	Média	84	80	83	92	272	287
	Inf5%	57	53	57	71	212	228
	Sup95%	110	107	110	114	336	347
σ^2_{gm}	Média	257	146	228	183	371	-
	Inf5%	166	2	126	104	276	-
	Sup95%	368	277	344	266	476	-
σ^2_T	Média	525	528	519	506	643	650
	Inf5%	443	436	436	439	583	588
	Sup95%	617	628	610	579	710	719
c	Média	0,40	0,66	0,47	0,55	-	-
	Inf5%	0,15	0,38	0,21	0,37	-	-
	Sup95%	0,64	0,99	0,72	0,73	-	-
h^2_M	Média	0,49	0,27	0,44	0,37	0,58	-
	Inf5%	0,31	0,004	0,24	0,23	0,46	-
	Sup95%	0,67	0,51	0,65	0,51	0,69	-
h^2_G	Média	0,84	0,84	0,83	0,81	-	0,56
	Inf5%	0,76	0,76	0,76	0,75	-	0,44
	Sup95%	0,90	0,91	0,90	0,87	-	0,66
λ	Média	-	-	-	78	-	-
	Inf5%	-	-	-	51	-	-
	Sup95%	-	-	-	124	-	-

σ^2_a variância do efeito do marcador; σ^2_u variância do efeito poligênico; σ^2_e variância do resíduo; σ^2_{gm} variância genética aditiva explicada pelos marcadores ($\sigma^2_{a\pi\sum p_i q_i}$); σ^2_T variância do drEBV; c é a proporção da variância genética aditiva não explicada pelos marcadores ($\sigma^2_u / (\sigma^2_{gm} + \sigma^2_u)$); $h^2_M = \sigma^2_{GM} / (\sigma^2_T)$; $h^2_G = (\sigma^2_{gm} + \sigma^2_u) / (\sigma^2_T)$.

Tabela 5 - Média, limite inferior (Inf5%) e superior (Sup95%) do intervalo de alta densidade da distribuição a *posteriori* das estimativas dos parâmetros para produção de gordura de acordo com cada modelo estudado e para o parâmetro λ do LASSO.

Parâmetro	Estatísticas	GBLUP	BayesC	BayesC π	LASSO	TRI-G	TRI-H
σ^2_a	Média	0,072	4,797	0,103	0,059	0,11	-
	Inf5%	0,048	2,198	0,055	0,037	0,08	-
	Sup95%	0,098	7,665	0,190	0,083	0,14	-
σ^2_u	Média	260	415	275	360	-	-
	Inf5%	77	185	99	203	-	-
	Sup95%	508	687	516	550	-	-
σ^2_e	Média	243	242	244	236	515	540
	Inf5%	177	173	178	180	402	436
	Sup95%	306	310	309	295	632	654
σ^2_{gm}	Média	500	327	467	417	740	-
	Inf5%	331	150	298	253	558	-
	Sup95%	680	524	658	579	935	-
σ^2_T	Média	1005	984	987	1009	1255	1269
	Inf5%	871	832	852	883	1139	1147
	Sup95%	1161	1157	1140	1146	1380	1404
c	Média	0,33	0,55	0,36	0,46	-	-
	Inf5%	0,11	0,29	0,15	0,28	-	-
	Sup95%	0,57	0,80	0,60	0,65	-	-
h^2_M	Média	0,50	0,33	0,47	0,40	0,59	-
	Inf5%	0,34	0,16	0,31	0,26	0,48	-
	Sup95%	0,65	0,51	0,63	0,54	0,70	-
h^2_G	Média	0,75	0,75	0,75	0,76	-	0,57
	Inf5%	0,66	0,65	0,65	0,68	-	0,46
	Sup95%	0,84	0,84	0,84	0,84	-	0,67
λ	Média	-	-	-	35	-	-
	Inf5%	-	-	-	24	-	-
	Sup95%	-	-	-	54	-	-

σ^2_a variância do efeito do marcador; σ^2_u variância do efeito poligênico; σ^2_e variância do resíduo; σ^2_{gm} variância genética aditiva explicada pelos marcadores ($\sigma^2_{a\pi\sum p_i q_i}$); σ^2_T variância do drEBV; c é a proporção da variância genética aditiva não explicada pelos marcadores ($\sigma^2_u / (\sigma^2_{gm} + \sigma^2_u)$); $h^2_M = \sigma^2_{gm} / (\sigma^2_T)$; $h^2_G = (\sigma^2_{gm} + \sigma^2_u) / (\sigma^2_T)$.

Tabela 6 - Média, limite inferior (Inf5%) e superior (Sup95%) do intervalo de alta densidade da distribuição *a posteriori* das estimativas dos parâmetros para produção de leite de acordo com cada modelo estudado e para o parâmetro λ do LASSO.

Parâmetro	Estatísticas	GBLUP	BayesC	BayesC π	LASSO	TRI-G	TRI-H
σ^2_a	Média	22	1781	44	21	37	-
	Inf5%	13	813	14	12	28	-
	Sup95%	32	2848	129	31	48	-
σ^2_u	Média	205627	246808	229076	235198	-	-
	Inf5%	99946	151763	120909	135789	-	-
	Sup95%	314804	349991	339529	342736	-	-
σ^2_e	Média	53333	49551	51615	46370	201920	211350
	Inf5%	31532	29279	30815	24995	161095	172500
	Sup95%	74525	70687	73199	67658	245500	253605
σ^2_{gm}	Média	152788	121736	132289	144385	257008	-
	Inf5%	92366	55574	64756	81795	192700	-
	Sup95%	228574	19468	211203	211891	327600	-
σ^2_T	Média	412659	417934	410700	427136	458960	463587
	Inf5%	350162	350834	348338	360720	417300	420800
	Sup95%	482359	492209	480341	498828	504100	511300
c	Média	0,57	0,67	0,62	0,61	-	-
	Inf5%	0,33	0,47	0,37	0,41	-	-
	Sup95%	0,76	0,85	0,82	0,80	-	-
h^2_M	Média	0,37	0,29	0,32	0,34	0,56	-
	Inf5%	0,21	0,14	0,16	0,18	0,45	-
	Sup95%	0,56	0,46	0,53	0,51	0,66	-
h^2_G	Média	0,87	0,88	0,87	0,89	-	0,54
	Inf5%	0,79	0,80	0,80	0,82	-	0,43
	Sup95%	0,93	0,94	0,93	0,95	-	0,65
λ	Média	-	-	-	0,10	-	-
	Inf5%	-	-	-	0,06	-	-
	Sup95%	-	-	-	0,17	-	-

σ^2_a variância do efeito do marcador; σ^2_u variância do efeito poligênico; σ^2_e variância do resíduo; σ^2_{gm} variância genética aditiva explicada pelos marcadores ($\sigma^2_{a\pi\sum p_i q_i}$); σ^2_T variância do drEBV; c é a proporção da variância genética aditiva não explicada pelos marcadores ($\sigma^2_u / (\sigma^2_{gm} + \sigma^2_u)$); $h^2_M = \sigma^2_{gm} / (\sigma^2_T)$; $h^2_G = (\sigma^2_{gm} + \sigma^2_u) / (\sigma^2_T)$.

O BayesC π também tendeu a apresentar estimativas maiores para as variâncias de marcador que os outros modelos, porém com magnitude menor que o BayesC e com algumas estimativas semelhantes as obtidas com os outros modelos pelo IAD de 95%. O valor de π foi em média igual à aproximadamente 0,32, 0,30 e 0,42 para PP, PG e PL respectivamente e a variância genética aditiva explicada pelos marcadores (σ^2_{gm}) foi de mesma magnitude que as estimadas pelo GBLUP. Isto sugere que possivelmente não havia QTLs de grande efeito próximos aos marcadores deste painel para as características estudadas. Esta explicação pode também ser sugerida observando-se o σ^2_{gm} , em que os modelos que consideraram heterogeneidade de variância de marcadores e seleção de covariáveis tenderam a apresentar σ^2_{gm} menores que o GBLUP (porém dentro do IAD) e o TRI-G (com IAD

diferentes). Para a variância residual, todos os modelos tenderam apresentar a mesma magnitude para as estimativas, com exceção do TRI-G e TRI-H, que apresentaram estimativas muito mais altas que os demais modelos pelo IAD de 95%. Já a variância total do pseudo-fenótipo (σ^2_T) foram menores com os modelos unicaracterística em relação aos multicaracterística, embora ao se verificar o IAD de 95 %, estas variâncias podem ser consideradas semelhantes. Esta diferença foi devido ao ajuste para a heterogeneidade de variância para o resíduo dos modelos unicaracterística.

A relação entre os efeitos dos marcadores e o poligênico pode ser melhor visualizada pela razão entre a variância do efeito poligênico e a variância genética aditiva total (poligênica mais a genética explicada pelos marcadores). Esta medida (c) se refere ao quanto os marcadores não explicaram da variância genética aditiva de uma característica. De uma forma geral, entre as características estudadas, os marcadores tenderam a explicar maior proporção da variância genética aditiva total para a PG, seguido da PP, e por último a PL, independente do modelo utilizado (Tabelas 4, 5, 6). Também foi observado que entre os modelos estudados, o GBLUP tendeu a explicar mais a variância genética aditiva. Isto indica mais uma vez, que um modelo com homogeneidade de variâncias sem seleção de covariáveis é a melhor alternativa para as avaliações genômicas destas características, considerando o painel de marcadores empregado.

A razão entre a variância genética explicada pelos marcadores (h^2_M) e a variância do drEBV foi semelhante entre os modelos unicaracterística, com o GBLUP tendendo a apresentar estimativas maiores. O TRI-G apresentou maiores valores de h^2_M que os modelos univariados, considerando um IAD de 95% (Tabela 4, 5, 6). Quando foi considerada a razão (h^2_G) entre a variância genética aditiva total ($\sigma^2_{gm} + \sigma^2_u$) e a variância do drEBV, os modelos univariados tenderam a estimativas maiores e melhores pelo IAD de 95% que o TRI-H. Este resultado indica que a combinação entre os coeficientes de parentesco obtidos pelo pedigree e o parentesco genômico em uma só matriz (H), é inferior que quando considerado os efeitos poligênicos e dos marcadores separadamente.

Habilidade e acurácia de predição

As estimativas das correlações de Pearson entre os GEBVs e o drEBV variaram de 0,36 a 0,43. Estes valores foram inferiores aos estimados por Colombani et al., (2013) para PL em uma população da raça Holandesa e superiores aos estimados pelos mesmos autores para uma população de bovinos da raça Montbéliarde. O GBLUP obteve os maiores valores de correlação (0,43) com drEBV para PP e PG, considerando tanto o $GEBV_M$ e o $GEBV_T$. Para a PL, o modelo TRI-G obteve maior correlação com $GEBV_M$, e os modelos de mistura (BayesC e BayesC π) maior correlação com o $GEBV_T$. Todos os modelos apresentaram resultados muito similares para a correlação, e somente o BayesC mostrou valores inferiores aos demais modelos quando foi considerado o $GEBV_M$ para PG (0,39) e PL (0,36). Erbe et al., (2012) reportaram maiores valores para estas correlações além de observarem estimativas maiores dos modelos Bayesianos em relação ao GBLUP. Porém estes autores utilizaram um painel imputado para 800 k. Isto possivelmente indica que um painel de maior densidade pode captar melhor os efeitos dos QTLs, de modo a obter melhores estimativas de correlação, principalmente pelos modelos bayesianos.

Pelo erro do quadrado médio da predição (MSE), diferenças maiores e mais nítidas foram observadas entre os modelos (Tabela 7). Menores valores de MSE estão associados a um melhor ajuste global, incluindo a escala entre GEBV e drEBV (NEVES et al., 2014). O GBLUP apresentou os melhores valores (menores) para MSE tanto quando $GEBV_M$ ou $GEBV_T$ foi considerado, independente da característica estudada. Os modelos TRI-G, TRI-H e o BayesC π apresentaram os piores valores (maiores) para este critério. Estes resultados estão de acordo aos reportados por Neves et al. (2014) que ao estudarem características de corte e fertilidade em Nelore também encontraram menor MSE para o GBLUP em relação aos modelos bayesianos.

Tabela 7 – Quadrado médio do erro de predição (MSE), coeficiente de regressão (b), acurácia empírica (r_i), estimada (r_c), esperada (r_e), e índice de desempenho dos modelos (IND_M e IND_T) utilizados para avaliar a produção de proteína (PP), gordura (PG) e leite (PL), considerando apenas o efeito de marcador (M) e o efeito de marcador mais o poligênico (T).

Característica	Modelo	MSE _M	MSE _T	b _M	b _T	r_{iM}	r_{iT}	$r_{cM}^{a/b}$	$r_{cT}^{a/b}$	r_e	IND _M	IND _T
PP	GBLUP	500	495	1,25	1,10	0,78	0,79	0,76/0,77	0,83/0,84	0,90	1,02	1,14
	BayesC	518	502	1,52	1,16	0,76	0,80	0,84/0,86	0,84/0,84		0,99	1,14
	BayesCπ	609	617	1,37	1,09	0,75	0,77	0,74/0,74	0,84/0,84		0,84	1,02
	LASSO	526	515	1,51	1,17	0,77	0,79	0,70/0,70	0,82/0,82		0,86	1,10
	TRI*	628	627	0,98	0,98	0,77	0,77	0,82/0,82	0,79/0,79		0,99	0,96
PG	GBLUP	987	985	1,19	1,07	0,72	0,73	0,80/0,80	0,81/0,82	0,88	1,00	1,04
	BayesC	1095	1077	1,24	0,87	0,65	0,70	0,71/0,72	0,81/0,81		0,78	0,95
	BayesCπ	1263	1259	1,35	1,11	0,71	0,73	0,79/0,79	0,81/0,81		0,80	0,90
	LASSO	1059	1035	1,39	1,07	0,70	0,73	0,76/0,76	0,82/0,82		0,85	1,03
	TRI*	1358	1357	0,96	0,96	0,71	0,71	0,82/0,83	0,80/0,80		0,85	0,83
PL	GBLUP	380523	373961	1,31	1,00	0,72	0,76	0,71/0,71	0,86/0,86	0,89	1,04	1,29
	BayesC	421179	408610	1,25	0,82	0,65	0,77	0,67/0,67	0,87/0,87		0,91	1,23
	BayesCπ	501291	501432	1,58	1,11	0,72	0,78	0,69/0,69	0,87/0,87		0,83	1,20
	LASSO	438040	418040	1,67	1,03	0,68	0,77	0,69/0,70	0,88/0,88		0,82	1,28
	TRI*	503066	502439	0,97	0,97	0,77	0,78	0,81/0,81	0,79/0,79		1,13	1,12

Valores em negrito refere-se aos melhores valores dentro de cada critério.

* No modelo TRI o índice M refere-se a matriz G e o índice T refere-se a matriz H.

^a média das acurácias individuais; ^b acurácia média obtida pela média dos PEVs.

$IND = r_i^{*2} + r_c^{*2} - [MSE/2(\sigma^2_T) + 2*(1-b^*)^2/(b^*)^2]$. * Refere-se aos GEBV_M ou GEBV_T.

O efeito de escala, inferido por meio do MSE, como esperado, foi proporcional as grandezas das características, ou seja, primeiramente para PL, seguido da PG, e depois para a PP (Tabela 7). O $GEBV_T$ apresentou menor efeito de escala com o $drEBV$ que o $GEBV_M$. Já em relação aos modelos multicaracterística, os dois modelos (TRI-G e TRI-H) apresentam maior efeito de escala que os demais modelos unicaracterística. A escala das predições é especialmente importante, para comparação do GEBV com o valor genético obtidos pela avaliação tradicional, e é dependente das diferenças inerentes aos dados analisados, ao pseudo-fenótipo estudado, da metodologia estatística empregada e aos efeitos ajustados no modelo, entre outros (NEVES et al., 2014).

Os coeficientes da regressão linear da $drEBV$ nas GEBVs variaram de 0,96 a 1,67 (Tabela 7). As estimativas do coeficiente de regressão menores que 1 indicam superestimação da variância das predições genômicas (inflação) enquanto que os coeficientes maiores que 1 indicam subestimação (deflação). Quando se considerou o efeito poligênico nas predições genômicas, o viés da predição tendeu a diminuir para todos os modelos estudados estando de acordo com os resultados reportados por Liu et al., (2011) utilizando o GBLUP em gado Holandês, e Duchemin et al., (2012) utilizando o BayesC π em ovelhas leiteiras. Os melhores valores (próximos a 1) foram observados para os modelos multicaracterística, sendo estes levemente inflacionados (menor que 1) (Tabela 7). Os demais modelos tenderam a mostrar valores deflacionados. Diferentes resultados foram observados por Gao et al., (2012), que observou tendência de inflação para as características PP, PG e PL em uma população de gado Holandês, utilizando diferentes metodologias de GBLUP. Também Erbe et al., (2012) utilizando o GBLUP, encontraram valores deflacionados para PG e PL em uma população da raça Holandesa e valores inflacionados, semelhantes aos observados neste trabalho para PP para mesma população e para as três características estudadas em uma população da raça Jersey. Estes autores também observaram tendência de inflação com os métodos bayesianos nestas duas populações com exceção das predições para PG e PL na população da raça Holandesa.

Colombani et al., (2013) utilizando o GBLUP, BayesC e Lasso também observaram valores inflacionados para as predições genômicas da PL. O baixa

deflação pode indicar a adequação das predições genômicas da população estudada já que a inflação pode favorecer erroneamente os touros jovens (sem informação de filhas) em detrimento aos touros mais velhos em teste (AGUILAR et al., 2010). Além do efeito poligênico (LIU et al., 2011) diferenças inerentes aos dados analisados, e a metodologia estatística empregada (AGUILAR et al., 2010) e ao pseudo-fenótipo utilizado pode explicar o viés das predições genômicas. Vitezica et al., (2011) observou que múltiplas etapas na avaliação podem causar mais viés nas predições que uma única etapa (*Single-Step*).

As predições genômicas foram melhoradas quando os efeitos poligênicos foram incluídos (Tabela 7), tanto pela acurácia empírica quanto pela acurácia estimada. Estes resultados mostram a vantagem de se considerar os efeitos poligênicos nas predições genômicas, principalmente quando as características são de menor herdabilidade. As acurácias empíricas variaram de 0,65 a 0,78 para $GEBV_M$, e de 0,70 a 0,80 para $GEBV_T$. Quando foi considerado o $GEBV_M$, o GBLUP mostrou os maiores valores de acurácia empírica para PP e PG, e o TRI-G para a PL.

Para o GEBVM o BayesC obteve os melhores valores de acurácia empírica para PP, o GBLUP, BayesC π e o Lasso obtiveram os melhores valores para PG e a TRI-H para PL. Os valores encontrados neste trabalho foram superiores aos reportados por Gao et al., (2012), que encontraram valores variando de 0,66 a 0,68 para PP, PG e PL utilizando o GBLUP. Já as acurácias estimadas variaram de 0,67 a 0,86 para $GEBV_M$ e de 0,79 a 0,88 para $GEBV_T$. Os modelos com melhores valores para a acurácia estimada, foram semelhantes aos observados para acurácia empírica, com exceção do BayesC para $GEBV_M$ da PP, TRI-G para $GEBV_M$ da PG, e Lasso para $GEBV_M$ PL. Em relação aos modelos multicaracterística, os valores de acurácia esperada tenderam a ser maiores para o TRI-G em relação ao TRI-H, diferentemente dos valores da acurácia empíricas que não alterou ou foi levemente maior para o TRI-H. Os resultados sugerem que as avaliações genômicas multicaracterística apresentou desempenho adequado além da predição para diferentes características em um processo único. Com base nisto, vale destacar que as avaliações genômicas multicaracterística devem ser melhor estudadas e consideradas como alternativa nas avaliações genômicas.

Os valores de acurácia de predição foram considerados adequados para utilização da seleção genômica na raça. Estas predições podem ser utilizadas para a escolha dos animais que irão participar do teste progênie, diminuindo o custo do teste com animais que apresentariam baixo desempenho genético.

De um modo geral, apesar dos valores da acurácia estimada tenderem a serem pouco maiores que os valores da acurácia empírica, as magnitudes destes valores foram semelhantes indicando que o método empírico pode ser usado para a avaliação da acurácia, quando um delineamento adequado de predição é empregado. No entanto, os dois tipos de estimativas da acurácia foram pouco menores que as acurácias esperadas adaptadas pela fórmula de Daetwyler et al., (2010), mostrando estimativas semelhantes para as características estudadas (média de 0,89). Isto indica que a intensidade de seleção para características estudadas é aumentada pela a utilização da GS, porém não atingiu seu potencial. A utilização de painéis de maior densidade e de pseudo-fenótipos provenientes avaliação genético-quantitativo mais acuradas, podem aproximar ainda mais os valores empírico e estimado do valore esperado.

Os desempenhos dos modelos genômicos foram muito semelhantes entre as metodologias, principalmente em relação às acurácias de predição. Desta forma, um índice (IND) foi proposto visando avaliar um ajuste global entre os critérios avaliados e para evidenciar a melhor metodologia entre as empregadas. Cada componente do IND pode variar de 0 a 1, sendo dois com ponderação positiva e dois negativas. Sendo assim, o IND pode variar de -2 a 2, e valores maiores está relacionado ao melhor ajuste global. Todos os critérios com a suas devidas ponderações variaram de 0 a 0,88, de modo que a penalização foi praticamente igual para todos os critérios. No entanto, este índice favoreceu mais a habilidade de predição dos modelos (predição para o conjunto de validação), por considerar duas medidas equivalentes: a acurácia empírica e a acurácia estimada.

Pela Tabela 7 é possível observar que o índice discriminou os desempenhos dos modelos estudados, facilitando a visualização de um ajuste global. Neste sentido, o modelo GBLUP obteve os maiores valores para o IND para todas as características, com exceção quando se considerou o GEV_M para PL, cujo melhor modelo foi o TRI-G. Desta forma, apesar da maioria dos

modelos terem apresentado estimativas com magnitudes semelhantes para os outros critérios, a metodologias que utilizaram a matriz G (genômica), seja por abordagens multivariadas ou principalmente univariada, foram superiores às demais metodologias, além de terem apresentado vantagens computacionais e facilidade de convergência (valores não mostrados). Esses resultados são diferentes dos sugeridos pela maioria dos estudos de simulação (MEUWISSEN; HAYES; GODDARD, 2001; SOLBERG et al., 2008; CLARK et al., 2011). Segundo Daetwyler et al., (2010) em estudo realizado com dados simulados, o desempenho do GBLUP em relação aos métodos que realizam seleção de covariáveis depende do número de QTL da característica. Quando o número de QTLs é maior que o número de segmentos de cromossomos efetivos, GBLUP apresenta desempenho igual ou melhor que os métodos que realizam seleção de covariáveis.

Em estudos utilizando dados reais em gado leiteiro, o GBLUP apresentou desempenho igual ou superior aos métodos baseados em algum tipo de seleção de covariáveis (HAYES et al., 2009; MOSER et al., 2009). Luan et al., (2009) observaram melhor desempenho (pelas acurácias) do GBLUP em relação aos modelos de mistura e BayesB para PP, PG e PL utilizando o mesmo painel deste estudo em uma população de bovinos da raça Norueguesa Vermelha. Já Erbe et al., (2012) observaram a superioridade de métodos de seleção de covariáveis sobre o GBLUP. Entretanto, estes autores usaram dados imputados para um painel de alta densidade (800 k) e sugeriram que os métodos bayesianos poderiam ser utilizados para tirar o máximo a vantagem do aumento da densidade do marcador. Legarra et al., (2011), estudando a raça Holandesa e Montbéliard observaram que o Lasso obteve alto incremento de acurácia para as predições em comparação ao GBLUP, principalmente em características que possuem o DGAT1, como o gene de efeito maior, como as características do estudo. No entanto, em populações zebuínas ainda não foi reportado o gene de efeito maior, como o DGTA1. Desta forma, neste estudo o desempenho do GBLUP em relação aos outros métodos pode ser explicado pela baixa densidade do painel utilizado, em que os marcadores não estão muito próximos aos QTLs de maiores efeitos e a variação devido a estrutura genético-aditiva é melhor explicada pela igual variação dos segmentos cromossômicos.

As estimativas da correlação de posto (Spearman) entre os modelos genômicos e o valor genético da avaliação tradicional foram iguais a um (1) ou próximas disso (Tabela 8), entre as metodologias genômicas unicaracterística. Isto indica a equivalência destas metodologias para a avaliação genômica desta população. Os modelos multicaracterística tenderam a estimar correlações de posto menores, porém altas, com os modelos unicaracterística. Já em relação à avaliação tradicional e genômicas, as correlações entre as predições foram em geral altas, e variaram de 0,76 a 0,87. Isto indica que pode haver mudanças na classificação dos animais pela implantação da seleção genômica em relação a avaliação tradicional.

Tabela 8 – Correlações de Spearman entre os valores genéticos obtido pela avaliação tradicional e entre os valores genéticos genômicos obtidos pelos modelos estudados considerando apenas o efeito do marcador (acima da diagonal) e o efeito de marcador mais o poligênico (abaixo da diagonal) para a produção de proteína (PP), gordura (PG) e leite (PL), obtidos por meio dos diferentes modelos.

Modelo	BLUP	GBLUP	BayesC	BayesCπ	Lasso	TRI*
PP						
BLUP	1,00	0,80	0,75	0,80	0,79	0,83
GBLUP	0,85	1,00	0,97	1,00	1,00	0,97
BayesC	0,85	1,00	1,00	0,98	0,98	0,93
BayesCπ	0,85	1,00	1,00	1,00	1,00	0,97
Lasso	0,86	1,00	1,00	1,00	1,00	0,96
TRI*	0,84	0,97	0,96	0,97	0,97	1,00
PG						
BLUP	1,00	0,85	0,82	0,85	0,84	0,87
GBLUP	0,89	1,00	0,98	1,00	1,00	0,98
BayesC	0,90	1,00	1,00	0,98	0,98	0,95
BayesCπ	0,89	1,00	1,00	1,00	1,00	0,98
Lasso	0,89	1,00	1,00	1,00	1,00	0,97
TRI*	0,87	0,99	0,98	0,99	0,98	1,00
PL						
BLUP	1,00	0,80	0,76	0,79	0,78	0,85
GBLUP	0,85	1,00	0,97	1,00	1,00	0,96
BayesC	0,84	1,00	1,00	0,98	0,98	0,92
BayesCπ	0,85	1,00	1,00	1,00	1,00	0,95
Lasso	0,84	1,00	1,00	1,00	1,00	0,95
TRI*	0,86	0,96	0,95	0,96	0,95	1,00

*Os valores genéticos considerando os marcadores foi obtido por meio da matriz G (acima da diagonal), enquanto o valor genético total foi obtido por meio da matriz H (abaixo da diagonal).

Para melhor detalhar o impacto da seleção genômica em relação a avaliação tradicional, a análise de coincidência de seleção foi realizada para diferentes grupos de sexo, número e proporções de animais selecionados (Tabela 9). De um modo geral as porcentagens de coincidências tenderam a diminuir quando diferentes proporções de indivíduos foram selecionadas, com exceção de quando se reduziu o número de 20 para os 10 melhores touros para PP e PG. Dentre os modelos estudados, em geral, o GBLUP unicaracterística seguido pelo tri-característica apresentou menor discrepância de seleção pelos valores genéticos (maior coincidência) em relação avaliação

tradicional e o modelo BayesC maior discrepância. A maioria dos valores das porcentagens de coincidências foram de média magnitude, porém variaram de baixa (0,40 para os cinco touros melhores com o BayesC para PL e PG) à alta (valores maiores que 0,80). Isto indica que há alguma divergência entre a seleção pelo EBV e GEBV. Considerando que os valores genômicos são mais acurados devido a maior quantidade de informações (marcadores) que a avaliação convencional (MEUWISSEN, 2007), as predições genômicas poderiam ser a melhor alternativa para a avaliação de touros jovens (baixa acurácia) nesta população.

Tabela 9 – Coincidência de seleção para diferentes números e proporções de animais classificados pelos valores genéticos obtidos pela avaliação tradicional e entre os valores genéticos genômicos com os modelos estudados, considerando apenas o efeito do marcador (primeiro valor) e o efeito de marcador mais o poligênico (segundo valor) para a produção de proteína (PP), gordura (PG) e leite (PL).

Modelo	Quantidade de touros selecionados			% de fêmeas selecionadas		
	20	10	5	50%	20%	10%
PP						
GBLUP	0,85/0,85	0,70/0,70	0,80/0,80	0,79/0,81	0,66/0,69	0,59/0,64
BayesC	0,80/0,85	0,60/0,70	0,40/0,80	0,76/0,81	0,60/0,68	0,53/0,65
BayesC π	0,85/0,85	0,70/0,70	0,80/0,80	0,78/0,82	0,65/0,68	0,58/0,65
Lasso	0,80/0,85	0,70/0,70	0,80/0,80	0,78/0,82	0,64/0,71	0,56/0,66
TRI*	0,80/0,80	0,60/0,60	0,60/0,60	0,81/0,81	0,71/0,70	0,57/0,57
PG						
GBLUP	0,80/0,80	0,80/0,80	0,80/0,80	0,83/0,85	0,72/0,77	0,69/0,70
BayesC	0,80/0,80	0,70/0,80	0,80/0,80	0,81/0,85	0,67/0,76	0,62/0,73
BayesC π	0,80/0,80	0,80/0,80	0,80/0,80	0,83/0,86	0,71/0,77	0,69/0,72
Lasso	0,80/0,80	0,70/0,80	0,80/0,80	0,82/0,86	0,69/0,77	0,65/0,72
TRI*	0,80/0,80	0,70/0,70	0,60/0,80	0,84/0,84	0,75/0,75	0,67/0,68
PL						
GBLUP	0,70/0,80	0,70/0,70	0,60/0,60	0,80/0,82	0,68/0,73	0,57/0,66
BayesC	0,70/0,80	0,80/0,70	0,40/0,60	0,78/0,81	0,65/0,73	0,50/0,67
BayesC π	0,65/0,80	0,70/0,80	0,60/0,60	0,80/0,82	0,68/0,73	0,56/0,69
Lasso	0,65/0,80	0,70/0,70	0,60/0,60	0,79/0,81	0,68/0,73	0,56/0,67
TRI*	0,85/0,85	0,70/0,70	0,60/0,60	0,82/0,83	0,75/0,75	0,60/0,62

*Os valores genéticos considerando os marcadores foi obtido por meio da matriz G (primeiro valor), enquanto o valor genético total foi obtido por meio da matriz H (segundo valor).

CONCLUSÃO

As predições genômicas foram adequadas para seleção de animais da raça Guzerá. Existem algumas divergências na classificação dos animais pelas avaliações tradicional e genômicas. A avaliação genômica é uma alternativa para a seleção de touros jovens (baixa acurácia) nesta população.

As metodologias de predição bayesianas, GBLUP e os modelos genômicos multicaracterística são equivalentes para a predição dos valores genômicos dos animais desta população. No entanto, o GBLUP é a melhor opção considerando um ajuste global, além da menor exigência computacional e facilidade em convergência em relação aos outros modelos.

REFERÊNCIAS BIBLIOGRÁFICA

- AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, Champaign, v. 93, n. 2, p. 743–752, 2010. Disponível em: <<http://dx.doi.org/10.3168/jds.2009-2730>>.
- BABIUK, L. A.; PHILLIPS, J. P.; MOO-YOUNG, M. **Animal biotechnology: comprehensive biotechnology, supplement**. Oxford: Pergamon Press 1989. 213 p.
- CLARK, S. A.; HICKEY, J. M.; VAN DER WERF, J. H. J. Different models of genetic variation and their effect on genomic evaluation. **Genetic Selection Evolution**, Paris, v. 43, p. 18, 2011. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-43-18>>.

COLOMBANI, C.; LEGARRA, A.; FRITZ, S.; GUILLAUME, F.; CROISEAU, P.; DUCROCQ, V.; ROBERTGRANIÉ, C. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC π methods for genomic selection in French Holstein and Montbéliarde breeds. **Journal of Dairy Science**, Champaign, v. 96, n. 1, p. 575–591, 2013. Disponível em: <<http://dx.doi.org/10.3168/jds.2011-5225>>.

DAETWYLER, H. D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS, J. A. Inbreeding in genome-wide selection. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 369-376, 2007. Disponível em: <<http://dx.doi.org/10.1111/j.1439-0388.2007.00693.x>>.

DAETWYLER, H. D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J. A. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, Bethesda, v. 185, n. 3, p. 1021–1031, 2010. Disponível em: <<http://dx.doi.org/10.1534/genetics.110.116855>>.

DUCHEMIN, S. I.; COLOMBANI, C.; LEGARRA, A.; BALOCHE, G.; LARROQUE, H.; ASTRUC, J. M.; BARILLET, F.; ROBERT-GRANIÉ, C.; MANFREDI, E. Genomic selection in the French Lacaune dairy sheep breed. **Journal of Dairy Science**, Champaign, v. 95, n. 5, p. 2723–2733, 2012. Disponível em: <<http://dx.doi.org/10.3168/jds.2011-4980>>.

- ERBE, M.; HAYES, B. J.; MATUKUMALLI, L. K.; GOSWAMI, S.; BOWMAN, P. J.; REICH, C. M.; MASON, B. A.; GODDARD, M. E. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. **Journal of Dairy Science**, Champaign, v. 95, n. 7, p. 4114–4129, 2012. Disponível em: <<http://dx.doi.org/10.3168/jds.2011-5019>>.
- GAO, H.; CHRISTENSEN, O. F.; MADSEN, P.; NIELSEN, U.S.; ZHANG, Y.; LUND, M. S.; SU, G. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. **Genetics Selection Evolution**, Londres, v. 44, p. 8, 2012. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-44-8>>.
- GARRICK, J. D.; TAYLOR, J. F.; FERNANDO, R. L. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetic Selection Evolution**, Londres, v. 41, p. 55, 2009. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-41-55>>.
- GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica**, Dordrecht, v. 36, n. 2, p. 245–257, 2009. Disponível em: <<http://dx.doi.org/10.1007/s10709-008-9308-0>>.
- GODDARD, M. E.; HAYES, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, New York, v. 10, p. 381–391, 2009. Disponível em: <<http://dx.doi.org/10.1038/nrg2575>>.

HARRIS, B. L.; JOHNSON, D. L. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. **Journal of Dairy Science**, Champaign, v. 93, n. 3, p. 1243–1252, 2010.

HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, p. 433-443, 2009. Disponível em: <<http://dx.doi.org/10.3168/jds.2008-1646>>.

LEGARRA, A.; RICARD A.; FILANGI O. **GS3 Genomic Selection — Gibbs Sampling — Gauss Seidel (and BayesCπ)**. Disponível em: <http://snp.toulouse.inra.fr/~alegarra/manualgs3_last.pdf>. Acesso em: 20 jan. 2015.

LEGARRA, A.; ROBERT-GRANIÉ, C.; CROISEAU, P.; GUILLAUME, F.; FRITZ, S. Improved lasso for genomic selection. **Genetics Research**, Cambridge, v. 93, n. 1, p. 77–87, 2011. Disponível em: <<http://dx.doi.org/10.1017/S0016672310000534>>.

LIU, Z.; SEEFRIED, F. R.; REINHARDT, F.; RENSING, S.; THALLER, G.; REENTS, R. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. **Genetics Selection Evolution**, Londres, v. 43, p. 19, 2011. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-43-19>>.

- LUAN, T.; WOOLLIAMS, J. A.; LIEN, S.; KENT, M.; SVENDSEN, M.; MEUWISSEN, T. H. E. The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. **Genetics**, Bethesda, v. 183, n. 3, p. 1119–1126, 2009. Disponível em: <<http://dx.doi.org/10.1534/genetics.109.107391>>.
- LUND, M. S.; DE ROOS, S. P. W.; DE VRIES, A. G.; DRUET, T.; DUCROCQ, V.; FRITZ, S.; GUILLAUME, F.; GULDBRANDTSEN, B.; LIU, Z.; REENTS, R. A common reference population from four European Holstein populations increases reliability of genomic predictions. **Genetics Selection Evolution**, Londres, v. 43, p. 43, 2011. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-43-43>>.
- MEUWISSEN, T. H. E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, Malden, v. 124, n. 6, p. 321-322, 2007. Disponível em: <<http://dx.doi.org/10.1111/j.1439-0388.2007.00708.x>>.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819–1829, 2001.
- MEYER, K. **WOMBAT** – a program for mixed model analyses by restricted maximum likelihood. User notes. Amidale: Animal Genetics and Breeding Unit, 2006.
- MISZTAL, I.; LEGARRA, A.; AGUILAR I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, Champaign, v. 92, n. 9, p. 4648–4655, 2009. Disponível em: <<http://dx.doi.org/10.3168/jds.2009-2064>>.

- MIZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA, Z. **Manual for BLUPF90 family of programs**. Athens: University of Georgia, 2014.
- MOSER, G.; TIER, B.; CRUMP, R. E.; KHATKAR, M. S.; RAADSMA, H. W. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. **Genetics Selection Evolution**, Londres, v. 41, p. 56, 2009. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-41-56>>.
- MUIR, B.; DOORMAAL, B. V.; KISTEMAKER, G. International genomic cooperation –North American perspective. **Interbull Bull**, Uppsala, v. 41, p. 71–76, 2010.
- NEVES H. H. R.; CARVALHEIRO, R.; PÉREZ O'BRIEN, A. M.; UTSUNOMIYA, Y. T.; CARMO, A. S.; SCHENKEL, F. S.; SÖLKNER, J.; MCEWAN, J. C.; VAN TASSELL, C. P.; COLE, J. B.; SILVA, M. V. G. B.; QUEIROZ, S. A.; SONSTEGARD T. S.; GARCIA, J. F. Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. **Genetics Selection Evolution**, Londres, v. 46, p. 17, 2014. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-46-17>>.
- OLSON, K. M.; VANRADEN, P. M.; TOOKER, M. E.; COOPER, T. A. Differences among methods to validate genomic evaluations for dairy cattle. **Journal of Dairy Science**, Champaign, v. 94, n. 5, p. 2613–2620, 2011. Disponível em: <<http://dx.doi.org/10.3168/jds.2010-3877>>.

- PEIXOTO, M. G. C. D.; SANTOS, G. G. dos; BRUNELI, F. A. T.; PENNA, V. M.; VERNEQUE, R. S.; PANETTO, J. C. C.; MACHADO, C. H. C.; MACHADO, M. A.; LOBO, R. B.; CARVALHO, M. R. S. **Programa Nacional do Guzerá para leite: resultados do teste de progênie, do programa de melhoramento genético de zebuínos da ABCZ e do núcleo MOET.** Juiz de Fora: Embrapa Gado de Leite, 2014. 72 p. (Documentos, 168).
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: a toolset for whole-genome association and population-based linkage analysis. **American Journal of Human Genetics**, Chicago, v. 81, p. 559-575, 2007.
- R CORE TEAM. **R: a language and environment for statistical computing.** Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>.
- SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, Malden, v. 123, n. 4, p. 218-223, 2006.
- SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **The American Journal of Human Genetics**, Chicago, v. 78, p. 629-644, 2006.
- SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Genomic selection using different marker types and densities. **Journal of Animal Science**, Champaign, v. 86, n. 10, p. 2447-2454, 2008. Disponível em: <<http://dx.doi.org/10.2527/jas.2007-0010>>.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Champaign, v. 91, n. 11, p. 4414-4423, 2008.

Disponível em: <<http://dx.doi.org/10.3168/jds.2007-0980>>.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: reliability of genomic predictions for North American Holstein bulls.

Journal of Dairy Science, Champaign, v. 92, n. 1 p. 16–24, 2009.

Disponível em: <<http://dx.doi.org/10.3168/jds.2008-1514>>.

VITEZICA, Z. G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. **Genetics Research**,

Cambridge, v. 93, n. 5, p. 357–366, 2011. Disponível em:

<<http://dx.doi.org/10.1017/S001667231100022X>>.

WEIGEL, K. A.; VAN TASSELL, C. P.; O' CONNEL, J. R.; VANRADEN, P. M.;

WIGGANS, G. R. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. **Journal of Dairy Science**,

Champaign, v. 93, p. 2229-2238, 2010. Disponível em:

<<http://dx.doi.org/10.3168/jds.2009-2849>>.

CAPÍTULO 4 – MAPEAMENTO AMPLO DO GENÔMA PARA A PROSPECÇÃO DE REGIÕES RELACIONADOS À PRODUÇÃO DE LEITE, GORDURA E PROTEÍNA EM BOVINOS DA RAÇA GUZERÁ

RESUMO - A importância de estudos de associação genômica ampla no âmbito da pecuária está na incorporação do QTL detectado nas avaliações genéticas, possibilitando maior acurácia na seleção e, portanto, maior progresso genético. Entretanto não existem muitos trabalhos considerando raças zebuínas em estudos desta natureza. Sabendo disso, objetivou-se com este trabalho identificar *loci* relacionados à produção de leite (PL), gordura (PG) e proteína (PP) através de um mapeamento amplo das regiões genômicas em uma população de bovinos da raça Guzerá. 909 animais da raça Guzerá foram genotipadas utilizando painel de 50k (“BovineSNP50 BeadChip”) e 777k (“BovineHD BeadChip”) da Illumina®, no entanto, foram utilizadas apenas informações moleculares referentes ao conjunto de marcadores presentes no painel de 50 k. Para o controle de qualidade das amostras foram adotados “call rate” de 0,90 e heterozigosidade de ± 3 desvios padrão da média. Já os critérios para qualidade do marcador foram adotados “call rate” superior à 0,98, MAF maior que 2%, *p-value* do equilíbrio de Hardy-Weinberg com até 10^{-6} , correlação entre marcadores até 0,998, além da eliminação de SNPs coincidentes e com possíveis erros de posicionamento físico em relação ao mapa de referência. O número de animais utilizado depois do controle de qualidade foi 897, sendo 45 touros e 852 vacas. O número de SNPs final foi de 25.024. As informações utilizadas referentes às características estudadas foram os valores genéticos desregressados. As análises estatísticas foram realizadas em duas etapas. Na primeira, as informações foram corrigidas para subestrutura e grau de parentesco entre os indivíduos pela obtenção da solução do efeito aditivo genômico dos indivíduos através da matriz genômica G (GBLUP). Em uma segunda etapa a heterogeneidade dos marcadores foram consideradas através da regressão genético-aditiva de cada marcador independentemente para este novo fenótipo. Foram calculadas as porcentagens de variâncias explicadas para cada SNP. Janelas composta por sete SNPs adjacentes que desviou substancialmente das demais regiões genômicas foram investigadas. As janelas deslizaram ao longo dos cromossomos a cada um (1) SNP. Foram observados dois sinais fortes relacionados à explicação da variância fenotípica nos cromossomos 8 e 28 para as três características estudadas. Foram encontrados 30 genes nestas regiões, sendo que alguns deles tinham alta expressão no tecido mamário, ou estavam relacionado à fisiologia de excreção e metabolismo glicídico. Dentro destas duas regiões, cinco candidatos à QTLs foram reportadas para PL e PP na literatura. No entanto, uma investigação mais aprofundada destas regiões e seus *loci* podem oferecer maior compreensão sobre a produção leiteira em zebuínos.

Palavras chaves: “GWAS”, SNPs, Zebu.

INTRODUÇÃO

Para a identificação de um *locus* controlador de característica quantitativa (QTL), diferentes abordagens podem ser empregadas, desde a varredura por mapas de ligação, até análises do desequilíbrio de ligação por delineamento em famílias (GEORGES et al., 1995; BARENDSE et al., 1997). Os painéis de baixa à moderada densidade de marcadores eram utilizados para tal varredura. Recentemente, painéis de genotipagem contendo milhares de marcadores estão sendo empregado nas análises de associação para identificação de QTLs (COLE et al., 2011). Esta técnica, chamada de análise de associação genômica ampla (GWAS), não necessita de delineamento experimental e explora o desequilíbrio de ligação dos milhares de marcadores que cobrem o genoma animal. No entanto, existem vários fatores que podem influenciar o GWAS, como a densidade do painel de marcadores, modelo estatístico empregado, herdabilidade das características, tamanho da amostra, número de testes múltiplos e estratificação da população (PRICE et al. 2006; PEARSON; MANOLIO, 2008).

Diferentes abordagens têm sido propostas para o GWAS de acordo com os fatores que o influenciam. As principais abordagens são: por meio de marcadores únicos por meio de testes de correlação ou da regressão genética aditiva (CLEVELAND; DEEB, 2009); modelos lineares mistos (LM) (YU et al., 2006; SAHANA et al., 2010); os modelos de haplótipos (HM); com utilização da genealogia baseada em modelos mistos (GenMix) (DASHAB et al., 2012); modelos com encolhimento dos efeitos dos marcadores e seleção de covariáveis como o Lasso (LEGARRA et al., 2011) e o BayesC (PETERS et al., 2013); além do single-step GWAS (ssGWAS) (WANG et al., 2012).

Em bovinos leiteiros existem vários trabalhos relacionando regiões genômicas às produções de leite (PL), gordura (PG) e proteína (PP), como Jiang et al. (2010), Guo et al., (2012) e Olsen et al., (2011). Para as características leiteiras a maioria dos estudos indicam a região onde está o DGTA1, no cromossomo 14, como a principal região associada (JIANG et al., 2010; WANG et al., 2013). Isto demonstra a viabilidade desta metodologia para identificação de regiões relacionadas à produção em gado leiteiro, já que este gene é comprovadamente o de efeito maior nas raças taurinas (GRISART et

al., 2004). Entretanto não existem muitos trabalhos considerando as raças zebuínas em estudos de associação, havendo possibilidade de existirem outras regiões polimórficas que expliquem grande parte da variação fenotípica da produção de leite e seus constituintes.

A importância de estudos desta natureza no âmbito da pecuária está na incorporação do QTL detectado nas avaliações genéticas, possibilitando maior acurácia na seleção e, portanto, maior progresso genético. Com isso, objetivou-se com este trabalho identificar *loci* relacionados à produção de leite, gordura e proteína através de um mapeamento amplo das regiões genômicas em uma população de bovinos da raça Guzerá.

MATERIAL E MÉTODOS

Foram utilizados animais da raça Guzerá genotipados com o painel de 50k (vacas) e HD (“Density Bovine SNP BeadChip”) (touro) da Illumina®. Entretanto para padronização da análise foram utilizadas apenas informações moleculares referente ao conjunto de marcadores presentes no painel de 50 k. Para o controle de qualidade das amostras foram adotados “call rate” de 0,90 e heterozigosidade de ± 3 desvios padrão da média. Já os critérios para qualidade do marcador foram adotados “call rate” superior à 0,98, MAF maior que 2%, *p-value* do equilíbrio de Hardy-Weinberg com até 10^{-6} , correlação entre marcadores até 0,998, além da eliminação de SNPs coincidentes e com possíveis erros de posicionamento físico em relação ao mapa de referência (UMD3.1). O número de animais utilizados depois do controle de qualidade foi 897, sendo 45 touros e 852 vacas, e o número de SNPs foi 25.024. Após isso, foram imputados os genótipos perdidos utilizando o software fastPHASE (SCHEET; STEPHENS, 2006).

A (sub) estratificação genética da população foi avaliada pela análise dos componentes principais obtidos pelo escalonamento multidimensional da matriz de dissimilaridade genômica transformada no quadrado das distâncias euclidianas utilizando o “software” PLINK (Purcell et al. 2007). A análise de agrupamento foi aplicada aos dois autovetores que explicaram a maior proporção da variância dos dados utilizando-se o algoritmo “k-means” (dois grupos). Um teste F (Fisher) para a comparação de médias das características estudadas entre os grupos de modo a verificar o efeito da sub-estratificação. O

nível de relacionamento entre os indivíduos da população foi avaliado pelas elementos superiores exteriores à diagonal da matriz de parentesco pelo pedigree (A), parentesco genômico (G) (VANRADEN, 2008) e de similaridade genômico (K) (LEUTENEGGER et al., 2003). As matrizes genômicas podem ser representadas da seguinte maneira:

$$G = \frac{ZZ'}{2 \sum p_i(1-p_i)} \quad \hat{k}_{l,k} = \frac{1}{NL} \sum \frac{(g_{i,l} - p_i)(g_{i,k} - p_i)}{p_i(1-p_i)}$$

em que $k_{l,k}$ é o parentesco genômico estimada entre indivíduos l e k, NL é o número total de *loci* utilizados para o cálculo, p_i é a frequência do alelo de referência para *locus* i, $g_{i,l}$ e $g_{i,k}$ são os *locus* i genótipos para indivíduos l e k, respectivamente (codificados como 0, 1 ou 2).

As informações utilizadas referentes às características estudadas, foram os valores genéticos desregressados da produção de leite (PL), gordura (PG) e proteína (PP) segundo Garrick, Taylor e Fernando (2009). As análises estatísticas foram realizadas em duas etapas e foi adaptada da metodologia proposta por Wang et al. (2012) para considerar a heterogeneidade de variância dos marcadores com o modelo GBLUP. Na primeira, as informações foram corrigidas para subestrutura e grau de parentesco entre os indivíduos pela obtenção da solução do efeito aditivo genômico dos indivíduos obtido com a matriz genômica G (GBLUP) utilizando-se os softwares da família *f90* (MIZTAL et al., 2015). Os resíduos foram ponderados por $(1-h^2)/[(c+(1-r_i^2)/r_i^2)h^2]$, em que r_i^2 é a “reliability” dos valores genéticos, e as estimativas de h^2 e c (proporção de variância genética aditiva não explicada pelos marcadores) utilizados foram respectivamente, 0,26 e 0,56 para PL, 0,28 e 0,33 para PG, e 0,27 e 0,40 para PP. Em uma segunda etapa a heterogeneidade dos marcadores foram consideradas por meio da regressão genético-aditiva de cada marcador independentemente para este novo fenótipo, diferentemente de Wang et al. (2012) que considerava a heterogeneidade pela ponderação da matriz G em duas ou mais etapas. Estas análises foram realizadas por meio de scripts em R.

Subsequentemente, uma estatística de qui-quadrado de Wald ($b_i / \text{var}(b_i^2)$) com foi usado para determinar se o SNP estava associado com as

características leiteiras considerados neste estudo. Neste teste é assumido uma distribuição qui-quadrado assintótica com um (1) grau de liberdade sob a hipótese nula. No entanto, nenhum marcador atingiu o nível de significância adotado de 0,05 corrigido para Bonferroni (1,998082e-06). Sendo assim, os valores das soluções da regressão genético aditiva (GEBV) foram utilizadas para calcular a variância explicada por cada SNP. “Janelas” composta por sete SNPs adjacentes que desviaram substancialmente das outras regiões genômicas foram investigadas. As janelas “deslizaram” ao longo dos cromossomos a cada um SNP. A percentagem de variância explicada por cada SNP foi calculada como:

$$\% \pi = \frac{2p_i q_i \hat{b}_i^2}{\sigma_T^2} \times 100$$

em que o índice i é relativo ao i -ésimo SNP, b_i é o efeito do alelo de substituição estimado, p_i e q_i são as frequências alélicas, e σ_T^2 é a variância total de cada característica (drEBV).

Adaptado de Utsunomya et al. (2014), foram investigadas as regiões acima de $11X IQR + Q_3$, em que o IQR é o intervalo interquartil e Q_3 é o terceiro quartil. Os genes presentes nestes regiões foram obtidos utilizando-se as respectivas coordenadas genômicas pelo “assembly” UMD 3.1 (ZIMIN et al., 2009) dos SNPs e o banco de informações do NCBI (<http://www.ncbi.nlm.nih.gov>). A base de dados da QTLdb (HU et al., 2013) também foi examinada para verificar se alguma região genômica identificada estava sobreposta com uma previamente descrita como *locus* de características quantitativas (QTL) para bovinos.

RESULTADOS E DISCUSSÃO

A identificação de estruturas na população é um dos principais fatores que influenciam o desempenho do GWAS. Neste sentido, pela análise dos componentes principais não foi identificado estrutura na população estudada (Figura 1). Entretanto, dentro do mesmo programa de melhoramento do Guzerá, existem linhas de seleção para leite e corte, porém de forma difusa (seleção simultânea para leite e corte) (PEIXOTO et al., 2014). Desta forma, foi

empregado a técnica de “k-means” nos dois primeiros autovetores, para a discriminação de dois subgrupos distintos (Figura 1). O grupo 1 conteve 457 animais e o grupo 2 conteve 440. As médias entre os dois grupos foram 522 e 234 para PL; 27 e 13 para PG; e 15 e 8 para PP. Estas médias foram significativas pelo teste F ($<2,2e^{-16}$) para todas as características estudadas. Os resultados indicaram que existem uma leve subestruturação da população. A correção dos fenótipos para esta subestrutura anteriormente às análises de GWAS devem ser considerados. No entanto, a utilização de metodologias que também fazem leve correção para subestrutura, como os modelos mistos, podem ser considerados.

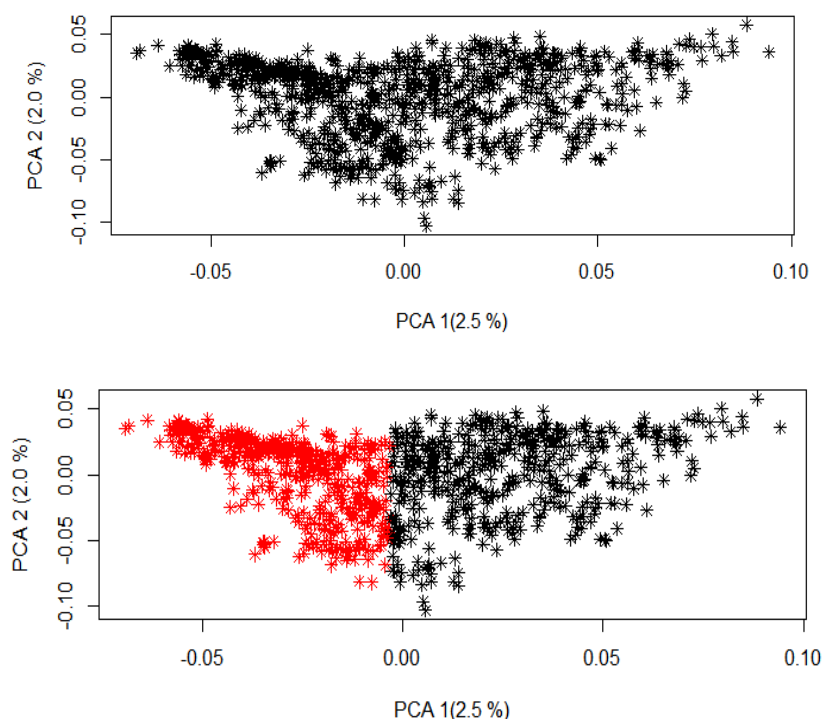


Figura 1 – Coordenadas do primeiro (PCA1) e segundo (PCA2) componente principal baseadas da matriz genômica de dissimilaridade. As porcentagens dentro do parêntese correspondem a variância explicada por cada autovetor. Na segunda figura as cores diferentes refere-se a clusterização de acordo com “k-means”.

A média dos coeficientes da matriz A foi 0,15 e variou de 0 a 0,66; da matriz G foi 0 e variou de -0,15 a 0,63; e o da matriz K foi 0,78 e variando de 0,75 a 0,90 (Figura 2). Na matriz A, uma grande parte dos indivíduos apresentou coeficiente de parentesco maior que 0, demonstrando que os indivíduos estavam relacionados pelo pedigree. O alto grau de parentesco entre muitos indivíduos, possivelmente é devido ao esquema MOET de

acasalamento entre algumas famílias. A matriz genômica (G) é centrada na variação genotípica de todos os *loci* considerados. Neste sentido, quase a metade dos animais apresentaram uma relação maior que a média para os *loci* que explicam a variabilidade da população (Figura 2). Já a matriz de similaridade indica o nível de semelhança dos indivíduos considerando todos *loci*, podendo ser estes de baixa ou alta variabilidade. Todos os animais apresentaram similaridade maior que 75 % indicando um alto nível de relacionamento dos indivíduos (Figura 2). Os resultados sugerem que o método para associação deve considerar os diferentes níveis de relação entre os indivíduos desta população. Sendo assim, o método GBLUP foi considerado adequado por considerar os níveis de relação e correção para subestrutura simultaneamente.

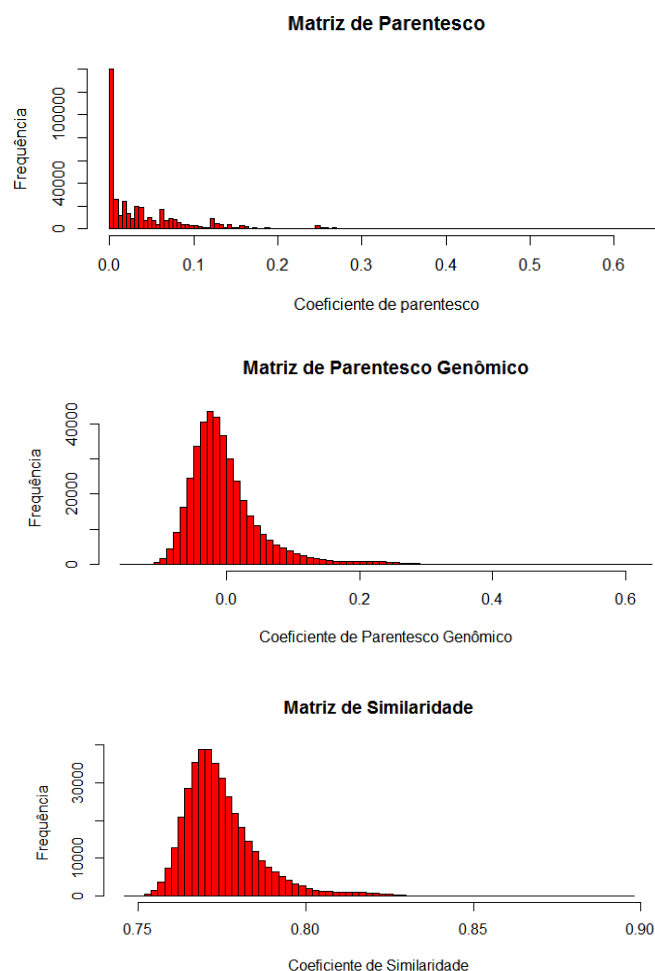


Figura 2 – Histograma dos elementos exteriores à diagonal da matriz de parentesco, parentesco genômico e da similaridade genômica da população da raça Guzerá utilizado neste estudo.

Pelo exame dos “Manhattan plots” foi possível observar dois sinais fortes relacionados à porcentagem de explicação da variância fenotípica (PEVF) no cromossomo 8 e 28 para as três características estudadas (Figura 3). Os perfis semelhantes dos gráficos para as três características indicam coerência das análises, e está de acordo com Jiang et al., (2010) que apesar de tere encontrado regiões diferentes em relação a este estudo, observaram resultados semelhantes entre as características PP, PG e PL estudando uma população da raça Holandesa. Seis janelas no cromossomo 8 e três no cromossomo 28 obteve a PEVF acima do nível crítico adotado ($11X IQR + Q_3$) para PL e PG (Tabela 1). Para PP os resultados foram similares, variando apenas no cromossomo 28, onde foi observado duas janelas com PEVF acima do nível crítico adotado. Para PL as janelas do cromossomo 8 as PEVF variaram 0,44 a 0,59, com média de 0,54 e no cromossomo 28 de 0,41 e 0,42 com média de 0,42 (Tabela 1). Para PG as janelas do cromossomo 8 obteve PEVF de 0,23 a 0,28 com média de 0,26 e as janelas do cromossomo 8 de 0,22 a 0,23 com média de 0,23. Já para PP as janelas do cromossomo 8 as PEVFs variaram de 0,22 a 0,29 com média de 0,26 e as janelas do cromossomo 8 de 0,22 a 0,23 com média de 0,22.

As janelas do cromossomo 8 tiveram média de r^2 e MAF menor que as janelas observadas no cromossomo 28 (Tabela 1). A média de r^2 está associado ao menor tamanho (em bp) da janela identificada nas análises. O maior valor de MAF está associado a maior explicação da variância fenotípica, por considerar a variância genotípica também (2pq), além do efeito de substituição alélica. No entanto, as explicações da variância foram maiores no cromossomo 8 que no 28 por possuírem maiores efeitos de substituição alélica. Métodos que consideram apenas o efeito de substituição alélica ou o seu *p-value* são mais sujeitos a erros de amostragem por não contabilizar a MAF. No entanto, o método utilizado identificou regiões de alta a baixa MAF.

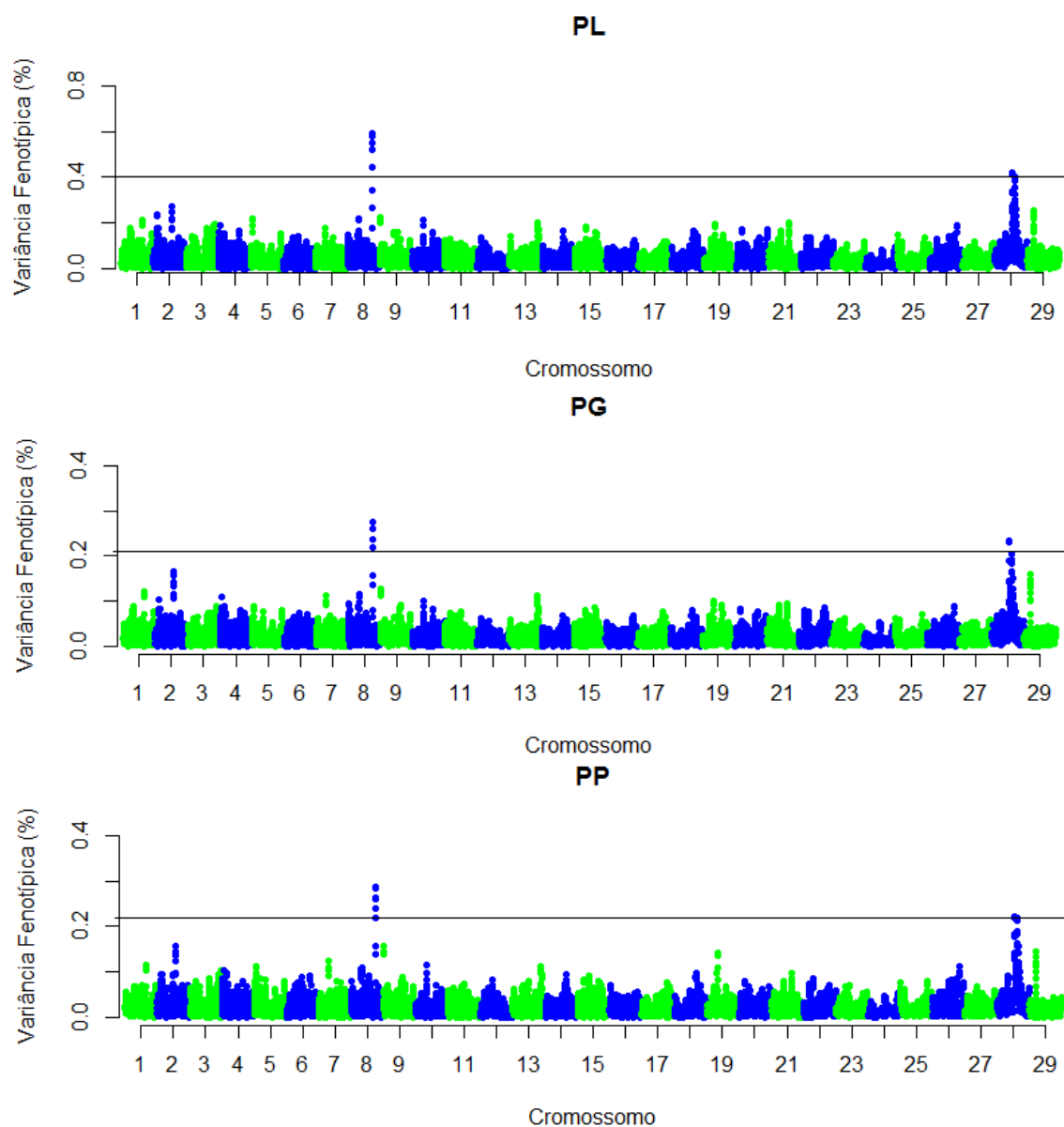


Figura 5 - Varredura do genoma bovino para produção de leite (PL), produção de gordura (PG) e produção de proteína (PP). Cromossomos são mostrados em cores alternadas (verde e azul). A linha horizontal representa o valor de $11X IQR + Q_3$, onde o IQR é o intervalo interquartil e Q_3 é o terceiro quartil.

Tabela 1 - Regiões do genoma apresentando fortes sinais de associação para cada característica estudada.

Característica	BTA	Posição Inicial	Posição Final	Comprimento	Média de r^2	Posição do pico	Média de MAF	Gene
PL,PP,PG	8	86128908	86501009	372101	0,22	86442366	5,3	FGD3,SUSD3, PHF2,
PL,PP,PG	8	85880449	86472562	592113	0,22	86442366	4,2	C9orf89,SNORA27,
PL,PP,PG	8	86283449	86560410	276961	0,30	86442366	7,0	NINJ1,WNK2,FAM120A,
PL,PP,PG	8	86205872	86538180	332308	0,22	86442366	6,2	ENSBTAG00000015218,
PL,PP,PG	8	86357827	86647495	289668	0,31	86442366	7,8	ENSBTAG00000009073
PL,PP,PG	8	85786900	86442366	655466	0,12	86442366	4,1	HERC4,MYPN,ATOH7, PBLD,HNRNPH3,RUFY2,
PL,PP,PG	28	24678582	25726782	1048200	0,53	25726782	11,0	DNA2,SLC25A16,TET1,
PL,PP,PG	28	24719677	25871892	1152215	0,53	25726782	11,0	CCAR1,STOX1,DDX21,
PL,PG	28	24900950	25938159	1037209	0,53	25726782	8,3	DDX50,KBP,SRGN, VPS26A,SUPV3L1, HKDC1,HK1,TACR2

Um total de 30 genes foram encontrados nas regiões candidatas (Tabela 1). Muitos genes possuíam funções diversificadas, sendo que para alguns deles não foram encontrada associação aparente com as características estudadas. Estes genes estavam envolvidos em processos no DNA como “splicing”, replicação e metilação (HNRNPH3, DNA2, TET1), processos no RNA nuclear (DDX21, DDX50, FAM120A) e mitocondrial (SUPV3L1), transcrição (ATOH7), inibição de outros genes (C9orf89), transportes e estrutura celular (HERC4), ativação de fosfolípidos e formação de filopodia (FGD3), respiração celular (mitocondria) (SLC25A16), metilação de proteínas (PHF2), funções no tecido nervoso (NINJ1, WNK2, TACR2) e no tecido muscular (MYPN). O gene MYPN também tem sido associado a características quantitativas da qualidade da carne (JIAO et al., 2010).

Os genes CCAR1, SRGN, VPS26A estão envolvidos na expressão proteína vacuolar que estão envolvidas no transporte de outras proteínas à partir de endossomas para serem utilizados no complexo de Golgi; no armazenamento de compostos em vesículas secretoras, e apoptose. As vesículas secretoras juntamente com o complexo de Golgi estão relacionado ao processo excreção, como no caso do tecido epitelial que está presente na glândula mamária (JUNQUEIRA; CARNEIRO, 2004). Já apoptose é um processo importante do desenvolvimento da função da glândula mamária e também prepara a glândula para a próxima lactação (WILDE; HENDERSON; KNIGHT, 1986, GREEN; STREULI, 2004).

Segundo ROSF JORD; DICKSON, (1999) a apoptose pode ocorrer em todos períodos da lactação. No início da lactação, existe equilíbrio entre a hiperplasia e apoptose na glândula mamária, porém a partir do pico de lactação o processo de apoptose das células alveolares aumenta gradualmente, reduzindo progressivamente a síntese de leite até a fase final da lactação (BOUTINAUD et al., 2004). Furth (1999) sugere que o manejo para aumentar a sobrevivência das células ao final da lactação, poderá aumentar a eficiência da produção de leite do animal. Sabendo que a raça Guzerá apresenta variabilidade na duração da lactação, possuindo muitos animais com lactações curtas (SANTOS et al., 2013), esta região pode explicar a menor duração da

lactação e ao menor nível de produção de leite da raça estudada e as raças taurinas.

O gene STOX1 tem alto nível de expressão no tecido mamário e codifica uma proteína de ligação que está relacionado a ocorrência de eclampsia e pré-eclâmpsia em mulheres (VAN DIJK et al., 2010). Os genes HKDC1 e HK1 são responsáveis pela síntese de quinases (como glicoquinase, manoqueinase frutuquinase e etc) que participam do metabolismo glicídico e da homeostase destes no organismo (GUO et al., 2015). O HK1 ainda expressa a hexoqueinase (produção de glicose na via glicolítica), que estão presentes em altos níveis no tecido mamário de ratas (KASELONIS; MCCABE; GRAY, 2009). De modo análogo ao mais influente gene para a expressão das características estudadas em taurinos, o DGTA1, está envolvido no metabolismo de lipídios (triglicerídios). No entanto, o cromossomo 14 não explicou a variância das características estudadas como as regiões discutidas acima. Segundo Rosse et al., (2014), o polimorfismo deste gene não é uma boa ferramenta para verificar os efeitos deste QTL em zebuínos, uma vez que a frequência do alelo DGAT1 é muito baixo nessas raças. Estes mesmos autores reportaram o primeiro polimorfismo DGAT1 na raça Guzará, que estava associado ao valor genético para a PL, PG e PP. Apesar de não ser considerado este polimorfismo nestas análises, é possível sugerir pelos resultados, que o metabolismo de glicídios pode ser mais na produção de leite e seus constituintes que o metabolismo de triglicerídios, dado que as raças tropicais armazenam menos gordura que as raças taurinas.

Pelo QTLdb, três QTLs e sete regiões associadas (candidatos à QTLs) à características quantitativas de cinco diferentes publicações foram reportados nas mesmas regiões discriminada por este trabalho. Destes, duas regiões no cromossomo 28 estavam relacionados diretamente à PP na raça Holandesa (DAETWYLER et al., 2008) e três regiões no cromossomo 8 estavam relacionados diretamente com a PL na raça Holandesa e Friesian (VEERKAMP et al., 2012). Os demais QTLs reportados estavam associados ao ganho de peso de carcaça, prenhez e ao teor de ferro. Apesar de não ter sido utilizado um painel de maior densidade, existe ainda a possibilidade de outras regiões explicarem parte das variâncias das características estudadas. No entanto, o resultado encontrado está de acordo com os reportados na literatura para estas

regiões, indicando que estas regiões possuem importantes QTLs que explicam a variação das características PL, PG e PL na raça Guzerá.

CONCLUSÃO

Foram identificados regiões nos cromossomos 8 e 28 que explicam uma parte diferenciada da variação fenotípica da produção do leite, gordura e proteína na população estudada. Essas regiões genômicas possuem genes candidatos com funcionalidades conhecidas e envolvidas no mecanismo de produção de leites e seus constituintes. Uma investigação mais aprofundada destas regiões com as novas tecnologias de sequenciamento e proteômica podem oferecer maior compreensão sobre o controle genético da produção leiteira em zebuínos.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARENDSE, W.; VAIMAN, D.; KEMP, S. J.; SUGIMOTO, Y.; ARMITAGE, S. M.; WILLIAMS, J. L.; SUN, H. S.; EGGEN, A.; AGABA, M.; ALEYASIN, S. A.; BAND, M.; BISHOP, M. D.; BUITKAMP, J.; BYRNE, K.; COLLINS, F.; COOPER, L.; COPPETTIERS, W.; DENYS, B.; DRINKWATER, R. D.; EASTERDAY, K.; ELDUQUE, C.; ENNIS, S.; ERHARDT, G.; FERRETTI, L.; FLAVIN, N.; GAO, Q.; GEORGES, M.; GURUNG, R.; HARLIZIUS, B.; HAWKINS, G.; HETZEL, J.; HIRANO, T.; HULME, D.; JORGENSEN, C.; KESSLER, M.; KIRKPATRICK, B. W.; KONFORTOV, B.; KOSTIA, S.; KUHN, C.; LENSTRA, J. A.; LEVEZIEL, H.; LEWIN, H. A.; LEYHE, B.; LIL, L.; MARTIN BURRIEL, I.; MCGRAW, R. A.; MILLER, J. R.; MOODY, D. E.; MOORE, S. S.; NAKANE, S.; NIJMAN, I. J.; OLSAKER, I.; POMP, D.; RANDO, A.; RON, M.; SHALOM, A.; TEALE, A. J.; THIEVEN, U.; URQUHART, B. G. D.; VAGE, D. I.; VAN DE WEGHE, A.; VARVIO, S.; VELMALA, R.; VILKKI, J.; WEIKARD, R.; WOODSIDE, C.; WOMACK, J. E.; ZANOTTI, M.; ZARAGOZA, P. A médium-density genetic linkage map of the bovine genome. **Mammalian Genome**, New York, v. 8, n. 1, p. 21-28, 1997.
- BOUTINAUD, M.; GUINARD-FLAMENT, J.; JAMMES, H. The number and activity of mammary epithelial cells, determining factors for milk production. **Reproduction, Nutrition and Development**, Wallingford, v. 44, n. 5, p. 499-508, 2004. Disponível em: <<http://dx.doi.org/10.1051/rnd:2004054>>.

- CLEVELAND, M.; DEEB, N. Evaluation of a genome-wide approach to multiple marker association considering different marker densities. **BMC Proceedings**, Londres, v. 3, suppl. 1, p. s5, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2654499/>>. Acesso em: 20 fev. 2015.
- COLE, J. B.; WIGGANS, G. R.; MA, L.; SONSTEGARD, T. S.; LAWLOR, T. J.; CROOKER, B. A.; VAN TASSELL, C. P.; YANG, J.; WANG, S.; MATUKUMALLI, L. K.; DA, Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics**, Londres, v. 12, p. 408, 2011. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-12-408>>.
- DASHAB, G. R.; KADRI, N. K.; SHARIATI, M. M.; SAHANA, G. Comparison of linear mixed model analysis and genealogy-based haplotype clustering with a Bayesian approach for association mapping in a pedigreed population. **BMC Proceedings**, Londres, v. 6, suppl. 2, p. s4, 2012. Disponível em: <<http://dx.doi.org/10.1186/1753-6561-6-S2-S4>>.
- DAETWYLER, H. D.; SCHENKEL, F. S.; SARGOLZAEI, M.; ROBINSON, J. A. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. **Journal of Dairy Science**, Champaign, v. 91, n. 8, p. 3225 – 32236, 2008. Disponível em: <<http://dx.doi.org/10.3168/jds.2007-0333>>.
- FURTH, P. A. Introduction: mammary gland involution and apoptosis of mammary epithelial cells. **Journal of Mammary Gland and Neoplasia**, New York, v. 4, n. 2, p. 123-127, 1999.
- GARRICK, D. J.; TAYLOR, J. F.; FERNANDO R. L. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetics Selection Evolution**, Londres, v. 41, p. 55, 2009. Disponível em: <<http://dx.doi.org/10.1186/1297-9686-41-55>>.

- GEORGES, M. D.; NIELSEN, M.; MACKINNON, A.; MISHRA, R.; OKIMOTO, A. T.; PASQUINO, L. S.; SARGEANT, A.; SORENSEN, M. R.; STEELE, X.; ZHAO, J.; WOMACK, E.; HOESCHELE, I. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. **Genetics**, Bethesda, v. 139, n. 2, p. 907-920, 1995.
- GREEN, K. A.; STREULI, C. H. Apoptosis regulation in the mammary gland. **Cellular and Molecular Life Sciences**, Berlin, v. 61, n. 15, p. 1867-1883, 2004.
- GUO, J.; JORJANI, H.; CARLBORG, Ö. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. **BMC Genetics**, Londres, v. 13, p. 82, 2012. Disponível em: <<http://dx.doi.org/10.1186/1471-2156-13-82>>.
- GUO, C.; LUDVIK, A. E.; ARLOTTO, M. E.; HAYES, M. G.; ARMSTRONG, L. L.; SCHOLTENS, D. M.; BROWN, C. D.; NEWGARD, C. B.; BECKER, T. C.; LAYDEN, B. T.; LOWE, W. L.; REDDY, T. E. Coordinated Regulatory Variation Associated with Gestational Hyperglycemia Regulates Expression of the Novel Hexokinase HKDC1. **Nature Communications**, London, v. 6, p. 6069, 2015. Disponível em: <<http://dx.doi.org/10.1038/ncomms7069>>.
- HU, Z. L.; PARK, C. A.; WU, X. L.; REECY, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, Londres, v. 41, 2013. Disponível em: <<http://dx.doi.org/10.1093/nar/gks1150>>.
- JIAO, Y.; ZAN, L. S.; LIU, Y. F.; WANG, H. B.; GUO, B. L. A novel polymorphism of the MYPN gene and its association with meat quality traits in *Bos taurus*. **Genetics and Molecular Research**, Ribeirão Preto, v. 9, n. 3, p. 1751-1758, 2010. Disponível em: <<http://dx.doi.org/10.4238/vol9-3gmr906>>.

- JIANG, L.; LIU, J.; SUN, D.; MA, P.; DING, X.; YU, Y.; ZHANG, Q. Genome wide association studies for milk production traits in chinese holstein population. **Plos one**, São Francisco, v. 5, 2010. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0013661>>.
- JUNQUEIRA, L. C.; CARNEIRO, J. **Histologia básica**. 10. ed. Rio de Janeiro: Guanabara Koogan, 2004. 488 p.
- KASELONIS, G. L.; MCCABE, E. R.; GRAY, S. M. Expression of hexokinase 1 and hexokinase 2 in mammary tissue of nonlactating and lactating rats: evaluation by RT-PCR. **Molecular Genetics and Metabolism**, São Diego, v. 68, n. 3, p. 371-374, 2009.
- LEGARRA, A.; ROBERT-GRANIÉ, C.; CROISEAU, P.; GUILLAUME, F.; FRITZ, S. Improved Lasso for genomic selection. **Genetics Research**, Cambridge, v. 93, n. 1, p. 77–87, 2011. Disponível em: <<http://dx.doi.org/10.1017/S0016672310000534>>.
- LEUTENEGGER, A. L.; PRUM, B.; GENIN, E.; VERNY, C.; LEMAINQUE, A.; CLERGET-DARPOUX, F.; THOMPSON, E. A. Estimation of the inbreeding coefficient through use of genomic data. **The American Journal of Human Genetics**, Chicago, v. 73, n. 3, p. 516–523, 2003. Disponível em: <<http://dx.doi.org/10.1086/378207>>.
- MIZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA, Z. **Manual for BLUPF90 family of programs**. Athens: University of Georgia, 2014. Disponível em: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf>. Acesso em: 22 fev. 2015.

- OLSEN, H. G.; HAYES, B. J.; KENT, M. P.; NOME, T.; SVENDSEN, M.; LARSGARD, A. G.; LIEN, S. Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. **Animal Genetics**, Oxford, v. 42, n. 5, p. 466-474, 2011. Disponível em: <<http://dx.doi.org/10.1111/j.1365-2052.2011.02179.x>>.
- PEARSON, T.; MANOLIO T. How to interpret a genome-wide association study. **JAMA**, Chicago, v. 299, n. 11, p. 1335–1344, 2008. Disponível em: <<http://dx.doi.org/10.1001/jama.299.11.1335>>.
- PEIXOTO, M. G. C. D.; SANTOS, G. G. dos; BRUNELI, F. A. T.; PENNA, V. M.; VERNEQUE, R. S.; PANETTO, J. C. C.; MACHADO, C. H. C.; MACHADO, M. A.; LOBO, R. B.; CARVALHO, M. R. S. **Programa nacional do Guzerá para leite**: resultados do teste de progênie, do programa de melhoramento genético de zebuínos da ABCZ e do núcleo MOET, Juiz de Fora: Embrapa Gado de Leite, 2014. 72 p. (Documentos, 168).
- PETERS, S. O.; KIZILKAYA, K.; GARRICK, D. J.; FERNANDO, R. L.; REECY, J. M.; WEABER, R. L.; SILVER, G. A.; THOMAS, M. G. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. **Journal of Dairy Science**, Champaign, v. 91, n. 2, p. 605-612, 2013. Disponível em: <<http://dx.doi.org/10.2527/jas.2012-5580>>.
- PRICE, A. L.; PATTERSON, N. J.; PLENGE, R. M.; WEINBLATT, M. E.; SHADICK, N. A.; REICH, D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, New York, v. 38, p. 904-909, 2006. Disponível em: <<http://dx.doi.org/10.1038/ng1847>>.

- ROSSE, I. C.; STEINBERG, R. S.; COIMBRA, R. S.; PEIXOTO, M. G.; VERNEQUE, R. S.; MACHADO, M. A.; FONSECA, C. G.; CARVALHO, M. R. Novel SNPs and INDEL polymorphisms in the 3'UTR of DGAT1 gene: in silico analyses and a possible association. **Molecular Biology Reports**, New York, v. 41, n. 7, p. 4555-4563, 2014. Disponível em: <<http://dx.doi.org/10.1007/s11033-014-3326-z>>.
- SAHANA, G.; GULDBRANDTSEN, B.; JANSS, L.; LUND, M. S. Comparison of association mapping methods in a complex pedigreed population. **Genetic Epidemiology**, New York, v. 34, p. 455–62, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.livsci.2012.12.012>>.
- SANTOS, D. J. A.; PEIXOTO, M. G. C. D.; BORQUIS, R. R. A.; VERNEQUE, R. S.; PANETTO, J. C. C.; TONHATI, H. Genetic parameters for test-day milk yield, 305-day milk yield, and lactation length in Guzarat cows. **Livestock Science**, Amsterdam, v. 152, n. 2, p. 114-119, 2013.
- ROSE JORD, E. C.; DICKSON, R. B. Growth factors, apoptosis and survival of mammary epithelial cells. **Journal of Mammary Gland Biology and Neoplasia**, New York, v. 4, n. 2, p. 229-237, 1999.
- SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **American Journal of Human Genetics**, Chicago, v. 78, p. 629-644, 2006.
- UTSUNOMIYA, U. T.; CARMO, A. S.; NEVES, H. R.; CARVALHEIRO, R.; MATOS, M. C.; ZAVAREZ, L. B.; ITO, P. K. R. K.; O'BRIEN, A. M. P. SÖLKNER, J.; PORTO-NETO, L. R.; SCHENKEL, F. S.; MCEWAN, J.; COLE, J. B.; SILVA, M. V. G. B.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; GARCIA, J. F. Genome-wide mapping of loci explaining variance in scrotal circumference in nellore cattle. **Plos ONE**, São Francisco, v. 9, 2014. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0088561>>.

- VAN DIJK, M.; VAN BEZU, J.; POUTSMA, A.; VEERHUIS, R.; ROZEMULLER, A. J.; SCHEPER, W.; BLANKENSTEIN, M. A.; OUDEJANS, C. B. The pre-eclampsia gene STOX1 controls a conserved pathway in placenta and brain upregulated in late-onset Alzheimer's disease. **Journal of Alzheimer's Disease**, Amsterdam, v. 19, n. 2, p. 673-679, 2010. Disponível em: <<http://dx.doi.org/10.3233/JAD-2010-1265>>.
- VANRADEN, P. M. Efficient methods to Compute Genomic predictions. **Journal of Dairy Science**, Champaign, v. 91, 4414-4423, 2008. Disponível em: <<http://dx.doi.org/10.3168/jds.2007-0980>>.
- VEERKAMP, R. F.; COFFEY, M.; BERRY, D.; DE HAAS, Y.; STRANDBERG, E.; BOVENHUIS, H.; CALUS, M.; WALL, E. Genome-wide associations for feed utilisation complex in primiparous Holstein-Friesian dairy cows from experimental research herds in four European countries. **Animal**, Cambridge, v. 6, n. 11, p. 1738-1749, 2012. Disponível em: <<http://dx.doi.org/10.1017/S1751731112001152>>.
- WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, Cambridge, v. 94, n. 2, p. 73-83, 2012. 10.1017/S0016672312000274>.
- WANG, H.; JIANG, L.; LIU, X.; YANG, J.; WEI, J.; XU, J.; ZHANG, Q.; LIU J.F. A Post-GWAS replication study confirming the *PTK2* gene associated with milk production traits in chinese holstein. **PLoS ONE**, São Francisco, v. 8, 2013. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0083625>>.
- WILDE, C. J.; HENDERSON, A. J.; KNIGHT, C. H. Metabolic adaptations in goat mammary tissue during pregnancy and lactation. **Journal of Reproduction and Fertility**, Cambridge, v. 76, n. 1, p. 289-298, 1986.

YU, J.; PRESSOIR, G.; BRIGGS, W. H.; BI, I. V.; YAMASAKI, M.; DOEBLEY, J. F.; MCMULLEN, M. D.; GAUT, B. S.; NIELSEN, D. M.; HOLLAND, J. B.; KRESOVICH, S.; BUCKLER, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, p. 203–208, 2006. Disponível em: <<http://dx.doi.org/10.1038/ng1702>>.

ZIMIN, A. V.; DELCHER, A. L.; FLOREA, L.; KELLEY, D. R.; SCHATZ, M. C.; PUIU, D.; HANRAHAN, F.; PERTEA, G.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; MARÇAIS, G.; ROBERTS, M.; SUBRAMANIAN, P.; YORKE, J. A.; SALZBERG, S. L. A whole-genome assembly of the domestic cow, *Bos taurus*. **Genome Biology**, Londres, v. 10, p. R42, 2009. Disponível em: <<http://dx.doi.org/10.1186/gb-2009-10-4-r42>>.

CAPÍTULO 5 - COMPARAÇÃO DE PAINÉIS DE SNPs DE ALTA DENSIDADE EM BOVINOS DA RAÇA GUZERÁ

RESUMO - Avanços nas tecnologias moleculares, como o sequenciamento do genoma e utilização de painéis de marcadores SNP, têm auxiliado no estudo e compreensão de processos genéticos. Considerando a constituição dos painéis de SNPs quanto aos genótipos raciais envolvidos e as vantagens de se empregar a seleção genômica ampla para a raça Guzerá, a avaliação preliminar destes painéis torna-se essencial. Com base nisto, os objetivos deste trabalho foram avaliar o desempenho do “Axiom® Genome-Wide BOS 1 Array” da Affymetrix e o “BovineHD BeadChip” da Illumina® para a raça, considerando quarenta e oito touros com representividade genética da população de Guzerá do Brasil. O conjunto UMD3.1 e Btau_4.2 foi usado como referências para os genes descritos. A área de cobertura do painel para o genoma foi avaliada usando análise de densidade de “Kernel”. A cobertura genética foi avaliada pelas estimativas das medidas do desequilíbrio de ligação (r^2). As análises indicaram que 156.997 SNPs (38%) e 142.890 (34%) do painel da Affymetrix são intragênicos (localizados dentro dos genes) pela referência UMD e BTAU respectivamente. Para o painel da Illumina esses valores são de 188.950 (41 %) e 162.105 (35 %) pela a referência UMD e BTAU, respectivamente. Com o painel da Affymetrix 57 % dos genes (15.492) da referência UMD e 58 % dos genes (14.613) do BTAU possuíram pelo menos um SNP intragênico. Já com o painel da Illumina os valores foram de aproximadamente de 62% dos genes da referência UMD (16.860) e Btau (15.683). Em relação à cobertura genética, foi encontrado valores de r^2 com média de 0,15 e 0,20 para distâncias de até 100 kb para os painéis da Affymetrix e Illumina, respectivamente. Os painéis apresentaram cobertura genômica adequada, possuindo alta densidade de marcadores espalhados pelo genoma após controle de qualidade, sendo que parte destes marcadores apresentou cobertura genética direta (dentro nos genes). Apesar das estimativas do r^2 terem sido maiores com o painel da Illumina, foram observados níveis altos e moderados de r^2 para pequenas distâncias com os dois painéis. Estes níveis podem ser adequados para auxiliar a obtenção de valores genéticos genômicos para animais da raça Guzerá. No entanto, a combinação destes dois painéis comerciais não indicou viabilidade.

Palavras chaves: Desequilíbrio de ligação, Seleção genômica, Zebu.

INTRODUÇÃO

Avanços nas tecnologias moleculares têm auxiliado no estudo e compreensão de processos genéticos. Entre essas tecnologias estão o sequenciamento com montagem do genoma e utilização de marcadores moleculares para polimorfismo de nucleotídeo único (SNP). O sequenciamento, ultimamente realizado por *BAC* (“bacterial artificial chromosome”) e WGS (“whole-genome shotgun”), possibilita a montagem do genoma de uma determinada espécie, permitindo, portanto, estabelecer o posicionamento (cromossomo e número de pares de base) de sequências menores e dos SNPs no genoma (MATUKUMALLI et al., 2009). Pela customização do processo de genotipagem, a tecnologia dos marcadores SNPs pode ser utilizada em larga escala para indivíduos de uma determinada população. Painéis contendo milhares de marcadores SNPs estão sendo empregados em genética humana bem como em espécie de interesse zootécnico para análise de associação de doenças, estudo de variabilidade genética e para seleção na pecuária (ZENGER et al. 2006; BARENDSE et al. 2007; BIFFI et al. 2010).

Os dois principais projetos genômicos realizados para a espécie bovina é o BTAU e o UMD, cujo sequenciamento utilizou apenas um animal da raça Hereford (BGSAC 2009; ZIMIN et al., 2009). Este sequenciamento contribuiu para grande parte da detecção de SNPs em *Bos taurus*. O “BovineHD SNP BeadChip” da Illumina e o “Axiom® Genome-Wide BOS 1 Array” da Affymetrix, contêm, além de muitos destes SNPs, também informações de SNPs baseados outras raças taurinas, raças africanas e raças zebuínas, das quais, pode-se destacar o Brahman, Nelore e Gir (ILLUMINA, 2012; AFFYMETRIX 2015).

O Guzerá é uma das raças da subespécie *Bos taurus indicus* mais antigas, sendo apontada como umas das raças que compunha o primeiro agrupamento étnico dos zebuínos (Primeiro Tipo Básico Zebuino), cujo tronco, originou grande parte das raças indianas atuais (JOSHI; PHILIPPS, 1954). A população de Guzerá está a pouco tempo exposta à seleção artificial, quando comparada aos taurinos, e os maiores ganhos genéticos foram observados nos últimos anos por meio dos programas de melhoramento genético, principalmente para leite (PEIXOTO et al., 2014).

Com o advento do painel de SNPs tornou-se possível a análise simultânea de muitas informações moleculares do genoma animal. Sua utilização em um programa de melhoramento visa acelerar o ganho genético, principalmente em rebanhos leiteiros. Considerando a constituição dos painéis de SNPs quanto aos genótipos raciais envolvidos, e a necessidade de empregar a seleção genômica ampla, a avaliação preliminar destes painéis torna-se essencial para a raça Guzerá.

A seleção genômica explora o desequilíbrio de ligação entre os marcadores, pois assume que os efeitos dos segmentos cromossômicos analisados representam também o desequilíbrio de ligação entre o marcador e o QTL (“Quantitative Trait Locus”) para toda a população. Os mapas de desequilíbrio também são importantes para analisar a diversidade genética entre raças bovinas e identificar regiões genômicas que estão sofrendo influência de seleção (MCKAY et al., 2007).

Com base nisto, os objetivos deste trabalho foram avaliar o desempenho do painel “Axiom® Genome-Wide BOS 1 Array” da Affymetrix e o “BovineHD BeadChip” da Illumina® quanto ao número de polimorfismos, cobertura física, distribuição nas regiões gênicas e desequilíbrio de ligação entre os marcadores dos painéis. Isto foi realizado de forma a fornecer subsídios para a definição de um painel comercial e de uma população de animais a serem genotipados (população referência) para um futuro emprego da seleção genômica na raça Guzerá.

MATERIAL E MÉTODOS

Genotipagem

Foram genotipados 48 touros pertencentes ao Programa de Melhoramento de Guzerá para Leite, sendo estes animais selecionados por serem os mais influentes dentro da população. A genotipagem foi realizada utilizando “BovineHD BeadChip” da Illumina e o “Axiom® Genome-Wide BOS 1 Array” da Affymetrix a partir de amostras de sangue dos animais. O “BovineHD BeadChip” contém mais de 770 mil marcadores espaçados a uma distância média de 3,43 kb entre os marcadores, enquanto o “Axiom® Genome-Wide BOS 1 Array” tem mais de 640 mil marcadores e uma distância média de 4,1

kb. As informações básicas analisadas consideraram, além dos dados dos genótipos dos animais, o nome do marcador, cromossomo em que se encontra e a localização em pares de base (bp).

Controle de qualidade dos genótipos nas amostras

Inicialmente, o painel da Affymetrix conteve 648.874 e o da Illumina 786.799 marcadores. Destes, foram eliminados marcadores sem posicionamento no genoma presentes no Y e DNA mitocondrial (QC1) resultando em 625.253 marcadores para o painel da Affymetrix e 781.770 marcadores para o painel da Illumina. Um terceiro painel formado pela combinação destes dois painéis também foi estudado. Este novo painel, após a eliminação citada acima, conteve 1.297.938 SNPs com espaçamento médio próximo de 2 kb. Para este número de SNPs (QC1) dos três painéis foram analisados o número de polimorfismos e no DNA.

Objetivando uma seleção de marcadores consistentes e que possivelmente serão utilizados para seleção genômica da população, foram considerados marcadores dos cromossomos autossômicos e X, que apresentaram “call rate” $\geq 0,95$ e MAF $\geq 0,05$ (frequência de alelos raros); e indivíduos com “call rate” $\geq 0,90$ (QC2). Após essa edição, o número de marcadores autossômicos foi igual a 415,133 SNPs para o painel da Affymetrix, 465.703 para o painel da Illumina, e 763.259 SNPs para os painéis combinados. Para este total foi realizado estudo de densidade de “Kernel” e quantificação dos SNPs por gene descrito. Para os SNPs autossômicos também foram realizadas análises do desequilíbrio de ligação (DL).

Referência de montagem do Genoma Bovino

Os conjuntos UMD3.1 e Btau_4.2 combinaram sequenciamento *BAC* (“bacterial artificial chromosome”) e WGS (“whole-genome shotgun”), os quais foram associados aos cromossomos utilizando métodos de mapeamento diferentes. Como os locais dos mapas dos SNPs baseiam-se no conjunto do UMD3.1, foram utilizadas as informações genéticas do “Cow Genome Resources” do “National Center for Biotechnology Information” (NCBI) (Recursos do Genoma Bovino) (<http://www.ncbi.nlm.nih.gov>). As informações

incluíram o símbolo do gene, a posição inicial, posição final, orientação no cromossomo, e a descrição do gene.

Estimação da Densidade dos marcadores ao longo de cada cromossomo

Objetivando estudar a área de cobertura dos três painéis para o genoma em estudo foi realizado também um estudo de densidade de “Kernel” (ROSENBLATT, 1956) utilizando o programa estatístico R (R CORE TEAM, 2013). Esta análise consiste em descrever o número de SNP ao longo do genoma bovino. Sendo $x_1, x_2, x_3 \dots, x_n \sim f$ amostras com distribuição independente e igualmente distribuída de uma variável aleatória X , onde x_i é o local de observação em bp (pares de base) do i -ésimo marcador. Então o estimador da densidade de “Kernel” será:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

em que h é o parâmetro de suavização, também chamado de largura de banda, e $K(\bullet)$ são núcleos de distribuição gaussiana com média zero e variância um (1), que pode ser representado como:

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - x_i)^2}{2h^2}}$$

em que h é um parâmetro livre, e tem uma forte influência nos resultados das estimativas de densidade de “Kernel” (gráficos). Para o presente estudo foi considerado $h=0,05$ (MICHELIZZI et al. 2011),.

Desequilíbrio de ligação

O DL entre dois marcadores foi avaliado usando o r^2 (HILL; ROBERTSON, 1968) e o valor absoluto de D' (LEWONTIN, 1988). O r^2 foi calculado da seguinte forma:

$$r^2 = \frac{(freq.AB * freq.ab - freq.Ab * freq.aB)^2}{(freq.A * freq.a * freq.B * freq.b)}$$

em que *freq.A*, *freq.a*, *freq.B* and *freq.b* são as frequências dos alelos A, a, B e b, respectivamente, e *freq.AB*, *freq.ab*, *freq.aB* e *freq.Ab* são as frequências dos haplótipos Ab, ab, aB e Ab na população, respectivamente.

RESULTADOS E DISCUSSÃO

Com “Axiom® Genome-Wide BOS 1 Array” foi realizada a genotipagem com sucesso (“call rate”>0,80) de aproximadamente 96% (626.150) dos marcadores, dos quais aproximadamente 80% (501.957) eram polimórficos. Para 20 marcadores não houve nenhuma genotipagem (“call rate”>0,80). Já com “BovineHD BeadChip” 99% dos marcadores (766.545) tiveram genotipagem com sucesso, com 78% (766.545) dos marcadores polimórficos. O número de marcadores sem reconhecimento no genoma foi de 186 com este painel.

A média das frequências do alelo de menor frequência (MAF) para os SNPs autossômicos foram praticamente iguais entre os painéis, sendo os valores iguais à 0,25 e 0,24 para o painel da Affymetrix e Illumina respectivamente. Estes resultados foram ligeiramente superiores à média de MAF para a raça Nelore relatada por Matukumalli et al. (2009) (0,19) e pelo Bovine Hapmap Consortium (GIBBS, et al., 2009) (0,20). Foi observado uma leve tendência de diminuição da frequência dentro das classes de MAF quando aumentou os valores de MAF (Figura 1). McKay et al. (2007) descreveram frequências e tendências semelhantes para bovinos da raça Brahman e Nelore às observadas neste trabalho. Entretanto, estes mesmos autores apresentaram diferentes frequências e tendência (aumentar a frequência com o aumento do MAF) para raças européias. Em relação aos cromossomos a média de MAF variou de 0,24 a 0,26 para o painel da Affymetrix e de 0,23 a 0,26 para o painel da Illumina. Khatkar et al. (2008), relatam média de MAF ligeiramente maiores (0,27 a 0,32) para os cromossomos em bovinos da raça Holandesa da Austrália. A maior MAF observada para esta raça pode ser devido ao fato que a maior parte dos SNPs utilizado nos painéis foram baseado em regiões polimórficas de raças taurinas.

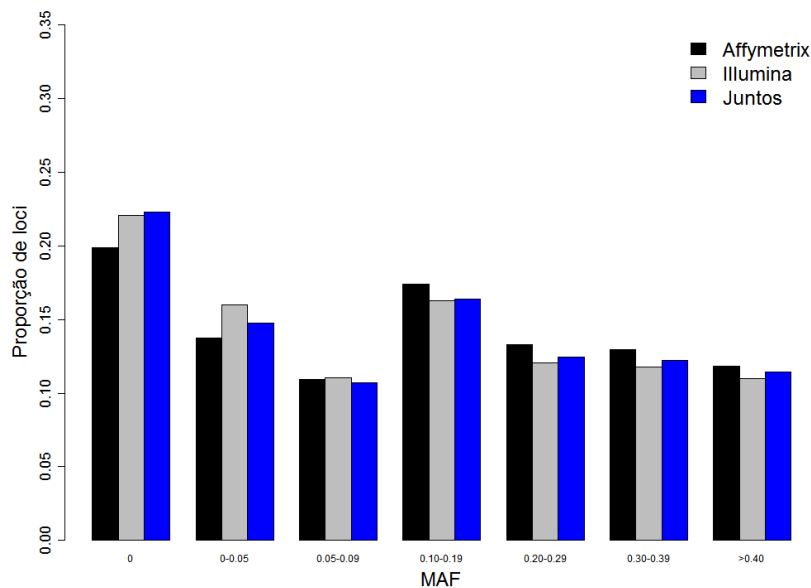


Figura 1 – Proporção média dos alelos de menor frequência (MAF) para os diferentes painéis estudados.

Após a imposição dos critérios do QC2, foi obtido o número de 415.133 SNPs com o painel da Affymetrix, sendo que 403.308 SNPs destes foram autossômicos, e 465.703 com painel da Illumina, dos quais 444.691 SNPs eram autossômicos. Já a combinação dos dois painéis resultou em 763.259 dos quais 735.647 eram autossômicos. O espaçamento médio entre os marcadores úteis (QC2) melhorou expressivamente quando os dois painéis foram combinados, desta forma o espaço médio diminui de 6,4 kb (Affymetrix) e 5,7 kb (Illumina) para 3,5 kb (juntos).

Pela descrição atualizada do genoma bovino para a referência UMD e Btau, o número de genes nos cromossomos bovinos variaram respectivamente de 299 e 304 no cromossomo 27 a 1.592 no BTA7 e 1.463 no BTA3. Já o número de SNPs editado variou de 7.507 no BTA 25 à 24.596 pelo painel da Affymetrix; de 7.758 no BTA8 a 28.444 BTA1 pelo painel da Illumina; e de 1.379 no BTA25 a 45.836 no BTA1 para os dois painéis combinados (Figura 2). Para todos os cromossomos, independente do painel ou a referência utilizada, há muito mais SNPs do que genes descritos. Isto indica que a cobertura dos painéis foi suficiente para todos os cromossomos. Em todos os cromossomos o número de SNPs presente no painel da Illumina foi maior que o da Affymetrix com a exceção do BTA23.

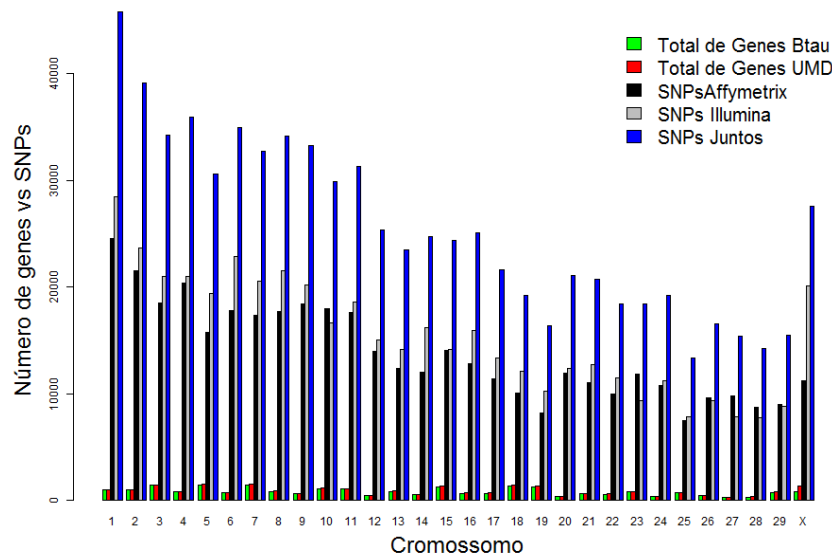


Figura 2 – Número de SNPs de diferentes painéis e genes de diferentes referências de montagem em cada cromossomo bovino.

As análises indicaram que 156.997 SNPs (38%) e 142.890 (34%) do painel da Affymetrix são intragênicos (localizados dentro dos genes) pela referência UMD e BTAU respectivamente. Para o painel da Illumina esses valores são de 188.950 (41 %) e 162.105 (35 %) e com os dois painéis juntos 300.586 (39%) e 265.404 (35%) de SNPs intragênicos pela a referência UMD e BTAU, respectivamente. Considerando que pelo UMD a proporção de DNA codificante é de aproximadamente 33% e pelo BTAU 34 %, todos os painéis apresentaram proporções semelhantes e levemente maiores que esperado. Com o painel da Affymetrix 57 % dos genes (15.492) da referência UMD e 58 % dos genes (14.613) do BTAU possuíram pelo menos um SNP intragênico (Figura 3). Com o painel da Illumina estes valores foram melhores, com aproximadamente 62% dos genes da referência UMD (16.860) e Btau (15.683) possuindo ao menos um SNP intragênico. A vantagem da combinação entre os dois painéis comerciais pode observado quando 70% dos genes com referência UMD (18.766) e 69% dos genes BTAU (17.314) tiveram ao menos um SNP intragênico além de apresentar maior número de SNPs dentro de cada gene.

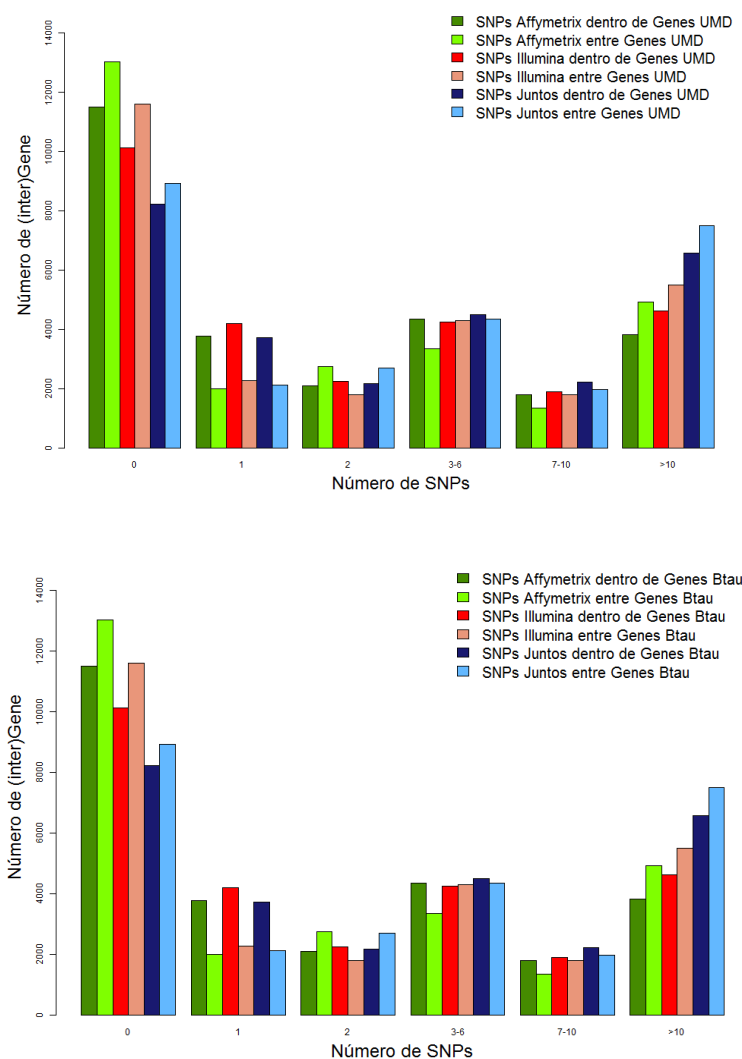


Figura 3 - Distribuição dos SNPs selecionados para a população da raça Guzerá de cada painel estudado pelos genes descritos na referência UMD e Btau.

Já em relação as regiões intergênicas, o painel da Affymetrix apresentou 14.390 regiões com ao menos um SNP pela referência UMD e 15.038 pela referência Btau (Figura 3). O painel da Illumina apresentou ao menos um SNP em 18.681 regiões intergênica da referência UMD e 15.711 Btau. O painel combinado não aumentou muito o número de regiões intergênica em relação ao Illumina: 18.681 regiões pela referência UMD e 17.252 regiões pela Btau. No entanto, este o número de SNPs contidos nestas regiões do painel combinado (juntos) aumentou. O número de SNPs dentro de

genes foi satisfatório, indicando uma grande porção de cobertura genética direta (dentro dos genes), ainda que a maioria dos SNPs estejam entre os genes. Apesar das informações codificantes estão contidas nos genes, a parte intergênica tem importante influência na regulação expressão gênica, replicação do DNA, montagem, estruturação e manutenção do cromossomo e auxílio a separação meiótica (LUDWIG 2002; CARROL; GOMPEL; BPRUDHOMME, 2008; SUBIRANA; MESSEGUER, 2010).

De modo geral há muito mais genes com SNP por cromossomo do que sem SNP (Figura 4). Com o painel da Affymetrix os BTA7, BTA29 e X pela referência UMD e Btau e o BTA25 pelo UMD tiveram número de genes com SNPs menor que o número de genes sem SNPs. Já com o painel da Illumina os BTA15 e X pelo UMD e BTAU tiveram número de genes com SNPs menor que o número de genes sem SNPs. Com os painéis combinados todos os cromossomos tiveram número de genes com SNPs menor que o número de genes sem SNPs.

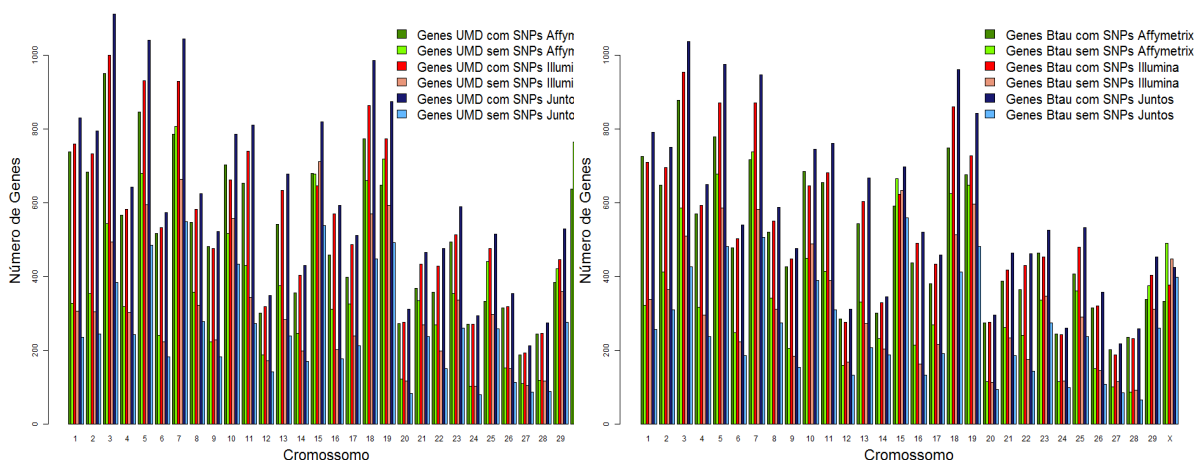
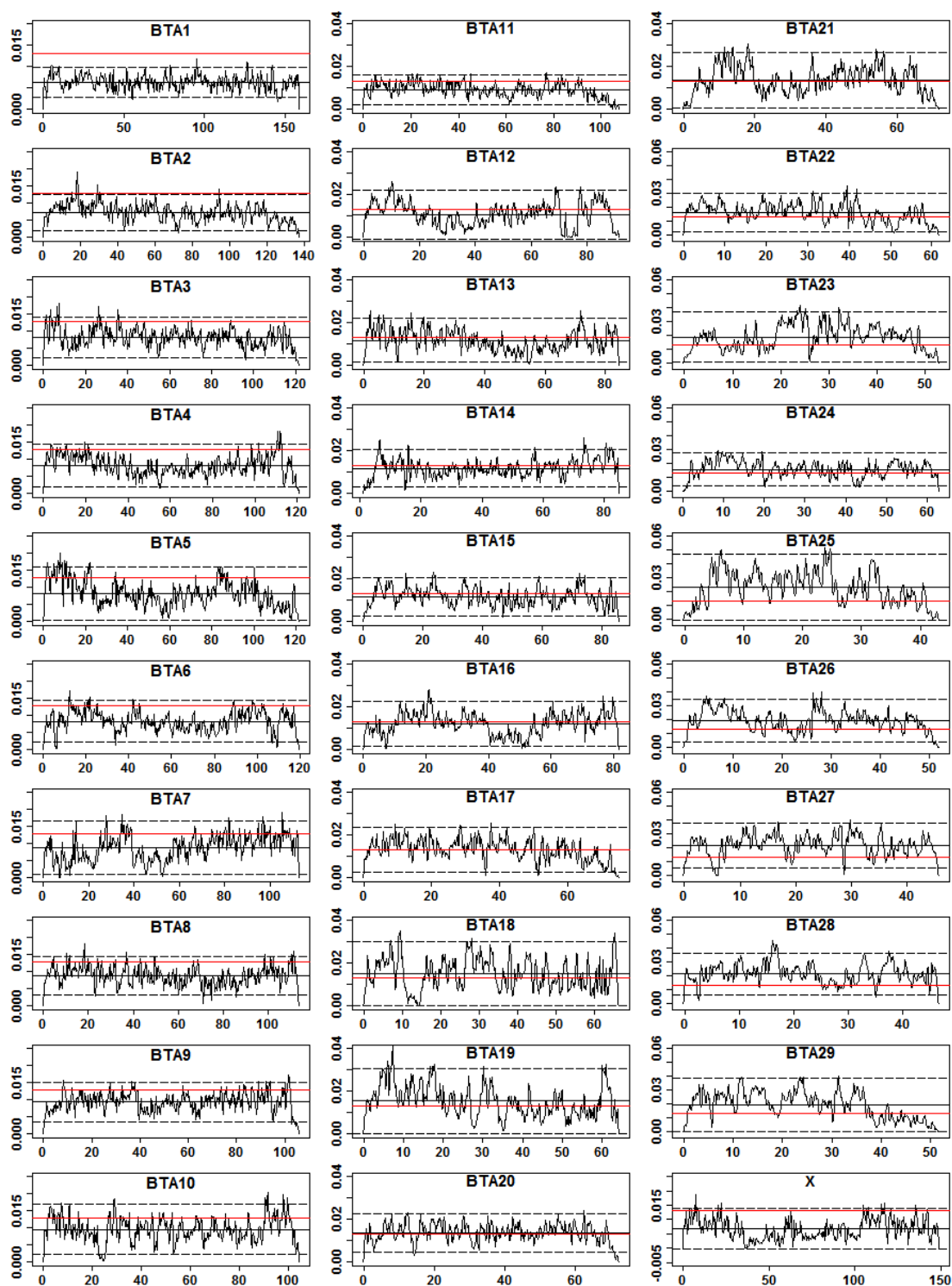


Figura 4 - Número de genes pela referência UMD e Btau com e sem SNPs de diferentes painéis estudados por cromossomo.

Os gráficos de densidade de "Kernel" mostraram gráficos com maior variação número de núcleos quando $h=0,05$ que quando $h>0,1$ (resultado não apresentado). Como o espaçamento entre marcadores dos dois painéis comerciais é de aproximadamente 4 Kb, foi testado $h=0,004$ não sendo observado diferenças marcantes em relação a 0,05. A adequação do $h=0,05$ também foi observada por Michelizzi et al. (2011), porém trabalhando com "Illumina Bovine50SNP BeadChip" em búfalos. Segundo estes autores, a

largura das bandas dos núcleos de densidade deveria ser aproximadamente a distância média dos SNPs, o que justificaria a utilização de $h=0,05$ (intervalo entre marcadores igual a aproximadamente 50 Kb). Quanto maior a densidade maior será a cobertura por marcadores para determinada área do cromossomo. Esta análise então vai indicar o quão homogêneo os marcadores se distribui ao longo do genoma e a escala é inerente ao painel utilizado.



— Média do cromossomo - - - 2 desvios padrão da média do cromossomo.
 — Média do genoma.

Figura 5 - Densidade de "Kernel" dos SNPs do "Axiom® Genome-Wide BOS 1 Array" da Affymetrix para cada cromossomo.

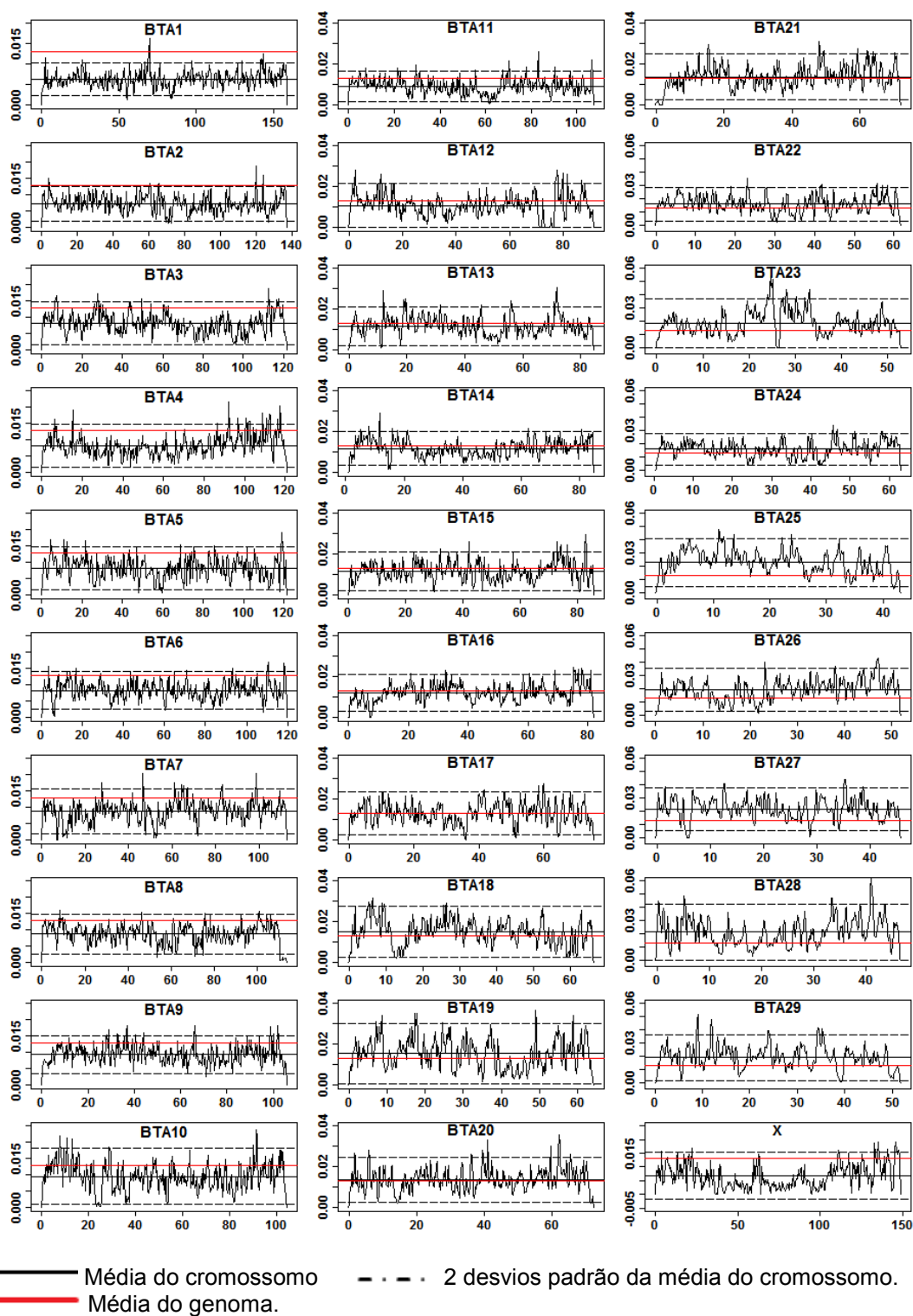


Figura 6 - Densidade de “Kernel” dos SNPs no “Illumina Bovine SNP770 BeadChip” para cada cromossomo.

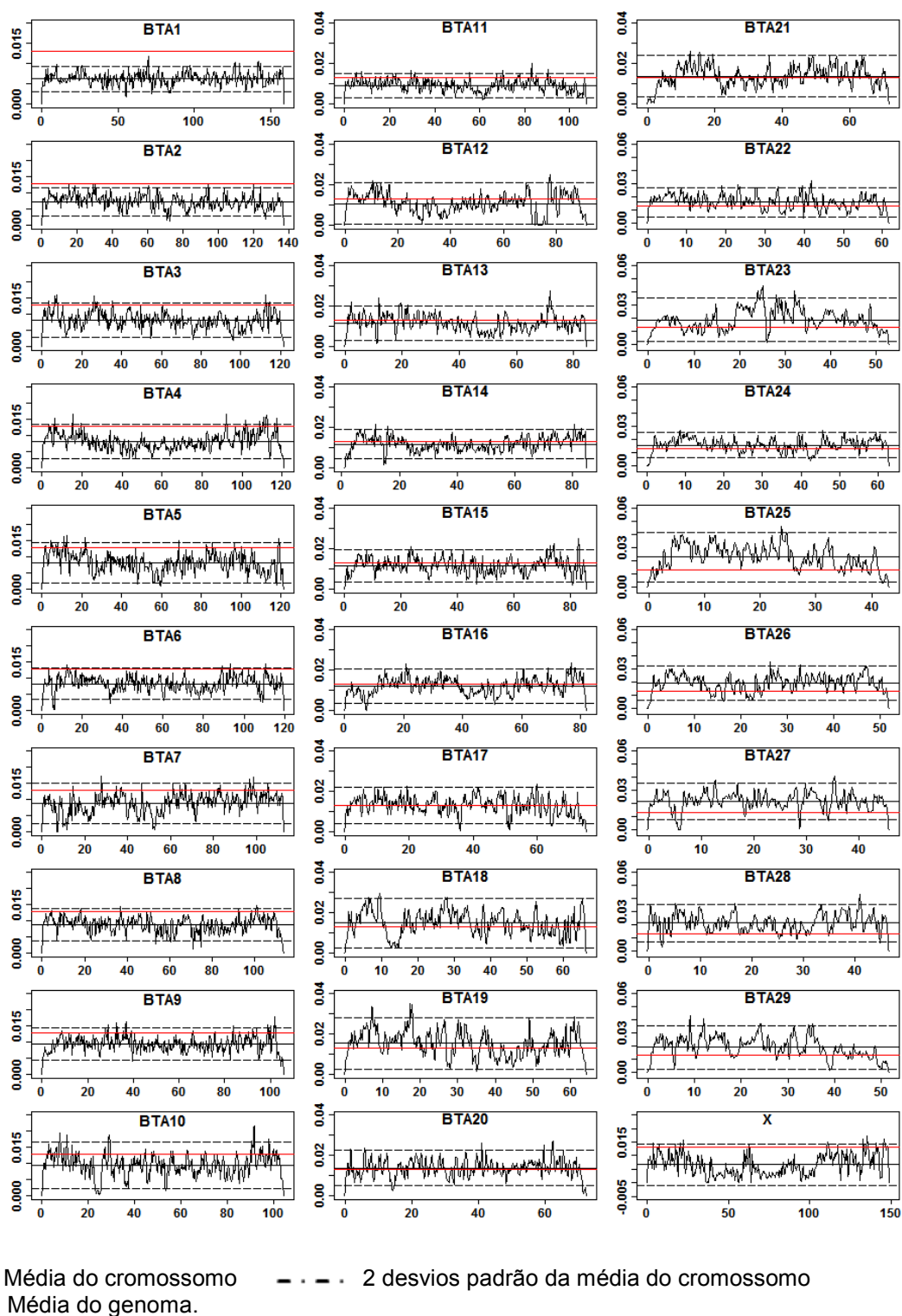


Figura 7 - Densidade de “Kernel” dos SNPs da combinação dos painéis (Juntos) para cada cromossomo.

Foi definida como região rica em SNPs aquela que se apresentou acima de dois desvios padrão da média, bem como região pobre aquela que se apresentou abaixo de dois desvios da média (Figura 7, 8 e 9). Quanto mais próximo de zero é o valor, menor é a cobertura para uma determinada região.

Como esperado, maior parte das regiões ficaram entre os dois desvios de densidade da média do cromossomo.

De um modo geral a distribuição dos SNPs foi regular pelos cromossomos, não apresentando muitas regiões ricas ou pobres. O painel da Illumina mostrou distribuição mais homogênea ao longo dos cromossomos quando comparado ao painel da Affymetrix. Sendo assim, a porção final do BTA29 apresentou um declínio na cobertura proporcionada pelo painel da Affymetrix, enquanto que com o painel da Illumina esta região se manteve em torno da média do cromossomo. Com os painéis combinados a cobertura desta região melhorou. No BTA7 e BTA16 foi observado um declínio súbito com o painel da Affymetrix (“gap”) que não foi observado com o painel da Illumina. No entanto, no BTA10 e BTA18 os “gaps” foram observados com os dois painéis. Foi observado que os BTA3, BTA4, BTA5, BTA10, BTA12, BTA14 e BTA23 com o painel da Affymetrix não apresentaram regiões ricas em SNP, apresentando inclusive algumas regiões pobres. O painel da Illumina apresentou este mesmo perfil com o BTA10, BTA12, BTA14 e BTA 23. Os cromossomos BTA7, BTA10, BTA13, BTA17, BTA 23, BTA 29 e X apresentaram maior dinâmica em relação à variação entre regiões ricas e pobres em SNPs nos dois painéis comerciais.

Em todos os painéis foi observado que no BTA1 os dois desvios superior foi menor que a média do genoma, e no cromossomo X esse desvio foi igual a média do genoma no painel da Illumina (Figura 5, 6 e 7), e um pouco maior no painel da Affymetrix. Nos BTA17, BTA20 e BTA21 foi observado, com os dois painéis, que a média do cromossomo foi próxima a média do genoma. Foi observado que nos BTA18, BTA19, e do BTA22 até o BTA29 a média do cromossomo foi maior que a média do genoma. Esta relação indica que os cromossomos maiores apesar de possuírem maior quantidade de marcadores SNPs ainda apresentam muitas regiões sem coberturas, diferentemente dos cromossomos menores.

Para avaliar a cobertura destes painéis de alta densidade por meio do DL, as distâncias curtas e os marcadores adjacentes foram estudados. Todos os pares de SNPs analisados a uma distância menor igual a 100 Kb produziram um total de 6.877.630, 8.291.614 e 22.019.133 combinações para a estimação do DL pelos painéis da Affymetrix, Illumina e com os painéis

combinados (Juntos), respectivamente (Tabela 1). A média geral foi de 0,15, 0,20 e 0,10 para r^2 com os painéis da Affymetrix, Illumina e Juntos. Pela Tabela 1 é possível observar a diferença nos padrões de DL pelos diferentes painéis, como também o efeito negativo na média do DL pela interação entre os painéis Affymetrix e Illumina (no painel Juntos).

Tabela 1 - Desequilíbrio de ligação (r^2) entre pares (N) de SNPs separados por diferentes distâncias até 100 kb para todos os autossomos em todos os painéis estudados.

Distância (kb)	Affymetrix			Illumina			Juntos		
	N	Média $r^2 \pm DP^1$	% $r^2 > 0,3^2$	N	Média $r^2 \pm DP^1$	% $r^2 > 0,3^2$	N	Média $r^2 \pm DP^1$	% $r^2 > 0,3^2$
0 – 1	74567	0,24 $\pm$ 0,25	30	37062	0,32 $\pm$ 0,29	43	208128	0,12 $\pm$ 0,21	14
1 – 2	74477	0,23 $\pm$ 0,24	29	72444	0,31 $\pm$ 0,29	41	224742	0,15 $\pm$ 0,23	17
2 – 3	72111	0,22 $\pm$ 0,24	28	80292	0,30 $\pm$ 0,28	40	226181	0,15 $\pm$ 0,23	18
3 – 4	70912	0,22 $\pm$ 0,23	27	90679	0,30 $\pm$ 0,28	39	232294	0,15 $\pm$ 0,23	18
4 – 5	70419	0,21 $\pm$ 0,23	26	89317	0,29 $\pm$ 0,28	38	230906	0,15 $\pm$ 0,22	17
5 – 10	349625	0,20 $\pm$ 0,22	24	435725	0,27 $\pm$ 0,27	36	1134878	0,14 $\pm$ 0,22	16
10 – 20	695320	0,18 $\pm$ 0,21	21	856734	0,25 $\pm$ 0,26	32	2243857	0,13 $\pm$ 0,20	14
20 – 30	695384	0,17 $\pm$ 0,20	19	845835	0,23 $\pm$ 0,25	28	2221353	0,12 $\pm$ 0,19	13
30 – 40	689768	0,15 $\pm$ 0,19	17	838262	0,21 $\pm$ 0,24	26	2207566	0,11 $\pm$ 0,18	11
40 – 50	687152	0,14 $\pm$ 0,18	15	834533	0,19 $\pm$ 0,23	24	2199231	0,10 $\pm$ 0,17	10
50 – 60	683206	0,14 $\pm$ 0,18	14	828954	0,18 $\pm$ 0,22	22	2189822	0,10 $\pm$ 0,16	10
60 – 70	681621	0,13 $\pm$ 0,17	13	823057	0,17 $\pm$ 0,22	20	2182601	0,09 $\pm$ 0,16	9
70–80	679678	0,12 $\pm$ 0,16	12	820545	0,16 $\pm$ 0,21	19	2174119	0,09 $\pm$ 0,15	8
80 – 90	676798	0,12 $\pm$ 0,16	11	820090	0,16 $\pm$ 0,20	18	2173650	0,08 $\pm$ 0,15	8
90 – 100	676592	0,11 $\pm$ 0,15	10	818085	0,15 $\pm$ 0,20	17	2169805	0,08 $\pm$ 0,14	7
0 – 100	6877630	0,15 $\pm$ 0,19	11	8291614	0,20 $\pm$ 0,24	24	22019133	0,10 $\pm$ 0,18	11

¹Desvio padrão; ²Percentage de pares de SNP com $r^2 > 0,3$.

Para qualificar a diferença entre os painéis, o teste de Tukey foi empregado para diferentes classes de distâncias até 100 kb, onde foi observado a significância entre todas as classes ($p < 2,2e^{-16}$). Dado isto, duas hipóteses podem ser consideradas: a primeira é devida diferenças no processo de genotipagem entre os painéis comerciais; a segunda é devido a alocação de marcadores diferenciadamente entre os painéis, ou seja, o painel Illumina pode ter alocado muitos SNPs dentro das mesmas regiões haplotípicas (baixa recombinação) enquanto o painel Affymetrix aloucou poucos SNPs dentro destas regiões e priorizou distribuir seus marcadores em diferentes regiões haplotípicas. Para verificar a primeira hipótese, o conjunto de 101.919 SNPs comuns aos dois painéis comercial foi analisado. Após a aplicação de um rigoroso controle de qualidade para amostra (“call rate”=1 e MAF $\geq$ 0,05), restaram 43.943 SNPs. O grau de concordância entre os painéis, dado pelo coeficiente de correlação de Pearson, foi de 0,53 para MAF e 0,25 para o r^2 .

O teste t (Student) pareado foi empregado para verificar as diferenças (Affymetrix-Illumina) na MAF entre os marcadores e r^2 entre os pares de marcadores. A diferença de MAF para os marcadores variou de -0,45 à 0,44 com média de $-0,008 \pm 13$ (Figura 8). No entanto, pelo teste t ($t = -13,165$ e 4394 graus de liberdade) a hipótese de igualdade entre os valores ($\alpha > 5\%$) foi descartada ($p\text{-value} < 2,2e-16$). A diferença de r^2 para os pares de marcadores variou de -1 a 1 com média de $0,010 \pm 0,29$. Este também pelo teste t ($t = 24,34$ e 541788 graus de liberdade) teve hipótese de igualdade (nula) entre os valores descartada ($p\text{-value} < 2,2e-16$). Apesar da média ter sido próximo a zero para os dois parâmetros estudados (MAF e r^2), os testes indicaram que existe diferenças na atribuição de genótipos pela leitura dos sinais ou pela clusterização destes sinais. Este resultado implica na impossibilidade de se combinar estes painéis, apesar do melhor desempenho observado para distribuição e cobertura física desta combinação. Resultados diferentes foram reportados por Rincon et al. (2011) que avaliando a performance destes dois painéis observaram que 89 % dos SNPs em comum aos dois painéis tiveram concordância de 100% para os genótipos. No entanto, diferentemente deste trabalho, estes autores estudaram duas raças taurinas (Jersey e Holstein). Possivelmente por estes painéis possuírem em sua constituição menor quantidade de polimorfismos específico para zebuínos pode ter contribuído para o resultado encontrado.

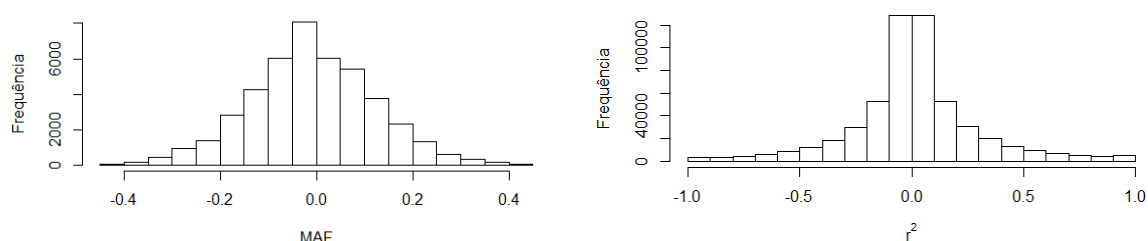


Figura 8 – Histogramas da diferença dos valores de MAF entre os marcadores e da diferença de r^2 para os pares de marcadores do painel Affymetrix e Illumina.

Admitindo que ainda existem diferenças entre os padrões de DL dos painéis devido a alocação de marcadores, avaliar qual o painel comercial que melhor genotipou é uma tarefa difícil. No entanto, pela literatura é possível

obter referências sobre os padrões de DL esperado. Silva et al. (2010) analisando o DL em 25 touros da raça Gir com um painel de 54.000 SNPs da Illumina obteve valor semelhante para média de r^2 (0,21) do painel da Illumina. Por cromossomo, estes autores também reportaram valores semelhantes de r^2 (0,20 a 0,24). Os resultados encontrados por McKay et al. (2007) para zebuínos também foram semelhantes aos encontrados neste trabalho com os dois painéis comerciais (0,15 a 0,20). Já valores menores e semelhantes aos encontrados pelo painel da Affymetrix foram encontrados por de Roos et al. (2008) para a raça Holandesa, Jersey e Angus ($r^2 \leq 0,15$) e Espigolan et al. (2013) para a raça Nelore (0,11). Entretanto todos estes autores utilizaram um painel da Illumina de 50 k ou 770 k, e não há estudos disponíveis com o painel da Affymetrix em bovinos ou a comparação destes painéis em outras raças. Além do processo de genotipagem e da possível diferença de alocação, o baixo número de animais amostrado pode também ter colaborado para a variação da medida de DL (YAN et al., 2009).

Para analisar o declínio das medidas de DL de acordo com a distância física entre os marcadores, os pares de SNPs foram classificados em intervalos de kb e a média do r^2 foram estimados para todo o genoma (Tabela 1). Como esperado os valores de DL diminuíram com o aumento da distância física entre os marcadores dos diferentes painéis. Apenas a primeira classe de distância, o painel combinado obteve menor estimativa que a segunda classe de menor distância. Conforme aumentou o intervalo, as estimativas diminuíram possibilitando um menor desvio padrão. Essa diminuição é esperada, uma vez que com o aumento do intervalo aumenta a frequência das estimativas iguais a zero, reduzindo a variação.

Níveis médios de r^2 altos ($\geq 0,30$) foram observados apenas com o painel da Illumina para distâncias físicas menores que 4 Kb. Níveis moderados (0,20 – 0,30) de r^2 foram observados para distâncias menor que 10 kb com o painel da Affymetrix, menor que 40 kb com o painel da Illumina e não foi observado este nível com o painel combinado. Apesar do pior resultado deste último painel para média e porcentagem de $r^2 > 0,30$, em termos absolutos com este painel observou-se maior número de pares de marcadores com $r^2 > 0,30$. Em geral para os três painéis, a porcentagem de r^2 maior que 0,3 foi baixo e não chegou a 50 % mesmo em distâncias curtas. Esses valores foram mais

baixos aos encontrados por Sargolzaei et al. (2008), que observaram que mais de 60 % dos marcadores apresentaram média maior que 0,30 em distâncias entre 0 e 1 Kb, estudando 821 touros Holandês e utilizando 5.564 SNPs.

Já analisando a média por cromossomo o r^2 para o painel da Affymetrix, Illumina e Juntos variou, respectivamente, de 0,11 a 0,16; 0,17 a 0,22; 0,09 a 0,11, com os menores valores observados no cromossomo 23 e os maiores no cromossomo 6. Foi observada tendência, embora pequena, da estimativa de r^2 diminuir com a diminuição do tamanho do cromossomo (não mostrado). Resultados semelhantes foram relatados por Bohamanova, Salgorzaei e Schenkel (2010), que estimaram DL para animais da raça Holandesa. Segundo Arias et al. (2009), a taxa de recombinação diminui com o aumento do comprimento do cromossomo. Desta forma, quanto menor o cromossomo espera-se que menor seja o DL.

As médias de r^2 entre os pares de marcadores adjacentes foram de 0,23, 0,30 e 0,14 para os painéis da Affymetrix, Illumina e Juntos, respectivamente. Por cromossomo, a média do r^2 foi maior que 0,22, 0,29 e 0,14 para os painéis da Affymetrix, Illumina e Juntos, respectivamente (Tabela 2). Pelo teste de Tukey, a média entre os marcadores adjacentes foram diferentes entre os três painéis. O pior resultado observado com a combinação dos painéis, mas uma vez implica que o processo atribuição de genótipos dos dois painéis possuem diferenças. Isto diminui as chances da hipótese de diferenças de alocação entre os painéis, já que o valor mínimo de r^2 deveria ser o mínimo um dos dois painéis em separados. A interlocação de marcadores que possuem atribuição de genótipos diferentes determina a menor correlação entre estes. Desta forma considerando o custo dos dois painéis e o benefício obtido pelo maior número de informações, a combinação dos painéis comercial não deve ser utilizada.

Tabela 2 - Distância média dos marcadores analisados (em bp) e média e desvio padrão (DP) de desequilíbrio de ligação (r^2) entre todos os pares de SNPs adjacentes para todos os autossomos com os diferentes painéis estudados.

BTA	Affymetrix		Illumina		Juntos	
	Distância média(bp)	Média r^2 $\pm$ DP ¹	Distância média	Média r^2 $\pm$ DP ¹	Distância média	Média r^2 $\pm$ DP ¹
1	7108	0,23 $\pm$ 0,24	7374	0,30 $\pm$ 0,29	4031	0,14 $\pm$ 0,23
2	6893	0,22 $\pm$ 0,24	7692	0,31 $\pm$ 0,29	4004	0,14 $\pm$ 0,23
3	7242	0,23 $\pm$ 0,24	7567	0,31 $\pm$ 0,29	4052	0,14 $\pm$ 0,23
4	6508	0,23 $\pm$ 0,24	7521	0,30 $\pm$ 0,28	3900	0,14 $\pm$ 0,23
5	8603	0,24 $\pm$ 0,25	8448	0,31 $\pm$ 0,29	4669	0,15 $\pm$ 0,23
6	7471	0,24 $\pm$ 0,25	6972	0,32 $\pm$ 0,29	4034	0,15 $\pm$ 0,24
7	7227	0,23 $\pm$ 0,24	7144	0,31 $\pm$ 0,29	3997	0,15 $\pm$ 0,23
8	6980	0,23 $\pm$ 0,24	6678	0,31 $\pm$ 0,28	3793	0,14 $\pm$ 0,23
9	6253	0,22 $\pm$ 0,24	6843	0,31 $\pm$ 0,28	3634	0,14 $\pm$ 0,22
10	6393	0,22 $\pm$ 0,24	8031	0,31 $\pm$ 0,29	3968	0,14 $\pm$ 0,22
11	6622	0,22 $\pm$ 0,24	7715	0,31 $\pm$ 0,29	3969	0,14 $\pm$ 0,23
12	7274	0,22 $\pm$ 0,24	7997	0,30 $\pm$ 0,28	4223	0,14 $\pm$ 0,22
13	7558	0,24 $\pm$ 0,24	7569	0,30 $\pm$ 0,28	4114	0,14 $\pm$ 0,23
14	7680	0,22 $\pm$ 0,24	6563	0,31 $\pm$ 0,28	3888	0,14 $\pm$ 0,23
15	6677	0,22 $\pm$ 0,24	7808	0,29 $\pm$ 0,28	3997	0,14 $\pm$ 0,22
16	6978	0,24 $\pm$ 0,25	6853	0,31 $\pm$ 0,29	3791	0,15 $\pm$ 0,24
17	7151	0,23 $\pm$ 0,24	7332	0,31 $\pm$ 0,29	3994	0,15 $\pm$ 0,23
18	7138	0,23 $\pm$ 0,25	7195	0,31 $\pm$ 0,29	4018	0,15 $\pm$ 0,24
19	8555	0,23 $\pm$ 0,25	8151	0,29 $\pm$ 0,28	4566	0,15 $\pm$ 0,23
20	6633	0,22 $\pm$ 0,24	7663	0,30 $\pm$ 0,29	3920	0,14 $\pm$ 0,22
21	7150	0,24 $\pm$ 0,25	7519	0,32 $\pm$ 0,29	4007	0,15 $\pm$ 0,24
22	6714	0,23 $\pm$ 0,25	7120	0,30 $\pm$ 0,29	3845	0,15 $\pm$ 0,23
23	4832	0,22 $\pm$ 0,24	6972	0,29 $\pm$ 0,28	3192	0,14 $\pm$ 0,22
24	6324	0,23 $\pm$ 0,24	7145	0,30 $\pm$ 0,28	3720	0,14 $\pm$ 0,22
25	6220	0,23 $\pm$ 0,24	7145	0,30 $\pm$ 0,29	3700	0,15 $\pm$ 0,23
26	5838	0,23 $\pm$ 0,24	7162	0,30 $\pm$ 0,29	3557	0,14 $\pm$ 0,23
27	5025	0,22 $\pm$ 0,24	7291	0,29 $\pm$ 0,28	3295	0,13 $\pm$ 0,21
28	5816	0,23 $\pm$ 0,25	7453	0,31 $\pm$ 0,29	3630	0,15 $\pm$ 0,23
29	6322	0,23 $\pm$ 0,24	7542	0,29 $\pm$ 0,28	3843	0,14 $\pm$ 0,22

Meuwissen et al. (2001) ao propor a ideia de seleção genômica usou marcadores adjacentes com $r^2 > 0,20$ indicando que tais marcadores, com esse DL com QTL, pode explicar a variação do QTL. Calus et al. (2008) usou dados simulados avaliou o efeito da média do r^2 entre pares de marcadores adjacentes na acurácia da seleção genômica (correlação de valores genéticos verdadeiros e GEBV para grupo de animais de validação). Eles observaram que a precisão de GEBV aumentou de 0,68 a 0,82 quando a média do r^2 entre os marcadores adjacentes aumentou de 0,1 para a 0,2. Com base nisto, os DLs entre marcadores foram suficientes em cada painel comercial estudado, podendo possibilitar a captura acurada do efeito do QTL pelos

marcadores, e permitindo assim, a aplicação de seleção genômica nesta população.

Os resultados deste estudo, sugerem que tanto o “Axiom® Genome-Wide BOS 1 Array” da Affymetrix *BeadChip* como o “BovineHD BeadChip”, proporcionou número suficiente de marcadores, com alta proporção de polimorfismos e alta taxa de genotipagem bem distribuídas no genoma de animais da raça Guzerá. A cobertura proporcionada pelos painéis foi considerada adequada com baixa regiões genômicas sem cobertura. Apresentou também cobertura para a maioria dos genes bovinos relatados. Mesmo após sua edição, os pares de SNPs tiveram desequilíbrio de ligação semelhantes à de muitas raças, inclusive taurinas, sendo seus níveis considerado suficiente para possível aplicação da seleção genômica de características quantitativas. No entanto, diferenças dos padrões de DL entre os painéis foram observados. A decisão sobre qual é o melhor painel comercial para a realização dos estudos genômicos ainda não é evidente, podendo qualquer um dos dois painéis serem utilizados, contanto que não sejam combinados.

CONCLUSÕES

O “Axiom® Genome-Wide BOS 1 Array” da Affymetrix e o “BovineHD BeadChip” proporcionaram cobertura genômica adequada para animais da raça Guzerá, possuindo marcadores bem distribuídos ao longo genoma, sendo que aproximadamente um terço destes, apresentaram cobertura genética direta (dentro dos genes). Existem diferenças no padrão de DL entre os painéis estudados. Porém com os dois painéis foi obtido níveis altos e moderados de desequilíbrio de ligação para pequenas distâncias. Estes níveis podem ser adequados para a obtenção de valores genéticos genômicos para os animais desta população da raça Guzerá. Qualquer um dos dois painéis comerciais pode ser utilizado para estudos genômicos. No entanto, a combinação destes painéis não indicou viabilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

- AFFYMETRIX. **Axiom® Genome-Wide Bos 1 Bovine Array**. Disponível em: <http://www.affymetrix.com/catalog/prod440002/AFFY/Axiom%26%23174%3B+Genome%26%2345%3BWide+BOS+1+Bovine+Array#1_1>. Acesso em: 05 mar. 2015.
- ARIAS, J. A.; KEEHAN, M.; FISHER, P.; COPPIETERS, W.; SPELMAN, R. A high density linkage map of the bovine genome. **BMC Genetics**, Londres, v. 10, p. 18, 2009. Disponível em: <<http://dx.doi.org/10.1186/1471-2156-10-18>>.
- BARENDSE, W.; REVERTER, A.; BUNCH, R. J.; HARRISON, B. E.; BARRIS, W.; THOMAS, M. B. A validated whole-genome association study of efficient food conversion in cattle. **Genetics**, Bethesda, v. 176, n. 3, p. 1893-1905, 2007. Disponível em: <<http://dx.doi.org/10.1534/genetics.107.072637>>.
- BIFFI, A.; ANDERSON, C. D.; DESIKAN, R. S.; SABUNCU, M.; CORTELLINI, L.; SCHMANSKY, N.; SALAT, D.; ROSAND, J. Genetic variation and neuroimaging measures in Alzheimer disease. **Archives of Neurology**, Chicago, v. 67, n. 6, p. 677-685, 2010. Disponível em: <<http://dx.doi.org/10.1001/archneurol.2010.108>>.
- BOHAMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, Londres, v.11, p. 421, 2010. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-11-421>>.
- BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM (BGSAC). The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, New York, v. 324, n. 5926, p. 522-528, 2009. Disponível em: <<http://dx.doi.org/10.1126/science.1169588>>.
- CALUS, M. P. L.; MEUWISSEN, T. H. E.; DE ROOS, A. P. W.; VEERKAMP, R. F. Accuracy of genomic selection using different methods to define haplotypes. **Genetics**, Bethesda, v. 178, n., p. 553-561, 2008.

- CARROL, S. B.; GOMPEL, N.; BPRUDHOMME, B. **Regulating evolution**: how gene switches make life. Washington: Scientific American, Division of Nature America, 2008. p. 60-67.
- DE ROOS, A. P. W.; HAYES, B. J.; SPELMAN, R. J.; GODDARD, M. E. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. **Genetics**, Bethesda, v. 179, n. 3, p. 1503–1512, 2008. Disponível em: <<http://dx.doi.org/10.1534/genetics.107.084301>>.
- ESPIGOLAN, R.; BALDI, F.; BOLIGON, A. A.; SOUZA, F. R. P.; GORDO, D. G. M.; TONUSSI, R. L.; CARDOSO, D. F.; OLIVEIRA, H. N.; TONHATI, H.; SARGOLZAEI, M.; SCHENKEL, F. S.; CARVALHEIRO, R.; FERRO, J. A.; ALBUQUERQUE, G. L. Study of whole genome linkage disequilibrium in Nellore cattle. **BMC Genomics**, Londres, v. 14, p. 305, 2013. Disponível em: <<http://dx.doi.org/10.1186/1471-2164-14-305>>.
- GIBBS, R. A.; TAYLOR J. F.; VAN TASSELL C. P.; BARENDSE, W.; EVERSOLE, K. A.; GILL, C. A.; GREEN, R. D.; HAMERNIK, D. L.; KAPPES, S. M.; LIEN, S. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, New York, v. 324, n. 5926, p. 528-532, 2009. Disponível em: <<http://dx.doi.org/10.1126/science.1167936>>.
- HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Berlin, v. 38, p. 226-231, 1968.
- ILLUMINA. **BovineHD genotyping BeadChip**. Disponível em: <http://www.illumina.com/Documents/%5Cproducts%5Cdatasheets%5Cdatasheet_bovineHD.pdf>. Acesso em: 05 mar. 2012.
- JOSHI, N. R.; PHILLIPS, R. **El ganado cebu de la India y del Pakistán**. Roma: FAO, 1954. 255 p.

- KHATKAR, M. S.; NICHOLAS, F. W.; COLLINS, A. R.; ZENGER, K. R.; CAVANAGH, J. A.; BARRIS, W.; SCHNABEL, R. D.; TAYLOR, J. F.; RAADSMA, H. W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high density SNP panel. **BMC Genomics**, Londres, v. 9, 2008. Disponível em: <<http://www.biomedcentral.com/content/pdf/1471-2164-9-187.pdf>>. Acesso em: 23 fev. 2015.
- LEWONTIN, R. C. On measures of gametic disequilibrium. **Genetics**, Bethesda, v. 120, n. 3, p. 849-852, 1988.
- LUDWIG, M. Z. Functional evolution of noncoding DNA. **Current Opinion in Genetics & Development**, Londres, v. 12, n. 6, p. 634-639, 2002.
- MATUKUMALLI, L. K.; LAWLEY, C. T.; SCHNABEL, R. D.; TAYLOR, J. F.; ALLAN, M. F.; HEATON, M. P.; O'CONNELL, J.; MOORE, S. S.; SMITH, T. P.; SONSTEGARD, T. S.; VAN TASSELL, C. P. Development and characterization of a high density SNP genotyping assay for cattle. **PLoS One**, São Francisco, v. 4, n. 4, 2009. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0005350>>.
- MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; DIAS NETO, E.; GILL, C.A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C.P.; WILLIAMS, J. L.; TAYLOR, J. F.; STEPHEN, S.; MOORE, S. S. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, Londres, v. 8, 2007. Disponível em: <<http://www.biomedcentral.com/1471-2156/8/74>>. Acesso em: 22 fev. 2015.
- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819–1829, 2001.

- MICHELIZZI, V. N.; WU, X.; DODSON, M. D.; MICHAL, J. J.; ZAMBRANO-VARON, J.; MCLEAN, D. J.; JIANG, Z. A Global view of 54,001 single nucleotide polymorphisms (snps) on the illumina bovineSNP50 beadchip and their transferability to water buffalo. **International Journal of Biological Sciences**, Lake Haven, v. 7, n. 1, p. 18-27, 2011.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>.
- RINCON, G.; WEBER, K. L.; VAN EENENNAAM, A. L.; GOLDEN, B. L.; MEDRANO, J. F. Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. **Journal of Dairy Science**, v. 94, p. 6116- 6121.
- ROSENBLATT, M. Remarks on some nonparametric estimates of a density function. **The Annals of Mathematical Statistics**, Beachwood, v. 27, p. 832, 1956.
- SARGOLZAEI, M.; SCHENKEL, F. S.; JANSEN, G. B.; SCHAEFFER, L. R. Extent of linkage disequilibrium in Holstein cattle in North America. **Journal of Dairy Science**, Champaign, v. 91, n. 5, p. 2106-2117, 2008. Disponível em:<<http://dx.doi.org/10.3168/jds.2007-0553>>.
- SILVA, C. R.; NEVES, H. H. R.; QUEIROZ, S. A.; SENA J. A. D.; PIMENTEL, E. C. G. 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, Germany. **Anais...**
- SUBIRANA, J. A.; MESSEGUER, X. The most frequent short sequences in non-coding DNA. **Nucleic Acids Research**, Londres, v. 38, n. 4, p.1172-1181, 2010.

- PEIXOTO, M. G. C. D.; SANTOS, G. G. dos; BRUNELI, F. A. T.; PENNA, V. M.; VERNEQUE, R. S.; PANETTO, J. C. C.; MACHADO, C. H. C.; MACHADO, M. A.; LOBO, R. B.; CARVALHO, M. R. S. **Programa Nacional do Guzerá para leite: resultados do teste de progênie, do programa de melhoramento genético de zebuínos da ABCZ e do núcleo MOET.** Juiz de Fora: Embrapa Gado de Leite, 2014. 72 p. (Documentos, 168).
- YAN, J.; SHAH, T.; WARBURTON, M. L.; BUCKLER, E. S.; MCMULLEN, M. D.; CROUCH, J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. **PLoS ONE**, São Francisco, v.4, n. 12, 2009. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0008451>>.
- ZENGER, K. R.; KHATKAR, M. S.; CAVANAGH, J. A. L.; HAWKEN, R. J.; RAADSMA, H. W. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. **Animal Genetics**, Oxford, v. 38, n. 1, p. 7-14, 2006. Disponível em: <<http://dx.doi.org/>>.
- ZIMIN, A. V.; DELCHER, A. L.; FLOREA, L.; KELLEY, D. R.; SCHATZ, M. C.; PUIU, D.; HANRAHAN, F.; PERTEA, G.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; MARÇAIS, G.; ROBERTS, M.; SUBRAMANIAN, P.; YORKE, J. A.; SALZBERG, S. L. A whole-genome assembly of the domestic cow, *Bos taurus*. **Genome Biology**, Londres, v. 10, p. R42, 2009. Disponível em: <<http://dx.doi.org/10.1186/gb-2009-10-4-r42>>.

CAPÍTULO 6 – CONSIDERAÇÕES FINAIS

A raça Guzerá, como os zebuínos em geral, foi exposta a uma seleção artificial de menor intensidade para características de interesse econômico que as raças taurinas. Entretanto, a especialização em produção dos zebuínos pode ser uma alternativa viável ao sistema de produção tropical e sustentável. A seleção artificial intensa sobre uma característica é resultado do tempo de exposição à seleção e a implementação de recursos alavancadores, como associações de registros genealógicos e programas de seleção. Os programas de seleção visam acelerar o ganho genético via utilização de metodologias científicas, que por sua vez podem ser incrementadas por tecnologias moleculares, como a recente seleção genômica. Os valores genéticos obtidos por meio das avaliações genômicas para touros jovens, por conter as informações moleculares, são mais acurados que os valores genéticos obtidos pela avaliação tradicional, e em algumas situações, podem permitir a redução do tempo de avaliação e com isso a diminuição do intervalo de geração.

Apesar dos benefícios advindos com a utilização da seleção genômica, a tecnologia necessita ser melhor estudada e aperfeiçoada quanto a modelagem estatística e quanto as próprias informações moleculares, seja pela qualidade (proximidade dos marcadores à mutação causal) e pela quantidade (números de marcadores). As informações moleculares já está sendo otimizadas com a utilização de painéis de maior densidade ou do sequenciamento completo do genoma. No entanto, estas duas tecnologias ainda são onerosas. A verificação da viabilidade dos estudos genômicos em uma população com tecnologias mais acessíveis é estrategicamente sensato. Sendo assim, foi realizado a genotipagem de grande parte da população por um painel de 50k e um ensaio geral sobre a adequação dos principais estudos genômicos foram avaliados utilizando grande parte dos animais da população do Programa de Melhoramento do Guzerá para Leite. Após isso, uma segunda genotipagem com painéis de alta densidade foi realizada em uma parte pequena referente aos touros mais influentes (representatividade genética) da população para avaliar o desempenho dos painéis de maior custo.

Com um painel de 50k foi possível estimar adequadamente o desequilíbrio de ligação entre os segmentos cromossômicos e o tamanho

efetivo da população. No entanto, o número de pares de marcadores com o desequilíbrio de ligação considerado útil para as análises genômicas foi baixo quando comparados aos observados nas raças taurinas. O desequilíbrio de ligação entre os marcadores adjacentes foi satisfatório para os 25 mil marcadores utilizados, porém também foi abaixo em relação níveis observados com os marcadores utilizados em programas de melhoramentos das raças taurinas. Painéis contendo maior número de marcadores devem ser utilizados. Talvez a imputação de marcadores do presente painel para um painel de maior densidade pode ser a melhor alternativa a ser considerada.

Mesmo com o painel de 50 k as predições genômicas foram adequadas e próximas das esperadas indicando a compatibilidade das informações fenotípicas coletadas no programa com as informações genômicas. Este resultado pode ser reflexo do alto grau de parentesco da população. Deste modo, alternativas que combine as avaliações genômicas e o esquema MOET, sem a exclusão deste último, pode ser uma alternativa vantajosa para as avaliações convencionais de touros jovens e deveria ser melhor estudada. Também com este painel foi possível discriminar regiões (genes e QTLs) relacionados a produção de leite e seus constituintes. No entanto, melhor prospecção de regiões poderia ser obtida com a utilização de um painel de alta densidade.

Com os painéis de alta densidade foi observado muitos de pares de marcadores com desequilíbrio de ligação útil e com níveis de moderados a altos entre os marcadores adjacentes. Desta forma, a utilização destes painéis poderiam potencializar os resultados obtidos nesta tese com as análises genômicas. No entanto, os painéis comerciais apresentaram padrões de desequilíbrio e de genotipagem diferentes. Isto também implica na inviabilidade da combinação destes painéis para potencialização das análises genômicas. Em teoria, um painel com alta densidade e com menor desequilíbrio de ligação entre os marcadores poderia ser o mais adequado, por contemplar diferentes segmentos haplotípicos do genoma. Discriminar qual o painel foi mais fidedigno para a genotipagem é uma tarefa difícil. Porém o painel da Illumina apresentou algumas vantagens em relação ao painel da Affymetrix, como maior quantidade de SNPs após o controle de qualidade, menor número de regiões com pouca cobertura e maior número de genes

cobertos. Além disso, considerando a possibilidade de imputação a partir do 50k para todos os animais utilizados nesta tese, o painel Illumina deve ser preferido devido interação negativa entre os processos de genotipagem Illumina X Affymetrix.