



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA GERAL /
BIOPROSPECÇÃO

ANDRÉ VIEIRA DO NASCIMENTO

GENÔMICA PARA PROSPECÇÃO DE GENES DE INTERESSE ECONÔMICO
EM BOVINOS NELORE

DOURADOS – MS

2016

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS BIOLÓGICAS E AMBIENTAIS

GENÔMICA PARA PROSPECÇÃO DE GENES DE INTERESSE ECONÔMICO
EM BOVINOS NELORE

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Biologia Geral/Bioprospecção da Faculdade de Ciências Biológicas e Ambientais da Universidade Federal da Grande Dourados, por André Vieira do Nascimento, para obtenção do título de Mestre em Biologia Geral/Bioprospecção, sob orientação da Prof^a Dr^a Alexéia Baruffatti Grisolia.

DOURADOS – MS

2016

Dados Internacionais de Catalogação na Publicação (CIP).

N244g Nascimento, Andre Vieira Do
Genômica para prospecção de genes de interesse econômico em bovinos Nelore /
Andre Vieira Do Nascimento -- Dourados: UFGD, 2016.
84f. : il. ; 30 cm.

Orientadora: Alexeia Barufatti Grisolia

Dissertação (Mestrado em Biologia Geral/Bioprospecção) - Faculdade de Ciências
Biológicas e Ambientais, Universidade Federal da Grande Dourados.
Inclui bibliografia

1. GWAS. 2. Bos indicus. 3. haplótipos. 4. precocidade sexual. 5. bovino de corte. I.
Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

©Direitos reservados. Permitido a reprodução parcial desde que citada a fonte.

"GENÔMICA PARA PROSPECÇÃO DE GENES DE INTERESSE ECONÔMICO EM BOVINOS NELORE".

POR

ANDRÉ VIEIRA DO NASCIMENTO

DISSERTAÇÃO APRESENTADA À UNIVERSIDADE FEDERAL DA GRANDE DOURADOS (UFGD), COMO PARTE DOS REQUISITOS EXIGIDOS PARA OBTENÇÃO DO TÍTULO DE MESTRE EM BIOLOGIA GERAL - ÁREA DE CONCENTRAÇÃO: "BIOPROSPECÇÃO".



PROF^a. DR^a ALEXEIA BARUFATTI GRISOLIA
ORIENTADORA – UFGD



PROF. DR. RUSBELRAUL ASPILCUETA BORQUIS
MEMBRO TITULAR – UFGD



PROF. DR. ADAM TAITI HARTH UTSUNOMIYA
MEMBRO TITULAR – UNESP/ CÂMPUS ARAÇATUBA



PROF. DR. ANIRENE GALVÃO TAVARES PEREIRA
MEMBRO TITULAR – UNESP/ CÂMPUS ARAÇATUBA

Aprovado em 28 de julho de 2016.

AGRADECIMENTOS

Agradeço à minha família, especialmente a minha mãe Lícia e minha irmã Amanda, por sempre acreditarem em mim e me apoiarem. São exemplos de força e solidez pra mim e tenho uma enorme admiração por vocês.

À minha incrível namorada Ândrea por todo o carinho, amor, incentivo e companheirismo. Além da contribuição direta por dias e noites na construção deste trabalho.

Aos meus grandes amigos (quase irmãos) Anderson, Rafael e Rodrigo, pela amizade, fidelidade e inúmeros momentos de descontração.

Aos amigos do Laboratório de Biotecnologia Aplicada a Produção Animal (Adrielle, Alexandre, Allana, Ândrea, Bruno, Dani, Jéssica, Joyce, Juliana F., Juliana S.), por todo o conhecimento adquirido, crescimento profissional, amizade e companheirismo. Admiro muito vocês, são exemplos de profissionais.

À Prof^a orientadora Alexéia pela orientação e oportunidade de conviver e aprender constantemente. E em todos esses anos se mostrou uma pessoa cada vez mais incrível, presente em momentos bons e ruins de minha vida (acadêmica e pessoal), sou muito grato por tudo.

Ao Prof^o José Fernando Garcia e sua fantástica equipe do Laboratório de Bioquímica e Biologia Molecular Animal (Adam, Anirene, Cárís, Fernanda, Flávia, Ludmilla, Marco, Pier, Rafaela, Silvana, Tamiris, Yuri), que me acolheram, dedicaram tempo a me transmitir conhecimentos que tomo como inestimáveis e fizeram-me evoluir em diferentes aspectos. Além da disponibilidade dos dados do ZGC para imputação. Tenho grande admiração e lisonjeio por ter convivido com todos.

À DeltaGen (Daniel Biluca e Cassiano Roberto) e Fazenda Jatiúca- MS (Ney Padilha) pelo suporte técnico e fornecimento do material biológico.

À Universidade Federal da Grande Dourados pelo apoio logístico e oportunidade de cursar a graduação em Biotecnologia.

Aos órgãos de fomento Fundect e CAPES, pelo apoio financeiro à pesquisa e concessão de bolsa de pós-graduação.

A todos que não citei, mas que de alguma forma contribuíram para o meu desenvolvimento acadêmico e possibilitou a realização deste trabalho.

DEDICATÓRIA

*Dedico este trabalho a minha mãe
Lícia, por todo o amor a mim
dedicado e por sempre confiar,
compreender e apoiar minhas
decisões.*

RESUMO GERAL

A bovinocultura de corte exerce grande impacto sobre a economia brasileira, trazendo à necessidade de animais mais produtivos, capazes de manter crescente a linha de produção, sem a conversão de importantes biomas em áreas de pastagem. Assim, justifica-se o estudo de associação genômica ampla (GWAS), destinado à identificação de marcadores associados aos fenótipos de idade ao primeiro parto (IPP) e peso em diferentes idades para bovinos Nelore. O grupo amostral foi decomposto em G1, contendo 2273 animais (995 machos e 1278 fêmeas) genotipados para mais de 777.000 marcadores e G2 (522 fêmeas) genotipados em painel com densidade superior a 50.000 marcadores, que tiveram os genótipos imputados para painel mais denso. Os dados genômicos foram submetidos ao controle de qualidade, visando remoção de marcadores ou amostras viesados, antes e após a imputação. Os marcadores foram organizados em blocos haplotípicos, contendo 5 marcadores cada. Para IPP o GWAS foi realizado a partir de duas metodologias de metanálise, a primeira considerou as populações independentes entre modelagens e a segunda a existência de correlação entre as populações. Para os fenótipos de peso, o estudo de associação foi baseado em modelo linear misto. Os resultados foram expostos em *Manhattan plot*. Aqueles haplótipos cujos efeitos estimados apresentaram-se maiores e/ou significativos, foram explorados para examinar a proximidade com genes e QTLs bovinos catalogados. Para características reprodutivas foram encontrados 9 genes e 45 QTLs para metanálise 1 enquanto para metanálise 2, 11 genes e 129 QTLs. Foi possível identificar para as características de peso ao nascimento, peso a desmama, peso à sobreano, ganho de peso do nascimento ao desmame e ganho de peso do desmame ao sobreano, respectivamente, 7, 9, 4, 4 e 5 genes. Quanto aos QTLs, 25 apresentaram função biológica associada à característica de peso. O GWAS utilizando a abordagem de alelos haplotípicos, mostrou-se eficiente na busca por variantes genéticas associadas aos fenótipos de interesse.

Palavras- chave: GWAS, *Bos indicus*, haplótipos, precocidade sexual, bovino de corte, IPP, CPM.

ABSTRACT

Cattle breeding have a major impact on the Brazilian economy, bringing to the need for more productive animals, capable of maintaining a growing production line, without the conversion of important biomes into pasture areas. Thus, the genome wide association study (GWAS) it is justified, for the identification of markers associated with age at first calving (AFC) and weight at different ages for Nelore cattle. The sample group was decomposed into G1, containing 2273 animals (995 males and 1278 females) genotyped to more than 777,000 panel-markers and G2 (522 females) genotyped with high density than 50,000 markers, which had the genotypes imputed to a more dense panel. The genomic data were submitted to quality control, aiming the removal of markers or biased samples, before and after the imputation. The markers were organized in haplotypic blocks, containing 5 markers each. For AFC the GWAS was performed using two meta-analysis methodologies, the first one considered the independent populations between modeling and the second the existence of correlation between the populations. For weight phenotypes, the association study was based on a linear mixed model. The results were exposed in Manhattan plot. Those haplotypes whose estimated effects were major and/or significant were explored to examine the proximity to cataloged bovine QTL and genes. For reproductive traits, 9 genes and 45 QTLs were found for 1 analysis, while for meta-analysis 2, 11 genes and 129 QTLs. It was possible to identify the characteristics of birth weight, weaning weight, yearling weight, weight gain from birth to weaning and weight gain from weaning to yearling, respectively, 7, 9, 4, 4 and 5 genes. Regarding the QTLs, 25 presented biological function associated to the weight characteristic. The GWAS using the haplotype allele approach proved to be efficient in the search for genetic variants associated with the phenotypes of interest.

Keywords: GWAS, *Bos indicus*, haplotypes, sexual precocity, cattle, AFC.

SUMÁRIO

LISTA DE ABREVIATURAS.....	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
REVISÃO BIBLIOGRÁFICA.....	11
Pecuária no Brasil	12
Avanços do melhoramento genético.....	13
Estudo de associação genômica ampla	17
Haplótipos.....	18
Imputação.....	19
Características Reprodutivas.....	20
Características Produtivas.....	22
Referências.....	24
HIPÓTESE	32
OBJETIVOS.....	33
Geral.....	33
Específicos	33
CAPÍTULO I.....	34
Introdução	35
Material e métodos.....	36
Grupo amostral e genotipagem	36
Grupos Fenotípicos	37
Controle de qualidade.....	37
Análise de componentes principais (PCA)	38
Imputação	38
Genótipos em fase	38
Desequilíbrio de ligação.....	38
Haplótipos	39
Determinação de pesos residuais.....	39
Análise estatística.....	39
Metanálise	40
Exploração das regiões genômicas.....	41
Resultados	41
Discussão	48
Conclusão.....	51

Referências.....	51
CAPÍTULO II.....	56
Introdução	57
Material e métodos.....	58
Grupo amostral.....	58
Características em estudo	58
Controle de qualidade.....	59
Análise de componentes principais (PCA)	59
Imputação	59
Desequilíbrio de ligação.....	60
Genótipos em fase	60
Haplótipos	60
Análise estatística.....	61
Exploração das regiões genômicas.....	61
Resultados	61
Discussão	68
Conclusão.....	71
Referências.....	71
CONCLUSÃO GERAL	77
ANEXOS	78

LISTA DE ABREVIATURAS

ABIEC	Associação Brasileira das Indústrias Exportadoras de Carne
BTA	<i>Bos taurus</i> autossomo
<i>Chr</i>	Cromossomo
CR	<i>Call rate</i>
DNA	Ácido desoxirribonucleico
EBV	Estimated breeding value; Valor genético predito
FAO	<i>Food and Agriculture Organization of the United Nations</i>
GC	Grupo de contemporâneos
GCTA	<i>Genome-wide Complex Trait Analysis</i>
GWAS	<i>Genome Wide Association Study; Estudo de associação genômica ampla</i>
HD	<i>High Density</i> , Alta densidade
HL	Hormônio luteinizante
HWE	<i>Hardy Weinberg Equilibrium</i> ; Equilíbrio de HardyWeinberg
IBS	<i>Identity by state</i> ; Identidade por estado
ICMS	Imposto sobre circulação de mercadorias e serviços
IPP	Idade ao primeiro parto
LD	<i>Linkage disequilibrium</i> ; Desequilíbrio de ligação
LHR	<i>Luteinizing Hormone Receptor</i> ; Receptor do hormônio luteinizante
MAF	<i>Minor Allele Frequency</i> ; Frequência do alelo menor
MAPA	Ministério da Agricultura Pecuária e Abastecimento
<i>meta-SS</i>	<i>Meta-analysis of Selection Signals</i>
PA	Peso adulto
PCA	<i>Principal Components Analysis</i> ; Análise de componentes principais
PCR-RFLP	Polymerase chain reaction - <i>restriction fragment length polymorphism</i> ; <i>Reação da cadeia da polimerase – Polimorfismos no comprimento de fragmentos de restrição</i>
QC	Controle de qualidade
QTL	<i>Quantitative Trait Loci</i> ; <i>Loci</i> de características quantitativas
SNP	<i>Single nucleotide polymorphism</i> ; polimorfismo de nucleotídeo único
TEC	Toneladas Equivalente Carcaça
ZGC	<i>Zebu Genome Consortium</i>

LISTA DE FIGURAS

- Figura 1.** Representação esquemática de um SNP (Polimorfismo de nucleotídeo de sítio único), demonstrando a substituição de um nucleotídeo na cadeia do DNA 14
- Figura 2.** Esquema da troca de bases nitrogenadas. As setas cinza e preta representam a ocorrência de transição e transversão, respectivamente. 14
- Figura 3.** Esquema representativo da associação indireta entre SNP e fenótipo quando o mesmo está em desequilíbrio de ligação (LD) com um *locus* causal 18
- Figura 4.** Representação de SNP e haplótipos. a) SNP: polimorfismos de nucleotídeo de sítio único em uma região do DNA de múltiplos indivíduos. b) Haplótipos: formação de haplótipos de acordo com a combinação dos polimorfismos fisicamente próximos.. ... 19
- Figura 5.** Cenário de imputação. As barras no DNA representam os marcadores (SNP). Os números (genótipos observados) e interrogação (? - genótipos não observados). Genótipos não obtidos no processo de genotipagem (?) no painel de densidade inferior são estimados com base nos genótipos do painel de densidade superior da população. 20
- Figura 6.** Representação gráfica da análise de componentes principais (PCA). Cada ponto representa, (A) distribuição dos animais de G1 e (B) distribuição dos animais G2. 42
- Figura 7.** *Manhattans plot* para estudo de associação genômica ampla entre blocos haplotípicos e características de IPP1 (A) e IPP2 (B). Cada ponto do gráfico representa um haplótipo e as cores cinza e preto alternadas, os cromossomos. As setas indicam as regiões coincidentes..... 43
- Figura 8.** *Manhattans plot* para estudo de associação entre $-\log_{10}(P\text{-value})$ dos alelos haplotípicos da metanálise 1 (A) e metanálise 2 (B) e característica de idade ao primeiro parto. Cada ponto do gráfico é um alelo haplotípico. Os cromossomos foram destacados em cores alternadas de cinza e preto, e o limiar de Bonferroni representado pela linha pontilhada. 44
- Figura 9.** *Manhattan plot* para estudo de associação genômica ampla entre blocos haplotípicos e característica de peso ao nascimento (A), peso à desmama (B) e peso ao sobreano (C). 62
- Figura 10.** *Manhattan plot* para estudo de associação genômica ampla entre blocos haplotípicos e ganho de peso do nascimento à desmama (A) e ganho de peso da desmama ao sobreano (B)..... 63

LISTA DE TABELAS

Tabela 1. Principais painéis de genotipagem de SNPs disponíveis no mercado.....	16
Tabela 2. <i>Loci</i> para características quantitativas mapeados por marcadores polimorfismo de nucleotídeo único (SNP) e microssatélites (MS) para fenótipos relacionados à reprodução.	22
Tabela 3. Informações das diferentes metanálises (1 e 2), demonstrando a localização cromossômica, o tipo de metanálise utilizado, a posição de início e final do haplótipo, com seu respectivo SNP, e valor de P para cada região associada.....	45
Tabela 4. Sumário dos genes presentes em janelas de 1 Mb centralizadas aos haplótipos mais significativos para metanálise 1.	46
Tabela 5. Sumário dos genes presentes em janelas de 1 Mb centralizadas aos haplótipos mais significativos para metanálise 2.	47
Tabela 6. QTLs relacionados à características reprodutivas encontradas em janelas de 1Mb nas proximidades dos haplótipos para ambas metanálises.	48
Tabela 7. Dados referentes aos cinco alelos haplotípicos mais significativos para cada característica PN (peso ao nascimento), PD (peso a desmama), (PS) peso ao sobreano, GPND (ganho de peso do nascimento a desmama) e GPDS (ganho de peso da desmama ao sobreano).....	64
Tabela 8. Sumário dos genes presentes em janelas de 1 Mb centralizadas aos haplótipos mais significativos. PN (peso ao nascimento), PD (peso à desmama), (PS) peso ao sobreano, GPND (ganho de peso do nascimento a desmama) e GPDS (ganho de peso da desmama ao sobreano).	65
Tabela 9. QTLs relacionados à características produtivas encontradas em janelas de 1Mb nas proximidades dos haplótipos mais significativos.	68

REVISÃO BIBLIOGRÁFICA

Pecuária no Brasil

A bovinocultura de corte é uma atividade econômica de grande importância nacional, representou no ano de 2014 cerca de 30% do total de R\$ 1,1 trilhão gerados do agronegócio (MAPA, 2014). Este impacto sob a economia nacional é resultado do investimento de 167 milhões de hectares de áreas de pastagens e rebanho comercial superior a 209 milhões de cabeças (ABIEC, 2016).

A expansão dos investimentos na pecuária promoveu a substituição da vegetação nativa em pastagens, desestabilizando a diversidade biológica das localidades em que isso ocorre. No Brasil, há registros de conversão de importantes biomas, como a Amazônia e o Cerrado, em pastagens (MCALPINE et al., 2009; CARVALHO, 2010). Dados registrados pelo Ministério do Meio Ambiente indicaram devastação de 983.348 Km² no Cerrado, 377.037 Km² na Caatinga, 96.289 do Pampa, 23.160 Km² do Pantanal e 837.865 da Mata Atlântica (MMA, 2011). Além do desmatamento, a criação destes ruminantes proporciona a emissão de gases responsáveis pelo efeito estufa (MCMANUS et al., 2016; OPIO et al., 2013).

Com o aumento da população humana mundial, torna-se necessário o aumento produtivo dos rebanhos. De acordo com estimativas da FAO (2009), haverá aumento de 73% na demanda mundial de carne bovina entre os anos de 2010 à 2050. Dessa forma, uma alternativa para aliar a conservação da biodiversidade e garantir o aumento produtivo é investir na busca de recursos genéticos que possibilitem elevar a capacidade de produção de alimento de origem animal, mantendo as mesmas quantidades de áreas de pastagens (TILMAN et al., 2002).

A grande maioria dos rebanhos brasileiros é constituída por animais zebuínos (*Bos taurus indicus*), indivíduos de origem indiana facilmente distinguível do gado europeu (*Bos taurus taurus*) (LIRA, 2010). Os zebuínos são extensivamente utilizados nacionalmente, representando cerca de 80% do rebanho de corte. Dentre esses, 90% pertencem à raça Nelore. A expansão da raça no Brasil é explicada por estes apresentarem características adaptativas favoráveis para condições ambientais, maior capacidade de aproveitamento de alimento, rusticidade e resistência a parasitas (ACNB, 2016).

Assim sendo, torna-se um desafio ao produtor de carne produzir alimento em quantidade e qualidade. Para alcançar bons resultados, é necessário gerenciamento eficaz do rebanho de maneira a garantir que a produtividade e lucratividade sejam consequência do melhoramento e manejo eficiente aplicado à realidade da propriedade

(DURHAM, 2011).

Avanços do melhoramento genético

O melhoramento genético fundamenta-se na seleção de indivíduos com frequência de alelos favoráveis. Empregando como base genética o conjunto de alelos das raças domesticadas, programas de melhoramento genético foram conduzidos intuitivamente, considerando características produtivas e estéticas de animais que apresentavam fenótipos superiores (COUTINHO et al., 2010). Entretanto, algumas características genéticas são difíceis ou onerosas de serem mensuradas com acurácia em programas de melhoramento tradicional, tornando o incremento nas taxas de melhoramento dessas características limitado pela necessidade de medidas mais efetivas. Essa necessidade conduziu os estudos ao uso de técnicas moleculares, que proporcionaram maior eficiência, permitindo que a seleção realizada garanta maior ganho genético com número de gerações menor, melhores chances de previsão por apresentarem neutralidade fenotípica, além da possibilidade desses marcadores estarem associados com características fenotípicas de interesse econômico (ALZATE-MARIN et al., 2005; POCZAI et al., 2013).

A possibilidade de utilizar informações de marcadores moleculares, caracterizados como regiões portadoras de variações na sequência no DNA, torna-se um aliado ao melhoramento. Estas variações genéticas, ou polimorfismos, ocorrem mais frequentemente em regiões não codificadoras, que por não causarem alterações fenotípicas, são estáveis por não sofrerem pressão seletiva, são transmitidas aos descendentes e determinam a fixação dos polimorfismos em determinada população (VIGNAL et al., 2002).

Progressos na tecnologia de sequenciamento do DNA e bioinformática foram importantes na área da biologia molecular e contribuíram para elucidação da funcionalidade desta molécula, abrindo oportunidade ao surgimento de novas tecnologias na identificação de polimorfismos (CAETANO, 2009).

Os marcadores moleculares constituem ferramenta auxiliar ao melhoramento animal. Estas são capazes de identificar polimorfismos para características de interesse, selecionando animais geneticamente superiores. Essa identificação genotípica pode ocorrer logo após o nascimento do animal, determinando maior intensidade de seleção e aumento do ganho genético com intervalo de gerações menores (ANEJA et al., 2012).

Dentre os marcadores, destacam-se os SNPs (do inglês, *single nucleotide*

polymorphism) ou polimorfismo de nucleotídeo único, e são responsáveis pela diversidade genômica (Figura 1). Estes são concebidos na permuta de uma única base nitrogenada na sequência de DNA (do inglês, *Deoxyribonucleic Acid*).

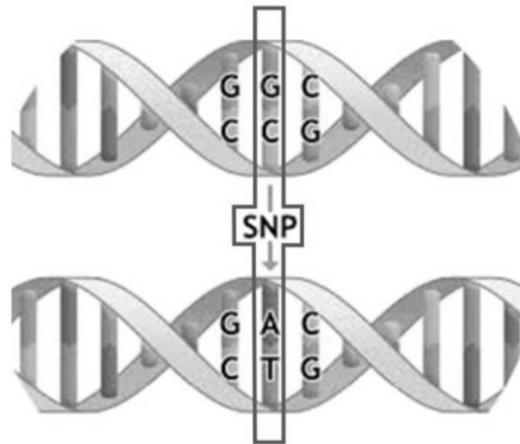


Figura 1. Representação esquemática de um SNP (Polimorfismo de nucleotídeo de sítio único), demonstrando a substituição de um nucleotídeo na cadeia do DNA (Adaptado de Society for Mucosal Immunology (2014)).

Quando as trocas ocorrem pelo mesmo tipo de base nitrogenada, ou seja, pirimidina por uma pirimidina (citosina por timina ou vice-versa) ou purina por uma purina (adenina por guanina ou vice-versa), chamam-se transições. As substituições por base de tipo diferente, pirimidina por purina ou vice-versa, são conhecidas como transversões (Figura 2) (MALYARCHUK, 2004).

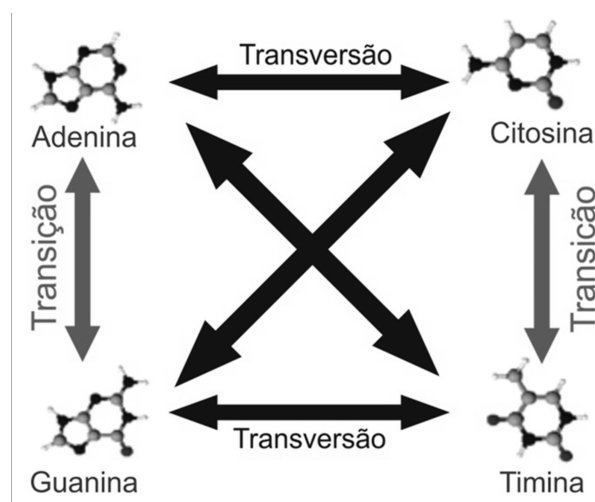


Figura 2. Esquema da troca de bases nitrogenadas. As setas cinza e preta representam a ocorrência de transição e transversão, respectivamente. (Adaptado de Carr (2014)).

A aplicação desses polimorfismos no melhoramento teve êxito com

desenvolvimento dos painéis de genotipagem de alta densidade. Estes constituem lâminas de vidro contendo na sua superfície milhares ou milhões de sondas (fragmentos específicos de DNA) capazes de interrogar polimorfismos pontuais, associado ao princípio de hibridação molecular (PASSOS-BUENO e MOREIRA, 2004).

Os painéis de SNP promoveram revolução na área de genômica animal, pois permitem varredura do genoma para milhares de marcadores ao mesmo tempo, com menor custo e maior rapidez, quando comparado a técnicas de genotipagem como PCR-RFLP (do inglês, *polymerase chain reaction - restriction fragment length polymorphism*) e detecção de microssatélites. O desenvolvimento de metodologias para genotipagem massal de dezenas de milhares até milhões de SNP em único ensaio, reduziu os custos de geração de dados genotípicos (CAETANO, 2009).

O uso da tecnologia dos painéis de genotipagem de SNP possibilita a cobertura uniforme de todo o genoma, fornecendo mapas genéticos densos, permitindo a realização de estudos de associação genética, desequilíbrio de ligação, avaliação do mérito genético do indivíduo, a identificação de *loci* de características quantitativas ou QTLs (do inglês, *Quantitative Trait Locus*), ensaios de confirmação de paternidade e identificação individual (DAW et al., 2005; DAETWYLER et al., 2008; LI et al., 2008).

Atualmente, existem diversas tecnologias de genotipagem com capacidades distintas de geração de dados genéticos, isso varia de acordo com o número de sondas no painel (chip) e espécie para qual é destinado (Tabela 1). Dentre as empresas fabricantes, a Illumina e Affymetrix se destacam no mercado mundial (LI et al, 2008).

Tabela 1. Principais painéis de genotipagem de SNPs disponíveis no mercado.

Nome comercial do painel	Espécie	Número de sondas	Empresa
HumanOmni5-4 v1.1 BeadChip	Humano	4.3 Milhões	Illumina
BovineHD BeadChip	Bovino	777,000	Illumina
Axiom [®] Genome-Wide Chicken Genotyping Array	Frango	580,000	Affymetrix
Axiom [®] Soybean Genotyping Array	Soja	180,961	Affymetrix
CanineHD Whole-Genome Genotyping BeadChip	Canino	170,000	Illumina
Axiom [®] Salmon Genotyping Array	Salmão	130,000	Affymetrix
Axiom [®] Strawberry Genotyping Array	Morango	95,062	Affymetrix
Axiom [®] Buffalo Genotyping Array	Bubalino	90,000	Affymetrix
PorcineSNP60 BeadChip	Suíno	65,000	Illumina
Axiom [®] Trout Genotyping Array	Truta	57,501	Affymetrix
BovineSNP50 BeadChip	Bovino	54,609	Illumina
EquineSNP50 BeadChip	Equino	54,602	Illumina
OvineSNP50 BeadChip	Ovino	54,241	Illumina
MaizeSNP50 BeadChip	Milho	50,000	Illumina
GeneChip [®] Rice 44K SNP Genotyping Array	Arroz	44,100	Affymetrix
Axiom [®] Cotton Genotyping Array	Algodão	35,550	Affymetrix

Fonte: Affymetrix e Illumina.

Os painéis de SNP para humanos são os mais robustos e, conseqüentemente, reportam maior quantidade de informações genômicas, podendo interrogar cerca de 4.3 milhões de marcadores e, alguns permitem avaliar até 500 mil SNP adicionais customizados, possibilitando ao pesquisador adequar o painel às suas necessidades. Esses painéis são voltados à área da saúde em estudos oncológicos e exoma (HEDGES et al., 2009; ILLUMINA, 2016).

Na agropecuária, existem painéis destinados às espécies animais e vegetais (AFFYMETRIX, 2016; ILLUMINA, 2016). Nesta área, sua aplicabilidade está associada à busca de indivíduos geneticamente superiores ou polimorfismos associados à características fenotípicas de importância econômica, visando maior produtividade e

custos reduzidos com melhoramento genético (RESENDE et al., 2008; CORNELIS et al., 2010).

O painel Illumina® BovineSNP50 foi desenvolvido por pesquisadores de área animal. Este apresenta sondas para interrogar 54,609 SNP com intervalo médio de 49.4 e mediana de 36.9 kb (kilobase, 1kb = 1,000 pares de bases) entre marcadores. Para sua construção foram utilizadas populações de animais predominantemente de origem taurina, dentre as raças (número de animais utilizados): Angus (22), Beefmaster (24), Jersey (9), Hereford (20), Holstein (22), Brahman (20), Gir (17), Nelore (5), Santa Gertrudis (7), entre outros (VAN TASSEL et al., 2008; ILLUMINA, 2012).

Posteriormente, a mesma empresa desenvolveu o Illumina® BovineHD BeadChip 777k (intervalo médio entre marcadores de 3.4 kb), capaz de cobrir SNP validados em bovinos de corte e leite. O chip de alta densidade apresenta genoma de espécies taurinas, zebuínas e híbridos adaptados aos climas temperado e tropical, sendo atualmente o mais adequado para uso em zebuínos.

Estudo de associação genômica ampla

As informações genóticas provenientes dos painéis de genotipagem podem ser utilizadas para estudo de associação genômica ampla ou GWAS (do inglês, *Genome Wide Association Study*), pois permitem realizar a associação massal dos marcadores com fenótipos de interesse (HIRSCHHORN e DALY, 2005; RESENDE et al., 2010). Isso proporciona a detecção de vários polimorfismos associados às características de interesse em diferentes localidades do genoma. Várias das características quantitativas (exemplo: peso, musculosidade, idade ao primeiro parto, entre outros) apresentam padrão de herança poligênica, isto é, o fenótipo é resultado da soma dos efeitos de diversos genes, desta forma, o GWAS permite verificar as diferentes regiões que influenciam a expressão do fenótipo (LEISERSON et al., 2013).

A identificação de variantes genéticas relacionadas à fenótipos de interesse pelo GWAS pode proporcionar a seleção de animais com genótipos superiores para determinada característica de forma antecipada. A realização da genotipagem do indivíduo logo após o nascimento será suficiente para detectar se o mesmo apresenta genótipos para caracteres que muitas vezes serão expressos somente na fase adulta (HIRSCHHORN e DALY, 2005; RESENDE et al., 2010).

O SNP pode ser denominado de acordo com a localização no genoma como: exônico, intrônico ou inter-gênico. A ocorrência de cada tipo varia bastante no genoma

bovino, sendo mais comum os SNPs localizados em regiões intergênicas (63,74%), seguidos pelos das regiões intrônicas (34,9%) e exônicas (1,35%) (THE BOVINE HAPMAP CONSORTIUM, 2009). Desta forma, apesar de substituições nucleotídicas proporcionarem trocas aminoacídicas e consequente modificação proteica, é rara a existência de SNP em regiões codantes (ORR et al., 2006). Entretanto, a informação massiva de SNPs proporciona a associação indireta com regiões funcionais (COLLINS et al., 1998). Assim, a associação de marcadores SNP pode ocorrer de duas maneiras, de forma direta, quando o próprio SNP exerce influência na regulação e/ou expressão gênica, ou indireta, no caso do marcador estar ligado por desequilíbrio de ligação (LD) a uma mutação causal (Figura 3).

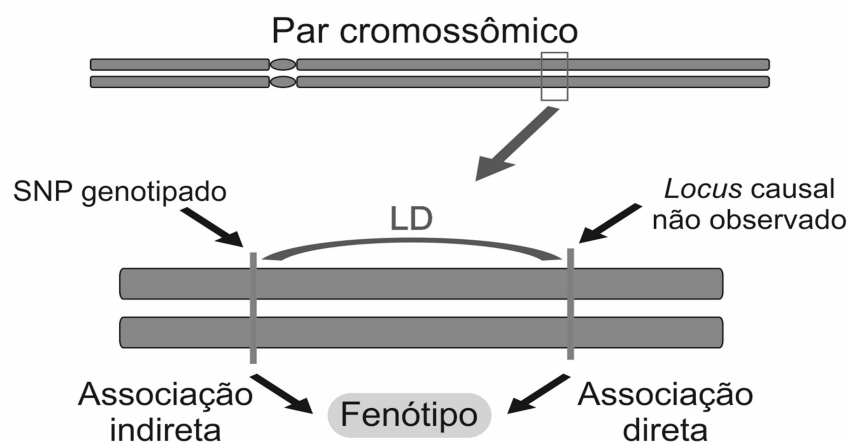


Figura 3. Esquema representativo da associação indireta entre SNP e fenótipo quando o mesmo está em desequilíbrio de ligação (LD) com um *locus* causal (Adaptado de Astle e Balding (2009)).

O LD representa o grau de associação ocorrente de forma não aleatória entre dois ou mais *locus*. Em uma população o tamanho do LD é influenciado pela proximidade existente entre marcadores e número de gerações entre indivíduos (ARDLIE et al., 2002). Padrões de LD em animais de produção comumente ocorrem pela existência de seleção e endogamia e são alterados por recombinação e mutações (KIM e KIRKPATRICK, 2009).

Haplótipos

Variantes genéticas com distância física entre si curta tendem a serem herdadas juntas. Indivíduos que possuam o nucleotídeo T em vez de C em uma determinada posição cromossômica podem compartilhar identidade para vários outros SNPs nas

proximidades de T. Assim, haplótipos (Figura 4) são definidos como regiões genômicas com presença de polimorfismos ligados e que segregam em conjunto dentro da população (INTERNATIONAL HAPMAP CONSORTIUM, 2003).



Figura 4. Representação de SNP e haplótipos. a) SNP: polimorfismos de nucleotídeo de sítio único em uma região do DNA de múltiplos indivíduos. b) Haplótipos: formação de haplótipos de acordo com a combinação dos polimorfismos fisicamente próximos. (Adaptado de International Hapmap Consortium (2003)).

Imputação

A imputação de genótipo é um processo estatístico computacional de predição de genótipos que não foram observados na amostra genotipada do indivíduo. As características genéticas estimadas são preditas conforme informações de uma população (haplótipos) referência (Figura 5) (MARCHINI e HOWIE, 2010). Este procedimento aplica-se na predição de marcadores moleculares e proporciona a imputação de painéis de SNP de baixa densidade para painéis de alta densidade (WEIGEL et al., 2010; MULDER et al., 2012). Assim, compilando dados, reduzindo custo por marcador, agregando melhoria na acurácia e estimando melhores efeitos de marcadores com características fenotípicas de interesse.

Cenário típico de imputação

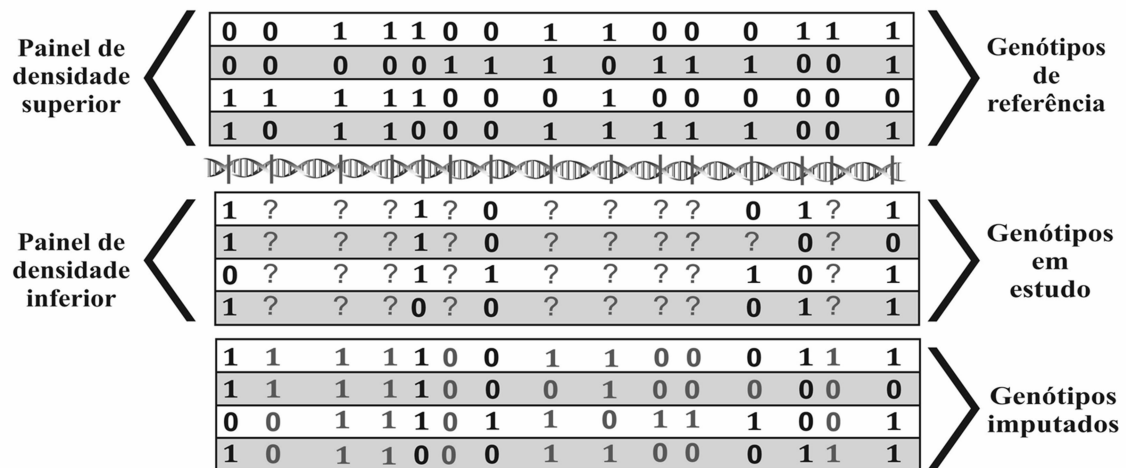


Figura 5. Cenário de imputação. As barras no DNA representam os marcadores (SNP). Os números (genótipos observados) e interrogação (? - genótipos não observados). Genótipos não obtidos no processo de genotipagem (?) no painel de densidade inferior são estimados com base nos genótipos do painel de densidade superior da população.

Características Reprodutivas

As características reprodutivas são consideradas de heranças complexas ou poligênicas. A eficiência reprodutiva de um rebanho pode ser mensurada por meio de idade à puberdade (IP), idade ao primeiro parto (IPP), intervalo entre partos (IEP), duração da gestação (PG), dias para o parto, habilidade de permanência no rebanho e probabilidade de prenhez (PP). Dentre estas, a IP, IPP e PP são características que indicam precocidade sexual.

A maioria das novilhas zebuínas apresenta o primeiro parto tardiamente (por volta dos 24 meses), em relação às raças taurinas (aproximadamente 14 meses). Estudos evidenciam que há diferenças nos mecanismos de regulação hormonal (sistema hipotalâmico-hipofisário) relativos ao hormônio luteinizante e formação do folículo ovariano, que proporciona a maior eficiência reprodutiva, ou seja, maior precocidade em taurinos quando comparados aos zebuínos (ABEYGUNAWARDENA e DEMATAWEWA, 2004; BARUSELLI et al., 2007; MILAZZOTO et al., 2008).

No entanto, devido a melhor adaptação dos zebuínos às condições ambientais do Brasil, tornou-se preponderante sua criação nos rebanhos brasileiros, em especial a raça Nelore (ABIEC, 2016). Essa adequação ocorre pela presença de alguns mecanismos fisiológicos existentes em raças de origem indiana, como maior tamanho e quantidade

de glândulas sudoríparas (reguladoras de temperatura), pigmentação escura da pele (conferindo resistência ao calor do ambiente) e pelagem curta e clara (diminuindo o fluxo interno de calor) (NAY, 1956; FINCH, 1986; LANDAETA-HERNÁNDEZ et al., 2011).

O desempenho reprodutivo das fêmeas na bovinocultura é de suma importância para a reposição de novilhas e produção de bezerros, influenciando na quantidade e qualidade do rebanho, proporcionando maior disponibilidade de gado para abate em menor tempo e a substituição de matrizes mais velhas, que não apresentam índices satisfatórios de prenhez, podendo-se obter benefícios de ganho genético, pela seleção desses animais (WATHES et al., 2008).

A idade ao primeiro parto (IPP) é uma das maneiras utilizadas para mensurar a eficiência reprodutiva das novilhas e apresenta herdabilidade de baixa a moderada, variando entre 0,01 a 0,27 (PEREIRA et al., 2001; PEREIRA et al., 2002; DIAS et al., 2004; BOLIGON et al., 2007; CAMMACK et al., 2009). Esse critério tem sido adotado pela facilidade de coleta de dados e sua relação direta com características voltadas a reprodução (BOLIGON et al., 2011).

Além da precocidade sexual das fêmeas, determinada pela IPP, é necessário considerar outros aspectos como, por exemplo, índices de mortalidade e distocia, para que se obtenha sucesso na seleção de fêmeas fisiologicamente preparadas para reprodução em idade reduzida. Van Melis et al. (2010) validaram esse quadro, obtendo a diminuição da IPP para períodos de até 16 meses (fêmeas Nelore), dando origem a prole saudável. Assim sendo, torna-se viável a obtenção de fêmeas zebuínas que apresentem IPP reduzida determinada pela concepção de números satisfatórios de bezerros saudáveis (BONGIORNI et al., 2012).

Apesar do reconhecimento da importância do melhoramento genético para obtenção de animais precoces, este objetivo se depara com entraves como: dificuldade na aquisição de informação, baixa herdabilidade e manifestação fenotípica tardia. Logo, o uso de métodos convencionais de seleção não são eficientes o bastante para melhorar a característica precocidade (BUDDENBERG et al., 1990; CAMMACK et al., 2009). Assim sendo, torna-se necessário a busca de novas tecnologias de seleção.

O mapeamento de QTLs bovinos, utilizando marcadores microssatélites, têm fornecido informações referentes aos fenótipos ligados à fertilidade (Tabela 2). No entanto, devido à baixa densidade de marcadores utilizados, os resultados não eram satisfatórios para detecção das variações genotípicas intrínsecas as características

complexas. Sendo assim, a utilização de GWAS para esta finalidade tem se demonstrado favorável, distinguindo-se das demais técnicas pela maior cobertura do genoma, fornecendo informações das bases genéticas que envolvem os processos reprodutivos.

Tabela 2. *Loci* para características quantitativas mapeados por marcadores polimorfismo de nucleotídeo único (SNP) e microssatélites (MS) para fenótipos relacionados à reprodução.

Cromossomo	Número de marcadores	Característica	Raça	Referencia
1-3, 5, 8-10, 13-15, 19, 21, 27.	44,333 SNP	Idade a puberdade	Brahman e Composto Tropical	Hawken et al., 2012
18	45,878 SNP	Facilidade de parto	Holstein	Cole et al., 2011
6, 10, 16	221 MS	Facilidade de parto	Holstein	Schnabel et al., 2005
1, 3, 7, X	45,878 SNP	Taxa de Prenhez das filhas	Holstein	Cole et al., 2011
		Índice de partos		
7, 11	145 MS	Natimorto	Swedish dairy	Holmberg e Andersson-Eklund, 2006
8, 10, 18	263 MS	Natimorto	Holstein	Kuhn et al., 2003
2, 5, 18, 25, 29	38,650 SNP	Taxa de Concepção	Holstein	Peñagaricano et al., 2012
4, 5, 7, 9, 10, 20, 24	36,387 SNP	Fertilidade	Holstein	Sahana et al., 2010
1, 4-13, 15-23, 25-28	390 MS	Circunferência escrotal	Angus	McClure et al., 2010

Adaptado de Fortes et al., 2013

Características Produtivas

O Brasil utiliza o sistema de classificação de carcaça, que verifica e atribui valores de acordo com parâmetros como: sexo, acabamento, idade, conformação, peso, entre outros. Este sistema tem como finalidade facilitar o comércio por padronização de carcaça, bonificando financeiramente por arroba (1@ = 15 kg) produtos homogêneos de qualidade. Isso incentiva o produtor a produzir carcaça de excelência (BRASIL, 1989; SAINZ e ARAUJO, 2001).

No caso do estado do Mato Grosso do Sul, o governo instituiu o programa novilho precoce, cujo objetivo é incentivar os produtores a criação de animais que possam apresentar-se terminados antecipadamente. Os produtores que possuem no rebanho animais que atendam aos padrões exigidos, garantem a redução das alíquotas

do imposto sobre circulação de mercadorias e serviços (ICMS), que podem chegar a 67% para animais com apenas dentes de leite (MATO GROSSO DO SUL, 1995). Além da padronização e bonificação estabelecida pelos diferentes estados brasileiros, cada frigorífico pode oferecer bônus de acordo com a qualidade da carcaça.

A avaliação da produtividade animal pode ser obtida através da análise de bovinos com potencial superior de produção de carne, dentre os fenótipos que permitem essa avaliação destaca-se a característica de peso.

Fenótipos relacionados ao peso são de importância primária nos sistemas de produção de gado para corte, uma vez que permitem avaliar o potencial de crescimento do animal, sendo também um dos primeiros critérios mensurados (ao nascer) do bovino (GUNAWAN e JAKARIA, 2011). Essas características de crescimento são facilmente mensuradas, relatadas como de média a alta herdabilidade, sendo de 0,25; 0,33; 0,34 para peso ao nascimento, desmama e sobreano, respectivamente (LÔBO e FILHO, 2003).

Outro fator que torna o peso um critério relevante no processo de seleção é a sua correlação com outros caracteres de interesse. Como exemplo, a circunferência escrotal (CE), tem sido relatada como indicador de fertilidade em touros. Boligon et al. (2011) descreveram a relação entre os fenótipos de peso e CE, que apresentou correlação positiva e moderada com circunferência escrotal ($0,15 \pm 0,03$ a $0,38 \pm 0,01$), sugerindo que a seleção para ganho de peso pode trazer ganhos em aspectos produtivos e reprodutivos.

Análises de GWAS para características produtivas têm identificado diversos genes e regiões do genoma, que afetam essa característica quantitativa em bovinos. Sendo assim, o estudo de associação demonstrou-se promissor para o melhoramento do rebanho assistido por marcadores. Esse fato contribuiu para o reconhecimento de regiões associadas, ampliando o conhecimento de genes que apresentam correlação negativa com o fenótipo de interesse. Para características de peso, o estabelecimento de correlação negativa é de grande importância, uma vez que busca-se animais que não tenham peso ao nascer elevados proporcionando dessa forma menores chances de dificuldades no parto, mas que sejam capazes de alcançarem desenvolvimento satisfatório em períodos posteriores (desmama e sobreano) (GROSZ e MACNEIL, 2001).

Referências

- ABEYGUNAWARDENA, H.; DEMATAWEWA, C. Pre-pubertal and postpartum anestrus in tropical Zebu cattle. **Animal Reproduction Science**, v. 82, p. 373-387, 2004.
- ABIEC. Associação Brasileira de Indústrias Exportadoras de Carne. São Paulo, SP, 2016. Disponível em: < <http://www.abiec.com.br> >. Acesso em: 25 Jan. 2016.
- ACNB. Associação de Criadores de Nelore do Brasil. São Paulo, SP, 2016. Disponível em: < <http://www.nelore.org.br/raca> >. Acesso em: 10 Mar. 2016.
- AFFYMETRIX. Product literature. Santa Clara, CA, 2016. Disponível em: < <http://www.affymetrix.com/estore> >. Acesso em: 16 Fev. 2016.
- ALZATE-MARIN, A. L.; CERVIGNI, G. D.; MOREIRA, M. A.; BARROS, E. G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v. 30, n. 4, p. 333-342, 2005.
- ANEJA, B.; YADAV, N. R.; CHAWLA, V.; YADAV, R. C. Sequence-related amplified polymorphism (SRAP) molecular marker system and its applications in crop improvement. **Molecular breeding**, v. 30, n. 4, p. 1635-1648, 2012.
- ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Nature Reviews Genetics**, v. 3, n. 4, p. 299-309, 2002.
- ASTLE, W.; BALDING, D. J. Population structure and cryptic relatedness in genetic association studies. **Statistical Science**, v., n., p. 451-471, 2009.
- BARUSELLI, P. S.; GIMENES, L. U.; SALES, J. N. D. S. Fisiologia reprodutiva de fêmeas taurinas e zebuínas. **Rev Bras Reprod Anim**, v. 31, n. 2, p. 205-211, 2007.
- BOLIGON, A. A.; BALDI, F.; ALBUQUERQUE, L. G. D. Genetic parameters and relationships between growth traits and scrotal circumference measured at different ages in Nelore cattle. **Genetics and Molecular Biology**, v. 34, n. 2, p. 225-230, 2011.
- BOLIGON, A. A.; RORATO, P. R. N.; ALBUQUERQUE, L. D. Correlações genéticas entre medidas de perímetro escrotal e características produtivas e reprodutivas de fêmeas da raça Nelore. **Revista Brasileira de Zootecnia**, v. 36, n. 3, p. 565-571, 2007.
- BONGIORNI, S.; MANCINI, G.; CHILLEMI, G.; PARISET, L.; VALENTINI, A.

Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. **PloS one**, v. 7, n. 12, p. e50137, 2012.

BRASIL. MINISTÉRIO DA AGRICULTURA. Sistema Nacional de Tipificação de Carcaças Bovinas. Portaria n. 612, de 05 de outubro de 1989: 1-4 p. 1989.

BUDDENBERG, B.; BROWN, C.; BROWN, A. Heritability estimates of calving date in Hereford cattle maintained on range under natural mating. **Journal of Animal Science**, v. 68, n. 1, p. 70-74, 1990.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, n. 8, p. 64-71, 2009.

CAMMACK, K.; THOMAS, M.; ENNS, R. Reproductive traits and their heritabilities in beef cattle. **The Professional Animal Scientist**, v. 25, n. 5, p. 517-528, 2009.

CARR, S. M. Transversions vs Transitions. 2014. Disponível em: <http://www.mun.ca/biology/scarr/Transitions_vs_Transversions.html>. Acesso em: 23 Mar. 2016.

CARVALHO, J. L. N.; AVANZI, J. C.; SILVA, M. L. N.; MELLO, C. R. D.; CERRI, C. E. P. Potencial de sequestro de carbono em diferentes biomas do Brasil. **Revista Brasileira de Ciência do Solo**, v. 34, n. 2, p. 277-290, 2010.

COLE, J. B.; WIGGANS, G. R.; MA, L.; et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. **BMC Genomics**, v. 12, n. 1, p. 1, 2011.

COLLINS, F. S.; BROOKS, L. D.; CHAKRAVARTI, A. A DNA polymorphism discovery resource for research on human genetic variation. **Genome research**, v. 8, n. 12, p. 1229-1231, 1998.

CORNELIS, M. C.; AGRAWAL, A.; COLE, J. W.; et al. The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. **Genetic Epidemiol**, v. 34, n. 4, p. 364-372, 2010.

COUTINHO, L.; JORGE, E.; ROSÁRIO, M.; REGITANO, L. A genômica na bovinocultura de corte. **Bovinoicultura de corte. Piracicaba: Fealq**, v. 2, n., p. 813-826, 2010.

- DAETWYLER, H. D.; SCHENKEL, F. S.; SARGOLZAEI, M.; ROBINSON, J. A. B. 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**, v. 91, n. 8, p. 3225-3236, 2008.
- DAW, E. W.; HEATH, S. C.; LU, Y. Single-nucleotide polymorphism versus microsatellite markers in a combined linkage and segregation analysis of a quantitative trait. **BMC Genet**, v. 6, n. Suppl 1, p. S32, 2005.
- DIAS, L. T.; EL FARO, L.; ALBUQUERQUE, L. G. D. Estimativas de herdabilidade para idade ao primeiro parto de novilhas da raça Nelore. **Revista Brasileira de Zootecnia**, v. 33, n. 1, p. 97-102, 2004.
- DURHAM, S. Beef cattle: improving production efficiency and meat quality. **Agricultural Research**, v. 59, n. 1, p. 18-19, 2011.
- FAO. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. How to Feed the World in 2050. 2009. Disponível em: < <http://www.fao.org/wsfs/forum2050/wsfs-forum/en/> >. Acesso em: 12 Fev. 2016.
- FINCH, V. Body temperature in beef cattle: its control and relevance to production in the tropics. **Journal of Animal Science**, v. 62, n. 2, p. 531-542, 1986.
- FORTES, M. R.; DEATLEY, K. L.; LEHNERT, S. A.; BURNS, B. M.; REVERTER, A.; HAWKEN, R. J.; BOE-HANSEN, G.; MOORE, S. S.; THOMAS, M. G. Genomic regions associated with fertility traits in male and female cattle: advances from microsatellites to high-density chips and beyond. **Animal Reproduction Science**, v. 141, n. 1, p. 1-19, 2013.
- GROSZ, M.; MACNEIL, M. Putative quantitative trait locus affecting birth weight on bovine chromosome 2. **Journal of Animal Science**, v. 79, n. 1, p. 68-72, 2001
- GUNAWAN, A.; JAKARIA. Genetic and Non-Genetics effect on birth, weaning, and yearling weight of Bali cattle. **Media Peternakan-Journal of Animal Science and Technology**, v. 34, n. 2, p. 93-98, 2011.
- HAWKEN, R.; ZHANG, Y.; FORTES, M.; et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. **J Anim Sci**, v. 90, n. 5, p. 1398-1410, 2012.
- HEDGES, D. J.; BURGESS, D.; POWELL, E.; et al. Exome sequencing of a

- multigenerational human pedigree. **PloS one**, v. 4, n. 12, p. e8232, 2009.
- HIRSCHHORN, J. N.; DALY, M. J. Genome-wide association studies for common diseases and complex traits. **Nat Rev Genet**, v. 6, n. 2, p. 95-108, 2005.
- HOLMBERG, M.; ANDERSSON-EKLUND, L. Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. **J Dairy Sci**, v. 89, n. 9, p. 3664-3671, 2006.
- ILLUMINA. BovineHD Genotyping BeadChip. San Diego, CA, 2012. Disponível em: < www.illumina.com >. Acesso em: 26 Jan. 2014.
- ILLUMINA. San Diego, CA, 2016. Disponível em: < <http://www.illumina.com/> >. Acesso em: 27 Fev. 2016.
- INTERNATIONAL HAPMAP CONSORTIUM. The International HapMap Project. **Nature**, v. 426, n. 6968, p. 789-96, 18 Dez. 2003.
- KIM, E. S.; KIRKPATRICK, B. Linkage disequilibrium in the North American Holstein population. **Animal Genetics**, v. 40, n. 3, p. 279-288, 2009.
- KUHN, C.; BENNEWITZ, J.; REINSCH, N.; et al. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. **J Dairy Sci**, v. 86, n. 1, p. 360-368, 2003.
- LANDAETA-HERNÁNDEZ, A.; ZAMBRANO-NAVA, S.; HERNÁNDEZ-FONSECA, J. P.; et al. Variability of hair coat and skin traits as related to adaptation in Criollo Limonero cattle. **Tropical animal health and production**, v. 43, n. 3, p. 657-663, 2011.
- LEISERSON, M. D.; ELDRIDGE, J. V.; RAMACHANDRAN, S.; RAPHAEL, B. J. Network analysis of GWAS data. **Current opinion in genetics & development**, v. 23, n. 6, p. 602-610, 2013.
- LI, M.; LI, C.; GUAN, W. Evaluation of coverage variation of SNP chips for genome-wide association studies. **Eur J Hum Genet**, v. 16, n. 5, p. 635-643, 2008.
- LIRA, J. Revisión sobre la genética del origen del ganado vacuno y las aportaciones del ADN antiguo. **Munibe Antropologia-Arkeologia**, v. 61, n., p. 153-170, 2010.
- LÔBO, R. N. B.; FILHO, R. M. Estimates of genetic parameters for weights and maturing traits derived by growth functions in Nellore cattle. **Ciência Animal**, v. 13, n.

1, p. 7-15, 2003.

MALYARCHUK, B. Similarity of mutation spectra of the mitochondrial DNA hypervariable segment 1 in Homo and Pan species. **Molecular Biology**, v. 38, n. 3, p. 370-375, 2004.

MAPA. MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. Produto Interno Bruto da agropecuária deve ser de R\$ 1,1 trilhão. 2014. Disponível em: < <http://www.agricultura.gov.br/comunicacao/noticias/2014/12/produto-interno-bruto-da-agropecuaria-deve-ser-de-rs-1-trilhao> >. Acesso em: 13 Fev. 2016.

MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. **Nat Rev Genet**, v. 11, n. 7, p. 499-511, Jul. 2010.

MATO GROSSO DO SUL. Decreto nº 8.421 de 28 de dezembro de 1995. Reformula o Programa de Apoio à Criação de Gado para o Abate Precoce (Novilho Precoce) e consolida a legislação a ele referente. Diário Oficial do Estado, Campo Grande, MS, 28 dez. 1995.

MCALPINE, C. A.; ETTER, A.; FEARNSIDE, P. M.; et al. Increasing world consumption of beef as a driver of regional and global change: A call for policy action based on evidence from Queensland (Australia), Colombia and Brazil. **Global Environmental Change**, v. 19, n. 1, p. 21-33, 2009.

MCCLURE, M.; MORSCI, N.; SCHNABEL, et al. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. **Animal Genetics**, v. 41, n. 6, p. 597-607, 2010.

MCMANUS, C.; BARCELLOS, J. O. J.; FORMENTON, B. K.; et al. Dynamics of Cattle Production in Brazil. **PloS one**, v. 11, n. 1, p. e0147138, 2016.

MILAZZOTTO, M.; RAHAL, P.; NICHI, M.; et al. New molecular variants of hypothalamus–pituitary–gonad axis genes and their association with early puberty phenotype in *Bos taurus indicus* (Nellore). **Livestock Science**, v. 114, n. 2, p. 274-279, 2008.

MMA. MINISTÉRIO DO MEIO AMBIENTE. Controle e Prevenção do Desmatamento - Monitoramento dos Biomas período 2008-2009. Brasília, 2011. Disponível em: <<http://www.mma.gov.br/florestas/control-e-preven%C3%A7%C3%A3o-do-desmatamento>>. Acesso em: 15 Jan. 2016.

MULDER, H.; CALUS, M.; DRUET, T.; SCHROOTEN, C. Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. **J Dairy Sci**, v. 95, n. 2, p. 876-889, 2012.

NAY, T.; HAYMAN, R. Sweat glands in Zebu (*Bos indicus* L.) and European (*B. taurus* L.) cattle. I. Size of individual glands, the denseness of their population, and their depth below the skin surface. **Crop and Pasture Science**, v. 7, n. 5, p. 482-492, 1956.

OPIO, C.; GERBER, P.; MOTTET, A.; FALCUCCHI, A.; TEMPIO, G.; MACLEOD, M.; VELLINGA, T.; HENDERSON, B.; STEINFELD, H. **Greenhouse gas emissions from ruminant supply chains – A global life cycle assessment**. Rome: Food and Agriculture Organization of the United Nations (FAO). 2013.

ORR, N.; BEKKER, V.; CHANOCK, S. Genetic association studies: marking them well. **Journal of Infectious Diseases**, v. 194, n. 11, p. 1475-1477, 2006.

PASSOS-BUENO, M. R. S.; MOREIRA, E. S. Ferramentas básicas da genética molecular humana. In: Mir, L. (Ed.). **Genômica**. São Paulo: Atheneu, 2004, p.43-69.

PEÑAGARICANO, F.; WEIGEL, K.; KHATIB, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. **Animal Genetics**, v. 43, n. s1, p. 65-71, 2012.

PEREIRA, E.; ELER, J. P.; FERRAZ, J. B. S. Análise genética de características reprodutivas na raça Nelore. **Pesquisa Agropecuária Brasileira**, v. 37, n. 5, p. 703-708, 2002.

PEREIRA, E.; ELER, J.; COSTA, F.; FERRAZ, J. Análise genética da idade ao primeiro parto e do perímetro escrotal em bovinos da raça Nelore. **Arq. bras. med. vet. zootec**, v. 53, n. 1, p. 116-121, 2001.

POCZAI, P.; VARGA, I.; LAOS, M.; CSEH, A.; BELL, N.; VALKONEN, J. P.; HYVÖNEN, J. Advances in plant gene-targeted and functional markers: a review. **Plant Methods**, v. 9, n. 1, p. 1-32, 2013.

RESENDE, M. D. V. D.; AGUIAR, A. M.; ABAD, J. I. M.; MISSIAGGIA, A. A.; SANSALONI, C.; PETROLI, C.; GRATTAPAGLIA, D.; RESENDE JÚNIOR, M. F. R. Computação da seleção genômica ampla (GWS). **Embrapa Florestas**, v., n., p., 2010.

RESENDE, M. D. V.; LOPES, P. S.; DA SILVA, R. L.; PIRES, I. E. Seleção genômica

ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v., n. 56, p. 63-77, 2008.

SAHANA, G.; GULDBRANDTSEN, B.; BENDIXEN, C.; LUND, M. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. **Animal Genetics**, v. 41, n. 6, p. 579-588, 2010.

SAINZ, R. D.; ARAUJO, F. R. Tipificação de carcaças de bovinos e suínos. CONGRESSO BRASILEIRO DE CIÊNCIA E TECNOLOGIA DE CARNES, 2001. Centro de Tecnologia de Carnes do Instituto de Tecnologia de Alimentos Campinas. p.26-55.

SCHNABEL, R.; SONSTEGARD, T.; TAYLOR, J.; ASHWELL, M. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US holstein families. **Animal Genetics**, v. 36, n. 5, p. 408-416, 2005

SOCIETY FOR MUCOSAL IMMUNOLOGY. Single Nucleotide Polymorphism (SNP) Allele Frequency DNA Pools. 2014. Disponível em: <<http://www.socmucimm.org/single-nucleotide-polymorphism-snp-allele-frequency-estimation-dna-pools-using-pyrosequencing/#prettyPhoto>>. Acesso em: 23 Mar. 2016.

THE BOVINE HAPMAP CONSORTIUM; GIBBS, R. A.; TAYLOR, J. F.; VAN TASSELL, C. P.; et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, v. 324, n. 5926, p. 528-532, 2009.

TILMAN, D.; CASSMAN, K. G.; MATSON, P. A.; NAYLOR, R.; POLASKY, S. Agricultural sustainability and intensive production practices. **Nature**, v. 418, n. 6898, p. 671-677, 2002.

VAN MELIS, M.; ELER, J.; ROSA, G.; et al. Additive genetic relationships between scrotal circumference, heifer pregnancy, and stayability in Nelore cattle. **Journal of Animal Science**, v. 88, n. 12, p. 3809-3813, 2010.

VAN TASSELL, C. P.; SMITH, T. P.; MATUKUMALLI, L. K.; et al. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nature methods**, v. 5, n. 3, p. 247-252, 2008.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, v. 34, n. 3, p. 275-306, 2002.

WATHES, D.; BRICKELL, J.; BOURNE, N.; et al. Factors influencing heifer survival and fertility on commercial dairy farms. **Animal**, v. 2, n. 8, p. 1135-1143, 2008.

WEIGEL, K.; VAN TASSELL, C.; O'CONNELL, J.; et al. Prediction of unobserved single nucleotide polymorphism genotypes of Jersey cattle using reference panels and population-based imputation algorithms. **J Dairy Sci**, v. 93, n. 5, p. 2229-2238, 2010.

HIPÓTESE

Estudo de associação genômica ampla utilizando alelos haplotípicos permitem encontrar associações com fenótipos de idade ao primeiro parto, peso ao nascimento, peso ao desmame, peso ao sobreano, ganho de peso do nascimento a desmama e ganho de peso da desmama ao sobreano.

OBJETIVOS

Geral

Uso de ferramenta genômica para identificação de alelos haplotípicos envolvidos na expressão de características de interesse econômico em bovinos Nelore.

Específicos

- Ampliar informação genômica utilizando imputação de genótipos de painel de média densidade para alta densidade;
- Associar alelos haplotípicos com características produtiva e reprodutiva;
- Prospectar genes candidatos associados à idade ao primeiro parto, peso ao nascimento, peso ao desmame, peso ao sobreano, ganho de peso do nascimento a desmama e ganho de peso da desmama ao sobreano em bovinos da raça Nelore.

CAPÍTULO I

Estudo de associação genômica ampla utilizando alelos haplotípicos para característica reprodutiva em bovinos Nelore

Artigo em elaboração a ser submetido:

PloS One

ISSN: 1932-6203

Fator impacto: 2.806

Qualis para área de biodiversidade: A1

Estudo de associação genômica ampla utilizando alelos haplotípicos para característica reprodutiva em bovinos Nelore

Introdução

Animais zebuínos prevalecem nos rebanhos brasileiros por apresentarem melhor adaptação ao clima predominantemente tropical [1, 2]. No entanto, as fêmeas apresentam considerável retardo no processo de maturidade sexual, ou seja, têm idade a puberdade tardia (~24 meses) quando comparadas à novilhas de raças taurinas (~14 meses). Assim sendo, a compreensão das bases genéticas envolvidas nos aspectos reprodutivos em Nelore, pode proporcionar melhora no desempenho reprodutivo das fêmeas, influenciando na produção de bezerros, reposição de matrizes e interferindo na quantidade e qualidade do rebanho [3-5].

Uma das maneiras utilizadas para aferir a precocidade sexual de fêmeas bovinas, é a idade ao primeiro parto, característica fundamental para obtenção de dados de eficiência reprodutiva do animal. Esse critério é comumente utilizado devido a facilidade de obtenção dos dados e apresenta baixo custo. A antecipação da entrada a puberdade das novilhas interfere no aumento da disponibilidade de animais tanto destinados ao corte, quanto destinados para outra finalidade no rebanho. De forma, que se as novilhas estiverem preparadas fisiologicamente, estarão aptas a gerarem bezerros viáveis. A idade ao primeiro parto é de suma importância para o sucesso do rebanho por ser um fenótipo de caráter reprodutivo envolvido diretamente com aspectos voltados a produtividade [6, 7].

O avanço nas tecnologias de genotipagem de alta densidade proporcionou a captura de informações genômicas de polimorfismos com espaçamento entre marcadores cada vez menor. Esse fato possibilitou a determinação de mecanismos genéticos subjacentes de forma eficiente e confiável, contribuindo para a precisão de estudos posteriores, como exemplo, o GWAS [8]

O GWAS (do inglês, *Genome Wide Association Study*) realiza a associação massal dos marcadores com fenótipos de interesse, permitindo o estudo de características complexas (poligênicas). Esta abordagem detecta diferentes regiões genômicas relacionadas à expressão de um fenótipo de interesse. A elucidação dos genótipos desejáveis possibilita a realização da seleção de animais precocemente. A prática da genotipagem do indivíduo no rebanho logo após o nascimento permitirá identificar a possibilidade do mesmo apresentar genótipo favorável para características

que muitas vezes serão expressas apenas na fase adulta [9, 10]. Desta forma, é possível obter maior ganho genético no rebanho por pressão de seleção superior e com menor intervalo de gerações [11].

Haplótipos são conjuntos específicos de polimorfismos presentes em determinada região cromossômica, que devido à proximidade física são herdados juntos e apresentam baixas taxas de recombinação em indivíduos de mesmo grupo contemporâneo. Estudos relatam, que análises realizadas com haplótipos são superiores às técnicas com SNPs individuais, proporcionando maior robustez aos estudos de associação. Assim sendo, a maior eficiência do uso de blocos haplóticos está relacionada à sua condição multi-alélica e minimização de associações espúrias por erros de assembly [8, 12].

O trabalho teve como objetivo realizar a associação entre alelos haplóticos distribuídos no genoma bovino e característica reprodutiva de IPP em animais Nelore e realizar a busca de genes e QTLs envolvidos na expressão deste fenótipo.

Material e métodos

Grupo amostral e genotipagem

Para o estudo foram utilizados bovinos da raça Nelore (*Bos indicus*) subdivididos em dois grupos de animais de acordo com a densidade do painel utilizado para genotipagem.

O primeiro grupo (G1) contém 2273 animais (995 machos e 1278 fêmeas), nascidos entre os anos de 1963 a 2008, provenientes de diferentes rebanhos. Esses indivíduos foram genotipados em painel Illumina BovineHD (>777.000 marcadores) e fazem parte do banco de dados do ZGC (*Zebu Genome Consortium*).

O segundo grupo (G2) contém 522 fêmeas, nascidas entre os anos de 2010 e 2011, pertencentes a Fazenda Jatiúca (Água Clara/MS). Esses indivíduos, submetidos às mesmas condições de manejo, foram genotipados em painel Illumina BovineSNP50 v2 (>50.000 marcadores).

Os procedimentos de genotipagem foram realizados seguindo o protocolo do fabricante. Os dados foram obtidos na plataforma iScan® (Illumina, San Diego, CA, USA) e posteriormente analisados por meio do software GenomeStudio v2011.1 (Illumina Inc., EUA).

Grupos Fenotípicos

As diferentes modelagens do fenótipo foram avaliadas da seguinte forma:

- IPP1 (1107 animais) - valor genético predito desregredido (dEBV), com base em fêmeas que pariram, com acurácia superior a 50%;
- IPP2 (1189 animais) - dEBV com acurácia superior a 50%. Além dos animais de IPP1, inclui informação para as fêmeas que não pariram e foram criadas nas mesmas condições (grupo de contemporâneos - GC) das fêmeas com informação para IPP1. No caso das fêmeas que não pariram, IPP2 é igual ao máximo de IPP1 no respectivo grupo de contemporâneos + 2 (penalização de 2 meses);

A expressão da característica de idade ao primeiro parto é bastante sensível a diversos efeitos, que podem ser de origem genética e ambiental. Assim sendo, o dEBV foi usado como fenótipo devido o cálculo levar em consideração não somente aspectos do animal como também de seu parentesco.

Controle de qualidade

Os dados genômicos foram submetidos a um controle de qualidade (QC) inicial, onde os critérios de exclusão de marcadores foram: não mapeados, mapeados para a mesma coordenada genômica e não autossômicos, ou seja, os SNP presentes nos cromossomos X, Y e DNA mitocondrial. Em seguida, foi realizada a análise de IBS (do inglês, *Identity By State*) para avaliação de amostras possivelmente duplicadas, onde foi calculada a proporção de alelos compartilhados por estado entre os indivíduos, assim amostras com IBS > 95% foram removidas.

Com a intenção de excluir marcadores e amostras potencialmente viesados, foram removidos os SNP levando em consideração os seguintes critérios: 1) MAF (do inglês, *Minor Allele Frequency*), marcadores com frequência alélica menor ou igual a 2%; 2) HWE (do inglês, *Hardy Weinberg Equilibrium*), SNP com valor $P < 10^{-5}$ para o Teste Exato de Fisher; 3) *Call rate* SNP (CR_{SNP}), polimorfismo que não se encontra ao menos em 95% da população. Após a verificação dos SNP foi realizado os QC por indivíduo, *Call rate* IND (CR_{IND}), onde amostras com menos de 90% de genótipos determinados foram desconsideradas.

O procedimento de controle de qualidade dos genótipos foi realizado por meio de uso de scripts customizados e da biblioteca GenABEL v. 1.8-0 [13] do software R v. 3.2.1 [14].

Análise de componentes principais (PCA)

A PCA foi realizada com a finalidade de verificar a proximidade genética entre os animais dos diferentes grupos, ou seja, averiguar o parentesco genômico entre eles. De modo que, pois quanto mais próximos geneticamente os animais são, maior a acurácia na predição dos genótipos no processo de imputação. Para isso, foram utilizados os marcadores em comum (~ 25.000 SNPs). Realizado um controle de qualidade para marcadores e amostras, de acordo com os seguintes parâmetros de exclusão: marcadores não autossômicos, SNPs com mesma coordenada genômica, $CR_{SNP} < 95\%$, $MAF < 2\%$, $EHW < 10^{-5}$, $CR_{IND} < 90\%$.

Posteriormente, uma matriz IBS (*identity by state*) foi construída e utilizou-se a função “*cmdscale*” do pacote GenABEL para a realização da PCA.

Imputação

Para o processo de imputação foram utilizados dois tipos de grupos de indivíduos: população de referência e população de imputação. Sendo o primeiro o grupo constituído de animais genotipados em painel de maior densidade (G1). Estes foram utilizados como base para inferência dos genótipos para os marcadores não presentes no painel de menor densidade (G2).

A imputação foi realizada com o software FImpute v. 2.2 [15], para a busca de haplótipos semelhantes, utilizando algoritmo apenas de base populacional.

Genótipos em fase

Para realizar a montagem dos haplótipos é necessário conhecer em qual dos cromossomos (materno ou paterno) um determinado alelo se encontra. Para tanto, foi realizada a inferência de fase dos genótipos com o software SHAPEIT v. 2.r837 [16] usando como população base todos os animais (G1 e G2). Assim foi possível obter as sequências estimadas de SNPs por par cromossômico.

Desequilíbrio de ligação

Análise de desequilíbrio de ligação (LD) entre marcadores foi realizada com a finalidade de determinar o número de SNPs por bloco haplotípico, pois a utilização de marcadores com alto LD proporciona a minimização de ocorrência de recombinação dentro dos blocos [17]. Assim, as análises de correlação entre marcadores foram realizadas utilizando blocos com 5, 10 e 15 marcadores com janelas deslizantes a cada

um SNP. A análise de LD foi realizada com o software Plink v. 1.9 [18, 19].

Haplótipos

Os blocos haplotípicos foram determinados com base na maior correlação entre SNPs encontrada na análise de desequilíbrio de ligação, sendo fixado o número de 5 marcadores (correlação entre marcadores superior a 0.37) por bloco com janela deslizante (*Slide*) de 1 SNP. Haplótipos com frequência menor que 5% na população foram retirados. A montagem dos blocos foi realizada utilizando o pacote GHap [20]. Os dados foram exportados em formato Plink, onde os genótipos foram determinados como 2, 1 ou 0, correspondendo a presença de duas cópias, uma cópia ou ausência do alelo haplotípico, respectivamente. A posição genômica para o teste de associação foi atribuída com base na média entre a localização do primeiro e último SNP do bloco haplotípico.

Determinação de pesos residuais

O cálculo dos pesos foi realizado de acordo com Garrick et al. [21]:

$$W = (1 - h^2) / \left[\left(c + \frac{1 - r^2}{r^2} \right) * h^2 \right]$$

onde h^2 é a herdabilidade da característica antes da desregressão, r^2 é o quadrado da acurácia e c é uma constante que varia entre 0 e 1. Com intuito de obter contrastes equilibrados entre dEBV de alta e baixa acurácia, foi adotado $c = 0.5$.

Análise estatística

As análises estatísticas foram realizadas com uso das funções *ghap.lmm* e *ghap.assoc* do pacote GHap. Primeiramente, os componentes de variância foram estimados pelo método de máxima verossimilhança utilizando o seguinte modelo linear misto:

$$y = Xb + Zu + e$$

em y é o vetor de dEBV, X é a matriz de incidência que relaciona os elementos de y aos efeitos fixos b (intercepto), Z é uma matriz de incidência para os efeitos aleatórios u

(animal), u é um vetor de efeitos aleatórios $\sim N(0, K\sigma_u^2)$, onde K é uma matriz de relacionamento genômico por alelos haplotípicos e ε é o vetor aleatório residual com (co)variância igual a $W\sigma_e^2$, assumindo a independência dos resíduos.

Assim, os resíduos da análise do modelo linear misto foram consideradas como observações ajustadas para covariâncias e efeito poligênico, sendo utilizadas em regressão de mínimos quadrados para testar a associação de cada alelo haplotípico com o fenótipo.

Metanálise

A fim de identificar regiões que afetam a característica de idade ao primeiro parto para as diferentes modelagens do fenótipo (IPP1 e IPP2), foram combinados os resultados das diferentes estratégias, mas de forma complementar, com intuito de proporcionar busca acurada de genes candidatos. A complexidade da característica de IPP e a presença de animais em comum nos datasets determinou a adoção de duas abordagens de metanálise. A primeira considera a população entre modelagens independentes e a segunda a existência de correlação entre as populações.

Metanálise 1

Como todos os métodos aplicados tiveram valores de P (*P-value*) a partir de distribuições normais com os mesmos parâmetros (média 0 e variância 1), a versão ponderada do método de Stouffer foi adaptada para transformação-Z de *P-values*, tal como apresentado por Whitlock [22]. Para cada alelo haplotípico e cada teste, o respectivo valor de P foi transformado em um *Z-score* por $Z_i = \Phi^{-1}(1-P_i)$. Dentro de testes de população (IPP1 e IPP2) o ponderador ω_i foi definido como a raiz quadrada do número de indivíduos. Em seguida, a estatística de k testes combinada, a cada haplótipo em cada população, foi definida como:

$$meta - SS = \frac{\sum_{i=1}^k \omega_i Z_i}{\sqrt{\sum_{i=1}^k \omega_i^2}}$$

Os valores de meta-SS (do inglês, *Meta-analysis of Selection Signals*) foram encaminhados de volta para a distribuição normal padrão, a fim de obter valores de significância combinados, que foram destinados para atender a combinação de informações entre os diferentes testes. O nível de significância para os *P-values* na

meta-SS foi com base no limiar de Bonferroni ($\alpha = 0.05/\text{número de alelos haplotípicos}$).

Metanálise 2

O método de metanálise descrito por Bolormaa et al. [6] foi utilizado para resumir estatísticas de marcadores único para as duas modelagens de IPP. Esta metodologia considera a existência de correlação entre as populações estudadas.

Para cada alelo haplotípico, t foi considerado como $q \times 1$ vetor simples de t -valores ($b/SE(b)$) para cada q características, C foi a matrix de correlação $q \times q$ de t -valores de marcadores. A estatística $t'C^{-1}t$ foi aproximadamente distribuída como χ^2 com q graus de liberdades (df), sob a hipótese nulidade de não pleiotropia.

Exploração das regiões genômicas

A exploração das regiões genômicas próximas aos haplótipos mais significativos, foram determinadas com os seguintes parâmetros: 1) cologaritmo de valores de P superiores ao limiar de Bonferroni ou 10 regiões mais significativas (top 10); 2) proximidade (alelos haplotípicos com distâncias inferiores a 500 Kb foram considerados como mesma região e selecionado aquele com menor valor de P) e 3) janela de 1 Mb (500 Kb para cada lado da posição do haplótipo). Este procedimento foi realizado com uso do Bedtools v. 2.25.0 [23].

As regiões mais significativas foram exploradas com base no *assembly* UMD v3.1. A ferramenta *BioMart* [24] do banco de dados Ensembl foi utilizada para busca dos genes próximos e o banco de dados *Cattle QTLdb* [25] para examinar a proximidade com QTLs de espécie bovina previamente catalogados na literatura.

Os resultados dos genes foram revisados utilizando a literatura científica disponível. Os bancos de dados *Mouse Genome Informatics* [26] e *GeneCards: The Human Gene Database* [27], forneceram informações da participação e envolvimento de genes ortólogos com a característica de idade ao primeiro parto.

Resultados

A partir da análise de desequilíbrio de ligação, foram observadas correlações médias entre marcadores de 0.37, 0.32 e 0.3, respectivamente para blocos de 5, 10 e 15 SNPs.

A Figura 6 demonstra os resultados referentes a PCA realizada com as populações G1 e G2. Foi possível observar que os pontos se distribuem de forma

semelhante nas diferentes populações, o que representa a proximidade genética e parentesco genômico entre os mesmos, garantindo resultados satisfatórios para o processo de imputação. No entanto, a diferença do número amostral entre os grupos G1 e G2, determinou resultados diferentes, conforme demonstrado na representação gráfica (Figura 6).

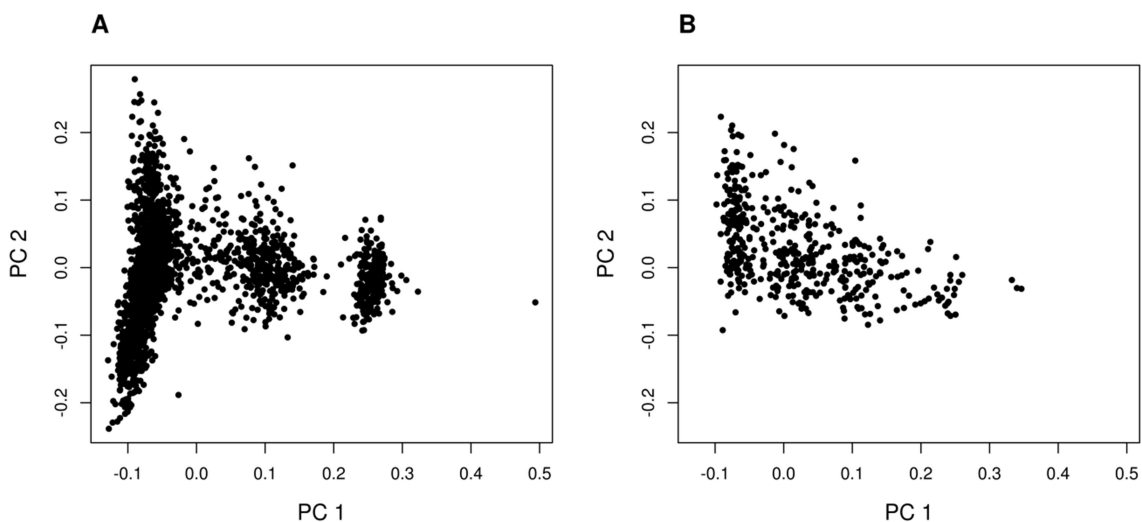


Figura 6. Representação gráfica da análise de componentes principais (PCA). Cada ponto representa, (A) distribuição dos animais de G1 e (B) distribuição dos animais G2.

Ao término do controle da qualidade e procedimento de imputação, restaram 511,375 SNPs e 2,560 indivíduos, sendo 2,107 do G1 e 453 G2. Como resultado da construção dos blocos haplotípicos foram obtidos 2.240,166 e 2.238,796 alelos, respectivamente para IPP1 e IPP2, que foram utilizados posteriormente para os testes de associação e metanálise.

Apesar da existência de discrepâncias em várias regiões cromossômicas entre os resultados dos testes de associação para os diferentes valores de IPP, foram observados locais similares entre os dois testes, como exemplo nos cromossomos 3, 5, 6, 21 e 29 (Figura 7). Após a realização da metanálise, as regiões coincidentes foram as que apresentaram maior probabilidade de associação (Figura 8).

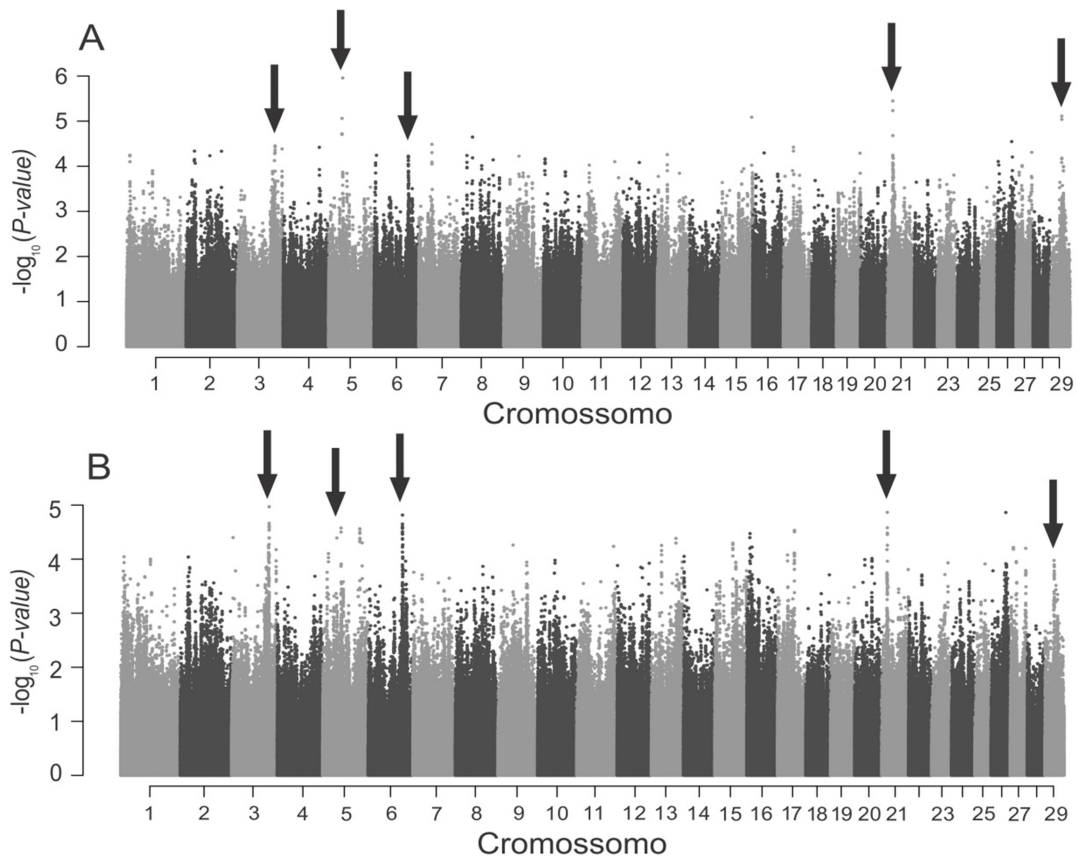


Figura 7. *Manhattans plot* para estudo de associação genômica ampla entre blocos haplotípicos e características de IPP1 (A) e IPP2 (B). Cada ponto do gráfico representa um haplótipo e as cores cinza e preto alternadas, os cromossomos. As setas indicam as regiões coincidentes.

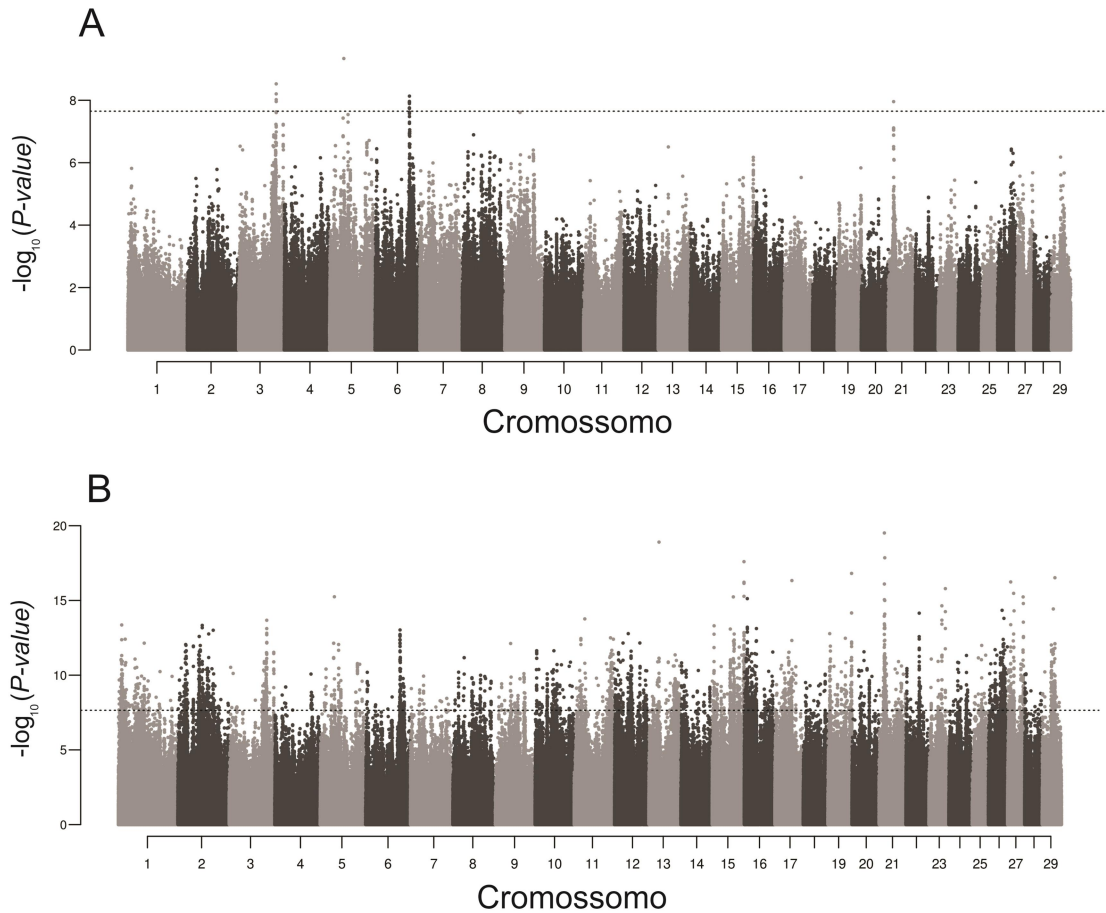


Figura 8. *Manhattans plot* para estudo de associação entre $-\log_{10}(P\text{-value})$ dos alelos haplotípicos da metanálise 1 (A) e metanálise 2 (B) e característica de idade ao primeiro parto. Cada ponto do gráfico é um alelo haplotípico. Os cromossomos foram destacados em cores alternadas de cinza e preto, e o limiar de Bonferroni representado pela linha pontilhada.

As regiões mais significativas mediante critérios previamente estabelecidos para as diferentes metodologias de metanálise são apresentadas na Tabela 3.

Tabela 3. Informações das diferentes metanálises (1 e 2), demonstrando a localização cromossômica, o tipo de metanálise utilizado, a posição de início e final do haplótipo, com seu respectivo SNP, e valor de P para cada região associada.

Crom.	Metan.	Início	Final	SNP1	SNP2	P-value
3	1	99936110	99959169	rs133273360	rs133299434	2.98 X 10 ⁻⁹
5	1	37960086	37979520	rs136867887	rs135623406	4.60 X 10 ⁻¹⁰
6	1	91649597	91663333	rs134333101	rs135909163	7.37 X 10 ⁻⁹
21	1	14204851	14219729	rs135923802	rs135841763	1.10 X 10 ⁻⁸
5	2	37960086	37979520	rs136867887	rs135623406	5.66 X 10 ⁻¹⁶
13	2	27305170	27337367	rs110668010	rs43712651	1.24 X 10 ⁻¹⁹
15	2	84681406	84708745	rs133085249	rs41792206	2.53 X 10 ⁻¹⁸
17	2	45241809	45248597	rs43328817	rs43328766	4.66 X 10 ⁻¹⁷
19	2	62958127	62983104	rs42702721	rs42702689	1.55 X 10 ⁻¹⁷
21	2	14304199	14317771	rs109985159	rs136536188	3.05 X 10 ⁻²⁰
23	2	43712068	43759190	rs137045977	rs43465627	1.61 X 10 ⁻¹⁶
27	2	8105417	8126082	rs42531087	rs109663354	5.77 X 10 ⁻¹⁷
27	2	15856634	15876873	rs42109271	rs42115916	3.34 X 10 ⁻¹⁶
29	2	34158773	34162753	rs135038561	rs133302510	3.02 X 10 ⁻¹⁷

Os resultados da metanálise 1 identificaram 45 QTLs e 27 genes candidatos. Dentre estes, 9 estão relacionados com aspectos reprodutivos previamente descritos (Tabela 4). Os genes *FAAH*, *BTC* e *PARMI*, apresentaram vínculos com implantação do embrião no útero, desenvolvimento de órgãos reprodutivos em fêmeas e controle do hormônio progesterona, respectivamente. Os genes *TAL1*, *DMBX1*, *STIL*, *PRICKLE1*, *CHD2* e *RGMA* estão envolvidos com o desenvolvimento embrionário normal.

Tabela 4. Sumário dos genes presentes em janelas de 1 Mb centralizadas aos haplótipos mais significativos para metanálise 1.

Gene	Ensembl ID	Posição (BTA:início-final)	Descrição
<i>TALI</i>	ENSBTAG00000005029	3:99610187-99618464	células-T de leucemia linfocítica aguda 1
<i>FAAH</i>	ENSBTAG00000007507	3:100317761-100338715	ácido graxo amida hidrolase
<i>DMBX1</i>	ENSBTAG00000009625	3:100217682-100223303	diencefalo/mesencefalo homeobox 1
<i>STIL</i>	ENSBTAG00000017844	3:99537145-99587496	SCL-interruptor de locus de proteína
<i>PRICKLE1</i>	ENSBTAG00000004237	6:91430305-91480129	prickle - proteína 1
<i>BTC</i>	ENSBTAG00000008596	5:38273723-38388071	betacelulina
<i>PARM1</i>	ENSBTAG00000015919	6:91597122-91723390	regulador-andrógeno de próstata mucina - proteína 1
<i>CHD2</i>	ENSBTAG00000017776	21:14390575-14483806	proteína 2 cromodomínio- helicase de ligação ao DNA
<i>RGMA</i>	ENSBTAG00000018603	21:14349793-14376704	membro A da família de proteínas de orientação repulsiva molecular

Os resultados da metanálise 2, de acordo com metodologia proposta por Bolormma, 27 genes e 129 QTLs associados ao fenótipo de IPP foram identificados. Dentre esses, 10 alelos haplotípicos com maiores índices de significância foram selecionados, considerando o limiar de Bonferroni (Tabela 3).

Os genes candidatos foram explorados pelas ferramentas de análise funcional, evidenciando 11 genes em 8 regiões distintas, que apresentaram associação com caracteres reprodutivos (Tabela 5). Os genes *PRICKLE1*, *GAD2*, *GNA13* e *CHD2* estão envolvidos em desenvolvimento embrionário; *GIF* e *TCN1* em transporte e concentração da vitamina B12; *EDN1*, *OPCML* e *PXMP2* relacionados a órgãos reprodutivos em fêmeas; e *GOLGA3* e *SPATA19* com infertilidade em machos.

Tabela 5. Sumário dos genes presentes em janelas de 1 Mb centralizadas aos haplótipos mais significativos para metanálise 2.

Gene	Ensembl ID	Posição (BTA:inicio-final)	Descrição
<i>PRICKLE1</i>	ENSBTAG00000008596	5:38273723-38388071	prickle - proteína 1
<i>GAD2</i>	ENSBTAG00000008056	13:26937414-26999001	glutamato descarboxilase 2
<i>GIF</i>	ENSBTAG00000012057	15:84361364-84380392	fator intrínseco gástrico
<i>TCN1</i>	ENSBTAG00000020580	15:84389636-84404937	transcobalamina 1
<i>PXMP2</i>	ENSBTAG00000007510	17:45501312-45511335	proteína de membrana do peroxissoma 2 golgi A3 golgi A3
<i>GOLGA3</i>	ENSBTAG00000009755	17:45418676-45455896	
<i>GNAI3</i>	ENSBTAG000000046174	19:62550515-62575992	proteína de ligação ao nucleótido guanina (proteínas G), alfa 13
<i>CHD2</i>	ENSBTAG00000017776	21:14390575-14483806	proteína 2 cromodomínio-helicase de ligação ao DNA
<i>EDNI</i>	ENSBTAG00000008096	23:44010444-44017425	endotelina 1
<i>SPATA19</i>	ENSBTAG00000014651	29:33755352-33760891	associado a espermatogênese 19
<i>OPCML</i>	ENSBTAG00000020769	29:34554780-35085038	opióide de ligação proteica/molécula de adesão celular

Dentre os QTLs encontrados nas diferentes abordagens de metanálise e associados à característica reprodutiva, destacaram-se os relacionados à facilidade no parto, taxa de prenhez das filhas e natimortos (Tabela 6).

Tabela 6. QTLs relacionados à características reprodutivas encontradas em janelas de 1Mb nas proximidades dos haplótipos para ambas metanálises.

Descrição	Posição (BTA:inicio-final)	QTLdb ID
Facilidade de parto	5:38351838-38351878	40879
	17:45145589-45145629	49226
	17:45403728-45403768	49239
	19:62805117-62805157	50256
	19:62805117-62805157	50263
	19:63376798-63376838	56421
	19:63376100-63376140	56422
	27:7833757-7833797	53331
	27:7833757-7833797	53343
	27:15637718-15637758	53392
	27:15637718-15637758	53404
	3:84669155-102513835	15173
	17:45403728-45403768	49240
27:7833757-7833797	53333	
Taxa de Prenhez das filhas	27:15637718-15637758	53394
	6:91677962-91678002	57130
Índice de partos	3:84669155-102513835	15172
Sobrevivência Embrionária	6:91677962-91678002	57089
Natimorto	3:84669155-102513835	15174
	3:99853456-101636723	30488
	17:45403728-45403768	49241
	19:62805117-62805157	50264
	27:7833757-7833797	53344
	27:15637718-15637758	53405
Taxa de Concepção	6:91677962-91678002	57143
	6:91677962-91678002	57167

Discussão

Quanto maior o parentesco entre os animais dentro de cada grupo (referência e

imputação) e número de indivíduos na população de referência, melhores são os resultados na predição de genótipos, implicando em maior acurácia dos resultados na imputação [28]. Assim, a eficiência do uso dos dados imputados foi determinada pela proximidade genética entre os grupos e o número de indivíduos na população de referência (superior a mil animais).

O GWAS utilizando blocos haplotípicos, tem capacidade superior de associação conferindo maior robustez aos resultados, quando comparados às análises utilizando SNPs, pois como é utilizado um conjunto de marcadores ocorre a minimização de associações espúrias por erros de *assembly*. A abordagem de haplótipo se sobressai ao uso de marcadores únicos, pelo fato dos efeitos de cada SNP separadamente serem considerados pequenos para ultrapassar os rigorosos limites de significância [8].

Para possibilitar maior robustez nos resultados referentes ao GWAS com a IPP, foram utilizadas diferentes metodologias de metanálise (1 e 2). Ambas detectaram genes, cuja função biológica está relacionada ao fenótipo estudado. No entanto, a metanálise 2 superestimou os resultados, ou seja, diversas regiões cromossômicas ultrapassaram o limiar de significância, isso pode ter ocorrido devido a ausência de alta correlação entre os modelos IPP1 e IPP2 (30%) [29]. A metanálise apresentada na Figura 8 mostrou o comportamento de um caractere de herança poligênica, sendo que diferentes regiões apresentaram associação com o fenótipo em estudo [30].

Dentre os genes encontrados na metanálise 1, os genes *TALI*, *DMBX1*, *STIL*, *RGMA* foram descritos como importantes no desenvolvimento cerebral embrionário em ratos. Apenas o *DMBX1* pode ser letal ou levar à infertilidade quando não são expressos em ratos [31-33].

FAAH, *BTC* e *PARMI*, são genes envolvidos com a formação de órgãos reprodutivos em fêmeas. *FAAH*, possui atividade no útero, mais especificamente em células epiteliais uterinas e miométrio celular durante o estro (período fértil) em ratos, sendo modulado por hormônios sexuais. Apesar de poucos estudos retratarem sua atuação, sabe-se que este gene influencia na implantação do embrião no útero [34], podendo explicar a ocorrência de IPP tardia por deficiência no processo de alocação do embrião na parede uterina.

O gene *BTC* é expresso no ovário e útero, relatado como codificador de importante proteína moduladora da fertilidade em ratas. Juntamente com o *PARMI*, está envolvido na taxa de concentração de progesterona, regulado por hormônio luteinizante, interferindo na função lútea e eficiência da ovulação. Ambos apresentam indícios de

serem potenciais reguladores de fertilidade em fêmeas [35, 36].

Acerca dos genes de relação direta com a idade ao primeiro parto para a metanálise 2, os genes *GIF*, *TCNI*, *EDNI* e *OPCML* apresentaram relação com o desenvolvimento normal de órgãos reprodutivos .

Dentre os elementos envolvidos na reprodução, os eventos que antecedem a liberação do óvulo, são dependentes de liberação hormonal, alterações moleculares e bioquímicas do ovário, levando a ruptura dos folículos e liberação do ovócito. O gene *EDNI* teve sua expressão identificada em folículos pré-ovulatórios, sendo relevante para estudos de caracteres reprodutivos [37].

A expressão do gene *OPCML* no útero e no ovário, sugere função reprodutiva, no entanto ainda é incerta a sua influência durante a ovulação. Sob aspecto patológico, a inativação ou decréscimo da expressão do *OPCML* apresenta envolvimento com o câncer epitelial de ovário, atuando como supressor tumoral [38].

O gene *GIF* codifica um membro da família de proteínas carreadoras da cobalamina (vitamina B12), sendo de grande importância para a absorção adequada desta vitamina. A sua deficiência em mulheres grávidas, está relacionada com alterações durante o período gestacional, tendo como consequência o aborto espontâneo [39]. O gene *TCNI*, assim como o *GIF*, exerce influência na concentração da cobalamina no organismo, além de ter sido apontado na atuação do metabolismo do folato (ácido fólico) em estudos de associação genômica em humanos, é essencial para o período gestacional promovendo desenvolvimento embrionário normal [40].

Os genes *PRICKLE1*, *GAD2*, *PXMP2*, *GNAI3* e *CHD2*, foram relacionados ao desenvolvimento embrionário. O *PRICKLE1*, *GNAI3* e *GAD2* foram associados à letalidade e problemas vasculares precoces do embrião [41, 42]. O *CHD2* e *PXMP2* similarmente aos anteriores, quando mutados, são letais ao embrião, além de desencadear o desenvolvimento incompleto ou anormal do óvulo em roedores [43].

O gene espermatogênese-associado-19 (*SPATA19*) é importante para diferenciação de células germinativas. Este é amplamente expresso durante o desenvolvimento dos testículos em ratos, considerado possível regulador de diferenciação celular em processos cancerígenos [44]. Clay et al. [45] e Raheja et al. [46] sugerem que, apesar da baixa herdabilidade, para melhorar a fertilidade das novilhas é importante a escolha dos touros, sendo assim, foram encontrados caracteres relativos à fertilidade em machos que podem ser relevantes na seleção em ambos os sexos para fenótipos reprodutivos.

Quanto aos genes em regiões próximas à marcadores significativos encontrados, *PRICKLE1* e *CHD2* foram identificados por ambas metodologias (Tabela 4 e 5), proporcionando evidências da associação dos genes com a característica estudada.

A presença de QTLs próximos às regiões associadas e relação destes com características reprodutivas (facilidade de parto, taxa de prenhez das filhas, índice de partos, sobrevivência embrionária, natimorto e taxa de concepção), evidenciaram e suportam as respostas obtidas para os genes. Assim sendo, pode-se atribuir maior confiabilidade aos resultados referentes a associações das regiões identificadas com característica de idade ao primeiro parto.

Estes resultados constituem subsídio para estudos posteriores, com o propósito de elucidar os mecanismos aos quais os genes relatados possam estar envolvidos na expressão da IPP. A maior compreensão poderá proporcionar o uso de genótipos específicos como norteadores em programas de melhoramento genético animal, além de possibilitar a construção de painéis de baixa densidade (de custo reduzido) com finalidade restrita de avaliar genótipos favoráveis à seleção.

Conclusão

A abordagem do uso de haplótipos utilizando as duas metodologias de metanálise permitiu associar os diferentes genótipos com idade ao primeiro parto. Os genes e QTLs identificados apresentaram relação com a IPP. A existência de genes candidatos fornecerá subsídios para estudos de validação possibilitando sua aplicabilidade em programas de melhoramento animal.

Referências

1. Finch V. Body temperature in beef cattle: its control and relevance to production in the tropics. *Journal of Animal Science*. 1986;62(2):531-542.
2. Landaeta-Hernández A, Zambrano-Nava S, Hernández-Fonseca JP, Godoy R, Calles M, Iragorri JL, et al. Variability of hair coat and skin traits as related to adaptation in Criollo Limonero cattle. *Tropical Animal Health and Production*. 2011;43(3):657-663.
3. Abeygunawardena H, Dematawewa C. Pre-pubertal and postpartum anestrus in tropical Zebu cattle. *Animal Reproduction Science*. 2004;82:373-387.
4. Baruselli PS, Gimenes LU, Sales JNdS. Fisiologia reprodutiva de fêmeas taurinas e zebuínas. *Rev Bras Reprod Anim*,. 2007;31(2):205-211.

5. Milazzotto M, Rahal P, Nichi M, Miranda-Neto T, Teixeira L, Ferraz J, et al. New molecular variants of hypothalamus–pituitary–gonad axis genes and their association with early puberty phenotype in *Bos taurus indicus* (Nellore). *Livestock Science*. 2008;114(2):274-279.
6. Boligon AA, Baldi F, Albuquerque LGd. Genetic parameters and relationships between growth traits and scrotal circumference measured at different ages in Nellore cattle. *Genetics and molecular biology*. 2011;34(2):225-230.
7. Wathes D, Brickell J, Bourne N, Swali A, Cheng Z. Factors influencing heifer survival and fertility on commercial dairy farms. *Animal*. 2008;2(8):1135-1143.
8. Wu Y, Fan H, Wang Y, Zhang L, Gao X, Chen Y, et al. Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. *PloS one*. 2014;9(10):e109330.
9. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005;6(2):95-108.
10. Resende MDVd, Aguiar AM, Abad JIM, Missiaggia AA, Sansaloni C, Petroli C, et al. Computação da seleção genômica ampla (GWS). Embrapa Florestas. 2010.
11. Aneja B, Yadav NR, Chawla V, Yadav RC. Sequence-related amplified polymorphism (SRAP) molecular marker system and its applications in crop improvement. *Molecular Breeding*. 2012;30(4):1635-1648.
12. Utsunomiya AT, Santos DJ, Boison SA, Utsunomiya YT, Milanese M, Bickhart DM, et al. Revealing misassembled segments in the bovine reference genome by high resolution linkage disequilibrium scan. *BMC genomics*. 2016;17(1):705.
13. Aulchenko YS, de Koning DJ, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*. 2007;177(1):577-585.
14. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2013.
15. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC genomics*. 2014;15(1):478.
16. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nature methods*. 2012;9(2):179-181.
17. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics*. 2002;3(4):299-309.

18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):1.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 2007;81(3):559-575.
20. Utsunomiya YT, Milanese M, Utsunomiya AT, Ajmone-Marsan P, Garcia JF. GHap: An R package for Genome-wide Haplotyping. *Bioinformatics*. 2016:btw356.
21. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*. 2009;41(1):1.
22. Whitlock M. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of evolutionary biology*. 2005;18(5):1368-1373.
23. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.
24. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)*. 2011;2011:30.
25. Hu ZL, Park CA, Wu XL, Reecy JM. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res*. 2013;41:871-879.
26. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Group MGD. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic acids research*. 2015;43(D1):D726-D736.
27. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*. 2016:1.30. 31-31.30. 33.
28. Carvalheiro R, Boison SA, Neves HH, Sargolzaei M, Schenkel FS, Utsunomiya YT, et al. Accuracy of genotype imputation in Nelore cattle. *Genetics Selection Evolution*. 2014;46(69):1-11.
29. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature,

- fatness and reproduction in beef cattle. *PLoS Genet.* 2014;10(3):e1004198.
30. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics.* 2011;88(1):76-82.
 31. Bennett H, Presti A, Adams D, Rios J, Benito C, Cohen D. A prenatal presentation of severe microcephaly and brain anomalies in a patient with novel compound heterozygous mutations in the STIL gene found postnatally with exome analysis. *Pediatric neurology.* 2014;51(3):434-436.
 32. Fitzgerald DP, Bradford D, Cooper HM. Neogenin is expressed on neurogenic and gliogenic progenitors in the embryonic and adult central nervous system. *Gene Expression Patterns.* 2007;7(7):784-792.
 33. Wong L, Weadick CJ, Kuo C, Chang BS, Tropepe V. Duplicate *dmbx1* genes regulate progenitor cell cycle and differentiation during zebrafish midbrain and retinal development. *BMC developmental biology.* 2010;10(1):100.
 34. Xiao A, Zhao Y, Duan E. Expression and regulation of the fatty acid amide hydrolase gene in the rat uterus during the estrous cycle and peri-implantation period. *Molecular human reproduction.* 2002;8(7):651-658.
 35. Gratao AA, Dahlhoff M, Sinowatz F, Wolf E, Schneider MR. Betacellulin overexpression in the mouse ovary leads to MAPK3/MAPK1 hyperactivation and reduces litter size by impairing fertilization. *Biology of reproduction.* 2008;78(1):43-52.
 36. Park JY, Jang H, Curry TE, Sakamoto A, Jo M. Prostate Androgen-Regulated Mucin-Like protein 1: A Novel Regulator of Progesterone Metabolism. *Molecular Endocrinology.* 2013;27(11):1871-1886.
 37. Choi D-H, Kim EK, Kim K-H, Lee K-A, Kang D-W, Kim HY, et al. Expression pattern of endothelin system components and localization of smooth muscle cells in the human pre-ovulatory follicle. *Human reproduction.* 2011;26(5):1171-1180.
 38. Fleming JS, McQuillan HJ, Millier MJ, Sellar GC. Expression of ovarian tumour suppressor OPCML in the female CD-1 mouse reproductive tract. *Reproduction.* 2009;137(4):721-726.
 39. Pardo J, Peled Y, Bar J, Hod M, Sela B, Rafael ZB, et al. Evaluation of low serum vitamin B12 in the non-anaemic pregnant patient. *Human Reproduction.* 2000;15(1):224-226.

40. Wang J, Zhao J-Y, Wang F, Peng Q-Q, Hou J, Sun S-N, et al. A genetic variant in vitamin B12 metabolic genes that reduces the risk of congenital heart disease in Han Chinese populations. *PloS one*. 2014;9(2):e88332.
41. Liu C, Lin C, Gao C, May-Simera H, Swaroop A, Li T. Null and hypomorph Prickle1 alleles in mice phenocopy human Robinow syndrome and disrupt signaling downstream of Wnt5a. *Biology open*. 2014;3(9):861-870.
42. Moon AM, Stauffer AM, Schwindinger WF, Sheridan K, Firment A, Robishaw JD. Disruption of G-protein γ 5 subtype causes embryonic lethality in mice. *PloS one*. 2014;9(3):e90970.
43. Kulkarni S, Nagarajan P, Wall J, Donovan DJ, Donell RL, Ligon AH, et al. Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. *American Journal of Medical Genetics Part A*. 2008;146(9):1117-1127.
44. Nourashrafeddin S, Ebrahimzadeh-Vesal R, Miryounesi M, Aarabi M, Zarghami N, Modarressi MH, et al. Analysis of SPATA19 gene expression during male germ cells development, lessons from in vivo and in vitro study. *Cell Biology International Reports*. 2014;21(1):1-7.
45. Clay J, McDaniel B, Brown C. Variances of and correlations among progeny tests for reproductive traits of cows sired by AI bulls. *Journal of dairy science*. 2004;87(7):2307-2313.
46. Raheja K, Nadarajah K, Burnside E. Relationship of bull fertility with daughter fertility and production traits in Holstein dairy cattle. *Journal of dairy science*. 1989;72(10):2679-2682.

CAPÍTULO II

Estudo de associação genômica ampla utilizando haplótipos para característica produtiva em bovinos Nelore

Artigo em elaboração a ser submetido:

BMC Genomics

ISSN: 1471-2164

Fator impacto: 3.986

Qualis para área de biodiversidade: A2

Estudo de associação genômica ampla utilizando haplótipos para característica produtiva em bovinos Nelore

Introdução

A raça Nelore destaca-se no rebanho de zebuínos utilizados na pecuária brasileira, atingindo alta dispersão no território nacional. Este fato pode estar relacionado à sua predisposição para características adaptativas favoráveis, que tornam esses animais capazes de se adequarem ao clima, regime alimentar à pasto, manifestando rusticidade e resistência a parasitas (ACNB, 2016). Acerca dos aspectos positivos de sua criação, os bovinos Nelore são amplamente utilizados em programas de melhoramento, que visam consolidar fenótipos de produtividade ainda não aperfeiçoados na raça (FILHO et al., 2009).

Diante da crescente demanda por carne (FAO 2009), a criação de bovinos para corte tem sido um desafio aos produtores, que vislumbram tornar seus rebanhos mais produtivos, porém de maneira que essa produção seja sustentável. Assim sendo, são desejados aumento da produtividade sem prejuízos para a qualidade da carne bovina. Para tanto, critérios são adotados para avaliar caracteres que influenciam na carcaça e também estão relacionados ao seu rendimento.

A caracterização da produtividade pode ser obtida por meio da análise animais com potencial superior de produção de carne. Dentre as características fenotípicas que permitem essa avaliação, destacam-se as relativas ao peso. Uma vez que permitem determinar o potencial de crescimento do animal, acompanhando seu desenvolvimento desde o nascimento do bovino a fase adulta (GUNAWAN e JAKARIA, 2011). Essas características de crescimento são facilmente mensuradas, e descritas como de moderada herdabilidade, sendo de 0,25; 0,33; 0,34 para peso ao nascimento, desmama e sobreano, respectivamente (LÔBO e FILHO, 2003).

Estudos de associação genômica ampla voltados ao melhoramento desses animais possibilitam maior produtividade utilizando menor número de animais no rebanho e conseqüentemente proporcionam manutenção das mesmas extensões de áreas de pastagens. A utilização de informações oriundas de SNP e analisadas pelo GWAS, possibilita explicar parte das variações fenotípicas, pela associação de numerosos marcadores à características de interesse, podendo melhorar a precisão ao selecionar animais para acasalamento direcionado, proporcionando uma pressão de seleção maior e aumento do ganho genético (BUZANSKAS et al., 2014, SANTANA et al., 2014).

Nesse contexto, torna-se relevante o estudo de associação entre alelos haplotípicos com características referentes ao peso buscando a seleção de genótipos favoráveis. Dentre esse caracteres destacam-se: peso ao nascimento (PN), peso a desmama (PD), peso ao sobreano (PS), ganho de peso do nascimento a desmama (GPND) e ganho de peso da desmama ao sobreano (GPDS) em fêmeas Nelore. Além disso, também é importante destacar a investigação de genes e QTLs relacionados à expressão destes fenótipos.

Material e métodos

Grupo amostral

Para o estudo foram utilizados 522 fêmeas da raça Nelore (*Bos indicus*), nascidas entre os anos de 2010 e 2011, pertencentes à Fazenda Jatiúca (Água Clara/MS) e submetidas às mesmas condições de manejo. O material biológico foi obtido por intermédio da empresa DeltaGen. Os animais foram genotipados em painel Illumina BovineSNP50 v2 (>50.000 marcadores) e imputados para painel de alta densidade, com base em população de referencia contendo 2273 animais Nelore genotipados com Illumina BovineHD (>777.000 marcadores).

Os procedimentos de genotipagem foram realizados seguindo o protocolo do fabricante, a leitura dos painéis foi realizada em plataforma iScan® (Illumina, San Diego, CA, USA) e posteriormente analisados usando o GenomeStudio v2011.1 (Illumina Inc., EUA).

Características em estudo

Para os estudos de associação foram utilizados dados de características de peso ao nascimento (PN), peso a desmama (PD), peso ao sobreano (PS), ganho de peso do nascimento a desmama (GPND) e ganho de peso da desmama ao sobreano (GPDS).

As medidas de peso foram aferidas do nascimento ao sobreano. As pesagens foram realizadas com os animais em jejum durante um período de 14 horas. Este procedimento foi conduzido em balanças eletrônicas previamente testadas.

Foram utilizados os fenótipos ajustados para PN, PD, PS, GPND e GPDS, fornecidos pela empresa GenSys, seguindo o procedimentos de modelagem padronizados pelos mesmo.

Controle de qualidade

Os dados genômicos foram submetidos a um controle de qualidade (QC) inicial, onde os critérios de exclusão de marcadores foram: não mapeados, mapeados para a mesma coordenada genômica e não autossômicos, ou seja, os SNP presentes nos cromossomos X, Y e DNA mitocondrial. Em seguida foi realizada a análise de IBS (do inglês, *Identity By State*) para avaliação de amostras possivelmente duplicadas, onde é calculada a proporção de alelos compartilhados por estado entre os indivíduos, assim amostras com $IBS > 95\%$ foram removidas.

Com a intenção de excluir marcadores e amostras potencialmente viesados, foram removidos os SNP levando em consideração os seguintes critérios: 1) MAF (do inglês, *Minor Allele Frequency*), marcadores com frequência alélica menor ou igual a 2%; 2) HWE (do inglês, *Hardy Weinberg Equilibrium*), SNP com valor $P < 10^{-5}$ para o Teste Exato de Fisher; 3) *Call rate* SNP (CR_{SNP}), polimorfismo que não se encontra ao menos em 95% da população. Após a verificação dos SNPs foi realizado os QC por indivíduo, *Call rate* IND (CR_{IND}), onde amostras com menos de 90% de genótipos determinados foram desconsideradas.

O procedimento de controle de qualidade dos genótipos foi realizado por meio de uso de scripts customizados e da biblioteca GenABEL v. 1.8-0 (AULCHENKO et al., 2007) do pacote estatístico R v. 3.2.1 (R CORE TEAM, 2008).

Análise de componentes principais (PCA)

Foi realizada a análise de componentes principais (PCA) com a finalidade de verificar a proximidade genética entre os animais dos diferentes grupos, ou seja, averiguar o parentesco genômico entre eles, pois quanto mais próximos geneticamente os animais são, maior a acurácia na predição dos genótipos no processo de imputação. Para isso, foram utilizados os marcadores em comum (~ 25.000 SNPs), realizado um controle de qualidade para marcadores e amostra de acordo com os seguintes parâmetros de exclusão: marcadores não autossômicos, SNPs com mesma coordenada genômica, $CR_{SNP} < 95\%$, $MAF < 2\%$, $EHW < 10^{-5}$, $CR_{IND} < 90\%$.

Posteriormente foi construída uma matriz IBS (*identity by state*) e a partir dela utilizando a função “*cmdscale*” do pacote GenABEL foi realizado o PCA.

Imputação

Para o processo de imputação foram utilizados dois tipos de grupos de

indivíduos: população de referência e população de imputação. Sendo o primeiro o grupo constituído de animais genotipados em painel de maior densidade, que foram utilizados como base para inferência dos genótipos para os marcadores não presentes no painel de menor densidade, o qual foi utilizado para genotipagem da população de imputação.

A imputação foi realizada com o software FImpute v. 2.2 (SARGOLZAEI et al. 2014), para a busca de haplótipos semelhantes, utilizando algoritmo apenas de base populacional.

Desequilíbrio de ligação

Análise de desequilíbrio de ligação (LD) entre marcadores foi realizada com a finalidade de determinar o número de SNPs por bloco haplotípico, pois a utilização de marcadores com alto LD proporciona a minimização de ocorrência de recombinação dentro dos blocos (ARDLIE et al., 2002). Assim, as análises de correlação entre marcadores foram realizadas utilizando blocos com 5, 10 e 15 marcadores com janelas deslizantes a cada um SNP. A análise de LD foi realizada com o software Plink v. 1.9 (CHANG et al., 2015; PURCELL et al., 2007).

Genótipos em fase

Para realizar a montagem dos haplótipos é necessária saber em qual dos cromossomos (materno ou paterno) um determinado alelo se encontra, para isso foi realizada a inferência de fase dos genótipos utilizando o software SHAPEIT v. 2.r837 (DELANEAU et al., 2012) usando todos os animais como população base. Assim foi possível obter as sequências estimadas de SNPs por par cromossômico.

Haplótipos

Os blocos haplotípicos foram determinados fixando o número de 5 marcadores por bloco com janela deslizante (*Slide*) de 1 SNP. A montagem dos blocos foi realizada utilizando o pacote GHap (UTSUNOMIYA et al., 2016) do R v. 3.2.1. Haplótipos com frequência menor que 5% na população foram retirados. Os dados foram exportados em formato Plink v.1.9 (CHANG et al., 2015; PURCELL et al., 2007), onde os genótipos foram determinados como 2, 1 ou 0, correspondendo a presença de duas cópias, uma cópia ou ausência do alelo haplotípico, respectivamente.

Análise estatística

A análise de associação baseada em modelo linear misto foi realizada no programa GCTA (*Genome-wide Complex Trait Analysis*) (YANG et al., 2011), onde foi aplicado o seguinte modelo:

$$y = \alpha + \beta x + g + \varepsilon$$

sendo y o fenótipo observado, α é média geral, β é o efeito aditivo (efeito fixo) do haplótipo candidato a ser testado para a associação, x é a variável indicadora do genótipo do bloco haplotípico, g é o efeito poligênico (efeito aleatório) ou seja, o efeito acumulado de todos os alelos haplotípicos, exceto o cromossomo onde o haplótipo candidato está localizado, evitando assim, que o marcador seja testado ao mesmo tempo como efeito fixo e aleatório (YANG et al., 2014) e ε é o resíduo.

Exploração das regiões genômicas

As regiões genômicas próximas aos alelos haplótipos mais significativos foram exploradas de acordo com os seguintes parâmetros: proximidade entre blocos (blocos haplotípicos com distâncias inferiores a 500 Kb foi considerado somente aquele com menor *P-value*), cinco regiões mais significativas e janela de 1 Mb (500 Kb para cada lado da posição do alelo haplotípo). Estes procedimentos foram realizados com uso do R v. 3.2.1 e Bedtools v. 2.25.0 (QUINLAN e HALL, 2010).

As cinco regiões mais significativas foram exploradas com base no *assembly* UMD v3.1. A ferramenta *BioMart* v.84 (KINSELLA et al., 2011) do banco de dados Ensembl foi utilizada para busca dos genes próximos e o banco de dados *Cattle QTLdb* v.29 (HU et al., 2013) para examinar a proximidade com QTLs de espécie bovina já catalogados na literatura.

Os resultados dos genes foram revisados utilizando a literatura científica disponível. Os bancos de dados *Mouse Genome Informatics* (EPPIG et al., 2015) e *GeneCards: The Human Gene Database* (STELZER et al., 2016), forneceram informações da participação e envolvimento de genes ortólogos com as características de PN, PD, PS, GPND e GPDS.

Resultados

Os resultados para os GWAS são apresentados em “Manhattan plot” para cada

característica em estudo (Figura 9 a Figura 10).

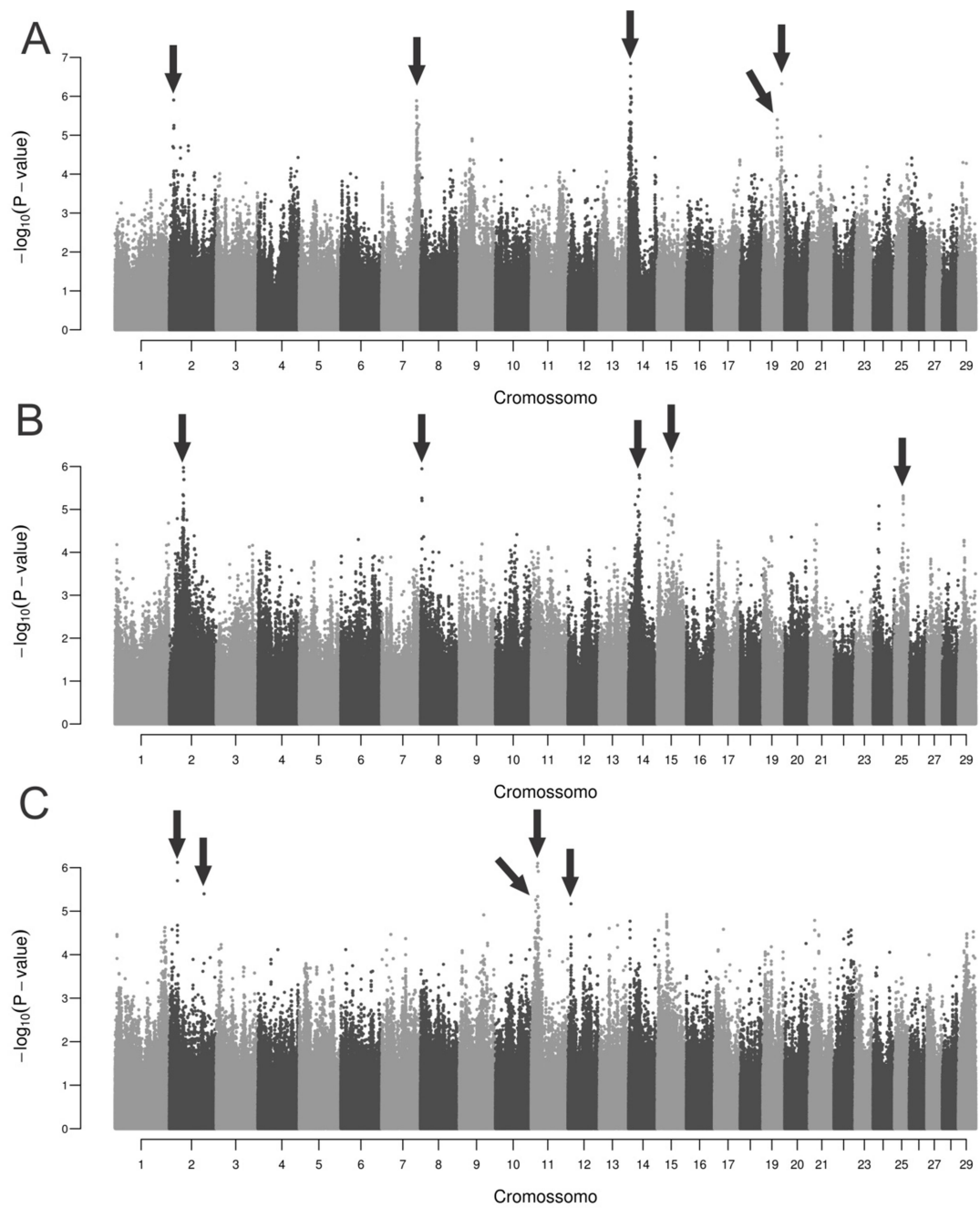


Figura 9. *Manhattan plot* para estudo de associação genômica ampla entre blocos haplotípicos e característica de peso ao nascimento (A), peso à desmama (B) e peso ao sobreano (C).

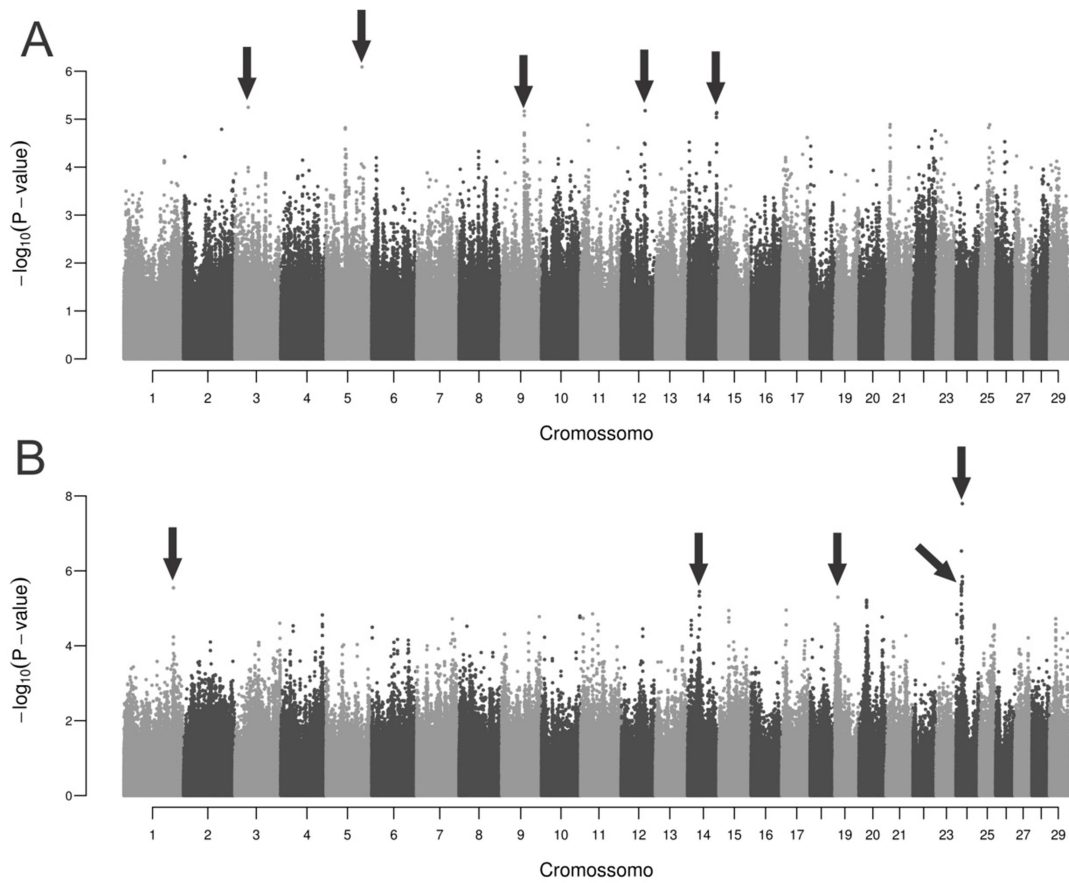


Figura 10. *Manhattan plot* para estudo de associação genômica ampla entre blocos haplotípicos e ganho de peso do nascimento à desmama (A) e ganho de peso da desmama ao sobreano (B).

As regiões selecionadas de acordo com os critérios metodológicos pré-estabelecidos para realizações das buscas de genes e QTLs estão representadas na Tabela 7.

Tabela 7. Dados referentes aos cinco alelos haplotípicos mais significativos para cada característica PN (peso ao nascimento), PD (peso a desmama), (PS) peso ao sobreano, GPND (ganho de peso do nascimento a desmama) e GPDS (ganho de peso da desmama ao sobreano).

Fenótipo	Cromo.	Início	Final	<i>P-value</i>
PN	2	12651384	12670793	1.24 X 10 ⁻⁶
PN	7	101738731	101772665	1.30 X 10 ⁻⁶
PN	14	7297092	7308013	1.43 X 10 ⁻⁷
PN	19	55449455	55466867	4.76 X 10 ⁻⁷
PN	19	42405531	42439413	4.00 X 10 ⁻⁶
PD	2	41352010	41382505	1.06 X 10 ⁻⁶
PD	8	4539462	4560109	1.13 X 10 ⁻⁶
PD	14	32603880	32628063	1.57 X 10 ⁻⁶
PD	15	42700929	42709013	6.24 X 10 ⁻⁷
PD	25	25784635	25811902	4.83 X 10 ⁻⁶
PS	2	23673399	23692687	7.56 X 10 ⁻⁷
PS	2	101284013	101331299	3.99 X 10 ⁻⁶
PS	11	19235848	19244834	7.94 X 10 ⁻⁷
PS	11	13644710	13654583	5.50 X 10 ⁻⁶
PS	12	8917519	8920940	6.76 X 10 ⁻⁶
GPND	1	131772277	131779646	2.83 X 10 ⁻⁶
GPND	14	33345858	33355738	3.56 X 10 ⁻⁶
GPND	19	8102774	8114162	5.04 X 10 ⁻⁶
GPND	24	17362683	17408023	1.60 X 10 ⁻⁸
GPND	24	14947586	14959962	2.95 X 10 ⁻⁷
GPDS	3	35359213	35377284	5.65 X 10 ⁻⁶
GPDS	5	95536136	95566971	8.13 X 10 ⁻⁷
GPDS	9	60199200	60223498	6.79 X 10 ⁻⁶
GPDS	12	64080207	64108644	6.65 X 10 ⁻⁶
GPDS	14	79149390	79169513	7.29 X 10 ⁻⁶

A análise funcional identificou genes que apresentam função biológica associada ao fenótipo de peso, acentuando as evidências de relação entre essas regiões encontradas e as características estudadas (Tabela 8).

Tabela 8. Sumário dos genes presentes em janelas de 1 Mb centralizadas aos haplótipos mais significativos. PN (peso ao nascimento), PD (peso à desmama), (PS) peso ao sobreano, GPND (ganho de peso do nascimento a desmama) e GPDS (ganho de peso da desmama ao sobreano).

Carc.	Gene	Ensembl ID	Posição (BTA:início-final)	Descrição
PN	<i>ST6GALNAC2</i>	ENSBTAG00000015255	19:55118592- 55153750	ST6 (alfa-N-acetilneuramínico-2,3-beta-galactosil)
PN	<i>KRT16</i>	ENSBTAG00000033766	19:42463453- 42465605	queratina 16, tipo I
PN	<i>SEPT9</i>	ENSBTAG00000002633	19:42650885- 42655321	septina 9
PN	<i>FKBP10</i>	ENSBTAG00000011454	19:42650885- 42655321	FK506 proteína de ligação 10
PN	<i>JMJD6</i>	ENSBTAG00000019153	19:55675264- 55681075	domínio contendo jumonji 6
PN	<i>UBE2O</i>	ENSBTAG00000020115	19:55928561- 55977738	enzima de conjugação ubiquitina E2O
PN	<i>CNP</i>	ENSBTAG00000025762	19:42774249- 42780091	2',3'- nucleotídeo cíclico 3' fosfodiesterase
PD	<i>KIAA0556</i>	ENSBTAG00000006129	25:25416302- 25588540	KIAA0556
PD	<i>KCNJ3</i>	ENSBTAG00000006159	2:41361137- 41491255	canal de potássio interno-retificador, subfamília J, membro 3
PD	<i>ATP2A1</i>	ENSBTAG00000006541	25:26188434- 26204655	sarcoplasma/retículo endoplasmático de cálcio ATPase 1
PD	<i>MYBL1</i>	ENSBTAG00000011981	14:32840049- 32862669	myb-relacionado a proteína A
PD	<i>AMPD3</i>	ENSBTAG00000015821	15:42759305- 42796464	adenosina monofosfato desaminase 3
PD	<i>TRIM55</i>	ENSBTAG00000021009	14:32164442- 32210852	tripartido motivo 55
PD	<i>ADM</i>	ENSBTAG00000021048	15:42911044- 42913326	adrenomedulina
PD	<i>CRH</i>	ENSBTAG00000033128	14:32213577- 32215193	hormônio liberador de corticotrofina

Continua...

Tabela 8. Continuação...

PD	<i>SH2B1</i>	ENSBTAG00000019219	25:26208972- 26214950	SH2B proteína adaptadora 1
PS	<i>TGFA</i>	ENSBTAG00000000783	11:13969792- 13984394	factor de crescimento transformante alfa
PS	<i>ATP6V1B1</i>	ENSBTAG00000010620	11:13463414- 13490532	ATPase transportador de H ⁺ V1 subunidade B1
PS	<i>PDK1</i>	ENSBTAG00000016836	2:24067475- 24107467	piruvato cinase desidrogenase, isozima 1
PS	<i>IKZF2</i>	ENSBTAG00000037804	2:101527602- 101706119	IKAROS família de dedo de zinco 2
GPND	<i>CA3</i>	ENSBTAG00000015214	14:79406494- 79416487	anidrase carbônica 3
GPND	<i>CA2</i>	ENSBTAG00000017733	14:79372712- 79388600	anidrase carbônica 2
GPND	<i>GUCY2C</i>	ENSBTAG00000019722	5:95605339- 95670749	guanilato ciclase 2C
GPND	<i>ATP6V0D2</i>	ENSBTAG00000021092	14:78908052- 78964199	ATPase transportador de H ⁺ V0 subunidade D2
GPDS	<i>COPS5</i>	ENSBTAG00000002784	14:33207508- 33226917	COP9 signalosome subunidade 5
GPDS	<i>PIK3CB</i>	ENSBTAG00000006909	1:131429580- 131530578	fosfatidilinositol-4,5-bisph
GPDS	<i>CLDN18</i>	ENSBTAG00000014589	1:132174474- 132191685	claudina 18
GPDS	<i>CSPP1</i>	ENSBTAG00000014689	14:33243588- 33342863	centrossoma associado ao polo do fuso, proteína 1
GPDS	<i>NOG</i>	ENSBTAG00000040282	19:7613376- 7614074	noggin

Para peso ao nascimento, foram identificados os genes: *CNP*, *FKBP10*, *JMJD6*, *KRT16*, *SEPT9*, *ST6GALNAC2*, *UBE2O*, com potencial associação com o fenótipo em estudo. Os genes *SEPT9*, *FKBP10* e *CNP* estão envolvidos com o desenvolvimento de patologias que interferem no tamanho corporal, enquanto o *JMJD6* e *UBE2O*, possuem atividade no tecido adiposo.

Sob o aspecto de peso ao desmame foram identificados os genes *KIAA0556*, *AMPD3*, *MYBL1*, *TRIM55*, *ADM*, *ATP2A1*, *CRH* e *KCNJ3*. Dentre os fenótipos que esses genes estabelecem relação, está a homeostase do tecido adiposo e o

desenvolvimento corporal normal.

Para peso ao sobreano, os genes *ATP6V1B1*, *IKZF2*, *PDK1* e *TGFA* foram descritos como atuantes no crescimento corporal e no tecido adiposo. Para o fenótipo de ganho de peso do nascimento à desmama, os genes *ATP6V0D2*, *CA2*, *CA3* e *GUCY2C*, foram os que forneceram maiores indícios de serem potenciais candidatos, por estarem envolvidos na estatura de ratos e obseidade.

Entre os genes candidatos associados à característica de ganho de peso da desmama ao sobreano, os *CLDN18*, *NOG*, *PIK3CB* e *COPS5*, foram relatados como relacionados com tamanho dos animais estudados. O gene *CSPP1* também apontado para esse fenótipo, está associado à patologias que causam impacto no peso em humanos.

As análises de enriquecimento funcional apontaram 333 QTLs próximos aos haplótipos significativos. Entretanto, 23 loci que apresentam relação com o fenótipo de peso. Dentre eles: 2 para PN, 10 para PD, 3 para PS, 3 para GPND e 5 relativos ao GPDS. Os QTLs presentes nas regiões em estudo relacionados às características Peso corporal ao nascimento, desmama e sobreano e ganho de peso são apresentados demonstrados na Tabela 9.

Tabela 9. QTLs relacionados à características produtivas encontradas em janelas de 1Mb nas proximidades dos haplótipos mais significativos.

Fenótipo	Posição do QTL (BTA:inicio-final)	Descrição	QTL ID
PN	2:12823349-12823389	Ganho de peso corporal	66872
PN	14:7646465-7646505	Peso corporal (nascimento)	31596
PD	2:41488021-41488061	Peso corporal (desmama)	66877
PD	2:41488021-41488061	Ganho de peso corporal	66878
PD	8:4998100-4998140	Ganho de peso corporal	67996
PD	15:42448009-42448049	Ganho de peso corporal	68790
PD	15:42715946-42715986	Peso corporal (nascimento)	23855
PD	15:42784233-42784273	Peso corporal (nascimento)	23856
PD	15:42840851-42840891	Peso corporal (nascimento)	23857
PD	15:42861342-42861382	Ganho de peso corporal	68789
PD	15:42886066-42886106	Ganho de peso corporal	68767
PD	15:43039762-43039802	Ganho de peso corporal	68757
PS	2:101636974-101637014	Peso corporal (desmama)	23647
PS	11:13870220-13870260	Peso corporal (nascimento)	68411
PS	11:13870220-13870260	Peso corporal (sobreano)	68412
GPND	1:131806575-131806615	Peso corporal (nascimento)	66815
GPND	1:132112072-132112112	Peso corporal (sobreano)	66705
GPND	24:15390589-15390629	Peso corporal (nascimento)	23922
GPDS	3:20-55074485	Peso corporal (desmama)	24711
GPDS	9:60465567-60465607	Peso corporal (sobreano)	68062
GPDS	9:60465567-60465607	Ganho de peso corporal	68063
GPDS	12:64002872-64983441	Taxa de crescimento relativo	35247
GPDS	12:63000336-63998631	Taxa de crescimento relativo	35248

Discussão

O melhoramento dos fenótipos estudados é de grande desafio aos produtores, uma vez que é interessante que os animais destinados ao corte atinjam peso satisfatório de terminação com menor idade. No entanto, essa condição possui correlação positiva com a característica de peso ao nascer e, este quando elevado pode acarretar em dificuldades durante o parto. Sendo assim, os esforços têm sido direcionados ao aumento do peso à desmama e sobreano, mantendo peso ao nascimento intermediário

(DECKER et al., 2012).

Quando comparados os conjuntos gênicos do peso ao nascimento e peso a desmama, estes mostraram similaridade nos cromossomos 2 e 14 (Figura 9), no peso a sobreano, apenas o cromossomo 2 permaneceu em evidência. A comparação entre conjuntos associados a fenótipos referentes ao ganho de peso (GPND e GPDS), apresentaram menor similaridade, o que pode ter sido ocasionado pela maior influência do ambiente em decorrência do tempo ao qual o animal foi exposto (PÉGOLO et al., 2009).

De fato, as características estudadas são de herança poligênica, sofrendo a influência de vários genes cada qual exercendo baixo efeito sobre o fenótipo. Apesar desse comportamento, os valores de herdabilidade previamente descritos em Nelore, sugerem que a seleção para esses fenótipos podem apresentar resultados satisfatórios. As estimativas de herdabilidade variam de acordo com diversos fatores, como por exemplo, o número de observações disponíveis, no entanto os valores até então encontrados são em média de 0.25; 0.33; 0.34 para peso ao nascimento, desmama e sobreano, respectivamente e, 0.31 para ganho de peso da desmama a sobreano (BOLIGON et al., 2009; FILHO et al., 2009).

Sob o aspecto produtivo, foram identificados sete genes com função biológica relacionada ao fenótipo de PN. O Keratin 16 (*KRT16*), que pertence à família de genes da queratina, quando mutado em ratos causa a morte logo após o nascer, aqueles que sobrevivem são caracterizados pelo menor tamanho em relação aos demais. Assim como *KRT16*, o *ST6GALNAC2* causa impacto no peso de roedores, sua atividade ainda é pouco explorada, no entanto, quando alterado proporciona pesos significativamente mais baixos (LESSARD et al., 2012; ORR et al., 2012).

Os genes *SEPT9*, *FKBP10* e *CNP* estão associados com patologias que acarretam à baixa massa corporal. O *SEPT9* causa amiotrofia nevrálgica hereditária em humanos, cujos portadores são caracterizados pela perda de massa muscular em membros anteriores (como braços) ou posteriores (pernas). Mutações em *FKBP10* resultam na osteogese imperfeita, que é caracterizada pela fragilidade óssea e morte de crianças ao nascer. As sobreviventes possuem decréscimo da massa óssea e não se desenvolvem normalmente. A doença Niemann-Pick, decorrente de mutações do gene *CNP*, causam o transporte deficiente de colesterol, gerando atraso no desenvolvimento de crianças e perda do tônus muscular. Este gene também participa da absorção intestinal e transporte de eletrólitos, apresenta-se envolvido com o fenótipo produtivo,

por interferir no processo de ganho de peso (ITABA et al., 2004; LESHINSKY-SILVER et al., 2013; LIETMAN et al., 2014).

Por desempenharem papel fundamental na diferenciação de adipócitos, os genes *JMJD6* e *UBE2O* foram descritos como importantes para manutenção da homeostase do tecido adiposo, interferindo no peso pela associação entre o aumento do número de adipócitos e obesidade em mamíferos (MASHTALIR et al., 2014; HU et al., 2015).

Quanto ao fenótipo de peso à desmama, foram identificados os genes *ADM*, *ATP2A1* e *CRH*, envolvidos com tecido adiposo. A influência da expressão do *ADM* na atividade metabólica dos adipócitos ainda não é bem esclarecida, no entanto sabe-se que atua como antagonista do *ET-1*, relatado em quadros de hipertensão e obesidade em humanos. Em relação ao gene *ATP2A1*, o mesmo foi identificado em estudo de GWAS realizado em europeus, buscando associação com adiposidade e circunferência da cintura, reforçando indícios de sua relação com a característica analisada. Mutações nos genes *CRH* e *SH2B1* foram relatados em ratos e humanos, ambos associados à obesidade (KNEER et al., 2005; HESTER et al., 2012; BENTLEY et al., 2013).

Diante da influência do desenvolvimento do músculo esquelético com a característica de peso, os genes *KIAA0556*, *AMPD3*, *MYBL1* e *TRIM55*, quando alterados acarretaram em indivíduos de baixa estatura, distorção nutricional e anormalidades no músculo esquelético (MAHNKE-ZIZELMAN et al., 1997; BOLCUN-FILAS et al., 2011; HE et al., 2015; SANDERS et al., 2015).

O gene *KCNJ3* foi relatado com atraso no desenvolvimento em roedores e desnutrição perinatal, uma vez que quando não expressos, os animais apresentaram baixo peso a desmama mesmo com reposição nutricional (AMARAL et al., 2015).

A análise funcional para peso a sobreano identificou o gene *ATP6V1B1*, cuja ausência de função afeta o crescimento e causam anomalias do esqueleto em mamíferos, ambos tendo como consequência a perda de índice de massa corporal. Outro gene envolvido no tamanho e peso é o *IKZF2*, quando genótipo heterozigoto os animais permanecem menores que os demais da mesma ninhada (CAI et al., 2009; DAVOLI et al., 2006).

O *PDK1* codifica uma proteína envolvida no atraso do crescimento, além de estar envolvido na homeostase da glicose e metabolismo da gordura em ratos e humanos. O gene *TGFA* também possui atuação no tecido adiposo, e desempenha papel no músculo esquelético. Ambos associados a patologias decorrentes da obesidade (BESSESENet al., 1995; BELGARDT et al., 2008).

Para característica de ganho de peso do nascimento à desmama, os genes *ATP6V0D2* e o *CA2* foram associados à defeitos na formação óssea e diminuição do crescimento, acarretando em indivíduos de baixa estatura. O GWAS também indicou a região do gene *CA3*, que em porcos contribui para animais com fenótipos de cortes magros e gordura intramuscular. O *GUCY2C* em camundongos prejudica a percepção de saciedade dos animais, resultando em animais obesos e diversas alterações metabólicas (FINBERG et al., 2005; LEE et al., 2006; ALSHARIDI et al., 2016; VALENTINO et al., 2011).

Dentre os genes candidatos relativos à característica de ganho de peso da desmama ao sobreano, os *CLDN18*, *NOG*, *PIK3CB* e *COPS5*, quando tiveram suas expressões anuladas resultaram em animais saudáveis, mas apresentaram tamanho corporal inferior por possuírem componentes ósseos menores, como o fêmur reduzido. O gene *CSPP1* presente em região identificado por GWAS, possui envolvimento com a síndrome de Joubert, dentre as características de portadores dessa síndrome estão: a diminuição do tônus muscular e desenvolvimento retardado, sugerindo a perda de peso desses indivíduos (TOMODA et al., 2004; TYLZANOWSKI et al., 2006; CIRAULO et al., 2008; KIM et al., 2013).

Quanto à análise de QTLs, ficou evidente nas regiões exploradas para as diferentes características de peso identificadas, a presença de *loci* já descritos anteriormente associados com: peso corporal ao nascimento, corporal ao desmama, corporal ao sobreano e ganho de peso. Reforçando os resultados dos genes encontrados no GWAS para os fenótipos estudados.

Conclusão

Os resultados podem contribuir para elucidar as bases genéticas envolvidas com o peso de bovinos. O estudo permitiu encontrar genes e regiões genômicas associadas com os fenótipos de peso ao nascimento (PN), peso a desmama (PD), peso ao sobreano (PS), ganho de peso do nascimento a desmama (GPND) e ganho de peso da desmama ao sobreano (GPDS). Apesar das características serem de natureza complexa, sua parcela de herança aditiva, pode contribuir para seleção de genótipos superiores, uma vez que possui herdabilidade moderada.

Referências

ACNB. Associação de Criadores de Nelore do Brasil. São Paulo, SP, 2016. Disponível

em: < <http://www.nelore.org.br/raca> >. Acesso em: 10 Mar. 2016.

Alsharidi A, Al-Hamed M, Alsuwaida A: Carbonic anhydrase II deficiency: report of a novel mutation. *CEN Case Reports* 2015:1-5.

Amaral A, Jakovcevski M, McGaughy J, Calderwood S, Mokler D, Rushmore R, Galler J, Akbarian S, Rosene D: Prenatal protein malnutrition decreases KCNJ3 and 2DG activity in rat prefrontal cortex. *Neuroscience* 2015, 286:79-86.

Ardlie KG, Kruglyak L, Seielstad M: Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* 2002, 3(4):299-309.

Aulchenko YS, de Koning DJ, Haley C: Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 2007, 177(1):577-585.

Belgardt BF, Husch A, Rother E, Ernst MB, Wunderlich FT, Hampel B, Klöckener T, Alessi D, Kloppenburg P, Brüning JC: PDK1 deficiency in POMC-expressing cells reveals FOXO1-dependent and-independent pathways in control of energy homeostasis and stress response. *Cell metabolism* 2008, 7(4):291-301.

Bentley L, Esapa CT, Nesbit MA, Head RA, Evans H, Lath D, Scudamore CL, Hough TA, Podrini C, Hannan FM: An N-ethyl-N-nitrosourea induced corticotropin-releasing hormone promoter mutation provides a mouse model for endogenous glucocorticoid excess. *Endocrinology* 2013, 155(3):908-922.

Bessesen DH, Rupp CL, Eckel RH: Trafficking of dietary fat in lean rats. *Obesity research* 1995, 3(2):191-203.

Bolcun-Filas E, Bannister LA, Barash A, Schimenti KJ, Hartford SA, Eppig JJ, Handel MA, Shen L, Schimenti JC: A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. *Development* 2011, 138(15):3319-3330.

Boligon AA, Albuquerque LGd, Mercadante MEZ, Lôbo RB: Herdabilidades e correlações entre pesos do nascimento à idade adulta em rebanhos da raça Nelore. *Revista Brasileira de Zootecnia* 2009:2320-2326.

Buzanskas ME, Grossi DA, Ventura RV, Schenkel FS, Sargolzaei M, Meirelles SL, Mokry FB, Higa RH, Mudadu MA, da Silva MVB: Genome-wide association for growth traits in Canchim beef cattle. *PloS one* 2014, 9(4):e94802.

Cai Q, Dierich A, Oulad-Abdelghani M, Chan S, Kastner P: Helios deficiency has minimal impact on T cell development and function. *The Journal of Immunology* 2009, 183(4):2303-2311.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015, 4(1):1.

Ciraolo E, Iezzi M, Marone R, Marengo S, Curcio C, Costa C, Azzolino O, Gonella C, Rubinetto C, Wu H: Phosphoinositide 3-kinase p110 β activity: key role in metabolism and mammary gland cancer but not development. *Science signaling* 2008, 1(36):ra3.

Davoli R, Fontanesi L, Braglia S, Nisi I, Scotti E, Buttazzoni L, Russo V: Investigation of SNPs in the ATP1A2, CA3 and DECR1 genes mapped to porcine chromosome 4: analysis in groups of pigs divergent for meat production and quality traits. *Italian Journal of Animal Science* 2006, 5(3):249-263.

Decker JE, Vasco DA, McKay SD, McClure MC, Rolf MM, Kim J, Northcutt SL, Bauck S, Woodward BW, Schnabel RD: A novel analytical method, Birth Date Selection Mapping, detects response of the Angus (*Bos taurus*) genome to selection on complex traits. *BMC genomics* 2012, 13(1):1.

Delaneau O, Marchini J, Zagury J-F: A linear complexity phasing method for thousands of genomes. *Nature methods* 2012, 9(2):179-181.

Eppig JT, Blake JA, Bult CJ et al: The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic acids research* 2015, 43(D1):D726-D736.

FAO. Food and Agriculture Organization of the United Nations. How to Feed the World in 2050. 2009. Disponível em: < <http://www.fao.org/wsfs/forum2050/wsfs-forum/en/>>. Acesso em: 23 Mar. 2016.

Finberg KE, Wagner CA, Bailey MA, Păunescu TG, Breton S, Brown D, Giebisch G, Geibel JP, Lifton RP: The B1-subunit of the H⁺ ATPase is required for maximal urinary acidification. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102(38):13616-13621.

Gunawan A, Jakaria: Genetic and Non-Genetics effect on birth, weaning, and yearling weight of Bali cattle. *Media Peternakan-Journal of Animal Science and Technology* 2011, 34(2):93-98.

He J, Quintana MT, Sullivan J, Parry T, Grevengeod T, Schisler JC, Hill JA, Yates CC, Mapanga RF, Essop MF: MuRF2 regulates PPAR γ 1 activity to protect against diabetic cardiomyopathy and enhance weight gain induced by a high fat diet. *Cardiovascular diabetology* 2015, 14(1):1.

Hester JM, Wing M, Li J, Palmer N, Xu J, Hicks P, Roh B, Norris J, Wagenknecht L, Langefeld C: Implication of European-derived adiposity loci in African Americans. *International journal of obesity* 2012, 36(3):465-473.

Hu Y-J, Belaghal H, Hsiao W-Y, Qi J, Bradner JE, Guertin DA, Sif S, Imbalzano AN: Transcriptional and post-transcriptional control of adipocyte differentiation by Jumonji domain-containing protein 6. *Nucleic acids research* 2015, 43(16):7790-7804.

Hu ZL, Park CA, Wu XL, Reecy JM: Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic acids research* 2013, 41:871-879.

Itaba S, Chijiwa Y, Matsuzaka H, Motomura Y, Nawata H: Presence of C-type natriuretic peptide (CNP) in guinea pig caecum: role and mechanisms of CNP in circular smooth muscle relaxation. *Neurogastroenterology & Motility* 2004, 16(3):375-382.

Kim H-Y, Alarcon C, Pourteymour S, Wergedal JE, Mohan S: Disruption of claudin-18 diminishes ovariectomy-induced bone loss in mice. *American Journal of Physiology-Endocrinology and Metabolism* 2013, 304(5):E531-E537.

Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A et al: Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database : the journal of biological databases and curation* 2011, 2011:30.

Knerr I, Schirl C, Horbach T, Stuppy A, Carbon R, Rascher W, Dötsch J: Maturation of the expression of adrenomedullin, endothelin-1 and nitric oxide synthases in adipose tissues from childhood to adulthood. *International journal of obesity* 2005, 29(3):275-280.

Koury Filho W, Albuquerque LGd, Alencar MMd, Forni S, Silva JAI, Lôbo RB: Estimativas de herdabilidade e correlações para escores visuais, peso e altura ao sobreano em rebanhos da raça Nelore. *Revista Brasileira de Zootecnia* 2009,

38(12):2362-2367.

Lee S-H, Rho J, Jeong D, Sul J-Y, Kim T, Kim N, Kang J-S, Miyamoto T, Suda T, Lee S-K: v-ATPase V0 subunit d2-deficient mice exhibit impaired osteoclast fusion and increased bone formation. *Nature medicine* 2006, 12(12):1403-1409.

Leshinsky-Silver E, Ginzberg M, Dabby R, Sadeh M, Lev D, Lerman-Sagie T: Neonatal vocal cord paralysis-an early presentation of hereditary neuralgic amyotrophy due to a mutation in the SEPT9 gene. *European Journal of Paediatric Neurology* 2013, 17(1):64-67.

Lessard JC, Coulombe PA: Keratin 16-null mice develop palmoplantar keratoderma, a hallmark feature of pachyonychia congenita and related disorders. *Journal of Investigative Dermatology* 2012, 132(5):1384-1391.

Lietman CD, Rajagopal A, Homan EP, Munivez E, Jiang M-M, Bertin TK, Chen Y, Hicks J, Weis M, Eyre D: Connective tissue alterations in Fkbp10^{-/-} mice. *Human molecular genetics* 2014, 23(18):4822-4831.

Lôbo RNB, Filho RM: Estimates of genetic parameters for weights and maturing traits derived by growth functions in Nelore cattle. *Ciência Animal* 2003, 13(1):7-15.

Mahnke-Zizelman DK, D'Cunha J, Wojnar JM, Brogley MA, Sabina RL: Regulation of rat AMP deaminase 3 (isoform C) by development and skeletal muscle fibre type. *Biochemical Journal* 1997, 326(2):521-529.

Mashtalir N, Daou S, Barbour H, Sen NN, Gagnon J, Hammond-Martel I, Dar HH, Therrien M, Affar EB: Autodeubiquitination protects the tumor suppressor BAP1 from cytoplasmic sequestration mediated by the atypical ubiquitin ligase UBE2O. *Molecular cell* 2014, 54(3):392-406.

Orr N, Bekker V, Chanock S: Genetic association studies: marking them well. *Journal of Infectious Diseases* 2006, 194(11):1475-1477.

Pégolo NT, Oliveira HN, Albuquerque LG, Bezerra LAF, Lôbo RB: Genotype by environment interaction for 450-day weight of Nelore cattle analyzed by reaction norm models. *Genetics and molecular biology* 2009, 32(2):281-287.

Purcell S, Neale B, Todd-Brown K et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007, 81(3):559-575.

Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26(6):841-842.

R Core Team: R: A language and environment for statistical computing. In: R Foundation for Statistical Computing. Vienna, Austria. ; 2013.

Sanders AA, de Vrieze E, Alazami AM, Alzahrani F, Malarkey EB, Soroush N, Tebbe L, Kuhns S, van Dam TJ, Alhashem A: KIAA0556 is a novel ciliary basal body component mutated in Joubert syndrome. *Genome biology* 2015, 16(1):1.

Santana MH, Utsunomiya YT, Neves HH, Gomes RC, Garcia JF, Fukumasu H, Silva SL, Junior GAO, Alexandre PA, Leme PR: Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. *BMC genetics* 2014, 15(1):1.

Sargolzaei M, Chesnais JP, Schenkel FS: A new approach for efficient genotype imputation using information from relatives. *BMC genomics* 2014, 15(1):478.

Stelzer G, Rosen N, Plaschkes I, et al: The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics* 2016:1.30. 31-31.30. 33.

Tomoda K, Yoneda-Kato N, Fukumoto A, Yamanaka S, Kato J-y: Multiple functions of Jab1 are required for early embryonic development and growth potential in mice. *Journal of Biological Chemistry* 2004, 279(41):43013-43018.

Tylzanowski P, Mebis L, Luyten FP: The Noggin null mouse phenotype is strain dependent and haploinsufficiency leads to skeletal defects. *Developmental dynamics* 2006, 235(6):1599-1607.

Utsunomiya YT, Milanese M, Utsunomiya AT, Ajmone-Marsan P, Garcia JF: GHap: An R package for Genome-wide Haplotyping. *Bioinformatics* 2016:btw356.

Valentino MA, Lin JE, Snook AE et al: A uroguanylin-GUCY2C endocrine axis regulates feeding in mice. *The Journal of clinical investigation* 2011, 121(9):3578-3588.

Yang J, Lee SH, Goddard ME, Visscher PM: GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* 2011, 88(1):76-82.

Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL: Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics* 2014, 46(2):100-106.

CONCLUSÃO GERAL

A partir da distribuição genotípica entre populações de referência e imputação, pode-se esperar que, as informações genéticas preditas por metodologia de imputação para painel de alta densidade foram consistentes. O uso de alelos haplotípicos foi eficiente para detecção de regiões genômicas associadas com características de idade ao primeiro parto, peso ao nascimento, peso ao desmame, peso ao sobreano, ganho de peso do nascimento a desmama e ganho de peso da desmama ao sobreano, proporcionando, a identificação de QTLs e genes que são candidatos à influência na manifestação dos fenótipos estudados. Os resultados oferecem subsídios para pesquisas futuras de validação de genes, viabilizando assim, o uso de novas estratégias em programas de melhoramento genético para precocidade sexual e maior produtividade.

ANEXOS

1. Plos One – Guidelines

Submission Instructions

These are step-by-step instructions for submitting a manuscript to our online submission system.

Do not resubmit your manuscript if you have otherwise already submitted it to *PLOS ONE*. If you are submitting a revised manuscript, read the revision guidelines. If you are responding to a technical check inquiry, follow the instructions you receive by email from PLOS staff.

Submit new manuscript

Click on the “Submit New Manuscript” link under the “New Submissions” heading. This will take you to a new screen.

Article type

Next, select your article type. Most submissions will be either a Research Article or a Clinical Trial.

Title

Enter the title and short title of your manuscript. Please enter your titles in sentence case (capitalize only the first word of the title, the first word of the subtitle, and any proper nouns and genus names). Do not enter titles in all capital letters.

Contributors

Next you will be asked to enter the names, email addresses, and institutional affiliations for each author. Each author should have made real and concrete contributions to the manuscript; you will have the opportunity to describe each author’s contribution in a further step. Each author should also be aware of the submission and approve the manuscript submitted. PLOS does not condone or accept guest authorship.

Read our authorship policy.

Corresponding author

Additionally, you will be asked to designate one author to act as corresponding author. Note that this is the person who will receive correspondence from the *PLOS ONE* editorial office, but does not need to be the person who will appear as corresponding author on the paper if accepted.

The corresponding author is responsible for ensuring that the author list and author contributions are accurate and complete. The person noted as the corresponding author should be the person who is actually responsible for ensuring that all the proper forms are submitted and accurate, and should be readily available for correspondence if *PLOS ONE* editors have questions or concerns.

The corresponding author must provide an ORCID iD at the time of submission by entering it in the user profile in the submission system. For more information about how to register for an ORCID iD and link it to Editorial Manager, watch this short video.

If a corresponding author will not be available for an extended period of time, please inform *PLOS ONE* at plosone@plos.org.

Read more information about designating corresponding authors and about corresponding author responsibilities.

Category

Next, you’ll receive a prompt about your paper’s Section/Category. If your paper describes primary research on human subjects, indicate that here by selecting the category Clinical; if not, select Other.

Abstract

This should exactly match the text of the Abstract in your manuscript file.

Keywords

These will help expedite the internal processing of your manuscript. You will not have another opportunity to edit these, so please make sure to add concise, accurate keywords at this point.

Competing interests

Please declare any affiliations or relationships that could be viewed as potentially competing interests. This information will be published with the final manuscript, if accepted, so please make sure that this is accurate and as detailed as possible.

Read our policy on declaring competing interests.

Financial disclosure

Please disclose your funders and the role they played in your manuscript. This information will be published with the final manuscript, if accepted, so please make sure that this is accurate and as detailed as possible.

Please include relevant grant numbers and the URL of any funder's website. Please also state whether any individuals employed or contracted by the funders (other than the named authors) played any role in: study design, data collection and analysis, decision to publish, or preparation of the manuscript. If so, please name the individual and describe their role. If no individuals employed or contracted by the funder(s), other than the authors, were involved in these tasks and the authors had unrestricted access to the data, please state, "The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."

Ethics statement

You must provide an ethics statement if your study made use of human or vertebrate animal subjects and/or tissue. The text in this field should exactly match the text of the ethics statement in the body of your paper; we recommend that you cut and paste your paper's ethics statement into this field. The information submitted here will be evaluated by the *PLOS ONE* editors to ensure that it meets our standards, so please include as much information as possible. See the *PLOS ONE* Criteria for Publication for more information about ethics requirements.

Previous interactions

If you are submitting your manuscript following a recommendation from another PLOS journal, or if you have had a previous interaction with *PLOS ONE* about the manuscript or closely related manuscripts, please indicate the extent of your previous interaction here.

Suggest Academic Editors

Please recommend 2-5 Academic Editors from our board who you feel are qualified to handle your submission based on their areas of expertise.

Collections submission

Please indicate if you are submitting your paper as a part of a collection. Read more information about PLOS Collections.

New taxon

If your paper describes a new taxon, please give its name. Read the guidelines for describing a new taxon.

Dual publications

Please explain whether any of the elements of your paper have ever been published anywhere before. If so, we may have to consider copyright issues, so it is very important that you provide this information in full.

Author contributions

This is a series of questions that will enable you to state the contributions of each

author. Each author listed on the manuscript should have made a real and concrete contribution to the submission, and each person who contributed to the manuscript should be listed. More information about authorship can be found in the Editorial Policies.

References

References should be listed after the main text, before the supporting information; References with more than six authors should list the first six author names, followed by “et al.”;

Please see the PLOS ONE guide lines for References here: <http://journals.plos.org/plosone/s/submission-guidelines#loc-references>.

2. BMC Genomics – Guidelines

Preparing main manuscript text

Quick points:

Use double line spacing;

Include line and page numbering;

Use SI units: Please ensure that all special characters used are embedded in the text, otherwise they will be lost during conversion to PDF;

Do not use page breaks in your manuscript.

File formats

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

Microsoft word (DOC, DOCX);

Rich text format (RTF);

TeX/LaTeX (use BioMed Central's TeX template).

Please note: editable files are required for processing in production. If your manuscript contains any non-editable files (such as PDFs) you will be required to re-submit an editable file if your manuscript is accepted.

Note that figures must be submitted as separate image files, not as part of the submitted manuscript file. For more information, see Preparing figures below.

Style and language

For editors and reviewers to accurately assess the work presented in your manuscript you need to ensure the English language is of sufficient quality to be understood. If you need help with writing in English you should consider:

Visiting the English language tutorial which covers the common mistakes when writing; in English;

Asking a colleague who is a native English speaker to review your manuscript for clarity;

Using a professional language editing service where editors will improve the English to ensure that your meaning is clear and identify problems that require your review. Two such services are provided by our affiliates Nature Research Editing Service and American Journal Experts.

Please note that the use of a language editing service is not a requirement for publication in the journal and does not imply or guarantee that the article will be selected for peer review or accepted.

Data and materials

For all journals, BioMed Central strongly encourages all datasets on which the conclusions of the manuscript rely to be either deposited in publicly available repositories (where available and appropriate) or presented in the main paper or additional supporting files, in machine-readable format (such as spread sheets rather

than PDFs) whenever possible. Please see the list of recommended repositories in our editorial policies.

For some journals, deposition of the data on which the conclusions of the manuscript rely is an absolute requirement. Please check the Instructions for Authors for the relevant journal and article type for journal specific policies.

For all manuscripts, information about data availability should be detailed in an 'Availability of data and materials' section. For more information on the content of this section, please see the Declarations section of the relevant journal's Instruction for Authors. For more information on BioMed Central's policies on data availability, please see our [editorial policies].

Formatting the 'Availability of data and materials' section of your manuscript

The following format for the 'Availability of data and materials' section of your manuscript should be used:

"The dataset(s) supporting the conclusions of this article is(are) available in the [repository name] repository, [unique persistent identifier and hyperlink to dataset(s) in http:// format]."

The following format is required when data are included as additional files:

"The dataset(s) supporting the conclusions of this article is(are) included within the article (and its additional file(s))."

BioMed Central endorses the Force 11 Data Citation Principles and requires that all publicly available datasets be fully referenced in the reference list with an accession number or unique identifier such as a DOI.

For databases, this section should state the web/ftp address at which the database is available and any restrictions to its use by non-academics.

For software, this section should include:

Project name: e.g. My bioinformatics project;

Project home page: e.g. <http://sourceforge.net/projects/mged>;

Archived version: DOI or unique identifier of archived software or code in repository (e.g. enodo);

Operating system(s): e.g. Platform independent;

Programming language: e.g. Java;

Other requirements: e.g. Java 1.3.1 or higher, Tomcat 4.0 or higher;

License: e.g. GNU GPL, FreeBSD etc;

Any restrictions to use by non-academics: e.g. licence needed.

Information on available repositories for other types of scientific data, including clinical data, can be found in our editorial policies.

References

See our editorial policies for author guidance on good citation practice.

All references, including URLs, must be numbered consecutively, in square brackets, in the order in which they are cited in the text, followed by any in tables or legends. The reference numbers must be finalized and the reference list fully formatted before submission. For further information including example references please read our reference preparation guidelines.

How to format your references

Examples of the BioMed Central reference style are shown below. Please ensure that the reference style is followed precisely; if the references are not in the correct style, they may need to be retyped and carefully proofread.

Web links and URLs: All web links and URLs, including links to the authors' own websites, should be given a reference number and included in the reference list rather than within the text of the manuscript. They should be provided in full, including both

the title of the site and the URL, as well as the date the site was accessed, in the following format: The Mouse Tumor Biology Database. <http://tumor.informatics.jax.org/mtbwi/index.do>. Accessed 20 May 2013. If an author or group of authors can clearly be associated with a web link, such as for weblogs, then they should be included in the reference.

Authors may wish to make use of reference management software to ensure that reference lists are correctly formatted.

Example reference style:

Article within a journal

Smith JJ. The world of science. *Am J Sci.* 1999;36:234-5.

Article within a journal (no page numbers)

Rohrmann S, Overvad K, Bueno-de-Mesquita HB, Jakobsen MU, Egeberg R, Tjønneland A, et al. Meat consumption and mortality - results from the European Prospective Investigation into Cancer and Nutrition. *BMC Med.* 2013;11:63.

Article within a journal by DOI

Slifka MK, Whitton JL. Clinical implications of dysregulated cytokine production. *Dig J Mol Med.* 2000; doi:10.1007/s801090000086.

Article within a journal supplement

Frumin AM, Nussbaum J, Esposito M. Functional asplenia: demonstration of splenic activity by bone marrow scan. *Blood* 1979;59 Suppl 1:26-32.

Book chapter, or an article within a book

Wyllie AH, Kerr JFR, Currie AR. Cell death: the significance of apoptosis. In: Bourne GH, Danielli JF, Jeon KW, editors. *International review of cytology*. London: Academic; 1980. p. 251-306.

OnlineFirst chapter in a series (without a volume designation but with a DOI)

Saito Y, Hyuga H. Rate equation approaches to amplification of enantiomeric excess and chiral symmetry breaking. *Top Curr Chem.* 2007. doi:10.1007/128_2006_108.

Complete book, authored

Blenkinsopp A, Paxton P. *Symptoms in the pharmacy: a guide to the management of common illness*. 3rd ed. Oxford: Blackwell Science; 1998.

Online document

Doe J. Title of subordinate document. In: *The dictionary of substances and their effects*. Royal Society of Chemistry. 1999. http://www.rsc.org/dose/title_of_subordinate_document. Accessed 15 Jan 1999.

Online database

Healthwise Knowledgebase. *US Pharmacopeia*, Rockville. 1998. <http://www.healthwise.org>. Accessed 21 Sept 1998.

Supplementary material/private homepage

Doe J. Title of supplementary material. 2000. <http://www.privatehomepage.com>. Accessed 22 Feb 2000.

University site

Doe, J: Title of preprint. <http://www.uni-heidelberg.de/mydata.html> (1999). Accessed 25 Dec 1999.

FTP site

Doe, J: Trivial HTTP, RFC2169. <ftp://ftp.isi.edu/in-notes/rfc2169.txt> (1999). Accessed 12 Nov 1999.

Organization site

ISSN International Centre: The ISSN register. <http://www.issn.org> (2006). Accessed 20 Feb 2007.

Dataset with persistent identifier

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, et al. Genome data from sweet and grain sorghum (*Sorghum bicolor*). GigaScience Database. 2011. <http://dx.doi.org/10.5524/100012>.

Preparing figures

When preparing figures, please follow the formatting instructions below.

Figures should be provided as separate files, not embedded in the main manuscript file; Each figure of a manuscript should be submitted as a single file that fits on a single page in portrait format;

Tables should NOT be submitted as figures but should be included in the main manuscript file;

Multi-panel figures (those with parts a, b, c, d etc.) should be submitted as a single composite file that contains all parts of the figure;

Figures should be numbered in the order they are first mentioned in the text, and uploaded in this order;

Figures should be uploaded in the correct orientation;

Figure titles (max 15 words) and legends (max 300 words) should be provided in the main manuscript, not in the graphic file;

Figure keys should be incorporated into the graphic, not into the legend of the figure;

Each figure should be closely cropped to minimize the amount of white space surrounding the illustration. Cropping figures improves accuracy when placing the figure in combination with other elements when the accepted manuscript is prepared for publication on our site. For more information on individual figure file formats, see our detailed instructions;

Individual figure files should not exceed 10 MB. If a suitable format is chosen, this file size is adequate for extremely high quality figures;

Please note that it is the responsibility of the author(s) to obtain permission from the copyright holder to reproduce figures (or tables) that have previously been published elsewhere. In order for all figures to be open access, authors must have permission from the rights holder if they wish to include images that have been published elsewhere in non open access journals. Permission should be indicated in the figure legend, and the original source included in the reference list.

Figure file types

We accept the following file formats for figures:

EPS (suitable for diagrams and/or images);

PDF (suitable for diagrams and/or images);

Microsoft Word (suitable for diagrams and/or images, figures must be a single page);

PowerPoint (suitable for diagrams and/or images, figures must be a single page);

TIFF (suitable for images);

JPEG (suitable for photographic images, less suitable for graphical images);

PNG (suitable for images);

BMP (suitable for images);

CDX (ChemDraw - suitable for molecular structures).

For information and suggestions of suitable file formats for specific figure types, please see our author academy.

Figure size and resolution

Figures are resized during publication of the final full text and PDF versions to conform to the BioMed Central standard dimensions, which are detailed below.

Figures on the web:

width of 600 pixels (standard), 1200 pixels (high resolution).

Figures in the final PDF version:

width of 85 mm for half page width figure;
width of 170 mm for full page width figure;
maximum height of 225 mm for figure and legend;
image resolution of approximately 300 dpi (dots per inch) at the final size.
Figures should be designed such that all information, including text, is legible at these dimensions. All lines should be wider than 0.25 pt when constrained to standard figure widths. All fonts must be embedded.

Figure file compression

Vector figures should if possible be submitted as PDF files, which are usually more compact than EPS files;

TIFF files should be saved with LZW compression, which is lossless (decreases file size without decreasing quality) in order to minimize upload time;

JPEG files should be saved at maximum quality;

Conversion of images between file types (especially lossy formats such as JPEG) should be kept to a minimum to avoid degradation of quality.

If you have any questions or are experiencing a problem with figures, please contact the customer service team at info@biomedcentral.com.

Preparing tables

When preparing tables, please follow the formatting instructions below.

Tables should be numbered and cited in the text in sequence using Arabic numerals (i.e. Table 1, Table 2 etc.);

Tables less than one A4 or Letter page in length can be placed in the appropriate location within the manuscript;

Tables larger than one A4 or Letter page in length can be placed at the end of the document text file. Please cite and indicate where the table should appear at the relevant location in the text file so that the table can be added in the correct place during production;

Larger datasets, or tables too wide for A4 or Letter landscape page can be uploaded as additional files. Please see [below] for more information;

Tabular data provided as additional files can be uploaded as an Excel spreadsheet (.xls) or comma separated values (.csv). Please use the standard file extensions;

Table titles (max 15 words) should be included above the table, and legends (max 300 words) should be included underneath the table;

Tables should not be embedded as figures or spreadsheet files, but should be formatted using 'Table object' function in your word processing program;

Color and shading may not be used. Parts of the table can be highlighted using superscript, numbering, lettering, symbols or bold text, the meaning of which should be explained in a table legend;

Commas should not be used to indicate numerical values.