

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**ASSOCIAÇÃO GENÔMICA AMPLA PARA
CARACTERÍSTICAS REPRODUTIVAS EM BOVINOS DA
RAÇA NELORE**

**Márcia Cristina Matos
Médica Veterinária**

2013

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**ASSOCIAÇÃO GENÔMICA AMPLA PARA
CARACTERÍSTICAS REPRODUTIVAS EM BOVINOS DA
RAÇA NELORE**

Márcia Cristina Matos

Orientador: Prof. Dr. Wilter Ricardo Russiano Vicente

Co-orientadores: Prof. Dr. José Fernando Garcia

Profa. Dra. Alexeia Barufatti Grisolia

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para obtenção do título de Doutor em Medicina Veterinária (Reprodução Animal).

2013

Matos, Márcia Cristina
M433a Associação genômica ampla para características reprodutivas em
bovinos da raça Nelore / Márcia Cristina Matos. – – Jaboticabal, 2013
xii, 114 p. : il. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de
Ciências Agrárias e Veterinárias, 2013

Orientador: Wilter Ricardo Russiano Vicente

Banca examinadora: Roberto Carvalheiro, Joaquim Mansano
Garcia, Fernando Sebastián Baldi Rey, Lenira El Faro
Bibliografia

1. *Bos taurus indicus*. 2. Idade ao Primeiro Parto. 3. Duração do
Período Gestacional. 4. Genômica Funcional. 5. Gene Candidato. 6.
SNP. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e
Veterinárias.

CDU 619:612.6:636.2

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

CERTIFICADO DE APROVAÇÃO

TÍTULO: ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERÍSTICAS REPRODUTIVAS EM BOVINOS DA RAÇA NELORE

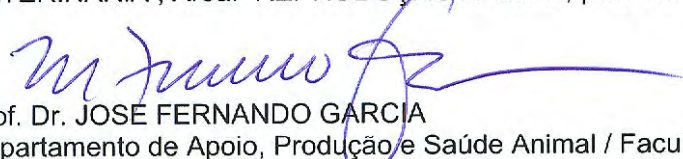
AUTORA: MÁRCIA CRISTINA MATOS

ORIENTADOR: Prof. Dr. WILTER RICARDO RUSSIANO VICENTE

CO-ORIENTADORA: Profa. Dra. ALEXEIA BARUFATTI GRISOLIA

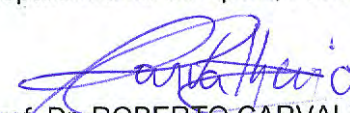
CO-ORIENTADOR: Prof. Dr. JOSE FERNANDO GARCIA

Aprovada como parte das exigências para obtenção do Título de DOUTOR EM MEDICINA VETERINÁRIA, Área: REPRODUÇÃO ANIMAL, pela Comissão Examinadora:



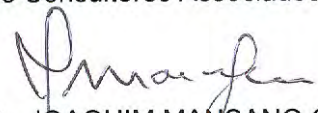
Prof. Dr. JOSE FERNANDO GARCIA

Departamento de Apoio, Produção e Saúde Animal / Faculdade de Odontologia de Araçatuba



Prof. Dr. ROBERTO CARVALHEIRO

GenSys Consultores Associados S/C Ltda / Porto Alegre/RS



Prof. Dr. JOAQUIM MANSANO GARCIA

Departamento de Medicina Veterinária Preventiva e Reprodução Animal / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal



Prof. Dr. FERNANDO SEBASTIÁN BALDI REY

Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal



Profa. Dra. LENIRA EL FARO

Agência Paulista de Tecnologia dos Agronegócios / Ribeirão Preto/SP

Data da realização: 01 de fevereiro de 2013.

DADOS CURRICULARES DA AUTORA

MÁRCIA CRISTINA MATOS – filha de Maria Regina Mobrizi Matos e Armando Fernando Matos, nasceu em 19 de janeiro de 1979 em Ourinhos, SP. Iniciou o curso de Medicina Veterinária em fevereiro de 1997 na Universidade Estadual de Londrina (UEL), onde se graduou em dezembro de 2001. Ingressou no curso de mestrado em março de 2002, junto ao Programa de Pós-graduação em Medicina Veterinária área de concentração Reprodução Animal da Faculdade de Ciências Agrárias e Veterinárias (FCAV), UNESP de Jaboticabal, onde foi bolsista FAPESP no período de agosto de 2002 a julho de 2004. Desenvolveu seu trabalho de mestrado sob orientação do Prof. Dr. José Luiz Moraes Vasconcelos e, em julho de 2004, obteve o título de Mestre. Durante o período de setembro de 2004 a novembro de 2008 trabalhou em laboratório comercial de produção *in vitro* de embriões bovinos. Em março de 2009 iniciou o curso de doutorado oferecido pelo Programa de Pós-graduação em Medicina Veterinária área de concentração Reprodução Animal da FCAV, UNESP de Jaboticabal, sendo bolsista CAPES no período de agosto de 2010 a fevereiro de 2013. Durante o período de setembro de 2009 a julho de 2010 foi docente responsável pelas disciplinas de Reprodução Animal do curso de Medicina Veterinária da Faculdade Anhanguera de Dourados, MS. Desenvolveu sua tese de doutorado sob orientação do Prof. Dr. Wilter Ricardo Russiano Vicente (FCAV, UNESP) e co-orientação do Prof. Dr. José Fernando Garcia (Faculdade de Medicina Veterinária de Araçatuba (FMVA), UNESP) e da Profa. Dra. Alexeia Barufatti Grisolia (Universidade Federal da Grande Dourados (UFGD)).

Sê

Se não puderes ser um pinheiro, no topo de uma colina,

Sê um arbusto no vale mas sê

O melhor arbusto à margem do regato.

Sê um ramo, se não puderes ser uma árvore.

Se não puderes ser um ramo, sê um pouco de relva

E dá alegria a algum caminho.

Se não puderes ser uma estrada,

Sê apenas uma senda,

Se não puderes ser o Sol, sê uma estrela.

Não é pelo tamanho que terás êxito ou fracasso...

Mas sê o melhor no que quer que sejas.

Pablo Neruda

AGRADECIMENTOS

Ao Prof. Wilter agradeço pela orientação, apoio e principalmente pela confiança depositada.

Aos meus estimados co-orientadores, Prof. Fernando e Profa. Alexeia, pela oportunidade, incentivo, atenção e por viabilizarem a execução deste trabalho.

À Fazenda São Jorge do Maracay agradeço por disponibilizarem animais e dados fenotípicos, além do suporte oferecido na execução deste trabalho.

Aos colaboradores Conexão Delta G e Zebu Genome Consortium pelos dados fenotípicos e genotípicos disponibilizados.

Ao Dr. Roberto Carneiro pela colaboração e atenção.

À CAPES pela concessão de bolsa de estudo.

A minha equipe de trabalho, essencialmente formada pelos meus queridos amigos Yuri e Adriana, que me ofereceu suporte e incentivo para iniciar meu aprendizado em genômica funcional e bioinformática.

Ao meu estimado amigo Marcelo Barbosa Bezerra pelo incentivo a pós-graduação.

A todos os amigos conquistados no Departamento de Reprodução Animal desde meu estágio curricular em 2001 até a conclusão deste trabalho. Aprendi algo com cada um de vocês.

À Profa. Cárís Maroni Nunes, aos colegas pós-graduandos do Laboratório de Bioquímica e Biologia Molecular (LBBMA) da FMVA - UNESP de Araçatuba e à equipe Deoxi Biotecnologia pela atenção e amizade.

Aos estimados amigos do Laboratório Aplicado à Produção Animal da FCA da UFGD: Bruno, Dani, Lara, Jussara, Joice e Jackeline pelo apoio e amizade.

Às minhas queridas amigas Flávia e Karen agradeço pelo apoio e carinho incondicionais.

A você Leonardo, meu maior incentivador, agradeço o amor, a paciência e o companheirismo.

À minha família agradeço e também dedico este trabalho. Obrigada pelo carinho, paciência e atenção ao longo desse período. Vocês são grandes exemplos de determinação.

Márcia Cristina Matos

Jaboticabal, fevereiro de 2013.

SUMÁRIO

	Página
RESUMO.....	iii
ABSTRACT.....	iv
LISTA DE ABREVIATURAS E SÍMBOLOS.....	v
LISTA DE TABELAS.....	viii
LISTA DE FIGURAS.....	ix
1. Introdução.....	1
2. Revisão de Literatura.....	3
2.1. Características Reprodutivas de Fêmeas.....	3
2.2. Marcadores SNP e Análise de SNP em Larga Escala.....	8
2.2.1. Aspectos Inerentes a Análise e Controle de Qualidade dos Dados de SNP.....	14
2.3. Análises do Tipo GWAS para Características Reprodutivas.....	16
3. Hipótese.....	18
4. Objetivos.....	18
4.1. Objetivo Geral.....	18
4.1.1. Objetivos Específicos.....	18
5. Material e Métodos.....	19
5.1. Animais.....	19
5.2. Extração de DNA e Análise de SNP.....	20
5.3. Dados Fenotípicos.....	21
5.3.1. Experimento 1.....	21
5.3.2. Experimento 2.....	22
5.3.3. Experimento 3.....	24
5.4. Controle de Qualidade dos Genótipos e Estratificação Populacional.....	25
5.5. Análise de Associação Genômica Ampla.....	26
5.6. Estratégias para a Identificação de Genes Candidatos.....	29

6. Resultados.....	30
6.1. Experimento 1.....	30
6.1.1. Controle de Qualidade e Estratificação Populacional.....	30
6.1.2. Análise de Associação Genômica Ampla.....	31
6.1.3. Identificação de Genes Candidatos.....	37
6.2. Experimento 2.....	41
6.2.1. Controle de Qualidade e Estratificação Populacional.....	41
6.2.2. Análise de Associação Genômica Ampla.....	44
6.2.3. Identificação de Genes Candidatos.....	50
6.3. Experimento 3.....	54
6.3.1. Controle de Qualidade e Estratificação Populacional.....	54
6.3.2. Análise de Associação Genômica Ampla.....	57
6.3.3. Identificação de Genes Candidatos.....	62
7. Discussão.....	67
7.1. Análises de Associação Genômica Ampla.....	67
7.2. Identificação de Genes Candidatos.....	70
8. Conclusões.....	78
9. Referências Bibliográficas.....	79
10. Apêndice.....	97
10.1. Tabelas Suplementares.....	97
10.2. Figuras Suplementares.....	105

ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERÍSTICAS REPRODUTIVAS EM BOVINOS DA RAÇA NELORE

RESUMO – Análises de associação genômica ampla (GWAS) para idade ao primeiro parto (IPP) e duração do período gestacional (PG) em bovinos da raça Nelore foram conduzidas utilizando painel de ~ 777.000 polimorfismos de nucleotídeo de sítio único (SNP). As análises de associação foram conduzidas em três grupos amostrais distintos com seus respectivos dados fenotípicos: 1) IPP em meses para 96 fêmeas, 2) valores genéticos preditos (EBVs) para IPP e PG para 831 touros, e 3) EBVs para IPP e PG para 1.278 fêmeas. Para cada conjunto de dados e para cada marcador, efeitos de substituição alélica foram estimados por meio do método dos quadrados mínimos ponderados e corrigidos para estratificação populacional. Os marcadores apresentando maiores efeitos estimados foram considerados como indicativos de evidência da associação da região cromossômica com o fenótipo investigado. Realizou-se a prospecção de genes candidatos nas regiões detectadas por diferentes metodologias (busca de SNP intragênicos, mapeamento de genes potencialmente em LD com SNP e construção de redes de termos funcionais, além de revisão manual para evidência em processo reprodutivo baseado em literatura científica). Grande número de regiões genômicas foram indicadas pelos efeitos de substituição alélica estimados, os quais apresentaram distribuição esperada para uma característica poligênica. Os resultados obtidos revelaram funções biológicas relacionadas a processos reprodutivos e sugeriram os genes *PRKCD* (protein kinase C delta type), *PRKCE* (protein kinase C epsilon type), *ITHI-1* (inter-alpha-trypsin inhibitor heavy chain H1 precursor), *ITHI-3* (inter-alpha-trypsin inhibitor heavy chain H3), *ITHI-4* (inter-alpha-trypsin inhibitor heavy chain H4 precursor), *BT.62377* (sarcoplasmic/endoplasmic reticulum calcium ATPase 3), *EFCAB6* (não caracterizado), *TPT1* (translationally-controlled tumor protein) e *AMH* (muellerian-inhibiting factor precursor); e os genes *ZBTB16* (zinc finger and BTB domain-containing protein 16) e *PIK3R1* (Phosphatidylinositol 3-kinase regulatory subunit alpha), como candidatos implicados nos fenótipos IPP e PG, respectivamente.

Palavras-chave: *Bos taurus indicus*, Idade ao Primeiro Parto, Duração do Período Gestacional, Genômica Funcional, Gene Candidato, SNP

GENOME-WIDE ASSOCIATION FOR REPRODUCTIVE TRAITS IN NELLORE BEEF CATTLE

ABSTRACT – Genome-wide scans for age at first calving (AFC) and gestation length (GL) in Nellore cattle were performed using a panel of over 777,000 single nucleotide polymorphisms (SNP). Association analyses were carried out in three distinct sample groups and their respective phenotypic data: 1) AFC, in months, for 96 cows, 2) AFC and GL estimated breeding values (EBVs) for 831 sires, and 3) AFC and GL EBVs for 1,278 cows. For each dataset and each marker, allele substitution effects were estimated via ordinary least squares and corrected for population stratification. Markers exhibiting high estimated effects were considered as indicative of evidence of association between the chromosomal region and phenotype investigated. Different methodologies were applied for the prospection of candidate genes within the detected regions (intragenic SNP mining, mapping of genes potentially in linkage disequilibrium with SNP, networking of functional terms, and manual review of genes previously found to be involved in reproductive processes). Large amounts of genomic regions were indicated by the estimated allele substitution effects, which presented a distribution consistent with a polygenic trait. The results obtained revealed biological functions related to reproductive processes and suggested *PRKCD* (protein kinase C delta type), *PRKCE* (protein kinase C epsilon type), *ITHI-1* (inter-alpha-trypsin inhibitor heavy chain H1 precursor), *ITHI-3* (inter-alpha-trypsin inhibitor heavy chain H3), *ITHI-4* (inter-alpha-trypsin inhibitor heavy chain H4 precursor), *BT.62377* (sarcolemmal/endoplasmic reticulum calcium ATPase 3), *EFCAB6* (uncharacterized gene), *TPT1* (translationally-controlled tumor protein) and *AMH* (muellerian-inhibiting factor precursor), and *ZBTB16* (zinc finger and BTB domain-containing protein 16) and *PIK3R1* (Phosphatidylinositol 3-kinase regulatory subunit alpha) as candidate genes related to AFC and GL, respectively.

Keywords: *Bos taurus indicus*, Age at First Calving, Gestation Length, Functional Genomics, Candidate Gene, SNP

LISTA DE ABREVIATURAS E SÍMBOLOS

AH	ácido hialurônico
AMH	anti-müllerian hormone; hormônio anti-mülleriano
<i>AMH</i>	muellerian-inhibiting factor precursor
ANCP	Associação Nacional dos Criadores e Pesquisadores
AR	androgen receptor; receptor de andrógeno
<i>BCL2L11</i>	bcl-2-like protein 11
BTA	<i>Bos taurus</i> autossomo
COC	<i>complexo cumulus oophorus</i> -oócito
<i>COLEC12</i>	Collectin-12
CNV	variação do número de cópias
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	deoxyribonucleic acid; ácido desoxirribonucleico
EBV	estimated breeding value; valor genético estimado
FSH	follicle-stimulating hormone
GC	genomic control; controle genômico
GCs	grupo de contemporâneos
GCscore	GenCall score
GnRH	gonadotropin-releasing hormone; hormônio liberador de gonadotrofinas
GO	Gene Ontology
GS	seleção genômica
GWAS	genome-wide association study; estudo de associação genômica ampla
HD	high-density; alta densidade
HWE	Hardy-Weinberg <i>equilibrium</i> ; equilíbrio de Hardy-Weinberg
IBD	identity-by-descent; identidade por descendência
IBS	identical-by-state; identidade por estado
IEP	intervalo entre partos
IGF1	insulin-like growth factor 1; fator de crescimento semelhante a insulina 1

IPP	idade ao primeiro parto
<i>ITHI</i>	inter-alpha-trypsin inhibitor heavy chain
<i>JMJD1C</i>	probable JmjC domain-containing histone demethylation protein 2C
K	1.000
kb	kilo base (10^3 pares de base)
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	linkage <i>disequilibrium</i> ; desequilíbrio de ligação
LH	luteinizing hormone; hormônio luteinizante
<i>LOXL2</i>	Lysyl oxidase homolog 2
MAF	minor allele frequency; frequência do menor alelo
MAS	seleção assistida por marcadores
Mb	mega base (10^6 pares de base)
μ l	microlitro (10^{-6} L)
miRNA	microRNA
MLM	modelos lineares mistos
pb	par de bases
PCA	principal-components analysis; análise de componentes principais
PCoA	análise de coordenadas principais
<i>PRKCD</i>	protein kinase C delta type
<i>PRKCE</i>	protein kinase C epsilon type
PG	período gestacional
<i>PIK3R1</i>	Phosphatidylinositol 3-kinase regulatory subunit alpha
PI3K	quinase de lipídeos
PKC	proteína quinase
PMGRN	Programa de Melhoramento Genético da Raça Nelore
PP	probabilidade de prenhez
PP14	probabilidade de prenhez aos 14 meses
PP16	probabilidade de prenhez aos 16 meses
QC	quality control; controle de qualidade
QTL	quantitative trait <i>loci</i> ; <i>loci</i> de característica quantitativa
RNA	ribonucleic acid; ácido ribonucleico

RNA _m	RNA mensageiro
<i>SEL1L</i>	sel-1 supressor of lin-12-like (<i>C. elegans</i>)
SERCA	ATPases transportadoras de íons
SNP	single nucleotide polymorphism; polimorfismo de nucleotídeo de sítio único
SnRNA	pequeno RNA nuclear
<i>TBX15</i>	T-box transcription factor TBX15
<i>TF</i>	serotransferrin precursor
<i>TPT1</i>	Translationally-controlled tumor protein
<i>U6</i>	U6 spliceosomal RNA
<i>ZBTB16</i>	zinc finger and BTB domain-containing protein 16
ZGC	Zebu Genome Consortium

LISTA DE TABELAS

Página

Tabela 1.	Sumário de relevantes análises do tipo GWAS para características reprodutivas na espécie bovina.....	17
Tabela 2.	Experimento 1. Estatística descritiva da Idade ao Primeiro Parto (IPP) dada em meses.....	22
Tabela 3.	Experimento 2. Estatística descritiva dos valores genéticos preditos (EBVs) das características Idade ao Primeiro Parto (IPP) e Período Gestacional (PG) em dias.....	24
Tabela 4.	Experimento 3. Estatística descritiva dos valores genéticos preditos (EBVs) das características Idade ao Primeiro Parto (IPP) e Período Gestacional (PG) em dias.....	24
Tabela 5.	Experimento 1. Potenciais genes candidatos para IPP.....	39
Tabela 6.	Experimento 2. Potenciais genes candidatos para EBV1 (IPP).....	51
Tabela 7.	Experimento 2. Potenciais genes candidatos para EBV2 (PG).....	51
Tabela 8.	Experimento 3. Potenciais genes candidatos para EBV1 (IPP).....	64
Tabela 9.	Experimento 3. Potenciais genes candidatos para EBV2 (PG).....	64

LISTA DE FIGURAS

Página

- Figura 1.** Ilustração da estrutura de SNP, haplótipos e tagSNP. a) SNP: quatro versões de sequências curtas de DNA de mesma região cromossômica de diferentes indivíduos. Há três bases onde ocorre variação (SNP). b) Haplótipos: um haplótipo é formado por uma particular combinação de SNP adjacentes. c) TagSNP: genotipagem de três tagSNP escolhidos a partir de um subgrupo pequeno de SNP, suficientes para identificar os quatro haplótipos descritos (Adaptado de THE INTERNATIONAL HAPMAP CONSORTIUM (2003)).....9
- Figura 2.** Ilustração do desequilíbrio de ligação (LD) na geração de uma associação fenotípica entre um marcador não causal genotipado ligado a um *locus* causal não genotipado (Adaptado de Astle e Balding (2009)).....10
- Figura 3.** Experimento 1. “Heatmap” da proporção de alelos idênticos por descendência (IBD) entre os pares amostrais. Valores de 0 (verde - amostras completamente não relacionadas) a 0,5 (vermelho - compartilhamento de metade dos alelos).....33
- Figura 4.** Experimento 1. Análise de escalonamento multidimensional (MDS) baseada no coeficiente IBS de parentesco genômico. (B) Os animais foram coloridos de acordo com o parentesco amostral (animais filhos de um mesmo pai e pontos amarelos indicam filhos de touros menos representativos). As porcentagens correspondem à variância explicada pelo respectivo autovetor.....34
- Figura 5.** Experimento 1. Gráfico de Quantil-Quantil (Q-Q) dos resultados do GWAS para IPP. A linha preta representa a distribuição esperada e os pontos pretos representam a distribuição observada da estatística χ^2 35
- Figura 6.** Experimento 1 (IPP). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,07 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.....36
- Figura 7.** Experimento 1. Distribuição dos efeitos estimados de substituição alélica das variantes SNP para IPP em meses.....37

- Figura 8.** Experimento 1. Redes de termos funcionais para IPP. Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.....40
- Figura 9.** Experimento 2. “Heatmap” da proporção de alelos idênticos por descendência (IBD) entre os pares amostrais. Valores de 0 (verde - amostras completamente não relacionadas) a 0,5 (vermelho - compartilhamento de metade dos alelos).....42
- Figura 10.** Experimento 2. Análise de escalonamento multidimensional (MDS) baseada no coeficiente de parentesco genômico. As porcentagens correspondem à variância explicada pelo respectivo autovetor.....43
- Figura 11.** Experimento 2. Gráficos de Quantil-Quantil (Q-Q) dos resultados dos GWAS para EBV1(A) e EBV2 (B). A linha preta representa a distribuição esperada e os pontos pretos representam a distribuição observada da estatística χ^2 46
- Figura 12.** Experimento 2 – EBV1 (IPP). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,16 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.....47
- Figura 13.** Experimento 2 – EBV2 (PG). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,17 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.....48
- Figura 14.** Experimento 2. Distribuição dos efeitos estimados de substituição alélica das variantes SNP para EBV1 (A) e EBV2 (B) em dias.....49
- Figura 15.** Experimento 2. Redes de termos funcionais para EBV1 (IPP). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.....52

- Figura 16.** Experimento 2. Redes de termos funcionais para EBV2 (PG). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.....53
- Figura 17.** Experimento 3. “Heatmap” da proporção de alelos idênticos por descendência (IBD) entre os pares amostrais. Valores de 0 (verde - amostras completamente não relacionadas) a 0,5 (vermelho - compartilhamento de metade dos alelos).....55
- Figura 18.** Experimento 3. Análise de escalonamento multidimensional (MDS) baseada no coeficiente de parentesco genômico. As porcentagens correspondem à variância explicada pelo respectivo autovetor.....56
- Figura 19.** Experimento 3. Gráficos de Quantil-Quantil (Q-Q) dos resultados dos GWAS para EBV1(A) e EBV2 (B). A linha preta representa a distribuição esperada e os pontos pretos representam a distribuição observada da estatística χ^2 58
- Figura 20.** Experimento 3 - EBV (IPP). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,07 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.....59
- Figura 21.** Experimento 3 - EBV (PG). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,06 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.....60
- Figura 22.** Experimento 3. Distribuição dos efeitos estimados de substituição alélica das variantes SNP para EBV1 (A) e EBV2 (B) em dias.....61
- Figura 23.** Experimento 3. Redes de termos funcionais para EBV1 (IPP). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.....65

Figura 24. Experimento 3. Redes de termos funcionais para EBV2 (PG). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.....66

1. Introdução

A reprodução é o componente-chave para a sobrevivência e evolução das espécies. Seu sucesso envolve a sincronização de vários e complexos mecanismos fisiológicos sob forte influência de fatores ambientais. Do ponto de vista produtivo, a reprodução é um dos fatores mais críticos do retorno econômico dos sistemas de produção das espécies domésticas. Em bovinos de corte, o desempenho reprodutivo da fêmea em particular é, provavelmente, o fator de maior relevância na viabilidade econômica do sistema de ciclo completo, devido ao seu impacto na eficiência produtiva e no ganho genético (CUSHMAN et al., 2008).

Para maximizar a produção, pesquisadores e produtores de bovinos de corte têm voltado sua atenção para as características reprodutivas mensuradas nas fêmeas. Dentre elas, a idade ao primeiro parto (IPP) é uma importante característica relacionada ao desempenho reprodutivo, particularmente das raças zebuínas, pois é a característica mais estudada como expressão de precocidade sexual. As fêmeas zebuínas são consideradas sexualmente tardias quando comparadas às raças de origem taurina e, embora a precocidade sexual seja desejável, ela é menos comumente observada nos rebanhos zebuínos, provavelmente devido às distintas condições ambientais, de alimentação e manejo, além da menor pressão de seleção desta subespécie (COUTINHO e ROSÁRIO, 2010; FORTES et al., 2011).

A redução da IPP tem sido objetivo da atividade para aumentar a produtividade dos rebanhos, pela oportunidade de elevar o número de bezerros produzidos durante a vida útil da matriz e por diminuir o custo de novilhas de reposição, possibilitando maior intensidade de seleção e a redução do intervalo de gerações. Desse modo, o manejo visando reproduzir fêmeas em idade mais jovem, a partir dos 14 meses de idade, reduz a IPP e evidencia a variação genética para a característica, favorecendo a identificação e seleção de animais sexualmente precoces.

Os programas de melhoramento genético têm incluído maior número de características reprodutivas nas avaliações genéticas das raças zebuínas ao longo dos últimos anos. Recentemente, alguns programas incorporaram a duração do período gestacional (PG) nessas avaliações rotineiras. A PG é considerada um

componente do intervalo entre partos (IEP) e, embora apresente pouca variação dentro das populações, possui importância econômica quando analisada em conjunto com outras características reprodutivas (ROCHA et al., 2005; MUCARI et al., 2011).

Fenótipos são a expressão física de genótipos influenciados pelos fatores ambientais e o relacionamento entre esses fatores sempre foi uma das questões centrais das ciências biológicas (SCHWARZENBACHER et al., 2012). As ferramentas genômicas atualmente disponíveis trouxeram nova visão para essa importante questão, contribuindo para a compreensão da base genética das variações fenotípicas para diversas características de interesse, tais como as de desempenho, conformação, saúde e reprodução. Desde 2001, quando a primeira versão da montagem do genoma humano foi concluída, o sequenciamento e montagem de genomas mamíferos progrediram muito rapidamente, levando a uma verdadeira revolução nas áreas de genética, biologia e bioinformática, pelos novos conhecimentos gerados e ferramentas construídas. Atualmente, os genomas de vários organismos são conhecidos, inclusive os de espécies-modelo para a espécie humana como, por exemplo, a bovina.

Após a conclusão do projeto de sequenciamento do genoma bovino grande número de marcadores do tipo polimorfismo de nucleotídeo de sítio único (SNP, "single nucleotide polymorphism") está disponível para a pesquisa genômica bovina (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009; SCHULMAN et al., 2011). A evolução tecnológica das plataformas de genotipagem de SNP possibilitou a criação de novas abordagens para investigar os componentes genéticos subjacentes a fenótipos complexos, genericamente denominadas de estudos de associação genômica ampla (GWAS, "genome-wide association studies").

A pesquisa genômica animal tem sido aplicada com o objetivo de aumentar a produtividade e compreender os mecanismos de resistência às doenças, com a finalidade de atender à crescente demanda mundial dos produtos de origem animal (SODHI e SCHOOK, 2011). Nesse contexto, os GWAS se tornaram abordagem frequente para descoberta e exploração de regiões genômicas associadas a variações existentes no âmbito das características economicamente relevantes aos

sistemas de produção de bovinos. Entretanto, a maior parte dessas características frequentemente apresenta base genética complexa, ou seja, são influenciadas por grande número de *loci* que, em sua maioria, exercem individualmente pequeno efeito.

Nesse contexto, devido à complexidade das características reprodutivas e o desconhecimento dos mecanismos genômicos subjacentes à reprodução da fêmea bovina, análises do tipo GWAS utilizando painel denso de marcadores SNP foram conduzidas em populações animais distintas, com o objetivo de identificar variantes SNP, genes candidatos e funções biológicas potencialmente associadas às características IPP e PG em bovinos da raça Nelore.

2. Revisão de Literatura

2.1. Características Reprodutivas de Fêmeas

A puberdade nas fêmeas mamíferas pode ser definida como o estágio de desenvolvimento que suporta a ciclicidade ovariana, ou seja, o desenvolvimento folicular e a manifestação do estro, seguidos da ovulação e da competência em se tornar gestante (WILLIAMS e AMSTALDEN, 2010). Sob o aspecto endócrino, no período da puberdade ocorre declínio progressivo da resposta hipotalâmica ao estradiol seguido do aumento da estimulação hipofisária e da secreção de altas concentrações do hormônio luteinizante (LH) (DAY et al., 1987). O aumento da pulsatilidade deste hormônio resulta no crescimento de folículos antrais e na síntese de estradiol. Este esteróide induz o comportamento estral e a liberação da onda pré-ovulatória de LH que, em última instância, induzirá a ovulação (RAWLINGS et al., 2003).

As alterações nas respostas aos neurotransmissores estimulatórios e inibitórios que ocorrem durante esse período dependem da idade e da disponibilidade de energia (RAWLINGS et al., 2003). Embora os processos e sinais metabólicos que regulam a maturação hipotalâmica não sejam compreendidos em toda sua complexidade, sua ocorrência se dá quando as taxas de crescimento e desenvolvimento corporal sejam adequadas para que o animal suporte os desafios

endocrinológicos que a desencadeiam (MAQUIVAR, 2011). Desse modo, existem mecanismos endócrinos que garantem a não ativação do sistema reprodutivo até que seu desenvolvimento somático seja compatível com a reprodução, o que ocorre quando o animal atinge cerca de 65 - 70% de peso adulto (SEMMELMANN et al., 2001).

As particularidades reprodutivas fisiológicas e também comportamentais da puberdade estão relacionadas à origem genética dos animais. As diferenças reprodutivas entre zebuínos e taurinos são bem conhecidas, pois implicam em diferentes manejos e respostas aos tratamentos hormonais como, por exemplo, superovulação e sincronização do ciclo estral e da ovulação (BARUSELLI et al., 2007). De modo geral, as diferenças fisiológicas entre estas subespécies estão associadas aos diferentes ambientes onde elas foram originadas e são atualmente exploradas. Fatores esses que levaram à menor pressão de seleção artificial observada nos zebuínos e, não menos importante, foi também o papel da seleção natural exercida pelo meio ambiente nessa subespécie (CHENOWETH, 1994).

As raças zebuínas são mais tolerantes aos parasitas e ao clima tropical, porém suas taxas reprodutivas são mais baixas quando comparadas aos animais de origem taurina, que são menos adaptados a essas regiões. A idade à puberdade, fenótipo que varia entre subespécies, raças e famílias, é considerada uma diferença marcante entre zebuínos e taurinos, pois novilhas zebuínas atingem a puberdade em idade mais avançada (RANDEL, 1984), em média entre 6 e 12 meses mais tarde e com peso corporal relativo ao peso adulto mais elevado quando comparadas às fêmeas de origem taurina (WILTBANK et al., 1969; DOBSON et al., 1986). De acordo com Pereira et al. (2002), a idade média ao primeiro parto em rebanhos da raça Nelore é de 34 meses de idade, o que corresponde idade à concepção de 24-25 meses.

A puberdade é um processo complexo que envolve diferentes sistemas e órgãos, fortemente influenciado por fatores de origem hereditária e de natureza ambiental. A maior tolerância dos zebuínos às pressões ambientais acarretou algumas consequências indesejáveis, tais como a puberdade tardia e à menor produtividade medida pelas características de carcaça. Entretanto, quando comparadas às raças europeias os rebanhos zebuínos apresentam maior

variabilidade genética devido a menor interferência humana no processo de seleção, possibilitando a existência de maior base para o melhoramento genético (NOGUEIRA, 2004).

O preceito de que a puberdade na subespécie zebuína esteja programada em função do alcance de determinado peso e/ou idade têm direcionado, até então, as ações de manejo reprodutivo empregado em novilhas. Entretanto, existe alta variabilidade genética nesta subespécie para a idade à puberdade e outros fenótipos reprodutivos e, desse modo, torna-se interessante, necessário e importante que animais mais precoces sejam selecionados (ELER, 2004).

A eficiência reprodutiva de um rebanho pode ser mensurada por meio de diferentes parâmetros, tais como: idade à puberdade, IPP, IEP, PG, dias para o parto, habilidade de permanência no rebanho e probabilidade de prenhez (PP). Tais características são consideradas complexas ou poligênicas e, dentre estas, a idade à puberdade, IPP e PP são características que também indicam precocidade sexual.

A precocidade sexual pode ser definida pela aptidão da fêmea bovina à reprodução em época antecipada em relação à média da raça ou população (TANAKA, 2010). De acordo com Albuquerque et al. (2006), essa é a característica de maior importância reprodutiva em zebuínos, uma vez que as novilhas que atingem mais cedo a maturação sexual são mais jovens ao primeiro parto e apresentam menor intervalo entre partos (LUNA-NEVAREZ et al., 2010).

Embora seja um fenótipo desejável, indivíduos sexualmente precoces são menos comuns nos rebanhos zebuínos em comparação aos animais taurinos (FORTES et al., 2011), provavelmente devido a menor pressão de seleção. Além disso, a seleção para precocidade sexual é dificultada pelas distintas condições ambientais, de alimentação e de manejo, pois a sazonalidade na disponibilidade e qualidade das pastagens prejudica o crescimento normal e atrasa o primeiro serviço (COUTINHO e ROSÁRIO, 2010). Desse modo, o manejo padrão dos sistemas de produção das raças zebuínas consiste no acasalamento das novilhas somente após os 24 meses de idade, fazendo com que a média de IPP permaneça próxima dos 36 meses. Tal estratégia de manejo perpetua a grande distância entre a realidade da pecuária extensiva e aquela concebida para a pecuária de corte ideal, na qual o primeiro parto ocorre próximo aos 24 meses, mantendo média de IEP de 365 dias.

Os critérios de seleção utilizados nos programas de melhoramento de zebuínos são baseados em características de crescimento e de desenvolvimento corporal e circunferência escrotal (ALBUQUERQUE et al., 2006). Até bem pouco tempo, a circunferência escrotal era a única característica amplamente utilizada como indicadora de fertilidade, tanto de machos quanto de fêmeas. Esta é uma característica de herdabilidade moderada (próxima de 0,20), fácil mensuração e que apresenta correlações genéticas favoráveis como, por exemplo, de - 0,22 com IPP e de 0,46 com probabilidade de prenhez aos 14 meses (PP14) (PEREIRA et al., 2000; SILVA et al., 2002).

Contudo, nos últimos anos têm crescido a atenção para as características reprodutivas mensuradas diretamente nas fêmeas devido à importância econômica que estas exercem sobre a pecuária de corte (SANTANA Jr et al., 2012). Assim, a redução das idades à puberdade e ao primeiro parto, com objetivo de aumentar a produtividade se tornou grande preocupação para os criadores e pesquisadores (NOGUEIRA, 2004). A herdabilidade estimada para idade à puberdade é alta ($\sim 0,5$), permitindo que o processo de seleção seja mais eficiente. No entanto, a obtenção de dados para o uso direto desta característica em programas de seleção genética é difícil e onerosa, principalmente em condições extensivas, pois envolve procedimentos mais complexos para a constatação da primeira ovulação (ALBUQUERQUE e FRIES, 1997; JOHNSTON et al., 2009).

A IPP marca o início da vida produtiva de uma fêmea de corte pois reflete a idade à puberdade, tratando-se de característica de fácil mensuração, cuja obtenção não implica em custo para o sistema (DIAS et al., 2004). Entretanto, as estimativas de herdabilidade para IPP são de baixas a moderadas, variando de 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). Desse modo, não se espera que a seleção direta para esta característica apresente resposta muito contundente.

O simples manejo visando à antecipação a exposição de novilhas aos machos tem favorecido a identificação de animais geneticamente superiores para precocidade reprodutiva (MONSALVES, 2008). Nesse sistema é que a característica PP está inserida, uma característica classificada como binária ou dicotômica, onde valor 1 assume sucesso para fêmeas gestantes e valor 0 assume fracasso para as

não gestantes após estação de acasalamento. A PP pode ser diferentemente abordada de acordo com a idade ao acasalamento como, por exemplo, PP14 ou 16 meses (PP16). Além disso, as herdabilidades estimadas para estas características são moderadas a altas, com valores em torno de 0,51 e 0,52 para PP16 e de 0,47 para PP14 (SILVA et al., 2005; MEIRELLES et al., 2009; SHIOTSUKI et al., 2009).

Atualmente alguns programas de melhoramento de raças de corte zebuínas têm incluído o PG em suas avaliações genéticas. Por ser considerado um dos componentes do IEP, o PG possui importância econômica quando analisado em conjunto com outras características reprodutivas (ROCHA et al., 2005; AZÊVEDO et al., 2006; MUCARI et al., 2011). Esta característica não apresenta grandes variações nas populações, entretanto as fêmeas com menores períodos gestacionais apresentam vantagens, tais como maior tempo para recuperação até o próximo acasalamento devido ao maior IEP (PEREIRA et al., 2002).

O PG também está correlacionado ao peso ao nascimento e facilidade ao parto, de modo que a seleção para PG mais curto diminuiria o peso ao nascimento e, conseqüentemente, a probabilidade de dificuldades ao parto (ROCHA et al., 2005). As implicações do PG em bovinos leiteiros têm sido ainda maiores, pelo aspecto econômico e também de manejo, pois influencia características relacionadas à sobrevivência do bezerro no período perinatal. Além disso, problemas no parto aumentam o risco de comprometimento da fertilidade da matriz (MALTECCA et al., 2011).

As médias de PG na raça Nelore variam de 290 a 295 dias e suas herdabilidades são frequentemente estimadas entre 0,12 e 0,26, valores indicativos de que a característica pode responder a processo de seleção genética (PEREIRA et al., 2002; CAVALCANTE et al., 2001; AZÊVEDO et al., 2006; BOLIGON et al., 2007).

O processo de seleção para precocidade sexual apresenta vantagens como o aumento da taxa desfrute, da vida produtiva e do progresso genético, pela redução do intervalo entre gerações e da taxa de reposição de novilhas (MEIRELLES et al., 2009). Assim, a atenção às características reprodutivas é fundamental na implantação e desempenho de programas de melhoramento genético, pois sua não consideração pode limitar a intensidade de seleção, além de interferir no intervalo

entre gerações, diminuindo a eficiência geral do processo (BOLIGON e ALBUQUERQUE, 2010). Apesar do reconhecimento de que medidas de eficiência e precocidade reprodutiva deveriam fazer parte dos objetivos de seleção, o melhoramento genético destas características tem sido prejudicado e lento devido a fatores como: dificuldade de obtenção de informações, baixa herdabilidade e expressão tardia de fenótipos sendo, portanto, características mais difíceis de serem melhoradas pelos métodos tradicionais de seleção (BUDDENBERG et al., 1990; CAMMACK et al., 2009).

2.2. Marcadores SNP e Análise de SNP em Larga Escala

A descoberta das bases genéticas implicadas na manifestação das diferenças fenotípicas requer a compreensão abrangente de todas as possíveis formas de variação genética (LIU e BICKHART, 2012). A diversidade da sequência genômica é representada principalmente pelos SNP ou por variações estruturais no tamanho dos segmentos genômicos. Os rearranjos cromossômicos podem levar a modificações na ordem (inversões e translocações) ou no número (duplicações e deleções) de regiões genômicas, modelando a variação fenotípica por meio de alterações na dosagem e regulação gênicas, assim como na estrutura transcrita (CLOP et al., 2012).

Os SNP constituem a classe mais abundante de sítios polimórficos em qualquer genoma. Esses polimorfismos são gerados durante a replicação do DNA por mutação espontânea ou induzida. Estima-se que existam aproximadamente 10 milhões destes polimorfismos tanto no genoma humano quanto no bovino, ou seja, um SNP a cada 1.000 bases (K) ou menos (WEINER e HUDSON, 2002; THE INTERNATIONAL HAPMAP CONSORTIUM, 2003; THE BOVINE HAPMAP CONSORTIUM, 2009). Este tipo de marcador apresenta taxa de mutação relativamente baixa, além de maior facilidade e relativo baixo custo de análise (genotipagem), quando comparado aos clássicos marcadores do tipo microsatélite.

Do ponto de vista molecular, um SNP é definido como o sítio no genoma onde somente um par de bases (nucleotídeos) difere, gerando dois alelos para um marcador específico, ou seja, uma mutação de ponto (**Figura 1**). Com relação a sua

natureza polimórfica, um SNP é considerado marcador bialélico embora, em situação rara, possa ser tri ou tetra-alelício.

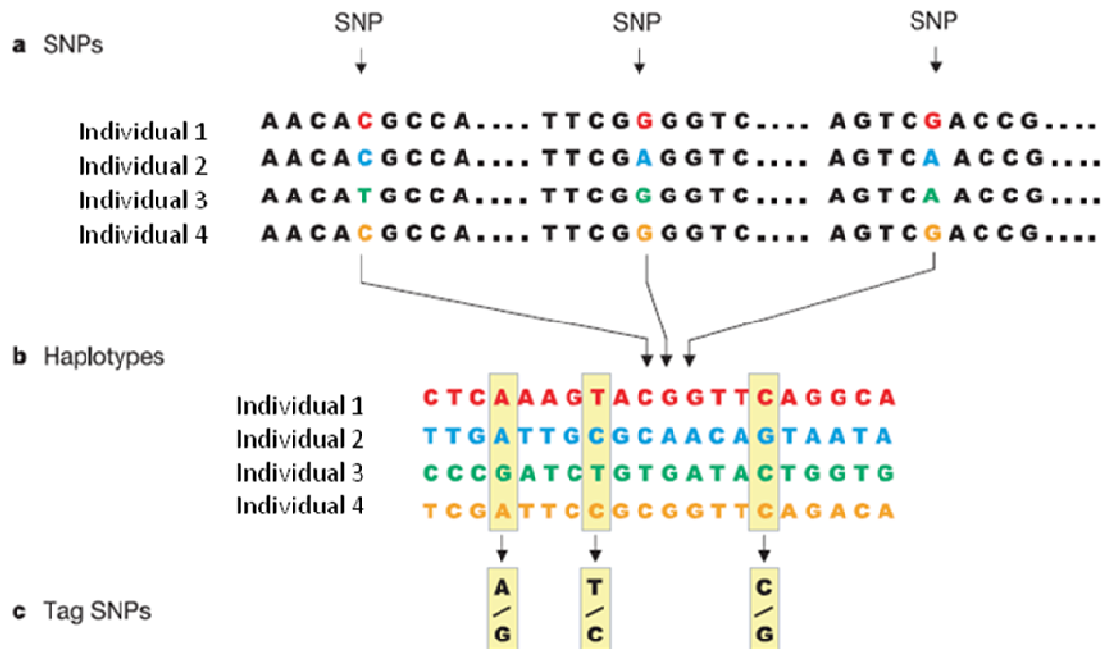


Figura 1. Ilustração da estrutura de SNP, haplótipos e tagSNP. a) SNP: quatro versões de sequências curtas de DNA de mesma região cromossômica de diferentes indivíduos. Há três bases onde ocorre variação (SNP). b) Haplótipos: um haplótipo é formado por uma particular combinação de SNP adjacentes. c) TagSNP: genotipagem de três tagSNP escolhidos a partir de um subgrupo pequeno de SNP, suficientes para identificar os quatro haplótipos descritos (Adaptado de THE INTERNATIONAL HAPMAP CONSORTIUM (2003)).

Dependendo da posição do SNP no genoma, o marcador pode ser classificado como: exônico (sinônimo e não sinônimo), intrônico e inter-gênico. De acordo com o The Bovine Hapmap Consortium (2009), os SNP estão distribuídos no genoma bovino da seguinte maneira: regiões intergênicas (63,74%) e, menor parte, em regiões intrônicas (34,9%) e exônicas (1,35%). Embora mudanças na sequência de nucleotídeos possam resultar em alteração de proteínas e contribuir para a variabilidade dos diferentes fenótipos observados, a maior parte das mutações não resulta em alterações aminoacídicas e, conseqüentemente, protéicas (ORR et al., 2006). No entanto, a informação de um grande número de SNP mapeados é de grande valor, pois podem funcionar como marcadores indiretos para SNP funcionais

(COLLINS et al., 1998). Dessa maneira, SNP podem estar associados ao fenótipo diretamente, nos casos em que o SNP é a mutação causal que altera a regulação e/ou expressão gênicas, ou indiretamente, quando o SNP está ligado à mutação causal por desequilíbrio de ligação (LD) (**Figura 2**).

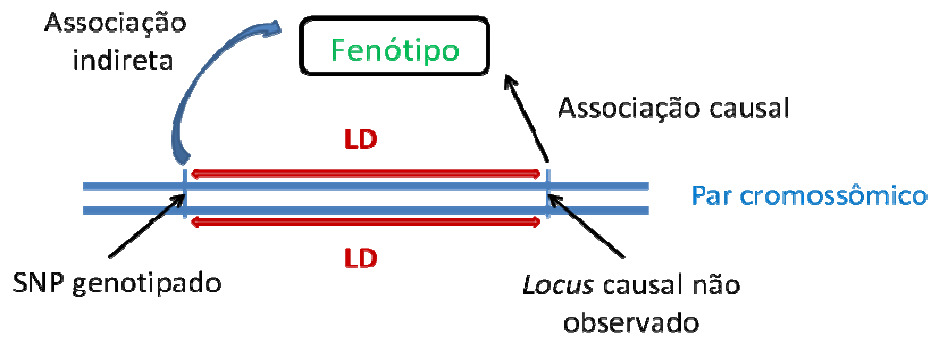


Figura 2. Ilustração do desequilíbrio de ligação (LD) na geração de uma associação fenotípica entre um marcador não causal genotipado ligado a um *locus* causal não genotipado (Adaptado de Astle e Balding (2009)).

O LD é a medida do grau de associação não aleatória entre dois ou mais *locus* no genoma. A extensão de LD em uma população ocorre principalmente devido à proximidade física entre os marcadores, no entanto a história e as forças evolutivas também a influenciam (ARDLIE et al., 2002). Nos animais de produção, o padrão de LD é gerado, principalmente, pela seleção e endogamia, decrescendo devido a recombinação e mutações (KIM e KIRKPATRICK, 2009).

As investigações iniciais sobre padrões de LD e de recombinação foram conduzidas na espécie humana, demonstrando que o genoma é caracterizado por estrutura de blocos altamente correlacionados com os SNP pelo LD (DALY et al., 2001; GABRIEL et al., 2002), com exceção de regiões que apresentam alta taxa de recombinação. Dessa maneira, alelos compõem blocos de SNP que estão em proximidade física e frequentemente correlacionados. Esses blocos são denominados haplótipos e segregam em conjunto dentro da população. Com frequência, somente um pequeno subgrupo de SNP, denominados de tagSNP, é suficiente para capturar toda a informação haplotípica (**Figura 2**).

A primeira aplicação de SNP em larga escala foi conduzida na espécie bovina em 2009, onde 37.470 SNP foram avaliados em 497 animais de 19 raças. Nesse estudo foram descritas as diferenças entre MAF (frequência do menor alelo) para as subespécies taurina e zebuína, sendo que os primeiros apresentaram MAF superiores às dos zebuínos (30% *versus* 19% de animais com MAF > 0.3, respectivamente). Além disso, foi observado que os blocos de LD eram superiores em taurinos, contemplando maiores distâncias genômicas (THE BOVINE HAPAMAP CONSORTIUM, 2009).

Com a conclusão do projeto de sequenciamento do genoma bovino, grande número de SNP foi descoberto e disponibilizado para pesquisa (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009). Paralelamente, a tecnologia de micro-arranjos de alta densidade foi desenvolvida e plataformas de análise (genotipagem) comerciais de SNP construídas. Quando comparado ao que era realizado anteriormente com os marcadores microssatélite, os micro-arranjos permitiram a varredura do genoma de modo rápido e eficiente, tornando possível a condução de análises do tipo GWAS.

Por definição, análises do tipo GWAS tipicamente avaliam grande número de marcadores genéticos SNP polimórficos em uma população com o objetivo de revelar associações entre SNP e fenótipo (HIRSCHHORN e DALY, 2005). Seu conceito surgiu a partir de estudos epidemiológicos humanos baseados na hipótese doença comum - variante comum, sugerindo que a ocorrência de doenças complexas comuns fosse influenciada por um número moderado de alelos, denominados de variantes causais. Assim, a variante (SNP) precisa ser comum para causar uma doença ou característica comum, ou seja, o SNP deve ser compartilhado em um grupo de indivíduos, presente em pelo menos 1 a 5% da população.

No âmbito da produção animal, análises do tipo GWAS podem ser definidas pelo uso de painéis densos de SNP para pesquisar em uma determinada população ou raça, quais as variações genéticas que exercem função sobre qualquer característica fenotípica de relevância econômica (FRAZER et al., 2009). Esses estudos estão em evidência na espécie bovina, pois se tornaram abordagem poderosa para a anotação de efeitos fenotípicos ou mapeamento de QTL

("quantitative trait *loci*") (COLE et al., 2011). Os painéis de SNP para genotipagem existentes na atualidade representam diferença significativa na densidade de marcadores para mapeamento de QTL, quando comparados às poucas dezenas de marcadores microssatélite que eram comumente genotipados por cromossomo nos estudos de ligação genética (SCHULMAN et al., 2011).

O BovineSNP50 BeadChip (Illumina Inc., USA) foi o primeiro painel de larga escala disponibilizado comercialmente para a espécie bovina. Ele permite a análise de mais de 54K SNP e foi desenvolvido por um conjunto de pesquisadores a partir de dados genômicos de populações das raças Holandesa, Angus e de animais cruzados de corte (Van TASSEL et al., 2008; MATUKUMALLI et al., 2009). Milhares de animais já foram genotipados com essa plataforma e muitos dos dados gerados estão sendo utilizados no processo de seleção genômica (GS) em bovinos de leite principalmente, nos EUA e Canadá (RINCON et al., 2011). Infelizmente, a seleção dos SNP que integram esse painel foi otimizada para oferecer MAF altas para raças taurinas, de modo que o teste é relativamente pouco informativo para as raças zebuínas.

As decisões de quais SNP serão inclusos nas plataformas de genotipagem comerciais são baseadas na uniformidade de cobertura em termos de distância física ou de LD entre pares de marcadores (WEIR e LAURIE, 2010). Na espécie bovina, foi estimado que o número de SNP necessários para análises do tipo GWAS deve ser maior do que 300K, pois a maior densidade de marcadores permite identificar aqueles marcadores em LD mais alto com os genes que afetam uma característica de interesse, otimizando as predições genômicas (KHATKAR et al., 2008; GODDARD e HAYES, 2009; SU et al., 2012).

Recentemente foi desenvolvido o Bovine High-Density BeadChip, que consiste em um painel de alta densidade de SNP para aplicação na análise do tipo GWAS nas mais variadas raças bovinas. Disponível desde 2010, o teste constitui em ferramenta de alta cobertura genômica, contendo mais de 777K SNP validados em bovinos de corte e leite, que incorporou populações taurinas, zebuínas e cruzadas, adaptadas aos climas temperado e tropical (ILLUMINA, 2011).

Além das análises do tipo GWAS, a genotipagem de SNP em larga escala possibilitou a aplicação da GS, um novo tipo de seleção assistida por marcadores

(MAS) que fornece a oportunidade de incorporar a informação genômica nas avaliações genéticas em programas de melhoramento. O método consiste na utilização da informação combinada de genótipos, fenótipos e genealogia para aumentar a acurácia dos valores genéticos preditos (EBVs) (MEUWISSEN et al., 2001). A GS em animais de produção está revolucionando o melhoramento genético, sendo uma realidade em países como Estados Unidos, Canadá, Austrália, Nova Zelândia e Holanda, pela potencial aceleração do melhoramento genético das raças leiteiras (HAYES et al., 2009; VanRADEN et al., 2009).

Segundo Johnston et al. (2012), com o uso da informação genômica também haverá a oportunidade de elevar o progresso genético em bovinos de corte, principalmente pela predição de EBVs com acurácia mais alta para touros jovens. Isto ocorre devido à limitada acurácia na seleção de animais jovens para várias características de importância econômica e, comumente, de baixa herdabilidade (Van EENENNAAM e DRAKE, 2012).

Os GWAS e a GS constituem-se em aplicações promissoras das ferramentas genômicas recentemente desenvolvidas para o melhoramento de características economicamente relevantes dos sistemas de produção (FAN et al., 2010). A aplicação da GS poderia permitir, por exemplo, a seleção acurada para fenótipos reprodutivos sem qualquer conhecimento dos mecanismos biológicos relacionados aos fenótipos avaliados (SNELLING et al., 2011). No entanto, a abordagem do tipo GWAS poderia auxiliar na identificação de vias biológicas responsáveis pela variação genética para o fenótipo reprodutivo de interesse e na elucidação da natureza genética dessas características. A compreensão dos padrões de variação genética associados a fenótipos de interesse constitui-se na base para o futuro dos programas de melhoramento, sendo as ferramentas genômicas disponíveis na atualidade uma oportunidade sem precedentes para descobrir *loci* ou variações genômicas para uso em seleção animal.

2.2.1. Aspectos Inerentes a Análise e Controle de Qualidade dos Dados de SNP

A recente popularidade das análises do tipo GWAS é reflexo dos avanços tecnológicos no campo da análise genética em larga escala, apresentando baixas taxas de erro e reduzido custo por marcador analisado (AMOS, 2007). Embora apresentem muitas vantagens, a principal fragilidade desses estudos é a vulnerabilidade quanto a possibilidade de obtenção de conclusões errôneas (GODDARD et al., 2009). São três as possíveis condicionantes para que uma associação entre um polimorfismo e um fenótipo seja significativa: (1) a descoberta é um falso positivo devido a uma associação espúria; (2) a associação é devida à proximidade de uma variante funcional em um *locus* adjacente (LD); e (3) o *locus* é funcional e diretamente afeta a expressão do fenótipo (LITONJUA e CELEDO, 2006). Dessa maneira, para que a evidência de uma associação seja verdadeira, a correlação observada não pode ser um viés de inferência dos genótipos, do delineamento experimental ou da estratégia de amostragem (ANDERSON et al., 2010; BELLENGUEZ et al., 2012).

Uma das etapas mais críticas de uma análise do tipo GWAS consiste no controle de qualidade (QC) dos dados genotípicos, pois este é o procedimento que deve remover marcadores e amostras potencialmente problemáticas, de baixa qualidade ou informatividade para minimizar falsas descobertas (PONGPANICH et al., 2010). Esta análise é rotineiramente dividida em duas categoriais principais, QC de SNP e QC das amostras. Comumente, no QC de SNP são utilizados filtros técnicos baseados, principalmente, nas variáveis: proporção de SNP faltantes ou "call rate", Equilíbrio de Hardy-Weinberg (HWE) e MAF.

Quando grande número de SNP é testado em um ensaio, vieses potenciais e associações falso-positivas ocorrem devido a artefatos técnicos. Nesse aspecto, o "call rate" é importante parâmetro de qualidade da genotipagem, pois representa a porcentagem de SNP que foram adequadamente determinados (AA, AB ou BB) em relação ao total disponível para análise. Esta medida é comumente utilizada como mensuração da integridade do procedimento e, valores superiores a 98% são

obtidos pelas tecnologias atualmente disponíveis para análise de SNP (MOORHEAD et al., 2006; PARE, 2010).

O teste para HWE é outra prática frequente na busca evidências de problemas com os dados genotípicos (WEIR e LAURIE, 2010). Erros de genotipagem comumente se manifestam como desvio das proporções esperadas de HWE, particularmente se a plataforma apresenta vieses na determinação do genótipo. Assim, uma estratégia de QC importante consiste no estabelecimento da conformidade com as proporções de HWE e eliminação de marcadores que desviam dessas proporções, antes que os testes de associação sejam conduzidos (GODDARD et al., 2009).

O princípio geral do HWE é que as frequências alélicas e genotípicas em uma população grande e com acasalamento aleatório, permanecem estáveis entre gerações, existindo relacionamento fixo entre as frequências alélicas e genotípicas (ZIEGLER e KÖNIG, 2006). As análises do tipo GWAS comumente excluem marcadores que mostram desvios para HWE, pois isto poderia ser um indicativo de erro de genotipagem. O princípio de HWE é utilizado há mais de um século para melhor compreensão das características genéticas das populações e está presente quando as frequências genotípicas de uma população permanecem em equilíbrio através das gerações ($p^2+2pq+q^2$), a menos que fatores específicos sejam introduzidos (DENG et al., 2000; WITTKE-THOMPSON et al., 2005). Assim, genótipos que desviam do HWE podem indicar endogamia, estratificação da população, seleção ou associação fenótipo-genótipo (BALDING, 2006).

Não está claro quantos destes desvios são devidos a artefatos de genotipagem e quantos são verdadeiros por desvios de frequência genotípica, contudo a inspeção dos gráficos de agrupamento poderia indicar que desvios extremos sejam devidos a ensaios deficientemente realizados (LAURIE et al., 2010). Quando presentes os erros de genotipagem repercutem negativamente, seja nos testes de associação como nas análises de LD ou seleção de tagSNP.

Como os polimorfismos geralmente são bi-alélicos, a frequência do menor alelo é denominada de MAF. Os ensaios comerciais utilizados para GWAS são construídos para fornecer excelente cobertura de SNP comuns, mas apresentam um potencial muito limitado para capturar variantes raras e de frequência baixa, ou seja,

aquelas cujas MAF estejam abaixo de 5% (ZEGGINI et al., 2005). Dessa maneira, o critério de filtragem para MAF pode variar de acordo com o tamanho amostral, mas limites entre 1 e 5% são comuns (ZIEGLER et al., 2008). Segundo TABANGIN et al. (2009), o descarte de SNP baseado em MAF abaixo de um determinado limite também está relacionado a fatores inerentes ao ensaio de genotipagem, como baixos "call rates" ou baixa acurácia na determinação de genótipos.

O QC aplicado às amostras basicamente se resume aos critérios "call rate" e amostras duplicadas. O "call rate" de uma amostra corresponde ao número de SNP que foi determinado dividido pelo total analisado. Com frequência são relatados limites superiores a 97%, no entanto valores acima de 90% são suficientes para garantir a confiabilidade do estudo (ZIEGLER, 2008; ANDERSON et al., 2010). Pares de amostras idênticas devem ser identificados e removidos do conjunto de dados, pois as análises de associação são sensíveis ao parentesco amostral e podem introduzir resultados falso-positivos.

2.3. Análises do Tipo GWAS para Características Reprodutivas

Muitos QTL utilizando marcadores microssatélite foram mapeados para características reprodutivas em bovinos nas décadas recentes. Atualmente existem 219 QTL identificados para características relacionadas à fertilidade dentre o total de 2.344 QTL para características economicamente relevantes de bovinos de corte e leite (CattleQTLdb, 2013). Os estudos de mapeamento de QTL alcançaram resultados notáveis, mas devido à baixa densidade de marcadores utilizados não foram suficientes para capturar toda a variação genética inerente às características complexas (ZHANG et al., 2012a). Nesse sentido, a abordagem do tipo GWAS tem apresentado resultados promissores na identificação de regiões genômicas implicadas nos processos reprodutivos. Na **Tabela 1** foram destacadas as principais análises do tipo GWAS para características reprodutivas realizadas até o presente momento na espécie, utilizando os dois principais painéis comerciais para análise de SNP.

Tabela 1. Sumário de relevantes análises do tipo GWAS para características reprodutivas na espécie bovina.

Painel	Cromossomo	Característica	Categoria/Raça	Referência
50k	BTA18	período gestacional	touros holandeses	Cole et al. (2009)
	BTA1,3-6,9,10-13,19,20,24,29	índice de fertilidade	touros holandeses	Sahana et al. (2010)
	BTA3-6,8,10,12 15,17,19,22,BTX	fertilidade	touros holandeses	Blaschek et al. (2011)
	BTA14	idade à puberdade e anestro pós-parto	novilhas Brahman	Hawken et al. (2011)
	BTA5	idade à puberdade	novilhas Tropical Composite	Hawken et al. (2011)
	BTA18	período gestacional	touros holandeses e Italian Brown	Maltecca et al. (2011)
	BTA1,2,11,19,20,21	idade à puberdade	novilhas Brahman e Tropical Composite	Fortes et al. (2012b)
	BTA5,25	taxa de concepção	touros holandeses	Penágaricano et al. (2012)
	BTA8	concepção ao primeiro serviço e prenhez	novilhas Brangus	Peters et al. (2012)
	BTX	taxa de prenhez	vacas holandesas	Cole et al. (2011)
777k	BTA18	facilidade de parto	vacas holandesas	Cole et al. (2011)
	BTA14,21	facilidade de parto	touros German Fleckvieh	Pausch et al. (2011)
	BTA14, BTX	circunferência escrotal e % de espermatozóides normais	touros Brahman	Fortes et al. (2012a)
	BTA14	idade ao primeiro corpo lúteo	novilhas Brahman	Fortes et al. (2012a)

3. Hipótese

Análises do tipo GWAS seguidas de métodos de análise funcional, a partir de diferentes populações de bovinos da raça Nelore, são capazes de detectar regiões cromossômicas, genes candidatos e funções biológicas associadas às características reprodutivas IPP e PG.

4. Objetivos

4.1. Objetivo Geral

Identificar polimorfismos de nucleotídeo de sítio único (SNP) e regiões genômicas associadas a características reprodutivas em bovinos da raça Nelore.

4.1.1. Objetivos Específicos

- Empregar análises do tipo GWAS em três populações de bovinos da raça Nelore, utilizando fenótipos relativos à idade ao primeiro parto (IPP) e duração do período gestacional (PG).
- Explorar as regiões genômicas identificadas pelas análises do tipo GWAS empregando métodos de análise funcional para detectar potenciais genes candidatos e funções biológicas associadas aos fenótipos estudados.

5. Material e Métodos

5.1. Animais

Três grupos amostrais independentes foram analisados e seguem descritos como Grupos 1, 2 e 3.

O Grupo 1 foi composto por 96 fêmeas da raça Nelore, nascidas entre os anos de 1996 e 2006, filhas de 19 touros e 94 matrizes. A seleção dos animais se deu a partir de um rebanho comercial de 500 fêmeas que possuíam dados relativos à IPP, tendo a paternidade conhecida como principal critério de seleção. A propriedade está localizada no município de Iguatemi - MS e adota manejo alimentar exclusivamente a pasto, com predomínio de *Brachiaria sp* e suplemento mineralizado *ad libitum*.

A propriedade participa do Programa de Melhoramento Genético da Raça Nelore (PMGRN) coordenado pela Associação Nacional dos Criadores e Pesquisadores (ANCP) sendo que, desde 1995, estabelece estação de acasalamento para novilhas denominada "estação de desafio", independentemente de peso e condição corporal. No início deste manejo, as fêmeas foram desafiadas aos 18 meses de idade e, a partir de 2003, o manejo começou a ser utilizado nas fêmeas entre 12 e 16 meses de idade. Essa "estação de desafio" ocorre por período de 45 dias entre os meses de janeiro e fevereiro, quando as novilhas são mantidas sob as mesmas condições de manejo e, em monta natural com um grupo de touros jovens, na proporção de um macho para 25 fêmeas.

Os Grupos 2 e 3 foram compostos, respectivamente, por 831 touros e 1.278 vacas da raça Nelore, pertencentes à Conexão Delta G, uma associação de pecuaristas voltada para o melhoramento genético das raças Nelore, Hereford e Braford, formada a partir de diferentes programas de melhoramento e mais de 250 fazendas distribuídas no Brasil e Paraguai. Além disso, dados fenotípicos e genotípicos destes animais integram o "Zebu Genome Consortium" (ZGC), projeto iniciado no ano de 2011 com a colaboração de pesquisadores e instituições do Brasil, Estados Unidos, Canadá, Áustria e Nova Zelândia.

5.2. Extração de DNA e Análise de SNP

DNA dos bovinos pertencentes ao Grupo 1 foi extraído a partir de amostras de sangue venoso colhido com anticoagulante EDTA por meio de "kit" de extração de DNA de sangue e tecido (DNeasy Blood & Tissue Kit, QIAGEN, Espanha) e avaliado quanto a sua concentração e pureza por meio de espectrofotometria de micro-volume (NanoDrop 2000, Thermo Fisher Scientific Inc., EUA). A concentração do DNA foi de ao menos 50 ng/μl e a razão 260:280 foi de 1,8 ($\pm$ 0,2). Posteriormente as amostras foram submetidas a análises de SNP em painel de alta densidade (BovineHD Genotyping BeadChip, Illumina Inc., EUA), sob o ensaio de genotipagem Infinium HD na plataforma iScan, de acordo com as instruções do fabricante (**Fig. supl. 1**).

Além dos dados de análise de SNP gerados a partir do Grupo 1, foram fornecidos pelo ZGC os genótipos das amostras dos Grupos 2 e 3 analisadas no BovineHD BeadChip. O DNA preparado a partir dos animais (touro) do Grupo 2 foi extraído a partir do sêmen com o uso de protocolo fenol:clorofórmio:álcool isoamílico (25:24:1) (SAMBROOK et al., 1989). O material genômico do Grupo 3 foi extraído a partir dos folículos pilosos utilizando "kit" de extração de DNA de tecidos (NucleoSpin[®] TissueKit, Macherey-Nagel, Alemanha).

O painel BovineHD BeadChip dispõe para análise 786.798 SNP, com intervalo médio de 3,402 e mediana de 2,693 kb entre marcadores (ILLUMINA, 2011). O ensaio de genotipagem Infinium HD envolve, resumidamente, as etapas de amplificação, fragmentação, precipitação, ressuspensão e hibridização do DNA genômico. Após o processo de hibridização, ocorre a extensão de uma única base nucleotídica da sequência que, marcada por corantes, é detectada pela fluorescência emitida.

As posições dos SNP dentro do ensaio foram baseadas na versão de montagem do genoma *Bos taurus* UMD 3.1 (acesso GenBank DAAA00000000.2). O painel de SNP testados continha 1.867 marcadores mapeados para "contigs" que não foram adequadamente posicionados nessa versão de montagem do genoma bovino, e que foram omitidos dos conjuntos de dados (**Tab. supls. 1 e 2**).

O agrupamento e a determinação dos genótipos foram conduzidos por meio do módulo de genotipagem do programa GenomeStudio v2011.1 (Illumina Inc., EUA). Para cada amostra, o programa converte as intensidades de fluorescência em genótipos de SNP. Nessa etapa, os algoritmos analisam os dados primários de normalização, agrupamento e determinação dos genótipos a partir dos dados de intensidade gerados na leitura do painel, procedimento onde um indicador de confiabilidade de cada genótipo denominado de “GenCall score” (GCScore) é estimado. O valor do GCScore é dado por sua posição em relação ao seu agrupamento, ou seja, seus genótipos hipotéticos (AA, AB e BB), variando entre 0 e 1. Segundo o fabricante, são considerados os genótipos com escores iguais ou superiores a 0,15, de modo que valores inferiores indicam genótipos que não foram determinados. No entanto, alguns pesquisadores são mais criteriosos com este parâmetro, estabelecendo limites superiores a 0,60 com o objetivo de evitar níveis intermediários de confiabilidade (TEO et al., 2007).

No presente estudo os genótipos determinados com sucesso foram definidos pelo GCScore > 0,70. Posteriormente, os dados genotípicos foram exportados como arquivos de texto e importados para o pacote estatístico R (versão 2.15.0), utilizando-se do programa Plink, funções customizadas em R e biblioteca GenABEL (versão 1.7-2), para que fossem procedidas as análises de controle de qualidade, verificação de estratificação populacional e testes de associações genômicas (AULCHENKO et al., 2007; PURCELL et al., 2007; AULCHENKO, 2011).

5.3. Dados Fenotípicos

Os Grupos amostrais 1, 2 e 3 constituíram os respectivos Experimentos 1, 2 e 3. Seguem descritas as informações sobre as características e os valores genéticos preditos (EBVs) que foram utilizados como fenótipos nas análises de associação.

5.3.1. Experimento 1

Os dados fenotípicos analisados neste experimento consistiram da característica IPP, que correspondeu à idade em meses em que ocorreu o primeiro

parto da novilha, calculada pela diferença entre a data do parto e sua data de nascimento.

Na **Tabela 2** estão apresentadas as estatísticas descritivas de IPP e na **Fig. supl. 2** pode ser verificada a distribuição desses dados. Além disso, informações detalhadas sobre registro de fenótipos dos rebanhos que integram o programa de precocidade sexual do PMGRN estão disponíveis em <http://www.ancp.org.br/>.

Tabela 2. Experimento 1. Estatística descritiva da Idade ao Primeiro Parto (IPP) dada em meses.

<i>N</i>	<i>Média</i>	<i>Desvio-padrão</i>	<i>Mediana</i>	<i>Min.</i>	<i>Máx.</i>
96	28,40	3,160	27,12	24,11	33,94

5.3.2. Experimento 2

Neste experimento foram utilizados como dados fenotípicos EBVs (fenótipos compostos) para as características IPP e PG de touros da raça Nelore, obtidos a partir da avaliação genética do ano de 2012 do Programa de Melhoramento Genético da Conexão Delta G.

A avaliação genética do presente programa de melhoramento genético foi conduzida a partir de dados de desempenho e pedigree da Base de Dados Aliança Nelore e realizada pela GenSys Consultores Associados S/S (ALIANÇA, 2012). Antes da avaliação, os dados foram submetidos à análise completa de conectabilidade entre grupo de contemporâneos (GCs), medida através das conexões genéticas de cada animal e seus ancestrais comuns, ponderadas pelo parentesco aditivo. Os GCs deveriam apresentar no mínimo dez ligações genéticas diretas para serem conectados à base de dados principal. A estimação dos EBVs foi realizada a partir de procedimentos robustos utilizando como modelo básico o modelo animal multivariado, resumidamente descrito por:

$$y_{ijkl} = [\mu + gci + aj + eijkl]\lambda_m$$

onde y_{ijkl} é a observação fenotípica do animal l pré-ajustada para os efeitos ambientais conhecidos, μ é a média geral da característica, gci é o efeito do grupo contemporâneo i (fixo), aj é o efeito genético direto do animal j (aleatório), $eijkl$ é efeito residual associado à observação $ijkl$ (aleatório) e λ_m é o fator de ajustamento para o nível m da variância residual do grupo contemporâneo i e resíduo da observação $ijkl$.

Os componentes de (co)variâncias foram previamente estimados pelo método de máxima verossimilhança restrita.

No cálculo do EBV para IPP foi considerado um ou mais desafios de acasalamento das filhas dos touros. Além disso, foi aplicada uma penalidade para as fêmeas que não ficaram gestantes ao desafio.

A característica PG é representada pelo número de dias entre a concepção e o nascimento. Seus valores foram preditos com base nos registros repetidos oriundos de inseminação artificial dos descendentes, considerando-o como característica da matriz. Informações detalhadas dos modelos e também do registro dos fenótipos IPP e PG estão disponíveis em <http://www.gensys.com.br/>.

As estatísticas descritivas dos EBVs para as características e suas respectivas acurácias seguem apresentadas na **Tabela 3**, sendo suas distribuições apresentadas na **Fig. supl. 3**. No presente experimento houve seleção dos touros baseado na disponibilidade de EBVs, pois parte dos touros não possuía filhas avaliadas para o fenótipo, o que acarretou na composição dos subgrupos amostrais EBV1 (n=309) e EBV2 (n=527), correspondentemente a IPP e PG.

O número médio de filhas para IPP foi de 46,14 (mín. de 1 e máx. de 697) e de 220,4 para PG (mín. de 1 e máx. de 4.229).

Tabela 3. Experimento 2. Estatística descritiva dos valores genéticos preditos (EBVs) das características Idade ao Primeiro Parto (IPP) e Período Gestacional (PG) em dias.

<i>Valor genético estimado (EBV)</i>	<i>Touros</i>	<i>Média</i>	<i>Desvio-padrão</i>	<i>Min.</i>	<i>Máx.</i>	<i>Acurácia média</i>
EBV1 (IPP)	309	-0,380	5,580	-48,06	17,00	0,739
EBV2 (PG)	527	0,045	2,406	-15,95	7,91	0,862

5.3.3. Experimento 3

Neste experimento também foram utilizados como fenótipos os EBVs para IPP e PG (EBV1 e EBV2, respectivamente) preditos para as matrizes na avaliação genética de 2012 do Programa de Melhoramento Genético da Conexão Delta G. Informações sobre as características e os aspectos metodológicos seguem conforme descrição anterior (**Item 5.3.2.**).

A **Tabela 4** apresenta as estatísticas descritivas dos EBVs e suas acurácias, sendo estas distribuições apresentadas na **Fig. supl. 4**. A seleção de fêmeas foi baseada na disponibilidade de EBVs calculados para as características, o que levou à formação dos subgrupos amostrais: EBV1 (n=1.076) e EBV2 (n=1.221).

Os EBVs das fêmeas foram preditos a partir do número médio de 1 filho para IPP (mín. de 0 e máx. de 7) e de 3,39 para PG (mín. de 0 e máx. de 11).

Tabela 4. Experimento 3. Estatística descritiva dos valores genéticos preditos (EBVs) das características Idade ao Primeiro Parto (IPP) e Período Gestacional (PG) em dias.

<i>Valor genético estimado (EBV)</i>	<i>Fêmeas</i>	<i>Média</i>	<i>Desvio-padrão</i>	<i>Min.</i>	<i>Máx.</i>	<i>Acurácia média</i>
EBV1 (IPP)	1.076	-2,481	4,462	-18,21	15,05	0,465
EBV2 (PG)	1.221	-0,206	1,715	-6,12	5,35	0,607

5.4. Controle de Qualidade dos Genótipos e Estratificação Populacional

Antes que as análises de associação fossem conduzidas, os dados genotípicos de cada grupo e/ou subgrupo amostral foram submetidos aos mesmos critérios de QC. Desse modo, foi realizada uma filtragem inicial com remoção de todos os marcadores não autossômicos, ou seja, os SNP dos cromossomos X, Y e DNA mitocondrial. Nessa etapa também foram identificados e removidos os SNP mapeados para a mesma posição genômica.

Com o objetivo de remover das análises os genótipos e amostras com baixa qualidade ou pouco informativos, os seguintes critérios de inclusão foram estabelecidos para as análises de QC: “call rate” > 98%, MAF > 2% e $p > 10^{-5}$ no Teste Exato de Fisher para HWE (WANG et al., 2005; THE WELLCOME TRUST CASE CONTROL CONSORTIUM, 2007; ZIEGLER et al., 2008; PARE, 2010). Amostras com mais de 10% de dados faltantes e com mais de 95% de identidade por estado (IBS) com outra amostra foram excluídas do respectivo conjunto de dados (ZIEGLER et al., 2008; PAUSCH et al., 2012). A análise de IBS é fundamentada na porcentagem de alelos compartilhados para cada par de indivíduos analisados com o objetivo de identificar amostras duplicadas. No presente estudo, a análise de IBS de cada conjunto de dados foi computada a partir de um subconjunto de 2.000 marcadores aleatórios.

A análise de escalonamento multidimensional (MDS), também denominada de análise de coordenadas principais (PCoA), foi a ferramenta de visualização utilizada para verificar a homogeneidade genética das populações. Todos os pares amostrais de cada grupo foram analisados com base na matriz IBS de parentesco genômico, estimada para um subgrupo de 100.000 marcadores aleatórios remanescentes das análises de QC, com o objetivo de reduzir o tempo computacional.

Para estimar o parentesco entre os pares amostrais de cada grupo, a análise de identidade por descendência (IBD) foi conduzida na dependência da informação de IBS e das frequências alélicas. Esta análise foi implementada no programa Plink, onde sua estimação seguiu as probabilidades de que 0, 1 e 2 alelos de cada marcador fossem idênticos por descendência, empregando as seguintes denotações: $P(Z=0)$, $P(Z=1)$ e $P(Z=2)$, respectivamente (PURCELL et al., 2007). A

proporção de alelos idênticos por descendência para cada par amostral, a partir dos marcadores que passaram pelo QC, foi estimada por:

$$\hat{\pi} = \frac{P(Z = 1)}{2} + P(Z = 2)$$

Para a visualização geral do parentesco presente em cada grupo amostral foram plotados "heatmaps" (mapas de calor) dos valores de IBD estimados, onde as amostras foram agrupadas de acordo com a quantidade de alelos compartilhados.

5.5. Análise de Associação Genômica Ampla

Análises do tipo GWAS independentes foram realizadas a partir do grupo (Experimento 1) ou subgrupos amostrais (Experimentos 2 e 3). A associação entre cada valor fenotípico e cada SNP foi conduzida em três etapas. Na primeira etapa, foi realizada uma análise de regressão linear múltipla utilizando o método dos quadrados mínimos ponderados, onde touro (pai) e GCs foram as variáveis preditoras no Experimento 1. Na formação do GCs foi considerado o ano e trimestre de nascimento dos animais (**Fig. supl. 5**). O modelo de regressão adotado para os Experimentos 2 e 3 foi ponderado para o quadrado da acurácia do EBV, ou seja, sua confiabilidade, com o objetivo de considerar a incerteza sobre sua estimativa. Nestas análises também foram incluídos os autovetores significativos da análise de MDS, ou seja, aqueles correlacionados com a variável resposta na correlação de Pearson ($p < 0,05$). Essa abordagem teve como objetivo controlar os potenciais fatores de confundimento por estratificação dos subgrupos amostrais. O modelo geral adotado pode ser descrito por:

$$y_i = \mu + \sum \beta_j X_{ij} + e_i \quad [1]$$

onde, y_i é fenótipo do indivíduo i , μ é a média geral, β é o efeito da variável j , X_{ij} é o valor do indivíduo i para a variável j e e_i é o resíduo do indivíduo i .

Na segunda etapa foram obtidos os resíduos estimados em [1] a partir de:

$$y_i^* = y_i - \hat{y}_i \quad [2]$$

Na última etapa foi conduzido teste de escore sobre os resíduos obtidos em [2]. Essa análise foi implementada em R utilizando a função *qtscore* da biblioteca GenABEL (AMIN et al., 2007; AULCHENKO et al., 2007). Esta análise foi realizada em dois passos:

$$R = \frac{Cov(y^*, x)}{\sqrt{Var(x)Var(y^*)}} \quad [A]$$

onde, x é o SNP e [A] representa a correlação de Pearson (R).

$$T^2 = N(R^2) \sim \chi_{1gl}^2 \quad [B]$$

onde, N é o número de indivíduos para o SNP.

A confiabilidade dos resultados foi verificada pela estimação do fator de inflação lambda (λ) pela regressão da característica sobre N loci:

$$\lambda = Median(T_1^2 + T_2^2, \dots, T_N^2) / 0,456 \quad [C]$$

onde, 0,456 é a mediana da distribuição χ_{1gl}^2 .

Se $1 < \lambda < 1.1$, então a inflação foi considerada aceitável e o método de ajustamento Genomic Control (GC) foi aplicado para corrigir o teste estatístico χ^2

pela divisão de todos os valores de T^2 observados pelo valor de λ estimado (DEVLIN e ROEDER, 1999).

Gráficos Quantil-Quantil (Q-Q) foram inspecionados para avaliar a qualidade dos testes aplicados, julgando a extensão dos sinais falso-positivos. Para cada modelo empregado foi visualizada a distribuição dos valores observados *versus* a distribuição dos valores esperados da hipótese nula da distribuição χ^2 .

Foi utilizado o método de Bonferroni para correção de testes múltiplos, onde α é definido por $0,05/N$, sendo que N é o número de marcadores SNP testados independentemente (BENJAMIN e HOCHBERG, 1995).

Considerando intervalo de confiança de 95% (IC), estimou-se o efeito de substituição alélica para todos os marcadores ($\hat{\beta}$) e o erro da estimativa (SE), como segue:

$$\hat{\beta} = \frac{Cov(y, x)}{Var(x)} \quad SE = \sqrt{\frac{\hat{\beta}^2}{T^2}}$$

considerando, $IC95\% = \hat{\beta} \pm t_{n-1, \alpha/2} SE$.

Os resultados dos testes foram apresentados separadamente em gráficos do tipo “Manhattan” (“Manhattan Plot”), sendo que no primeiro gráfico foi plotado o co-logaritmo dos valores de p corrigidos por GC e ajustados para testes múltiplos (Método de Bonferroni) e, no segundo, foi plotado o efeito de substituição alélica estimado de cada marcador para o fenótipo, ambos os gráficos contra a posição genômica do SNP.

5.6. Estratégias para a Identificação de Genes Candidatos

Com base na distribuição de efeitos de substituição alélica estimados para cada marcador, foram considerados SNP de maior efeito sobre as características estudadas (SNP “top”) aqueles com valor de β com dispersão da média superior a + 6 ou inferior a - 6 unidades de desvios-padrão, considerando-se de interesse tanto marcadores com influência positiva quanto negativa. A lista de SNP “top” de cada Experimento foi utilizada para a prospecção de genes candidatos aplicando estratégia constituída de duas etapas: análise sistematizada e análise manual.

A análise sistematizada constitui-se de dois procedimentos distintos. O primeiro compreendeu a busca de SNP “top” intragênicos, considerando-se como definição de domínio gênico da ontologia de sequências do banco de dados Ensembl (FLICEK et al., 2013). Desta forma, genes contendo SNP “top” localizados nos limites de 5 kb de suas regiões UTR 5' e 3' foram considerados candidatos. Esta análise foi realizada utilizando-se o banco de dados Ensembl Variation 69, através da ferramenta BioMart (KINSELLA et al., 2011). Para cada caso, foram obtidas informações sobre o posicionamento do SNP em relação ao gene (5 kb upstream/downstream, região UTR 5'/3', exon, íntron), potencial consequência da variação ao transcrito ou proteína, nome, descrição e termos de ontologia do gene.

O segundo procedimento envolveu o mapeamento de genes potencialmente em LD com SNP “top” e a construção de redes de termos funcionais. Neste caso, foi calculado o decaimento de LD em torno de cada SNP “top” por meio da correlação pareada (r^2) entre este e os marcadores vizinhos. Os cálculos foram realizados até os últimos marcadores em posições proximais e distais do cromossomo que exibiam $r^2 > 0,7$ ou distância menor que 500 kb do SNP “top”. O segmento cromossômico delimitado pelas posições destes últimos marcadores foi considerado como um único *locus*, e qualquer gene sobrepondo a suas posições foi considerado em LD com o SNP “top”.

Os genes potencialmente em LD com SNP “top” foram submetidos ao "Database for Annotation, Visualization and Integrated Discovery" (DAVID) v.6.7, para a anotação de termos funcionais, tendo como fonte os bancos de dados: "Gene Ontology" (GO) (categorias: componente celular, processo biológico e função

molecular), KEGG Pathway ("Kyoto Encyclopedia of Genes and Genomes") e InterPro (HUANG et al., 2009a; HUANG et al., 2009b). Redes de termos inter-relacionados foram construídas baseadas no número de genes compartilhados entre termos. Essa etapa foi conduzida com o auxílio do programa Cytoscape, utilizando o "plug-in Enrichment Map Cytoscape" (MERICO et al., 2010). Os termos foram definidos como nodos, sendo que as arestas que ligam os nodos representaram o compartilhamento de genes, e sua espessura o número de genes sobrepostos.

Os resultados das análises sistematizadas foram posteriormente revisados manualmente com o intuito de buscar evidências da participação de genes ou funções apontados pela análise prévia em processos reprodutivos. Para tanto foi acessada a literatura científica disponível, principalmente de artigos publicados em periódicos indexados no PubMed ou ScienceDirect. Finalmente, genes apoiados por múltiplas evidências foram destacados e preferencialmente discutidos.

6. Resultados

6.1. Experimento 1

6.1.1. Controle de Qualidade e Estratificação Populacional

O "call rate" médio dos marcadores antes do QC foi de 94% e o número de SNP excluídos de acordo o critério de filtragem aplicado pode ser verificado na **Tab. supl. 1**. As análises de QC removeram o total de 276.762 SNP do conjunto de dados, sendo que o critério MAF foi o responsável pelo maior número de marcadores removidos (**Tab.supl. 1 e Fig. supl. 6**).

Com base nos critérios amostrais nenhum animal foi excluído das análises, pois todos apresentaram "call rate" acima de 98% e não foram verificados indivíduos duplicados. Embora não tenha sido um critério de qualidade para exclusão de amostras, foram observadas as taxas de heterozigosidade amostral, antes e após o QC, sendo os respectivos valores aproximados de 0,22 ($\pm 0,006$) e 0,33 ($\pm 0,007$).

A proporção de IBD entre dois indivíduos pode ser utilizada para estimar em que proporção duas amostras são relacionadas, ou seja, quais as proporções

esperadas para os graus de parentesco no conjunto de dados. Neste estudo foi observada a proporção média IBD de 0,13 ($\pm 0,063$), ou seja, média de terceiro grau de parentesco amostral. O extenso agrupamento de indivíduos aparentados pode ser verificado no “heatmap” dos coeficientes de parentesco (**Figura 3**). As distribuições das proporções de IBD e também dos valores de IBS estão descritas na **Fig. supl. 7**.

Baseado na inspeção visual, as análises de MDS sugeriram estratificação populacional vinculada ao parentesco amostral (**Figura 4**).

O total de marcadores autossômicos e de alta qualidade foi de 466.093, correspondendo a aproximadamente 59,2% do total disponibilizado no ensaio empregado para análise. As informações dos SNP por cromossomo e dos dados de densidade do painel estão sumarizadas, correspondentemente, nas **Tab. supls. 2 e 3**.

6.1.2. Análise de Associação Genômica Ampla

Para a análise de associação foram selecionados 96 animais com genótipos para 466.093 SNP, distribuídos nos 29 autossomos (BTAs). O número de marcadores variou de 8.052 no cromossomo BTA25 a 29.768 no cromossomo BTA1 (**Tab. supl. 2**).

Os resíduos obtidos da regressão linear sobre as variáveis preditoras atenderam as pressuposições de normalidade (Teste Shapiro-Wilk) e homocedasticidade (Teste Breusch-Pagan Estudantizado) e, portanto, os mesmos foram utilizados como variável dependente nas análises de associação (**Fig. supl. 8**).

Uma divergência significativa entre as linhas de regressão da distribuição esperada e da observada foi constatada para os marcadores testados. O valor de lambda foi estimado em aproximadamente 0,84, evidenciando deflação generalizada do teste, de modo que o modelo adotado não removeu suficientemente os efeitos de fatores confundidores na análise (**Figura 5**).

A **Figura 6** mostra o "Manhattan Plot" do co-logaritmo dos valores de significância do teste de associação corrigidos por GC. Nenhuma associação significativa entre os SNP e IPP foi detectada ($\alpha = 1,07 \times 10^{-7}$).

O "Manhattan Plot" do efeito de substituição alélica estimado para cada marcador testado e sua estatística descritiva podem ser verificadas na **Figura 6** e na **Tab. supl. 4.**, respectivamente. As médias do efeito estimado em meses e de seu correspondente erro-padrão foram de, aproximadamente, 0,016 ($\pm 0,552$) e 0,552 ($\pm 0,233$), respectivamente (**Figura 7**).

A partir desses resultados foram selecionados 96 SNP "top" para as estratégias de identificação de genes candidatos, ou seja, os SNP com os maiores efeitos de substituição alélica estimados para IPP. Esses SNP estavam posicionados ao longo de 18 dentre os 29 cromossomos, com maior representatividade de SNP do cromossomo BTA10 (**Tab. supl. 5**).

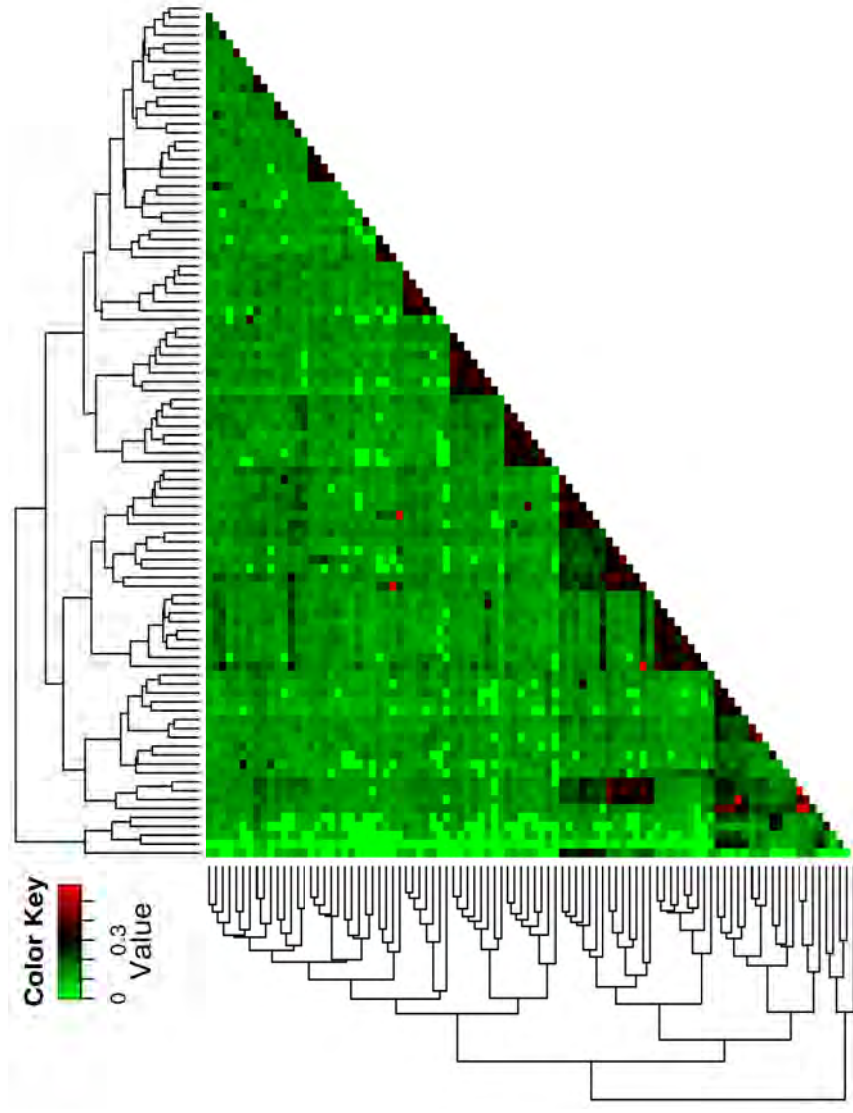


Figura 3. Experimento 1. “Heatmap” da proporção de alelos idênticos por descendência (IBD) entre os pares amostrais. Valores de 0 (verde - amostras completamente não relacionadas) a 0,5 (vermelho - compartilhamento de metade dos alelos).

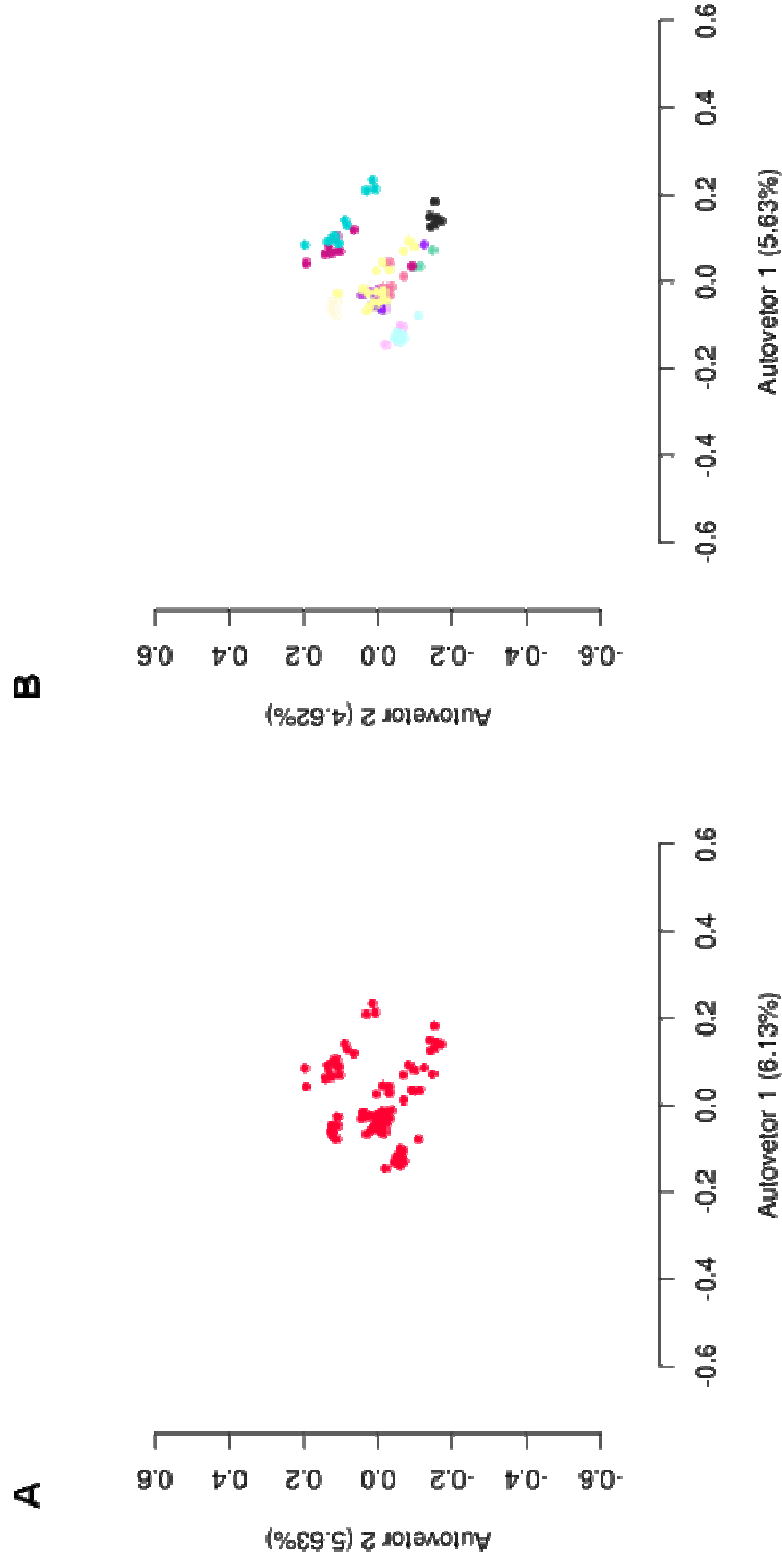


Figura 4. Experimento 1. Análise de escalonamento multidimensional (MDS) baseada no coeficiente IBS de parentesco genômico. (B) Os animais foram coloridos de acordo com o parentesco amostral (animais filhos de um mesmo pai e pontos amarelos indicam filhos de touros menos representativos). As porcentagens correspondem à variância explicada pelo respectivo autovetor.

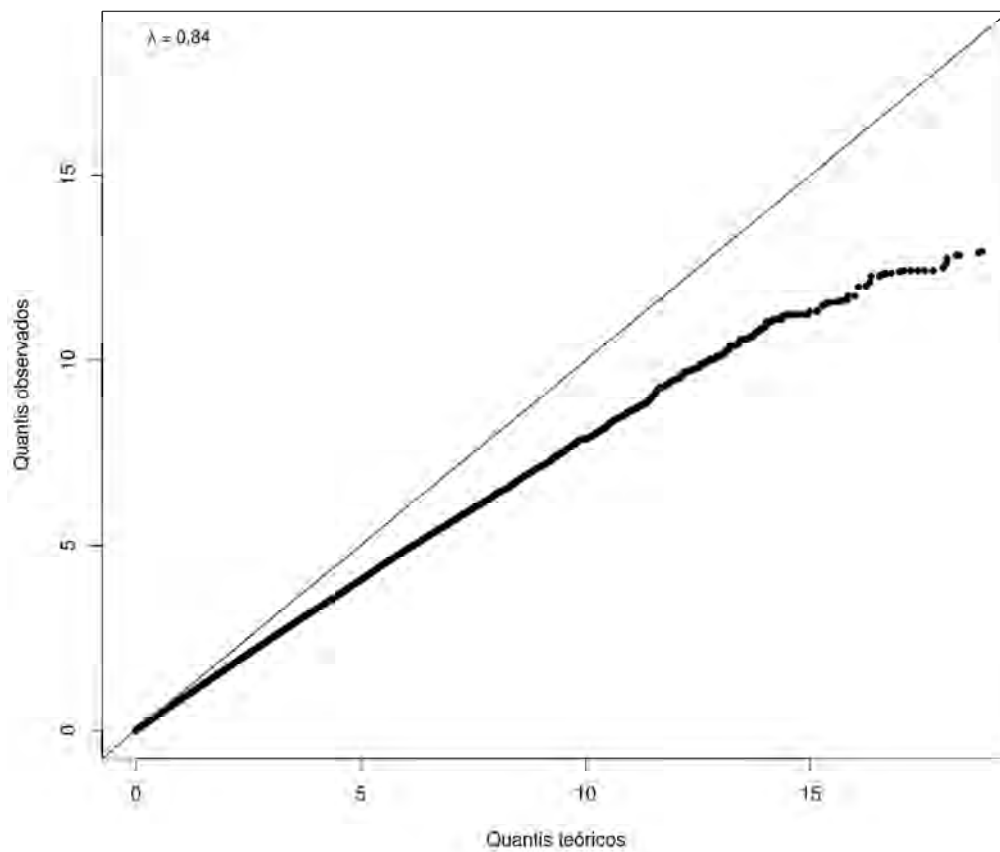
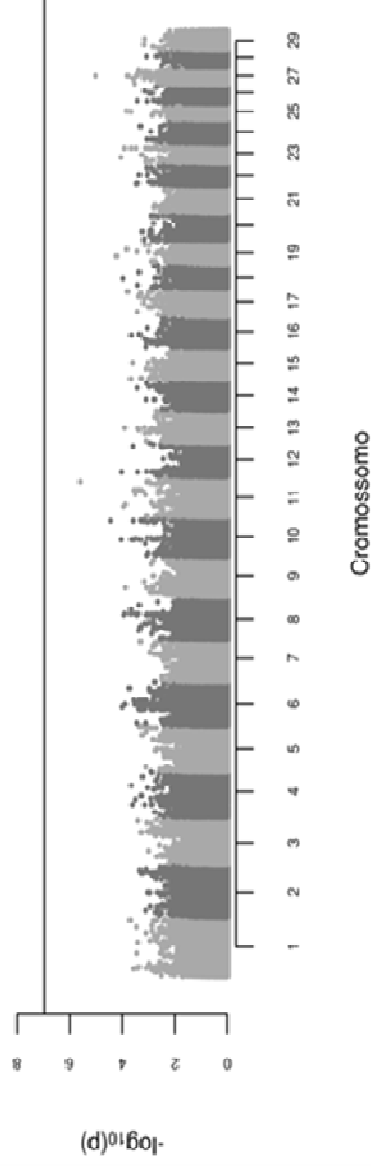


Figura 5. Experimento 1. Gráfico de Quantil-Quantil (Q-Q) dos resultados do GWAS para IPP. A linha preta representa a distribuição esperada e os pontos pretos representam a distribuição observada da estatística χ^2 .

A



B

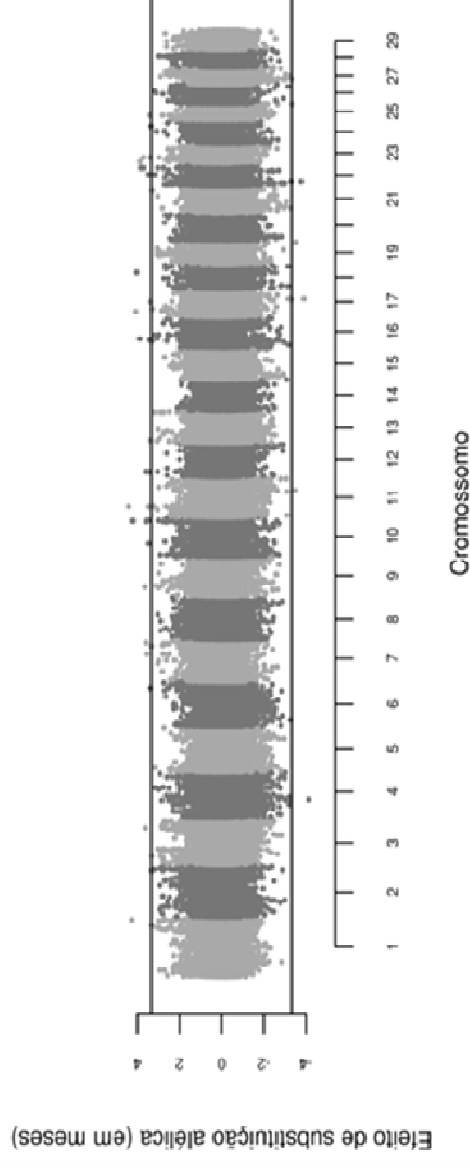


Figura 6. Experimento 1 (IPP). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,07 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.

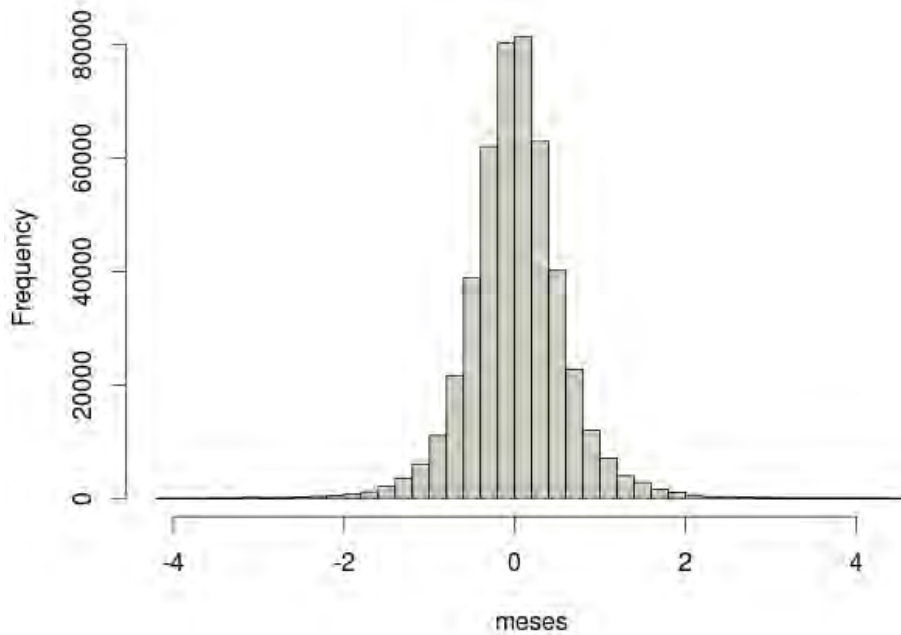


Figura 7. Experimento 1. Distribuição dos efeitos estimados de substituição alélica das variantes SNP para IPP em meses.

6.1.3. Identificação de Genes Candidatos

Foram identificados 38 SNP intragênicos dentre os 96 SNP "top" selecionados das análises de GWAS para IPP (**Tab. supl. 7**). Estes SNP foram localizados principalmente nos genes *SEL1L* (ENSBTAG00000008083), *BT.50439* (ENSBTAG00000020614) e *EPSTI1* (ENSBTAG00000019054), nos cromossomos BTA10, BTA11 e BTA12, respectivamente (**Tabela 5**).

Para a construção das redes de termos inter-relacionados foram utilizadas somente as informações de 53 dentre os 91 genes mapeados (**Tab. supl. 6**), gerando nove redes, contendo 151 nodos que representaram os termos funcionais conectados por 707 arestas que representaram o compartilhamento dos genes. Esses resultados estão resumidamente ilustrados na **Figura 8**. As redes foram relacionadas aos processos biológicos de morte celular programada e reparo de DNA e às funções moleculares de fosforilação e ativação/proliferação de células de defesa (leucócitos e linfócitos T). Algumas redes também evidenciaram as vias metabólicas do fósforo e glicosaminoglicanas. Os genes que se apresentaram

altamente compartilhados dentro da rede de glicosaminoglicanas foram: *BT.18317* (ENSBTAG00000007843), *ITIH3* (ENSBTAG00000007846) e *ITIH4* (ENSBTAG00000007850) todos no cromossomo BTA22; além do *BT.32313* (ENSBTAG00000020898) no cromossomo BTA25 (**Tabela 5**).

A atividade de endoproteases e de membros da família proteína quinase, mais especificamente da serina-treonina quinase, também foi ressaltada no compartilhamento dos genes *PRKCD* (ENSBTAG00000008719) no cromossomo BTA22 e *BT.50439* (ENSBTAG00000020614) no cromossomo BTA11 (**Tabela 5**).

Tabela 5. Experimento 1. Potenciais genes candidatos para IPP.

<i>Ensembl ID</i>	<i>BTA</i>	<i>Posição</i>	<i>Gene ID</i>	<i>Nome</i>	<i>Descrição</i>	<i>Termo(s) GO</i>
ENSBTAG00000008083	10	93.904,711 - 93.965,166	614959	<i>SEL1L</i>	sel-1 suppressor of lin-12-like (<i>C. elegans</i>)	proteína de ligação
*ENSBTAG00000020614	11	27.935,104 - 28.472,632	507041	<i>BT.50439 (PRKCE)</i>	protein kinase C epsilon type	transdução de sinal intracelular
*ENSBTAG00000008719	22	48.307,244 - 48.337,780	505708	<i>PRKCD</i>	protein kinase C delta type	transdução de sinal intracelular
ENSBTAG000000019054	12	13.063,081 - 13.157,473	614555	<i>EPSTI1</i>	não caracterizado	-
*ENSBTAG000000007843	22	48.655,027 - 48.669,271	508356	<i>BT.18317 (ITIHI1)</i>	inter-alpha-trypsin inhibitor heavy chain H1 precursor	metabolismo do ácido hialurônico
*ENSBTAG000000007846	22	48.633,831 - 48.647,777	508355	<i>ITIHI3</i>	inter-alpha-trypsin inhibitor heavy chain H3	metabolismo do ácido hialurônico
*ENSBTAG000000007850	22	48.615,397 - 48.630,387	513700	<i>ITIHI4</i>	inter-alpha-trypsin inhibitor heavy chain H4 precursor	metabolismo do ácido hialurônico
ENSBTAG000000020898	25	41.372,194 - 41.390,545	513030	<i>BT.32313</i>	carbohydrate sulfotransferase 12	resposta imune ligação de carboidrato

* genes relacionados com processos fisiológicos reprodutivos de acordo com a literatura científica disponível.

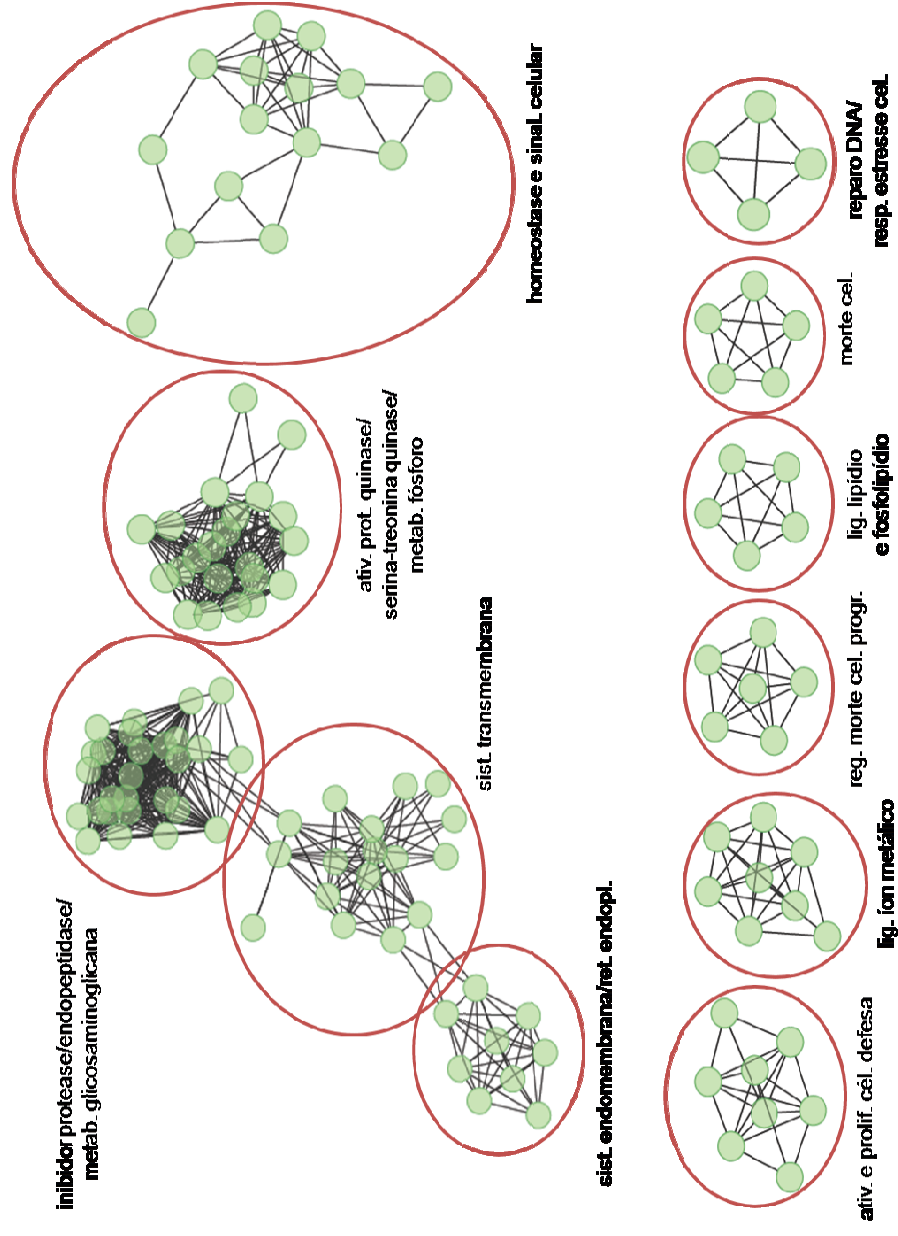


Figura 8. Experimento 1. Redes de termos funcionais para IPP. Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.

6.2. Experimento 2

6.2.1. Controle de Qualidade e Estratificação Populacional

O total de SNP removidos pelo QC foi de 307.468 e 309.067 para os subgrupos EBV1 e EBV2, respectivamente. A distribuição da MAF dos marcadores antes do QC e a descrição da quantidade de marcadores excluída em cada etapa do QC estão, respectivamente, na **Fig. supl. 9** e **Tab. supl. 1**.

Como a presença de amostras duplicadas no conjunto de dados pode introduzir resultados falso-positivos, 22 animais foram identificados e removidos com base na análise de IBS ($> 95\%$). Após essa filtragem, a média de IBS computada para todas as amostras foi de $0,76 (\pm 0,011)$ (**Fig. supl. 7**).

O baixo "call rate" ($< 90\%$) foi responsável pela exclusão de nove amostras, quatro do subgrupo EBV1 e cinco do EBV2. O "call rate" médio das amostras remanescentes de ambos os subgrupos foi de 99% . Além disso, as taxas de heterozigosidade para todas as amostras foram de $0,22 (\pm 0,011)$ e $0,29 (\pm 0,012)$ antes e após as análises de QC, respectivamente.

A proporção média de IBD foi estimada em $0,05 (\pm 0,065)$ (**Fig. supl. 7** e **Figura 9**). A análise de MDS deste grupo amostral também sugeriu estratificação populacional, o que influenciou nas análises de associação (**Figura 10**).

Contudo, o número de SNP que passou no QC foi de 427.770 e 426.171 para 299 e 510 amostras, dos respectivos subgrupos EBV1 e EBV2. Nas **Tab. supls. 2** e **3** estão resumidas as informações sobre a distribuição dos marcadores em cada cromossomo e dados de densidade do painel, respectivamente.

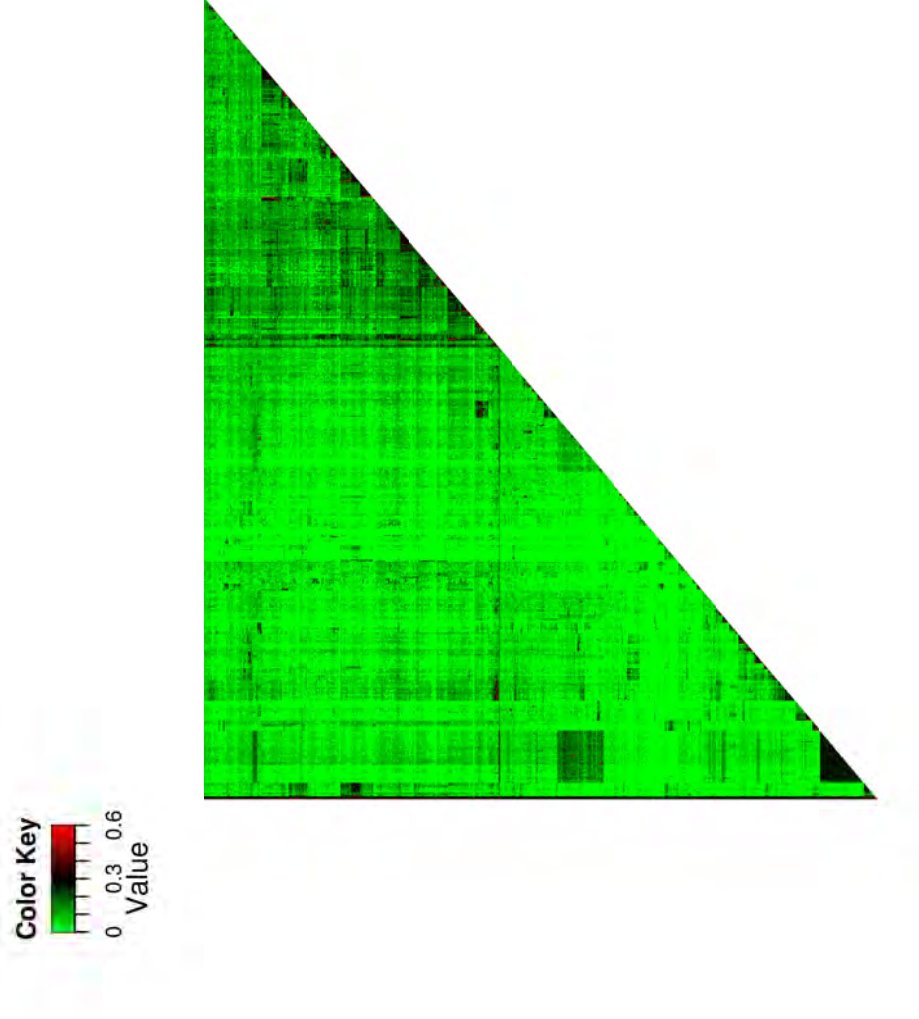


Figura 9. Experimento 2. “Heatmap” da proporção de alelos idênticos por descendência (IBD) entre os pares amostrais. Valores de 0 (verde - amostras completamente não relacionadas) a 0,5 (vermelho - compartilhamento de metade dos alelos).

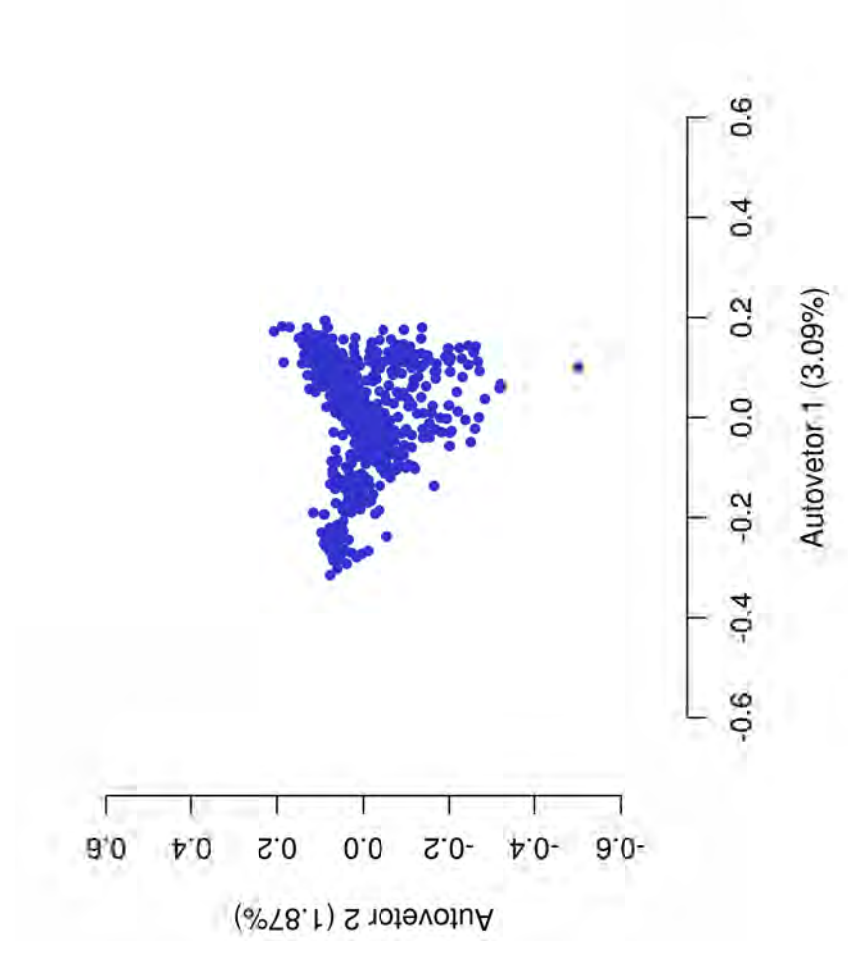


Figura 10. Experimento 2. Análise de escalonamento multidimensional (MDS) baseada no coeficiente de parentesco genômico. As porcentagens correspondem à variância explicada pelo respectivo autovetor.

6.2.2. Análise de Associação Genômica Ampla

Após o QC, permaneceram nas análises 299 (EBV1) e 510 (EBV2) amostras. Entretanto, foram excluídas duas amostras marginais de cada subgrupo com base nas distribuições dos EBVs (± 4 unidades de desvio-padrão) (**Fig. supl. 3**). Assim, as análises do tipo GWAS foram conduzidas separadamente para os correspondentes subgrupos, a partir de 427.770 e 246.171 SNP e 297 e 508 touros, respectivamente.

Um total de 13 e 25 autovetores das análises de MDS foram correlacionados significativamente com os fenótipos e explicaram 15,23 e 12,23% da variabilidade dos dados genotípicos dos respectivos subgrupos EBV1 e EBV2. Os resíduos obtidos da regressão linear sobre os autovetores significativos ponderada para a confiabilidade dos EBVs atenderam às pressuposições de normalidade (Teste Anderson-Darling) e homocedasticidade (Teste Breusch-Pagan Estudantizado) e, desse modo, os mesmos foram utilizados como variável dependente nas análises de associação (**Fig. supl. 11**).

Os gráficos Q-Q mostraram que os desvios dos testes estatísticos observados a partir dos quantis teóricos foram leves e aceitáveis ($\lambda = 1,06$ e $1,03$, respectivamente). Assim, os valores de T^2 que desviaram dos valores esperados foram interpretados como SNP potencialmente associados aos respectivos fenótipos (**Figura 11**).

Os resultados da associação genômica para os valores de significância corrigidos por GC estão apresentados nas **Figuras 12 e 13**. Nenhuma associação significativa entre SNP e o EBV testado foi detectada em ambas as análises do tipo GWAS (EBV1, $\alpha = 1,16 \times 10^{-7}$; EBV2, $\alpha = 1,17 \times 10^{-7}$).

Os resultados dos testes de associação para o efeito de substituição alélica estimado do SNP estão apresentados nas **Figuras 12 e 13** e na **Tab. supl. 4**. Os efeitos médios estimados e seus respectivos erros-padrão, foram de -0,012 dias ($\pm 0,499$) e 0,443 dias ($\pm 0,180$) para EBV1 e de 0,011 dias ($\pm 0,197$) e 0,180 dias ($\pm 0,074$) para EBV2 (**Figura 14**).

No total, 114 e 38 SNP “top” dos respectivos subgrupos EBV1 e EBV2 foram selecionados para as análises de identificação de genes candidatos. Para ambas as

características, os SNP relevantes estavam distribuídos em diversos cromossomos, com maior número de marcadores no cromossomo BTA26 nas análises do subgrupo EBV1 (**Tab. supl. 5**).

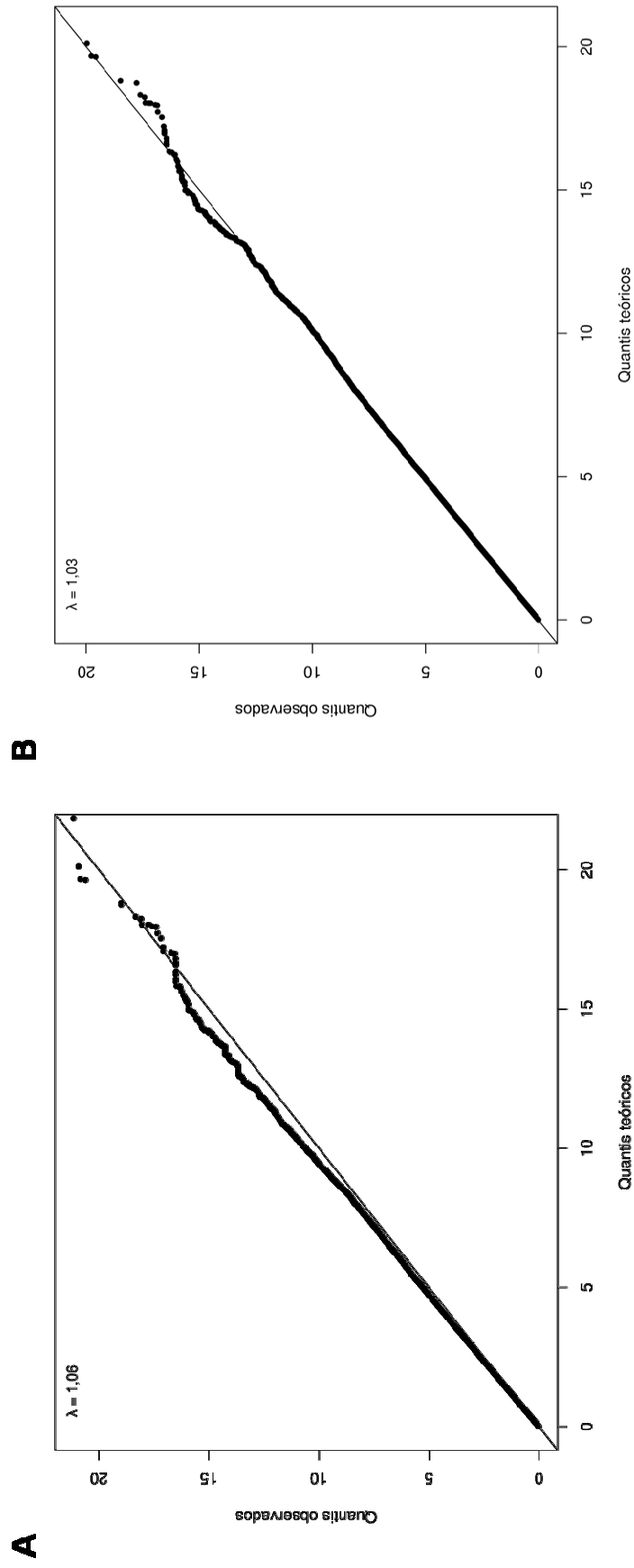


Figura 11. Experimento 2. Gráficos de Quantil-Quantil (Q-Q) dos resultados dos GWAS para EBV1(A) e EBV2 (B). A linha preta representa a distribuição esperada e os pontos pretos representam a distribuição observada da estatística χ^2 .

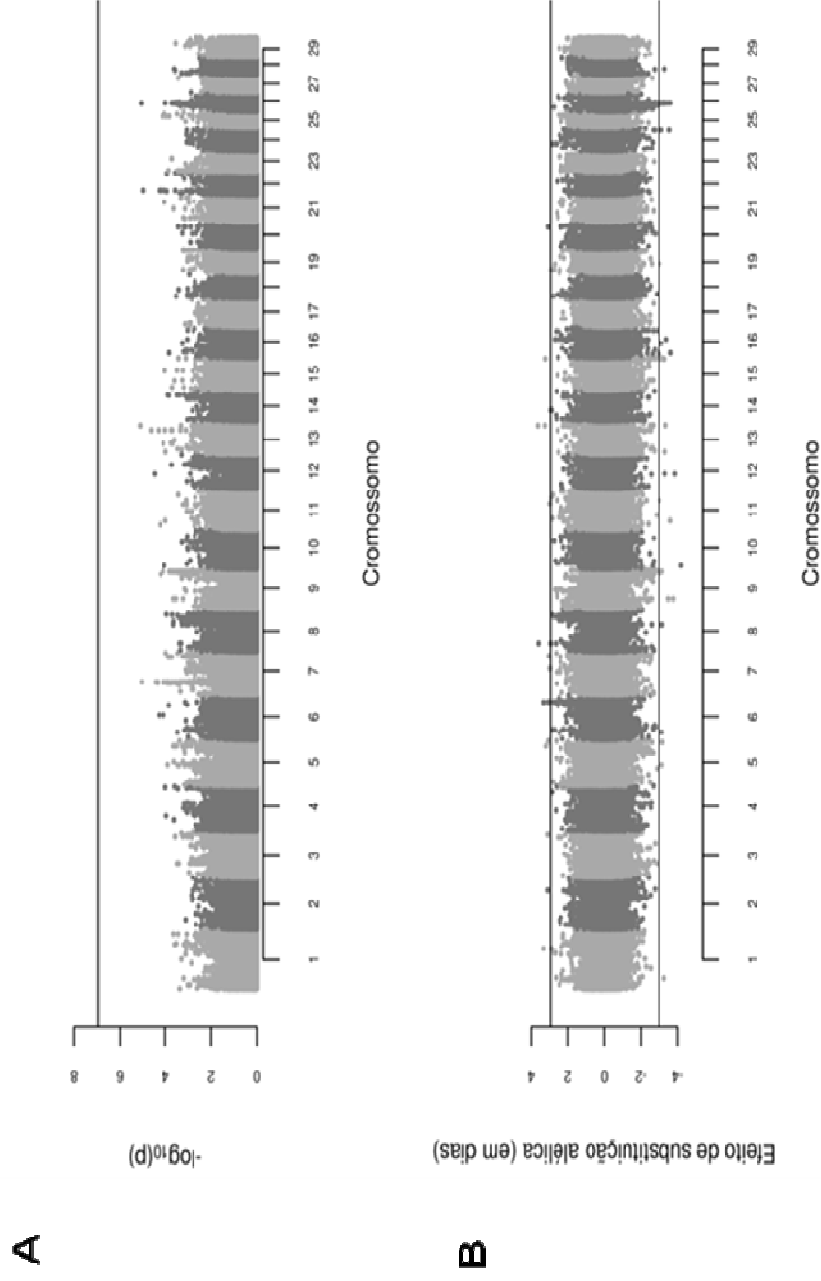


Figura 12. Experimento 2 – EBV1 (IPP). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,16 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.

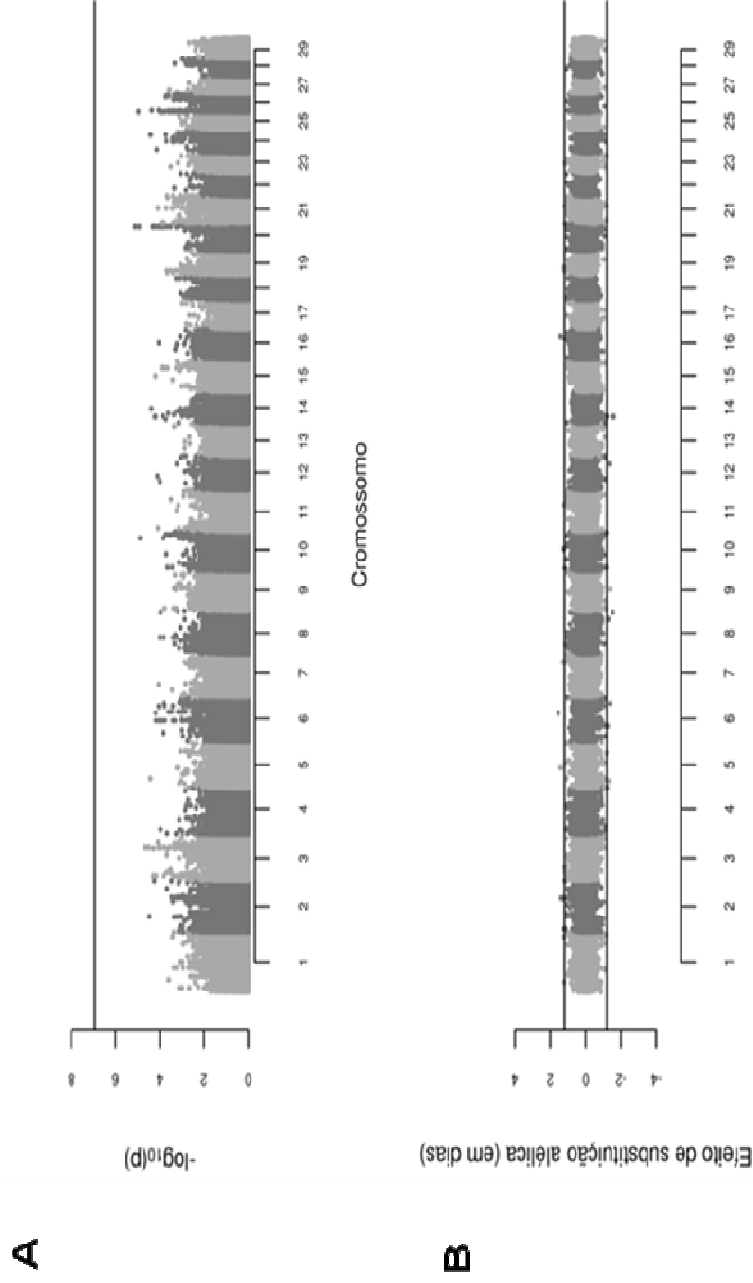


Figura 13. Experimento 2 – EBV2 (PG). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,17 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.

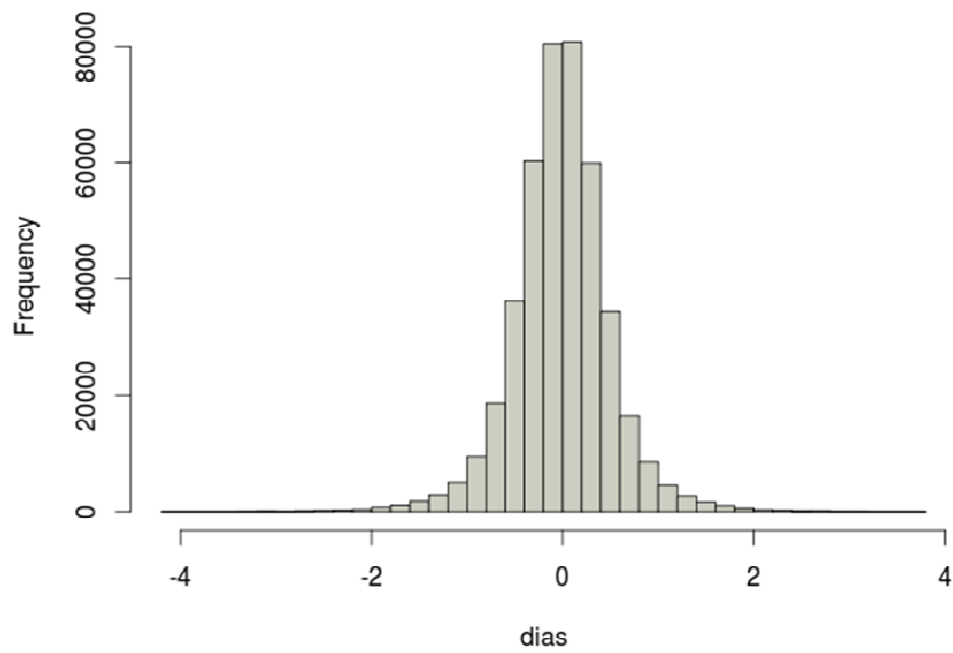
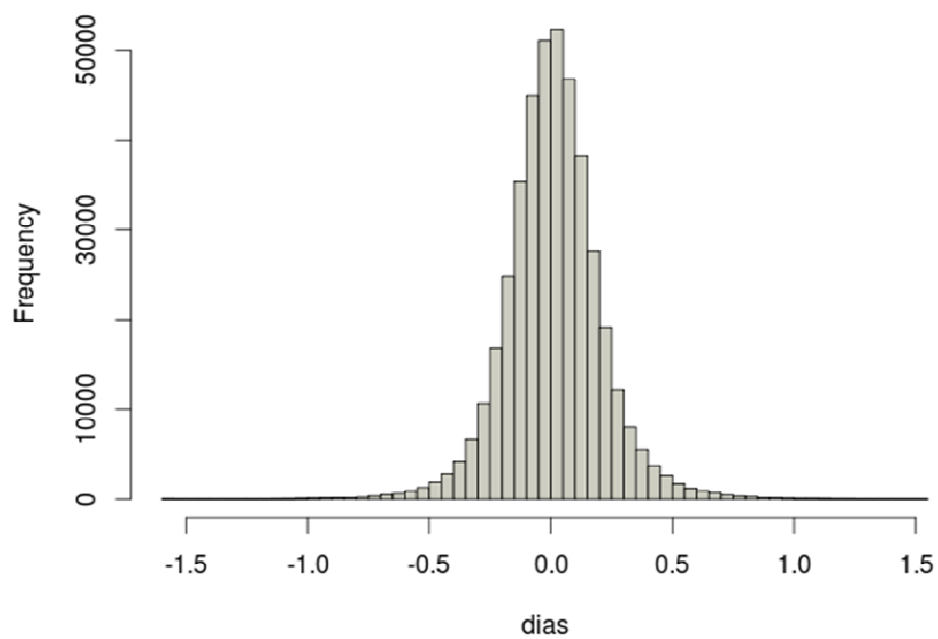
A**B**

Figura 14. Experimento 2. Distribuição dos efeitos estimados de substituição alélica das variantes SNP para EBV1 (A) e EBV2 (B) em dias.

6.2.3. Identificação de Genes Candidatos

Dentre os SNP "top" selecionados das análises de GWAS para EBV1, 73 foram localizados em regiões intragênicas (**Tab. supl. 7**). Genes possuindo SNP intrônicos se destacaram, tais como: *BT.38812* (ENSBTAG000000030189) no cromossomo BTA5; *LOXL2* (ENSBTAG000000006214) no cromossomo BTA8; *C10orf76* (ENSBTAG000000013120) e *BT.25271* (ENSBTAG000000007476), ambos no cromossomo BTA26. Também se enquadrando nesses aspectos duas variantes SNP foram sinônimas e localizadas nos genes *BT.62377* (ENSBTAG000000014806) no cromossomo BTA19 e *LDB1* (ENSBTAG000000005780) no cromossomo BTA26, respectivamente. Além disso, foram identificadas outras três variantes (intrônica, "downstream" e não-sinônima) no gene *FAM111A* (ENSBTAG000000002450) no cromossomo BTA15 (**Tabela 6 e Tab. supl. 8**).

A partir de 63 genes mapeados para EBV1 foram identificados termos e construídas cinco redes funcionais, que juntas apresentaram 148 nodos conectados por 528 arestas (**Tab. supl. 6 e Figura 15**). As redes foram correlacionadas aos componentes celulares: mitocôndrias, citoesqueleto e microtúbulos e, principalmente, às funções de glicosilação, fosforilação e metabolismo do fósforo; transporte e ligação de íons metálicos (cálcio e magnésio) e ligação de ATP, nucleosídeos e ribo(nucleotídeos). Além disso, algumas redes destacaram a atividade de peptidases e das proteínas: quinase, WAP com domínio 4-dissulfeto e de proteínas com domínios WD40 e FBOX.

No conjunto de SNP "top" de EBV2, apenas 10 dentre 37 SNP "top" foram variantes intragênicas (**Tab. supl. 7**). Dentre estas, três variantes foram posicionadas em dois genes codificadores de microRNA (miRNA): uma "upstream" no gene ENSBTAG000000046504 no cromossomo BTA9 e duas ("upstream" e "downstream") no gene ENSBTAG000000045178 no cromossomo BTA14 (**Tabela 7**).

Além disso, três redes de termos foram compostas para EBV2 contendo 16 nodos e 37 arestas, sendo sua construção fundamentada em apenas 13 genes (**Figura 16**). Essas redes foram associadas à transcrição gênica e ligação de íons metálicos, baseadas principalmente no compartilhamento do gene *BT.88023* (ENSBTAG000000026309) no cromossomo BTA14 (**Tabela 7**).

Tabela 6. Experimento 2. Potenciais genes candidatos para EBV1 (IPP).

<i>Ensembl ID</i>	<i>BTA</i>	<i>Posição</i>	<i>Gene ID</i>	<i>Nome</i>	<i>Descrição</i>	<i>Termo(s) GO</i>
*ENSBTAG000000030189	5	115.127,019 - 115.419,183	-	<i>BT.38812 (EFCAB6)</i>	não caracterizado	ligação de íon cálcio
ENSBTAG000000006214	8	71.278,583 - 71.349,786	532684	<i>LOXL2</i>	Lysyl oxidase homolog 2	transcrição celular
ENSBTAG000000013120	26	22.446,522 - 22.608,433	100336729	<i>C10orf76</i>	não caracterizado	proteína de ligação
ENSBTAG000000007476	26	22.001,775 - 22.172,725	534072	<i>BT.25271</i>	F-box/WD repeat-containing protein 1A	transcrição celular
*ENSBTAG000000014806	19	25.163,841 - 25.198,117	512313	<i>BT.62377</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 3	proteína de ligação
ENSBTAG000000005780	26	22.652,090 - 22.663,974	-	<i>LDB1</i>	não caracterizado	ligação de metais
ENSBTAG000000002450	15	83.602,863 - 83.605,727	-	<i>FAM111A</i>	não caracterizado	transcrição celular
						diferenciação de neurônios
						atividade catalítica

* genes relacionados com processos fisiológicos reprodutivos de acordo com a literatura científica disponível.

Tabela 7. Experimento 2. Potenciais genes candidatos para EBV2 (PG).

<i>Ensembl ID</i>	<i>BTA</i>	<i>Posição</i>	<i>Gene ID</i>	<i>Nome</i>	<i>Descrição</i>	<i>Termo(s) GO</i>
ENSBTAG000000046504	9		-	-	-	-
ENSBTAG000000045178	14		-	-	-	-
ENSBTAG000000026309	14	18.386,657 - 18.389,173	515040	<i>BT.88023</i>	zinc fingers and homeoboxes protein 2	regulação da transcrição celular

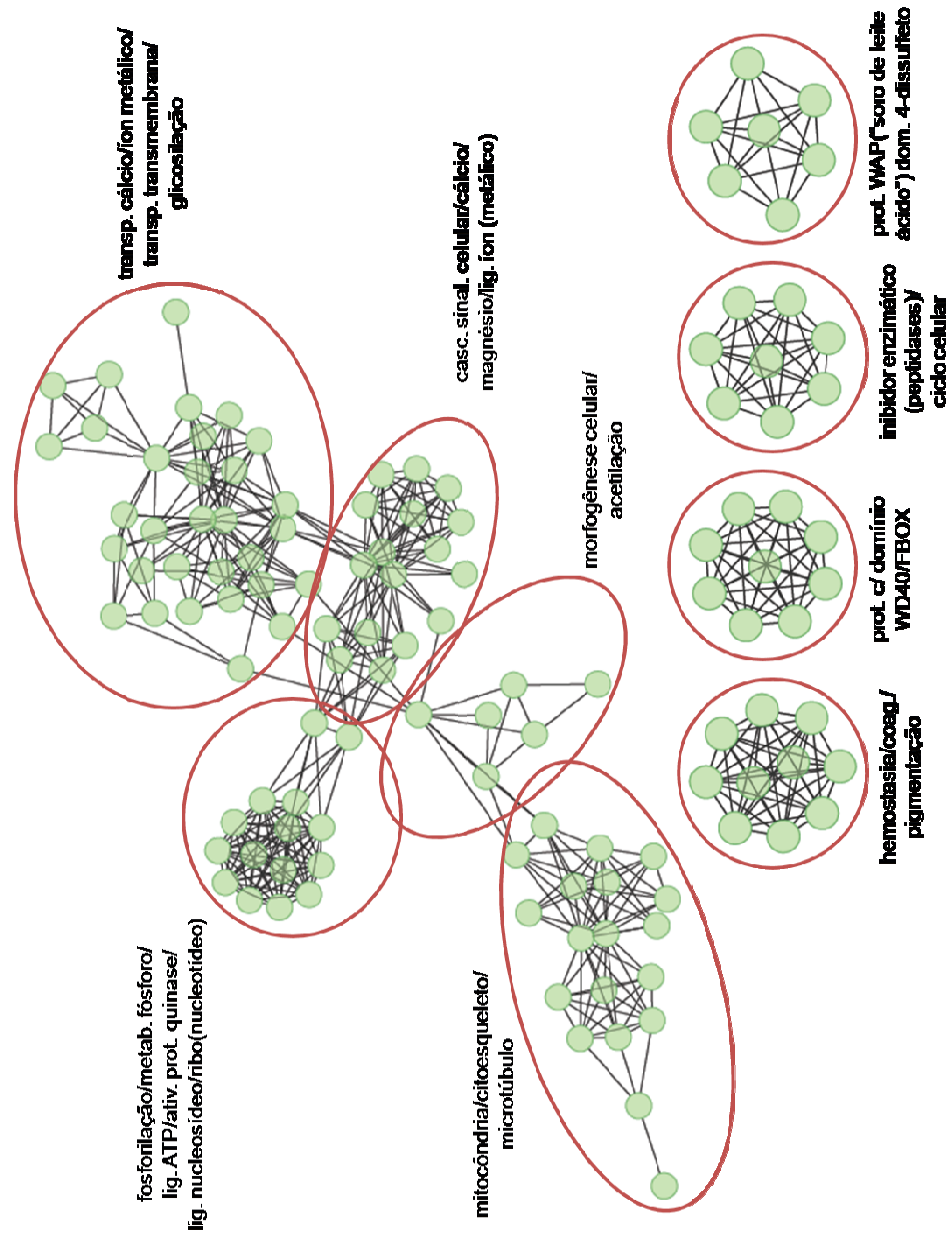


Figura 15. Experimento 2. Redes de termos funcionais para EBV1 (IPP). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.

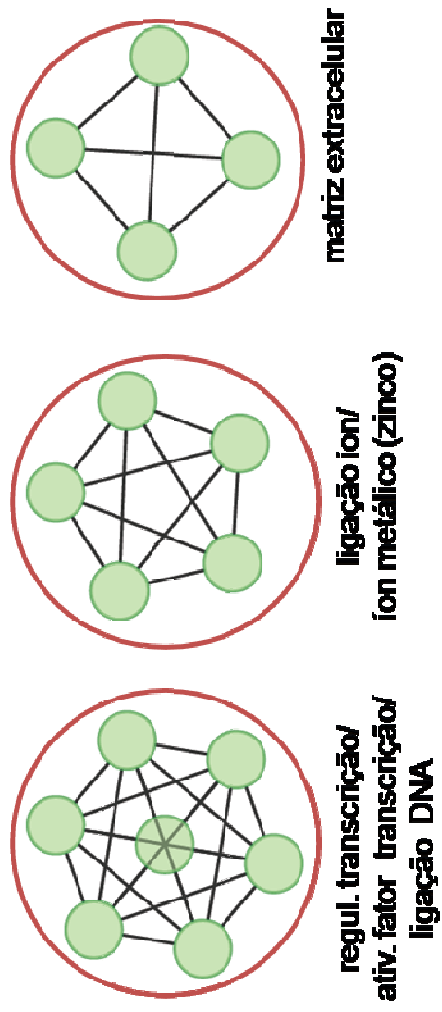


Figura 16. Experimento 2. Redes de termos funcionais para EBV2 (PG). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.

6.3. Experimento 3

6.3.1. Controle de Qualidade e Estratificação Populacional

As análises de QC removeram 268.278 e 265.893 SNP correspondentes aos subgrupos EBV1 e EBV2. Como o QC foi realizado sequencialmente a soma de marcadores excluídos em cada categoria tendeu a ser maior do que o número efetivamente removido, pois alguns SNP foram excluídos por mais de um critério. (**Tab. supl. 1** e **Fig. supl. 10**).

Nenhuma amostra foi removida do conjunto de dados pelos critérios de IBS e “call rate”. Todos os indivíduos apresentaram “call rate” superior a 90% e a média de IBS foi de 0,81 ($\pm 0,009$) (**Fig. supl. 7**). As taxas de heterozigosidade antes e após o QC foram, respectivamente, de 0,22 ($\pm 0,012$) e 0,29 ($\pm 0,008$).

A proporção média de IBD estimada entre os indivíduos foi de 0,04 ($\pm 0,054$) (**Figura 17** e **Fig. supl. 7**). Os resultados da análise de MDS sugeriram heterogeneidade genética da população equivalente à observada nas amostras do Experimento 2 (**Figura 18**).

O número de SNP que permaneceram nas análises foi de 427.770 e 426.171 para 1.076 e 1.221 amostras dos respectivos subgrupos. A distribuição dos SNP nos cromossomos e os dados sumarizados de densidade do painel estão, correspondentemente, nas **Tabs. supls. 2 e 3**.

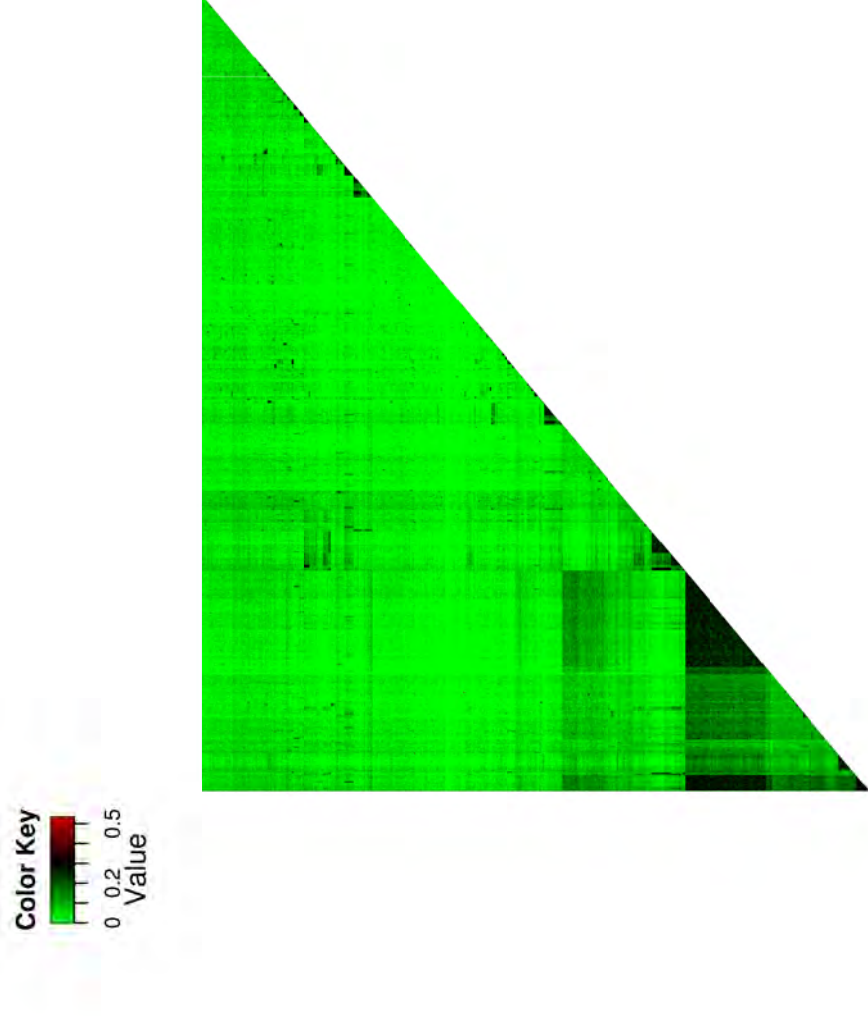


Figura 17. Experimento 3. “Heatmap” da proporção de alelos idênticos por descendência (IBD) entre os pares amostrais. Valores de 0 (verde - amostras completamente não relacionadas) a 0,5 (vermelho - compartilhamento de metade dos alelos).

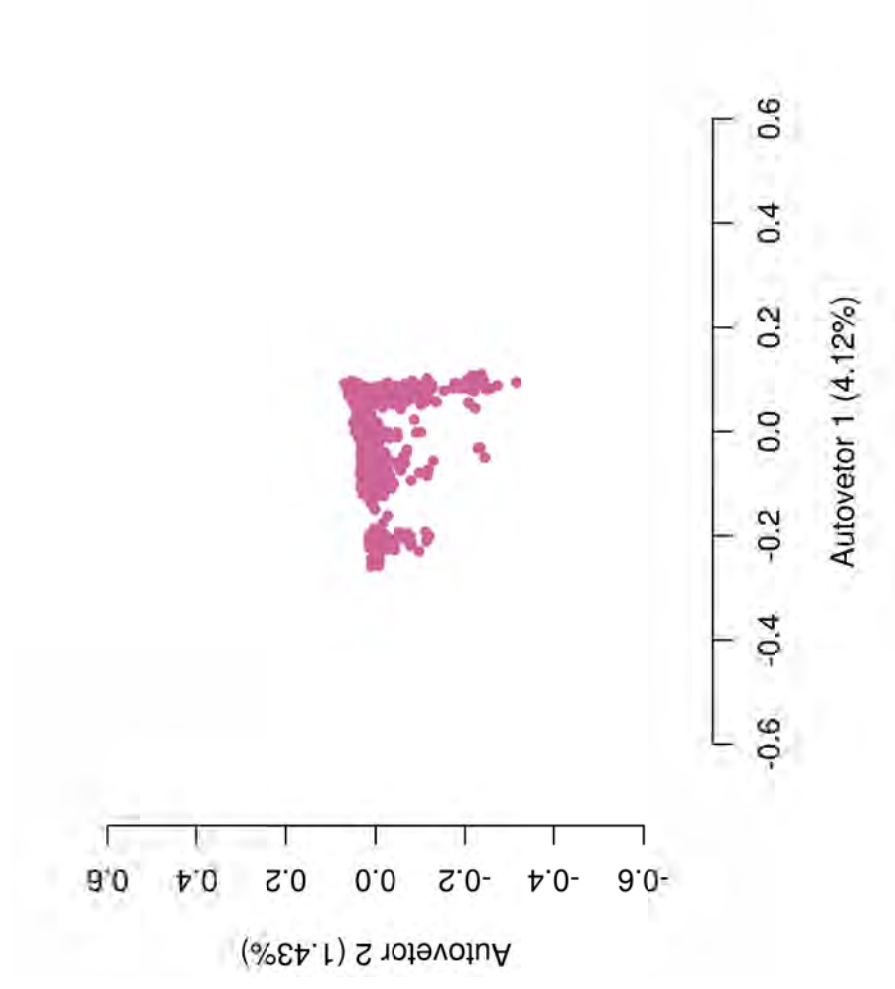


Figura 18. Experimento 3. Análise de escalonamento multidimensional (MDS) baseada no coeficiente de parentesco genômico. As porcentagens correspondem à variância explicada pelo respectivo autovetor.

6.3.2. Análise de Associação Genômica Ampla

Após as análises de QC, 1.076 e 1.221 amostras permaneceram nos subgrupos correspondentes, EBV1 e EBV2. Como critério de qualidade aplicado aos dados fenotípicos, oito amostras marginais foram excluídas da população EBV1 (± 4 unidades de desvio-padrão) (**Fig. supl. 3**). Desse modo, as análises do tipo GWAS, também conduzidas separadamente para os respectivos subgrupos, foram realizadas a partir de 466.960 e 469.345 SNP e 1.068 e 1.221 matrizes.

No total, 56 e 60 autovetores das análises de MDS foram significativamente correlacionados com EBV1 e EBV2, correspondendo a 14,56 e 14,73% da variabilidade genética dos fenótipos. Os resíduos obtidos da regressão linear sobre os autovetores significativos ponderada para a confiabilidade dos EBVs, atenderam as pressuposições de normalidade (EBV1 - Teste Anderson-Darling; EBV2 - Teste Shapiro-Wilk) e homocedasticidade (Teste Breusch-Pagan Estudantizado), sendo esses valores usados como variável dependente no teste de associação (**Fig. supl. 12**).

Os gráficos Q-Q e os fatores de inflação estimados (EBV1 = 0,99 e EBV2 = 1,03) justificaram os métodos de associação adotados, pois eliminaram o viés introduzido pela estratificação populacional (**Figura 19**).

Os "Manhattan Plots" dos resultados dos testes de associação ($-\log_{10}$ dos valores de p corrigidos por GC) estão apresentados nas **Figuras 20 e 21**. Em ambas as análises do tipo GWAS, nenhuma associação significativa entre SNP e o EBV foi observada (EBV1, $\alpha = 1,07 \times 10^{-7}$; EBV2, $\alpha = 1,06 \times 10^{-7}$).

Os resultados dos efeitos estimados dos marcadores estão apresentados nas **Figuras 20 e 21** e na **Tab. supl. 4**. Os efeitos médios estimados e seus respectivos erros-padrão para EBV1 foram de -0,011 dias ($\pm 0,258$) e 0,236 dias ($\pm 0,097$) e para EBV2 foi de -0,002 dias ($\pm 0,097$) e 0,088 dias ($\pm 0,036$) (**Figura 22**).

Foram selecionados 72 e 64 SNP "top" para as análises de identificação de genes candidatos para EBV1 e EBV2, respectivamente. Os SNP mais relevantes estavam distribuídos ao longo dos cromossomos e maior número de marcadores foi encontrado no cromossomo BTA14 (EBV1) e cromossomos BTA1 e 4 (EBV2) (**Tab. supl. 5**).

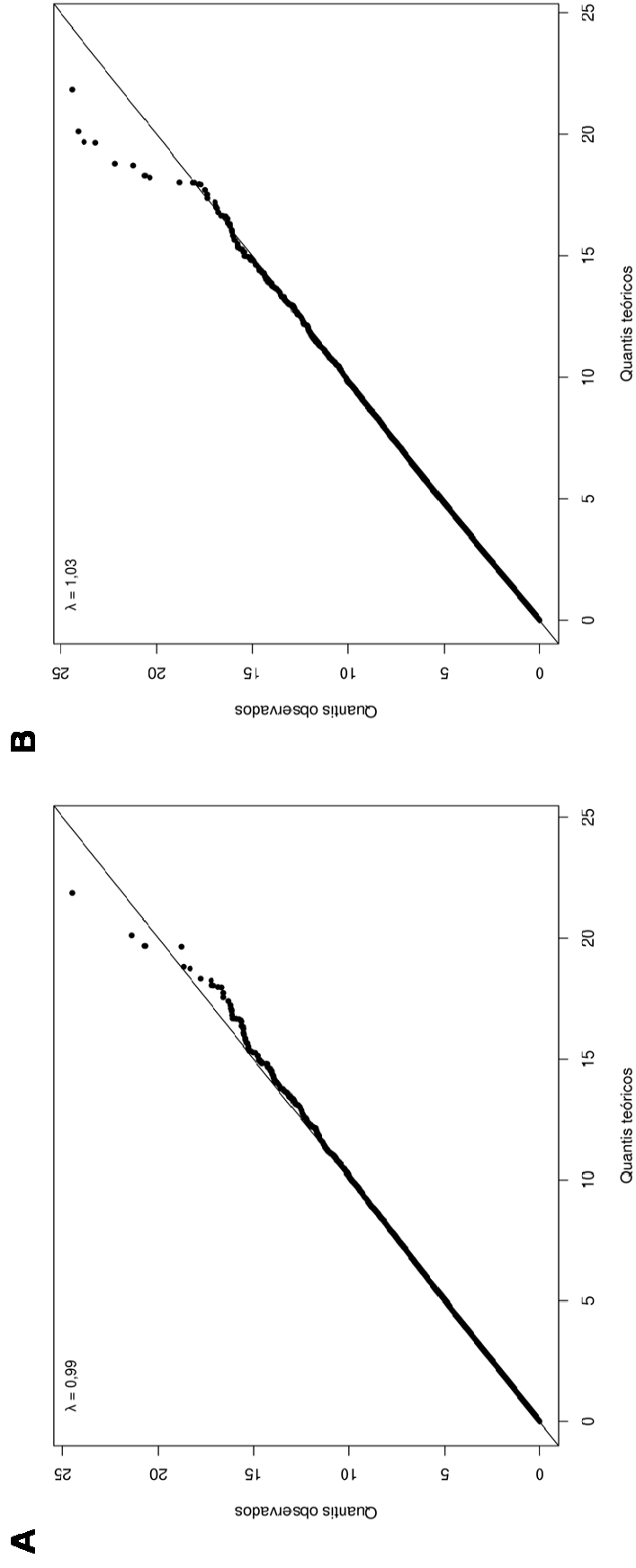


Figura 19. Experimento 3. Gráficos de Quantil-Quantil (Q-Q) dos resultados dos GWAS para EBV1(A) e EBV2 (B). A linha preta representa a distribuição esperada e os pontos pretos representam a distribuição observada da estatística χ^2 .

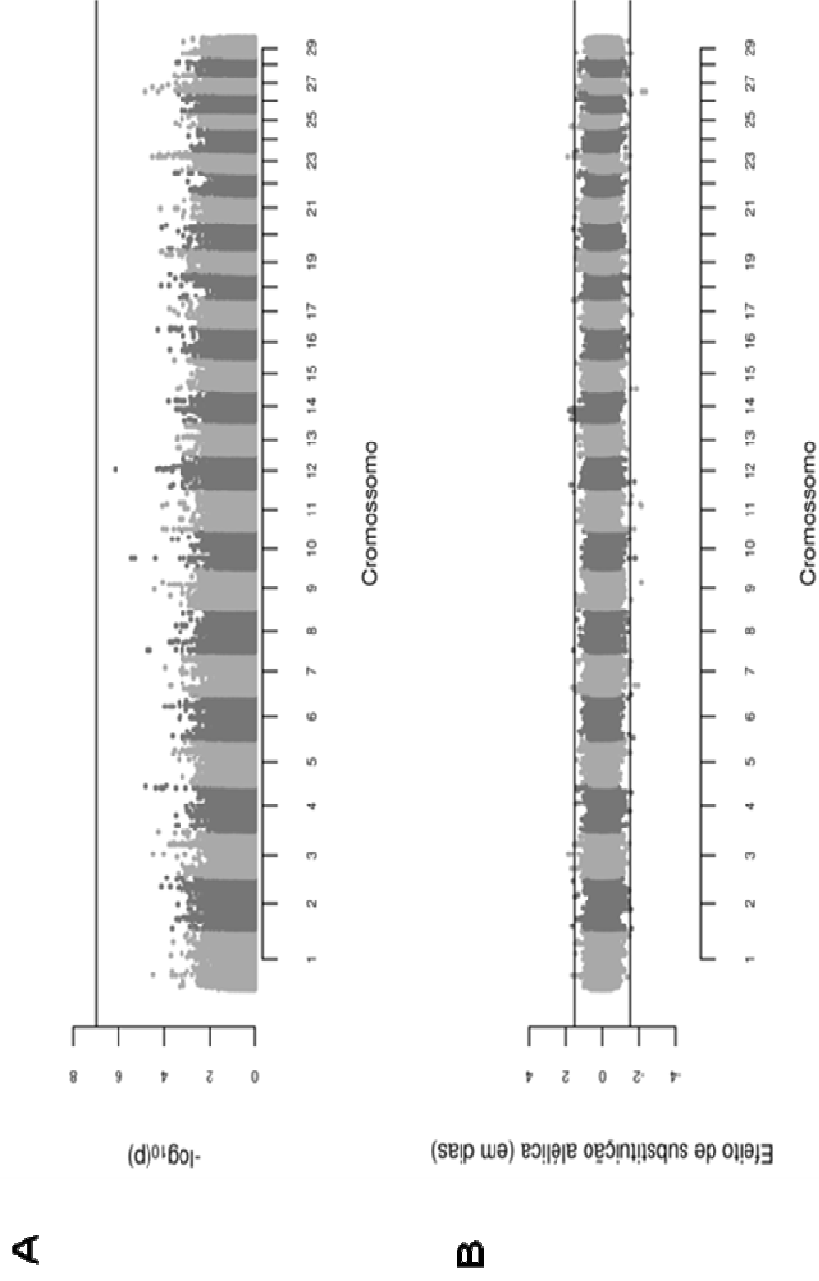


Figura 20. Experimento 3 – EBV1 (IPP). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,07 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.

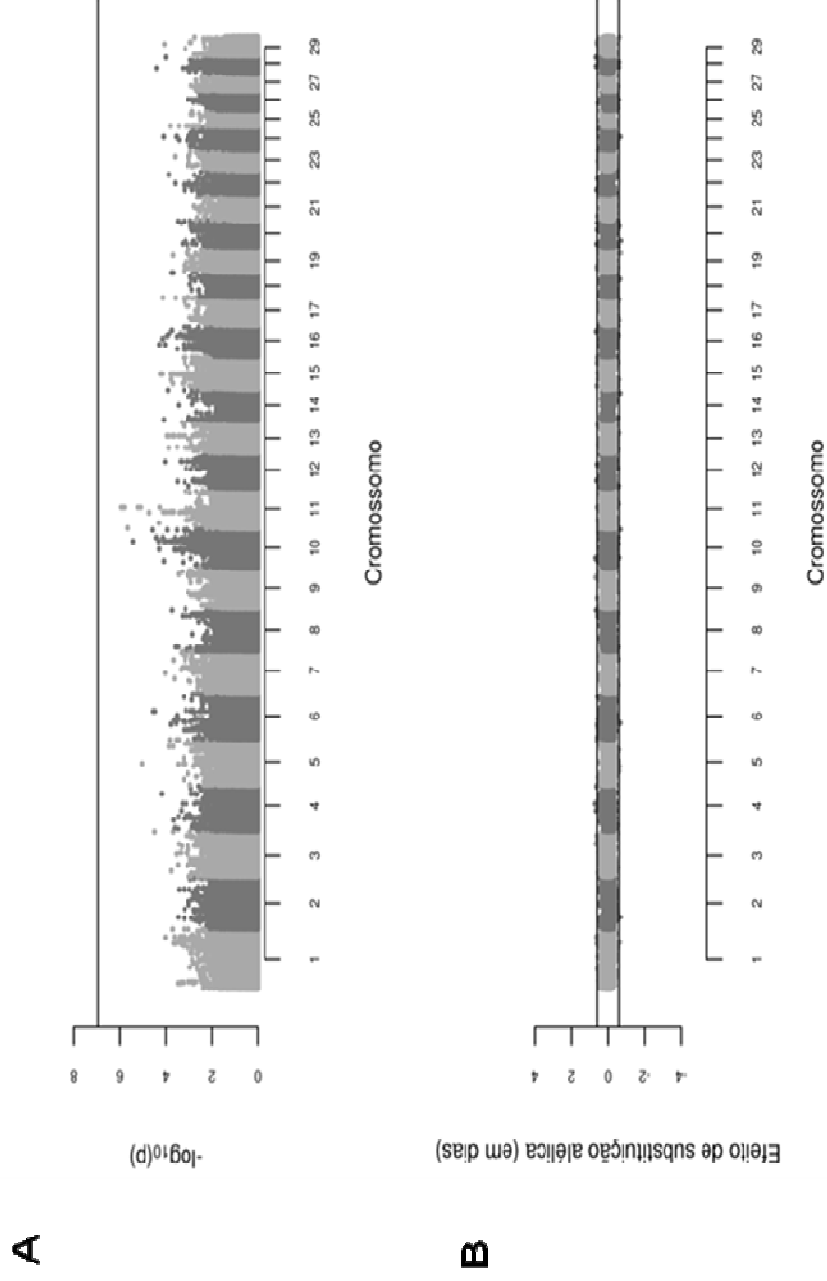


Figura 21. Experimento 3 – EBV2 (PG). (A) Manhattan Plot dos valores de significância (p) do teste de associação; a linha horizontal representa o limiar de significância de Bonferroni ($\alpha=1,06 \times 10^{-7}$). (B) Manhattan Plot da estimativa do efeito de substituição alélica das variantes SNP; linhas horizontais representam os limites de ± 6 desvios-padrão. Cromossomos de numeração par são mostrados em cinza escuro e cromossomos de numeração ímpar são mostrados em cinza claro.

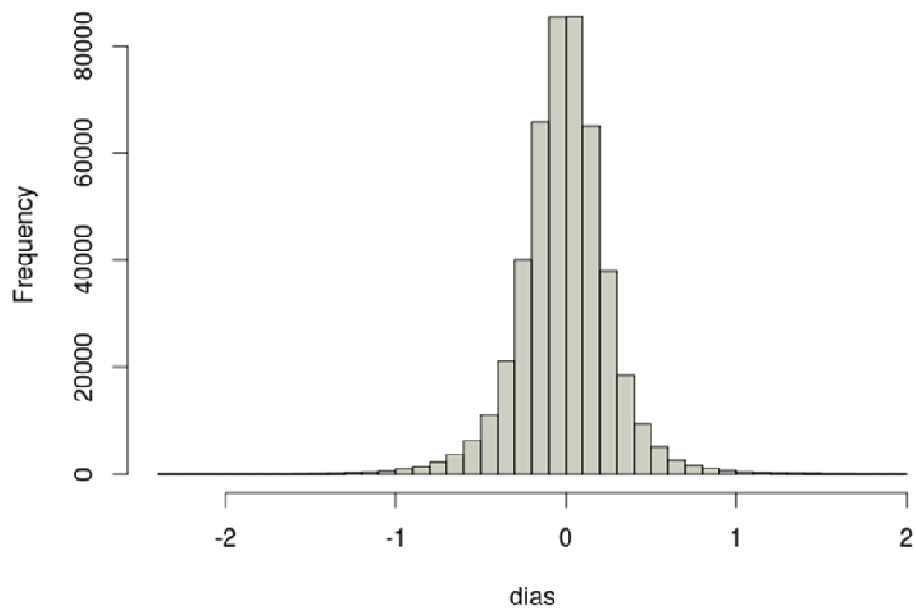
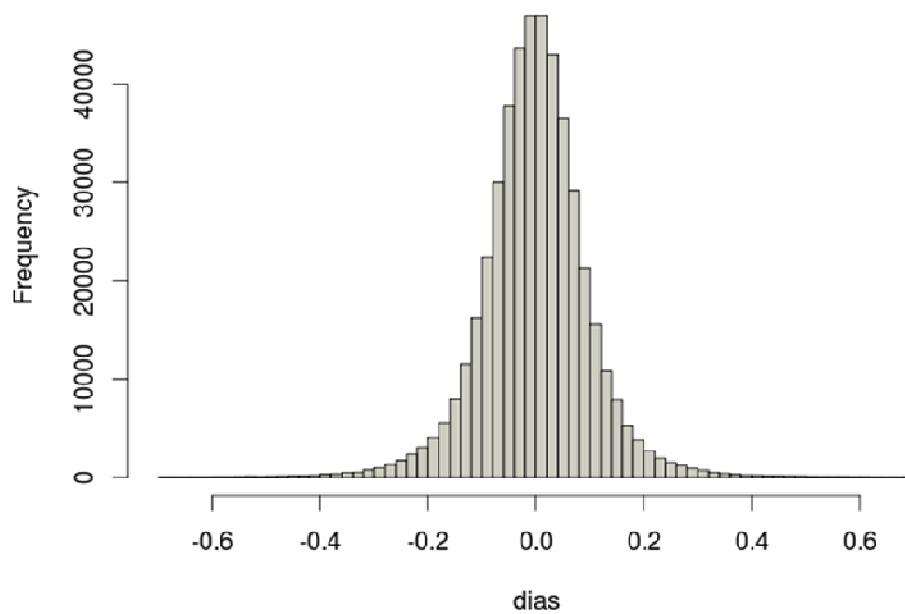
A**B**

Figura 22. Experimento 3. Distribuição dos efeitos estimados de substituição alélica das variantes SNP para EBV1 (A) e EBV2 (B) em dias.

6.3.3. Identificação de Genes Candidatos

Com base no conjunto de SNP "top" de EBV1, 29 marcadores intragênicos foram identificados (**Tab. supl. 7**), sendo a maior parte deles intrônicos. Relativamente àqueles SNP associados com genes potencialmente envolvidos no fenótipo analisado, dois SNP, um intrônico e um "upstream", estavam posicionados no gene *C16ORF73* (ENSBTAG00000020001) no cromossomo BTA25. Outras duas variantes SNP, também intrônica e "upstream", estavam posicionadas no gene *TPT1* (ENSBTAG00000048229) no mesmo cromossomo. Ademais, outro SNP "upstream" foi identificado no gene codificador de pequeno RNA nuclear (SnRNA), o gene *U6* (ENSBTAG00000043734) no cromossomo BTA7. Também foi detectada variável não sinônima no gene *KIAA0408* (ENSBTAG00000003919) no cromossomo BTA9 (**Tabela 8 e Tab. supl. 8**).

Termos identificados a partir de 50 genes mapeados a partir dos resultados das análises de GWAS para EBV1 foram utilizados na construção de três redes, que totalizaram 138 nodos e 545 arestas (**Tab. supl. 6 e Figura 23**). Nessas redes foram apontados os componentes celulares: núcleo, nucleoplasma, membrana plasmática e região extracelular. Diversas funções moleculares também foram observadas, tais como: ligação de íons metálicos e (ribo)nucleotídeos, fosforilação, acetilação e glicosilação. As vias de sinalização do cálcio e rodopsina também foram evidenciadas. Além disso, aspectos relacionados à biogênese dos ribossomos, ligação de DNA, processamento de RNAm e regulação da transcrição se destacaram nas redes.

Diversos processos biológicos relacionados à reprodução também foram apontados nas redes funcionais para EBV1 como, por exemplo, o desenvolvimento embrionário uterino. Nesta rede estavam representados os seguintes genes: *TBX15* (ENSBTAG00000007767) no cromossomo BTA3, *bta-mir-2882* (ENSBTAG00000019220) no cromossomo BTA7 e *BCL2L11* (ENSBTAG00000024105) no cromossomo BTA11. O *BCL2L11* se destacou na rede de termos relacionados aos processos de diferenciação gonadal e de estruturas reprodutivas e de desenvolvimento de características sexuais primárias, os quais

também foram relacionados ao gene *AMH* (ENSBTAG00000014955) localizado no cromossomo BTA7 (**Tabela 8**).

Com relação ao conjunto de SNP "top" para EBV2 foram identificados 35 SNP intragênicos (**Tab. supl. 7**). Os principais genes observados e associados a aspectos fisiológicos reprodutivos foram variações intrônicas localizadas no: *TF* (ENSBTAG00000007273), *COLEC12* (ENSBTAG00000007705) e *JMJD1C* (ENSBTAG00000001573), nos cromossomos BTA1, BTA24 e BTA28, respectivamente (**Tabela 9**).

Cinco redes para EBV2, contendo 83 nodos e 249 arestas, foram construídas a partir de termos relacionados a 30 genes mapeados (**Figura 24**). As redes foram correlacionadas aos componentes celulares Complexo de Golgi, endossomos, mitocôndrias e membrana plasmática. Os processos biológicos de diferenciação celular, morte celular programada e desenvolvimento do sistema imune foram evidenciados com base na destacada espessura de arestas da presente rede, associada aos genes *ZBTB16* (ENSBTAG00000011266) no cromossomo BTA15 e *PIK3R1* (ENSBTAG00000010989) no cromossomo BTA20 (**Tabela 9**).

Tabela 8. Experimento 3. Potenciais genes candidatos para EBV1 (IPP).

<i>Ensembl ID</i>	<i>BTA</i>	<i>Posição</i>	<i>Gene ID</i>	<i>Nome</i>	<i>Descrição</i>	<i>Termo (s) GO</i>
ENSBTAG000000020001	25	1.402,494 - 1.436,118	618487	<i>C16ORF73</i>	não caracterizado	meiose
*ENSBTAG000000048229	25	1.394,201 - 1.417,716	326599	<i>TPT1</i>	Translationally-controlled tumor protein	ligação de cálcio
ENSBTAG000000043734	7	85.090,720 - 85.090,826	-	<i>U6</i>	U6 spliceosomal RNA	-
ENSBTAG000000003919	9	24.120,228 - 24.131,671	100140348	<i>KIAA0408</i>	não caracterizado	proteína de ligação
ENSBTAG000000007767	3	24.236,815 - 24.359,735	521974	<i>TBX15</i>	T-box transcription factor TBX15	morfogênese embrionária
ENSBTAG000000019220	7	16.648,313 - 16.738,471	414274	<i>bta-mir-2882</i>	bta-mir-2882	transcrição celular
ENSBTAG000000024105	11	1.156,954 - 1.199,792	507390	<i>BCL2L11</i>	bcl-2-like protein 11	desenv. embrionário
*ENSBTAG000000014955	7	22.696,978 - 22.699,843	280718	<i>AMH</i>	muellerian-inhibiting factor precursor	transcrição celular
						desenv. embrionário, gonadal
						espermato gênese, apoptose
						desenvolvimento gonadal

* genes relacionados com processos fisiológicos reprodutivos de acordo com a literatura científica disponível.

Tabela 9. Experimento 3. Potenciais genes candidatos para EBV2 (PG).

<i>Ensembl ID</i>	<i>BTA</i>	<i>Posição</i>	<i>Gene ID</i>	<i>Nome</i>	<i>Descrição</i>	<i>Termo(s) GO</i>
ENSBTAG000000007273	1	13.6683,311- 136.722,309	280705	<i>TF</i>	serotransferrin precursor	proteína de ligação
ENSBTAG000000007705	24	35.630,929 - 35.816,269	504741	<i>COLEC12</i>	Collectin-12	homeostase do ferro
ENSBTAG000000001573	28	19.393,375 - 19.663,523	781067	<i>JMJD1C</i>	probable JmjC domain-containing histone demethylation protein 2C	resposta imune
*ENSBTAG000000011266	15	24.917,133 - 25.116,350	ZBTB16	<i>ZBTB16</i>	zinc finger and BTB domain-containing protein 16	ligação com carboidrato
						proteína de ligação
						transcrição celular
						morfogênese embrionária
						desenv. sistema nervoso
						transdução celular, apoptose
*ENSBTAG000000010989	20	11.329,104 - 11.409,885	282307	<i>PIK3R1</i>	Phosphatidylinositol 3-Kinase regulatory subunit alpha	ligação hormônios/fatores de cresc.

* genes relacionados com processos fisiológicos reprodutivos de acordo com a literatura científica disponível.

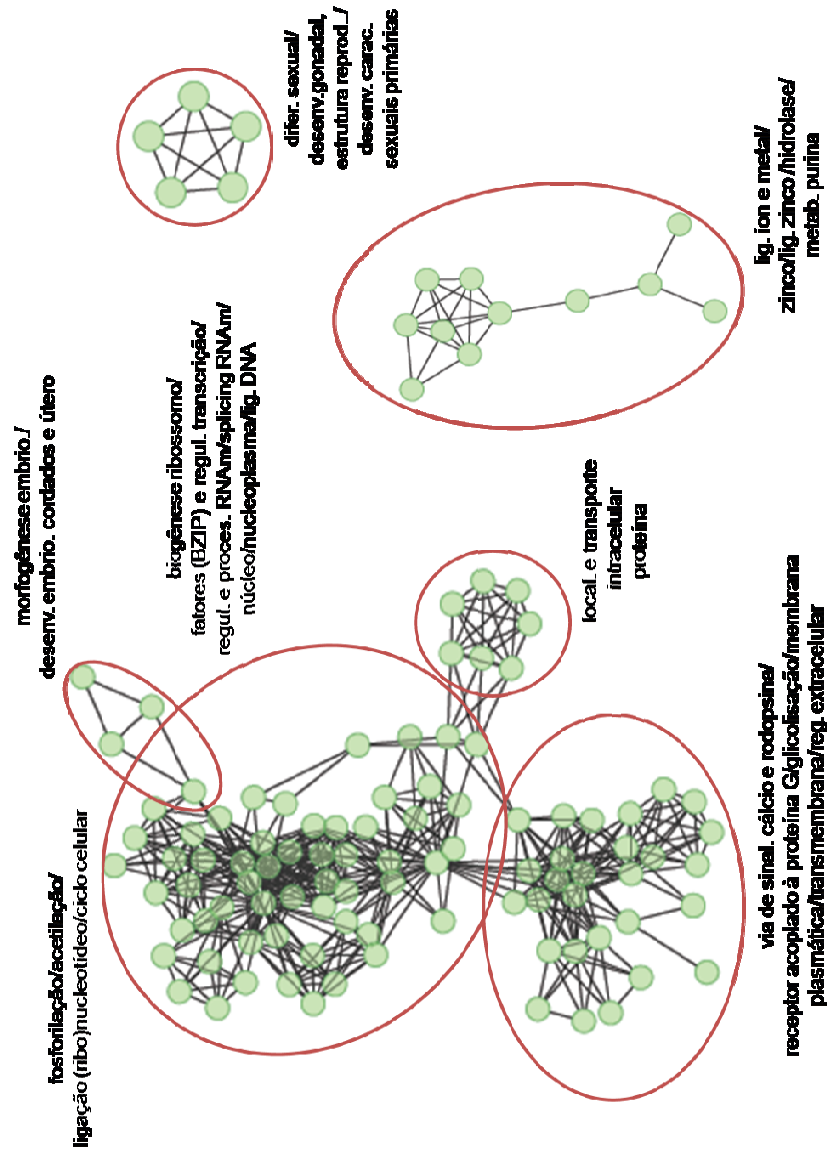


Figura 23. Experimento 3. Redes de termos funcionais para EBV1 (IPP). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.

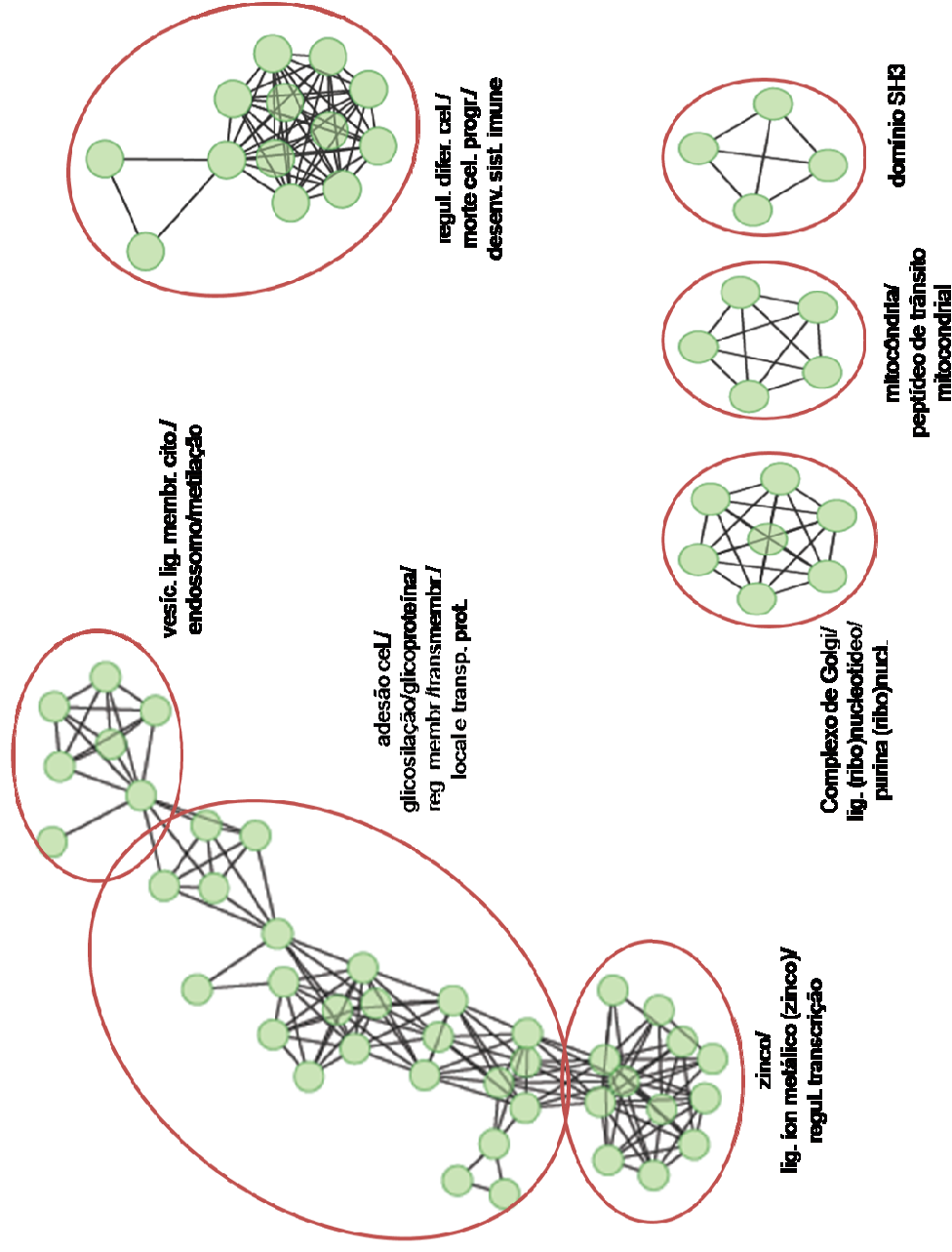


Figura 24. Experimento 3. Redes de termos funcionais para EBV2 (PG). Os termos foram minerados a partir de genes mapeados. Cada nodo representa um termo, cada aresta representa o compartilhamento de genes e sua espessura o número de genes sobrepostos.

7. Discussão

7.1. Análises de Associação Genômica Ampla

O conceito básico que permeia as análises do tipo GWAS é o da detecção de associações entre uma dada característica fenotípica e marcadores em uma população. Em geral painéis com alta densidade de marcadores SNP são utilizados para essas análises. No entanto, um problema comum no delineamento experimental baseado em GWAS refere-se à estratificação populacional, capaz de apontar para associações espúrias caso a mesma não seja detectada e devidamente corrigida (YAN et al., 2009). Quando múltiplos subgrupos dentro de uma dada população estão presentes, e caso exista diferença para os valores médios fenotípicos, a variação das frequências alélicas observadas inflacionará as taxas de erro Tipo I, aumentando as associações falso-positivas. A estratificação populacional aumenta quando as amostras diferem quanto à ancestralidade ou quando existe alto parentesco amostral (SMITH e NEWTON-CHEH, 2009). Antes que as análises de associação sejam conduzidas é recomendado que a estratificação seja verificada, o que comumente é realizado com técnicas de redução de dimensionalidade de dados complexos como, por exemplo, pela análise clássica de MDS. Esta técnica multivariada é utilizada para averiguar estratificação pela mensuração da similaridade alélica entre *loci* entre todas as amostras baseada na matriz IBS.

Juntamente com os resultados da análise de IBD, a análise de MDS conduzida no presente trabalho sugeriu heterogeneidade genética vinculada ao alto parentesco amostral dos animais do Experimento 1, fato indesejável em estudos baseados em população, pois estes assumem que todas as amostras deveriam ser não relacionadas, de relacionamento distante ou desconhecido, pois as estimativas das frequências alélicas são sensíveis à inflação ou deflação quando os dados genotípicos são derivados de indivíduos com relacionamentos familiares (STEVENS et al., 2011).

O parentesco reflete a história compartilhada de membros de uma mesma família ou de uma mesma população e, desse modo, afeta todos os caracteres que

apresentam componente genético (WEIR et al., 2006). Em análises do tipo GWAS é comum o uso do termo relacionamento não esperado (ou oculto), que se refere à ancestralidade comum recente entre pequenos grupos de indivíduos, frequentemente aos pares. Como a ancestralidade compartilhada corresponde ao relacionamento, a estratificação populacional também pode ser definida por padrões de parentesco entre grupos de indivíduos, fato claramente evidenciado nas análises do Experimento 1 (ASTLE e BALDING, 2009). É bem provável que o uso intensificado de touros geneticamente superiores pela inseminação artificial seja o principal colaborador para a formação de populações estratificadas (PAUSCH et al., 2011). Diante desse potencial confundidor, utilizou-se de variáveis preditoras no modelo linear com o objetivo removê-lo das análises do Experimento 1. No entanto, esse procedimento não foi eficiente já que os resultados indicaram deflação generalizada do teste (**Figura 5**).

Recentemente, os modelos lineares mistos (MLM) tem sido a ferramenta de escolha para corrigir o parentesco amostral (AMIN et al., 2007; ZHOU e STEPHENS, 2012). Nos métodos baseados em MLM, a estrutura populacional é considerada um efeito fixo, enquanto o parentesco entre os animais é assumido na estrutura de (co)variâncias do efeito aleatório para os indivíduos em um modelo poligênico (HENDERSON, 1984; ZHANG et al., 2010). No entanto, a estimação de componentes de (co)variância por métodos como máxima verossimilhança restrita requer grande tamanho amostral para que a estimativa seja não viesada e suficientemente acurada, o que não caberia ao Experimento 1. Contudo, mesmo diante da deflação dos resultados das análises do tipo GWAS, o conjunto de dados do Experimento 1 foi submetido às diferentes estratégias para prospecção de genes candidatos.

A correção da estratificação populacional pelas abordagens baseadas em análise de componentes principais (PCA) tem sido aplicada com sucesso nesse tipo de estudo, onde os autovetores da análise são utilizados como covariáveis nos modelos de regressão do teste de associação (PRICE et al., 2006). Dentro deste contexto, optou-se pela inclusão dos autovetores da análise de MDS como efeitos fixos no modelo linear das análises de associação dos Experimentos 2 e 3, pois a estratificação amostral também se apresentou como um potencial confundidor

nesses conjuntos de dados. Este procedimento eliminou, com sucesso, artefatos de associação resultantes de heterogeneidade genética (**Figuras 11 e 19**).

As associações das variantes genéticas com os fenótipos estudados foram avaliadas sob o modelo de herança aditivo, isto é, assumiu-se que o valor fenotípico dependia do número de cópias do alelo menor. Fundamentado neste modelo, os SNP testados não alcançaram o limiar de significância pelo ajustamento de Bonferroni em nenhum dos GWAS realizados.

Uma dificuldade fundamental na interpretação dos resultados dos estudos de associação em larga escala se deve ao problema conhecido como testes múltiplos (MOSKVINA e SCHMIDT, 2008). Diante deste fato, pesquisadores aplicam métodos de ajustamento como, por exemplo, a correção de Bonferroni, onde o limiar de significância é definido por $0,05/n$, sendo n o número de SNP testados independentemente (BENJAMIN e HOCHBERG, 1995). É possível que a correção adotada neste estudo tenha sido muito conservativa, de modo que algumas associações relevantes possam ter sido falsamente eliminadas. De fato, o problema dos testes múltiplos é um grande desafio nos GWAS, pois a dependência dos dados genéticos devido aos altos níveis de LD entre SNP faz com que Bonferroni seja, de fato, uma correção muito conservativa (FROMMLET et al., 2012). Este problema se intensifica, particularmente, quando a densidade de SNP aumenta e o tamanho amostral diminui (FINLAY et al., 2012).

Diferentemente do que vem sendo conduzido nas espécies domésticas, o tamanho amostral dos GWAS praticados na espécie humana frequentemente supera 10.000 indivíduos (SUN, 2012). Dessa maneira, o tamanho das populações analisadas no presente estudo pode ser apontado como fator limitante do poder de detecção de associações significativas ou de efeitos de substituição alélica moderados.

Os EBVs para IPP e PG previamente preditos foram utilizados como fenótipos nos Experimentos 2 e 3 com o objetivo de aumentar o poder das análises de associação. Estudos recentes têm utilizado, com sucesso, o EBV como fenótipo composto (COLE et al., 2011; PAUSCH et al., 2011; FINLAY et al., 2012). Os EBVs são predições do valor genético do indivíduo baseado em fenótipos mensurados em seus parentes, resumizando a informação de muitos descendentes avaliados. No

entanto, a magnitude da diferença observada entre os efeitos estimados dos SNP para os fenótipos compostos EBVs (IPP e PG) foi pequena, dentro de cada análise do tipo GWAS aplicada nesses experimentos. Tal fato era esperado uma vez que os EBVs utilizados como fenótipos não refletem diretamente as variações fenotípicas na mesma ordem de grandeza na qual foram medidas (nesse caso em dias).

Entretanto, quando os valores de significância obtidos nas análises de GWAS foram apresentados na forma de "Manhattan Plot", observou-se a presença de picos difusos e bem distribuídos por diversos cromossomos em cada análise, sinalizando para a existência de grande número de regiões genômicas com efeitos sobre as características analisadas (**Figuras 6, 12 e 20**). De fato, múltiplos genes contribuem para a variação fenotípica da maior parte das características produtivas, cada um com baixo efeito conforme o padrão de herança poligênica. Tal tendência também pôde ser observada pela distribuição dos efeitos de substituição alélica estimados em cada análise do tipo GWAS, onde muitos SNP não apresentaram efeitos sobre os fenótipos testados ou esses efeitos foram muito baixos (**Figuras 7, 13 e 21**).

Esses achados foram coerentes com os conhecimentos da biologia envolvida na manifestação dos fenótipos reprodutivos IPP e PG (características poligênicas), criando as bases para as explorações funcionais desenvolvidas no presente trabalho.

7.2. Identificação de Genes Candidatos

As abordagens de análise funcional, comumente implementadas após análises do tipo GWAS, baseadas em redes e vias biológicas têm se apresentado como métodos complementares a este tipo de análise, pois permitem a visualização das relações entre grande número de genes simultaneamente. Dada a complexidade dos mecanismos genéticos e dos efeitos sinérgicos entre os genes, espera-se que o efeito holístico de uma rede, ou de uma via metabólica, seja maior do que a soma dos efeitos individuais dos genes (SUN, 2012). Contudo, a interpretação biológica de um grande número de listas de SNP, genes ou termos biologicamente relacionados tem sido uma tarefa desafiadora (WANG et al., 2011).

As características estudadas no presente trabalho foram complexas e, portanto, influenciadas por muitos processos biológicos e número desconhecido de genes. Além dos SNP intragênicos identificados e da alta conectividade entre muitos genes evidenciada nas redes, os resultados funcionais para os fenótipos (próprios e compostos) revelaram, em última instância, potenciais genes candidatos para IPP (n= 8) no Experimento 1; EBV1-IPP (n= 7) e EBV2-PG (n= 3) no Experimento 2; EBV1-IPP (n= 8) e EBV2-PG (n= 5) no Experimento 3 (**Tabelas 5 a 9**). Com base na literatura científica existente, alguns dos genes candidatos apresentados pela combinação das análises do tipo GWAS e funcionais dos Experimentos 1, 2 e 3, direta ou indiretamente envolvidos em processos reprodutivos, foram abordados com mais profundidade.

A análise criteriosa dos genes listados como resultados experimentais do presente trabalho apontou para extensa rede de termos relacionados à atividade das proteínas quinases (PKCs) para IPP. Esta rede foi associada ao compartilhamento de dois genes principais, *PRKCD* e *PRKCE*. As PKCs são as principais transdutoras celulares e pertencem à família de múltiplos genes serina/treonina quinases. As redes de transdução de sinal celular são responsáveis por converter as informações ambientais em ações celulares. As PKCs influenciam eventos celulares de maneira abrupta e reversível, modificando as funções das proteínas. Estas são amplamente conservadas nos eucariotos e podem coexistir em uma única célula (ROSSE et al., 2010).

Sob o aspecto reprodutivo, as PKCs participam de eventos fisiológicos fundamentais desencadeados pela ação do GnRH nas células gonadotróficas. Basicamente, o modelo de interação se dá pela ligação do GnRH a receptores acoplados a proteína G. A proteína G ativada promove o aumento de fosfolipase C (PLC) que resulta na clivagem de fosfatidilinositol e geração de inositol 1,4,5-trifosfato (IP3) e diacilglicerol (DAG). O IP3 libera Ca^{2+} dos estoques intracelulares, ao passo que DAG ativa PKCs. Estes eventos culminam na síntese e liberação de hormônio folículo estimulante (FSH) e LH (RAMAKRISHNAPPA et al., 2005).

Vários estudos apontam para diferentes isoformas de GnRH e para o fato de que seus sítios de ação não se restringem ao hipotálamo e a hipófise. Assim, suas ações seriam direcionadas também a vários outros tecidos como ovários, placenta,

endométrio, oviduto, testículos, próstata e glândulas mamárias. Dessa maneira, é bastante provável que as vias de sinalização celular das ações destas isoformas sejam as mesmas que atuam no sistema nervoso central. Além disso, as PKCs já foram identificadas em espermatozóides e oócitos, sugerindo que a família das quinases exerça múltiplas funções na gametogênese, fertilização e início do desenvolvimento (KALIVE et al., 2009). As PRKCD e PRKCE são isoformas ativadas pela DAG e fosfatidilserina e, embora sejam poucos os seus relatos, são fortes as evidências de que a PRKCE exerça um papel crítico no controle de resposta das células luteais à $\text{PGF}_{2\alpha}$ (SEN et al., 2005).

Outra superfamília de genes evidenciada nos resultados para IPP foi a de inibidores de proteases *ITIH* (inter α trypsin inhibitor heavy chains). Estas proteases são formadas a partir de polipeptídeos distintos que incluem quatro proteínas precursoras de cadeia pesada (ITIH-1, 2, 3 e 4) que se ligam a integrinas, colágeno e proteoglicanas como, por exemplo, o ácido hialurônico (AH). Há evidências de que as ITIHs desempenhem um papel-chave na formação da matriz do *cumulus* no processo denominado de expansão ou mucificação. Este processo ocorre na progressão da maturação oocitária e é caracterizado pela síntese de uma grande quantidade de AH e consequente formação de uma matriz muco-elástica altamente hidratada. Portanto, a formação dos complexos ITIH-AH desempenha um papel na fertilidade, dada a importância da expansão do *cumulus* na ovulação e captura do complexo *cumulus oophorus*-oócito (COC) pelo oviduto.

As ITIHs foram previamente identificadas no fluido folicular bovino e no endométrio suíno e, provavelmente, contribuem para o controle da pressão osmótica intra-folicular e do processo de implantação embrionária, respectivamente (CLARKE et al., 2006; BALCELLS et al., 2011). Em estudo recente, a expressão do gene *ITIH3* no endométrio foi correlacionada com a prolificidade em suínos. Além disso, nesse mesmo estudo SNP para *ITIH1*, 3 e 4 foram significativamente associados ao número de leitões nascidos vivos (BALCELLS et al., 2011).

Os resultados das análises do tipo GWAS para IPP apontaram o gene *BT.62377* que codifica a isoforma SERCA 3 (sarcoplasmic/endoplasmic reticulum calcium ATPase 3). As proteínas SERCA pertencem a um subtipo de ATPases transportadoras de íons e desempenham importante papel na sinalização e

homeostasia do Ca^{2+} . Foi observado que algumas isoformas codificadas pelo gene *SERCA 3* agem na regulação de eventos reprodutivos fundamentais, tais como maturação oocitária e espermiática no epidídimo, capacitação espermiática, reconhecimento de gametas e fertilização, no entanto ainda não há informações sobre o papel direto de *SERCA 3* nesse contexto (PRASAD et al., 2004; BONI et al., 2007).

Outro gene apontado para IPP foi o *EFCAB6*, que atua nas vias de sinalização celular envolvendo o íon cálcio. O cálcio é amplamente utilizado na regulação das vias de sinalização, principalmente sob controle das proteínas ligadoras de cálcio (EF-hand calcium binding proteins) (CHAZIN, 2011). Estas proteínas participam de ampla gama de processos biológicos e apresentam notável homologia de sequência e similaridade estrutural (BHATTACHARYA et al., 2004). Mais precisamente, o gene *EFCAB6* codificador da proteína de mesmo nome *EFCAB6* (EF-hand calcium binding domain), está implicado na regulação das funções reprodutivas masculinas. Bovinos transgênicos para o gene *FAT1*, relacionado com o metabolismo de lipídeos, apresentaram baixa expressão de *EFCAB6* com supressão da síntese de receptores de andrógenos (AR), afetando a expressão de genes masculinos específicos e a fertilidade de espermatozoides maduros (GUO et al., 2011).

Outro gene destacado nos resultados funcionais para IPP foi o *TPT1*, também denominado *TCTP* (translationally-controlled tumor protein). Este gene é altamente conservado e abundantemente expresso nos eucariotos, com participação em muitos processos celulares como síntese protéica e crescimento celular. Sua expressão é essencial para o desenvolvimento embrionário inicial, proliferação de "stem cells" embrionárias e implantação do blastocisto (LI et al., 2011; ZHANG et al., 2012b). Além disso, *TPT1* é considerado um gene-chave para as primeiras etapas de reprogramação de células somáticas, pois sua proteína ativa a transcrição no núcleo transplantado de genes fundamentais nesse processo, o *oct4* e *nanog* (KOZIOL et al., 2007; TANI et al., 2007).

Os resultados também evidenciaram o gene anti-müllerian hormone (*AMH*) no processo de diferenciação das gônadas. O *AMH* pertence à família de fatores de crescimento (*TFG-β*) que está relacionada a muitas funções celulares essenciais, tais como proliferação, diferenciação e apoptose. Além do *AMH*, nessa família de

genes encontram-se aqueles que codificam as BMP (bone morphogenetic proteins), ativinas e inibinas. O *AMH* produz o hormônio de mesmo nome que, secretado pelas células de Sertoli no início da diferenciação sexual do trato reprodutivo, regride os ductos Müllerianos nos fetos masculinos.

Nas fêmeas o AMH é produzido pelas células da granulosa e do *cumulus* de folículos ovarianos em crescimento. As concentrações séricas deste hormônio têm sido utilizadas como teste de reserva ovariana em humanos devido a sua correlação com o número de folículos primordiais (VISSER et al., 2006; IRELAND et al., 2009). Em pesquisas recentes, revisadas por Ireland et al. (2011), foi demonstrado que a dosagem de AMH em bovinos é um método confiável para identificar fêmeas adultas jovens subférteis e também para controlar o impacto de fatores ambientais na fertilidade.

O AMH é um importante fator de crescimento intra-ovariano com ações sob o crescimento folicular inicial e final. O padrão de expressão de *AMH* em camundongos e mulheres sugere efeito inibitório no crescimento de folículos primordiais e também no crescimento folicular sob ação do FSH, sugerindo ainda que este hormônio seja um fator crítico na sensibilidade folicular ao FSH. Por exercer papel essencial na reprodução de mamíferos, o *AMH* é um gene candidato para desordens ovarianas e infertilidade, de modo que polimorfismos neste gene podem modificar as atividades biológicas do hormônio, com implicações no controle do recrutamento e crescimento folicular (RIGON et al., 2010).

As investigações para precocidade sexual e desempenho reprodutivo nas fêmeas zebuínas estiveram até pouco tempo fundamentadas nos estudos de genes candidatos como, por exemplo, nos genes FSHR, IGF1 e GH (MILAZZOTTO et al., 2008; LAUREANO et al., 2009; ANDREA et al., 2011). Atualmente, grande parte das análises do tipo GWAS é conduzida em animais taurinos das raças Holandesa e Angus e ainda não existem relatos de GWAS para animais da raça Nelore. No entanto, estudos recentes para fenótipos reprodutivos estão sendo realizados em animais zebuínos das raças Brahman e mestiços.

A partir da construção de uma matriz de associação para características reprodutivas em novilhas Brahman e Tropical Composite na Austrália, 1.272 genes foram correlacionados com idade à puberdade, havendo destaque para genes

candidatos e vias biológicas envolvidas na liberação pulsátil de GnRH. Os processos biológicos foram relacionados, principalmente, com adesão celular, orientação axônica, sinalização via família de receptores tirosina quinases (ErbB) e atividade de glutamato, além de importantes fatores de transcrição (HIVEP3, TOX, EYA1, NCOA2 e ZFHX4) (FORTES et al., 2010 e 2011). Outro estudo recente mapeou três QTL em uma região de 2 Mb do cromossomo BTA 8 para as características concepção ao primeiro serviço e prenhez de novilhas Brangus utilizando painel de SNP 50k. Nessa região foram identificados genes implicados em funções celulares e neuronais, como *PLIN2*, *ADAMTSL1*, *HAUS6*, *RRAGA*, *SH3GL2* e *CNTLN*. (PETERS et al., 2012). No presente trabalho não foram identificados genes, vias biológicas ou fatores de transcrição comuns aos estudos prévios mencionados.

Números menores de processos biológicos e de genes candidatos potenciais foram observados para a característica PG, cuja ênfase esteve no processo de transcrição gênica e genes codificadores de miRNA. Contudo, a maior parte desses genes não foi candidato evidente para qualquer processo reprodutivo. Apenas dois destes genes se apresentaram como candidatos plausíveis, o *ZBTB16* (zinc finger and BTB domain-containing protein 16) e o *PIK3R1* (phosphatidylinositol 3-kinase regulatory subunit alpha).

O *ZBTB16A* codifica a proteína de mesmo nome, também chamada de PLZF (promyelocytic leukemia zinc finger). Esta proteína pertence à família das zinc fingers proteins (ZFP), proteínas regulatórias que atuam na ligação de DNA e RNA e nas interações proteína-proteína. A expressão de *ZBTB16* está limitada às espermatogônias indiferenciadas e, desse modo, tem sido alvo de estudos de transplantação de espermatogônias “stem cells”, pois funcionam como marcador auxiliando na compreensão dos mecanismos que regulam a atividade biológica destas células (REDING et al., 2010).

O *PIK3R1* é um sinalizador do IGF1R (insulin-like growth factor I receptor), pertencente à família de transdutores celulares PI3K (enzyme phosphatidylinositol-3-kinase), quinases de lipídeos que, sob influência de fatores de crescimento e outros hormônios, controlam funções essenciais de crescimento, metabolismo e imunidade. Há evidências de que a sinalização via PI3K regule a ingestão de alimento e o balanço energético, funcionando como um mediador dos efeitos da leptina nos

neurônios hipotalâmicos (DONATO Jr. et al., 2010). Além disso, as vias PI3K são vitais para oogênese, foliculogênese e ovulação. Estudos recentes, inclusive com modelos camundongos geneticamente modificados, alguns genes (por exemplo, *Pten*) apontam a importância das PI3K na sobrevivência e ativação de folículos primordiais; proliferação e diferenciação das células da granulosa em resposta às gonadotrofinas e, ainda, que a sinalização PI3K no oócito seja um novo fator de efeito materno para o início do desenvolvimento (REDDY et al., 2008; ZHENG e LIU, 2011; ZHENG et al., 2012).

Um SNP localizado no *PIK3R1*, gene que também foi identificado no presente estudo, foi significativamente associado à idade à puberdade em novilhas da raça Brahman. Inclusive, esta variante estava próxima a vários genes que participam da via metabólica do IGF1 (FORTES et al., 2012b). De fato, a atividade das PI3Ks é um importante mecanismo na regulação do eixo hipotalâmico-hipofisário-gonadal por serem mediadores de modeladores neuroendócrinos-chave da secreção de GnRH, o IGF1, insulina e leptina (ACOSTA-MARTÍNEZ et al., 2009).

Embora seja esperado que a maior parte dos *loci* exerça pequeno efeito sobre uma determinada característica complexa, ocasionalmente pode existir *locus* com efeito moderado, devido a processo de seleção recente para um alelo mutante (KEMPER e GODDARD, 2012). Em análise do tipo GWAS utilizando touros holandeses, foi detectada variante SNP no gene *SIGLEC5* (sialic acid binding Ig-like lectin 5) do cromossomo BTA18, fortemente associada ao PG (MALTECCA et al., 2011). No presente estudo não foi observada evidência para este gene, porém suas variações poderiam atrasar o parto, aumentar o tamanho do bezerro e o risco de partos distócicos (COLE et al., 2009; MALTECCA et al., 2011).

Os mecanismos genômicos subjacentes ao desempenho reprodutivo da fêmea, inicialmente expresso pela ocorrência da puberdade e manifestação do estro, seguidos pelo sucesso na concepção, gestação, parto e, finalmente, pela repetição anual de ciclo estral-concepção-gestação-parto, são em grande parte desconhecidos. Todavia, as tecnologias genômicas têm proporcionando uma visão mais profunda das vias moleculares que estão envolvidas nestes mecanismos (HIENDLEDER et al., 2005; SNELLING et al., 2011). Os resultados obtidos no presente estudo ressaltaram que múltiplos genes associados influenciaram as

características IPP e PG, dada a diversidade de funções biológicas apontadas, pois o estabelecimento da gestação requer numerosos processos fisiológicos precisamente cronometrados entre ovário/corpo lúteo, oviduto, endométrio e embrião (SPENCER et al., 2008).

Compreender como os genes interagem para influenciar um determinado fenótipo é um grande desafio, muitas vezes dificultado pelo efeito confundidor dos fatores ambientais sobre os fenótipos. Contudo, as informações funcionais geradas neste estudo buscaram traduzir os resultados obtidos nas análises do tipo GWAS em hipóteses biológicas através de conexões funcionais. Estas análises priorizaram, portanto, a investigação de processos biológicos e poderiam contribuir para pesquisas futuras voltadas para a validação dos genes candidatos e elucidação dos mecanismos de interação gênica. Esses dados também poderiam ser disponibilizados em bancos para futuras meta-análises, onde a combinação de diferentes estudos pode aumentar efetivamente o tamanho amostral e, conseqüentemente, o poder de detecção de associações significativas.

8. Conclusões

1. As cinco análises do tipo GWAS conduzidas em três populações distintas de bovinos da raça Nelore para IPP e PG, não evidenciaram associações significativas para os fenótipos analisados quando utilizada a correção de Bonferroni como critério de ajuste para testes múltiplos.
2. Os efeitos de substituição alélica estimados para os marcadores nos três experimentos conduzidos no presente trabalho corroboram com a natureza poligênica para IPP e PG.
3. A análise funcional sistematizada empregada no presente trabalho permitiu a construção de redes de termos inter-relacionados de forma eficiente, pois evidenciou processos biológicos e funções celulares que, além de serem essenciais ao organismo, também se relacionam com diferentes processos reprodutivos.
4. Os genes *PRKCD* (protein kinase C delta type), *PRKCE* (protein kinase C epsilon type), *ITHI-1* (inter-alpha-trypsin inhibitor heavy chain H1 precursor), *ITHI-3* (inter-alpha-trypsin inhibitor heavy chain H3), *ITHI-4* (inter-alpha-trypsin inhibitor heavy chain H4 precursor), *BT.62377* (sarcolemmal/endoplasmic reticulum calcium ATPase 3), *EFCAB6* (não caracterizado), *TPT1* (translationally-controlled tumor protein) e *AMH* (muellerian-inhibiting factor precursor) constituem-se potenciais candidatos envolvidos na manifestação da IPP, com base nos métodos e bancos de dados utilizados no presente estudo.
5. Os genes *ZBTB16* (zinc finger and BTB domain-containing protein 16) e *PIK3R1* (Phosphatidylinositol 3-kinase regulatory subunit alpha) constituem-se potenciais candidatos envolvidos na manifestação do PG, com base nos métodos e bancos de dados utilizados no presente estudo.

9. Referências Bibliográficas

ACOSTA-MARTINEZ, M.; LUO, J.; ELIAS, C.; WOLFE, A.; LEVINE, J.E. Male-biased effects of gonadotropin-releasing hormone neuron-specific deletion of the phosphoinositide 3-kinase regulatory subunit p85alpha on the reproductive axis. **Endocrinology**, v.150, p.4203-4212, 2009.

ALBUQUERQUE, L. G.; FRIES, L. A. Precocidade: estratégias de seleção. In: XXI SIMPÓSIO O NELORE DO SÉCULO, 1997, Uberaba. **Anais...**Uberaba: ABCZ, 1997. p.164-179.

ALBUQUERQUE, L.G.; MERCADANTE, M.E.Z.; ELER, J.P. Recent Studies on the Genetic Basis for Selection of Bos Indicus for Beef Production. In: 8th WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 2006. **Proceedings...**Belo Horizonte, Brazil, 2006.

ALIANÇA: Sumário de touros 2012 Aliança Nelore. GenSys Consultores Associados, 2012, 97p.

AMIN N.; van DUIJN, C.M.; AULCHENKO, Y.S. A genomic background based method for association analysis in related individuals. **PLoS One**, v.2, e1274, 2007.

AMOS, C.I. Successful design and conduct of genome-wide association studies. **Human Molecular Genetics**, v.16, R220-R225, 2007.

ANDERSON, C.A.; PETTERSSON, F.H.; CLARKE, G.M.; CARDON, L.R.; MORRIS, A.; ZONDERVAN, K.T. Data quality control in genetic case-control association studies. **Nature Protocols**, v.5, p.1564-1573, 2010.

ANDREA, M.V.; GOMES, M.V.M.; MARCONDES, C.R.; OLIVEIRA, K.N.; RAMOS, E.S.; FONTELES, S.B. Relationship between polymorphism of growth hormone and precocity traits in Nelore heifers. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.63, p.153-157, 2011.

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.24, p.451-471, 2009.

AULCHENKO, Y.S.; RIPKE, S.; ISAACS, A.; VAN DUIJN, C.M. GenABEL: an R library for genome-wide association analysis. **Bioinformatics**, v.23, p.1294-1296, 2007.

AULCHENKO, Y.S. Abel Tutorial, 2011, 243p.

AZEVEDO, D.M.M.R.; MARTINS FILHO, R.; LÔBO, R.N.B.; MALHADO, C.H.M.; LÔBO, R.B.; MOURA, A.A.A.; FILHO, E.C.P. Desempenho reprodutivo de vacas Nelore no Norte e Nordeste do Brasil. **Revista Brasileira de Zootecnia**, v.35, n.3, p.988-996, 2006.

BALCELLS, I.; CASTELLÓ, A.; NOGUERAC, J.L.; FERNÁNDEZ-RODRÍGUEZ, A.; SÁNCHEZ, A.; TOMÁS A. Sequencing and gene expression of the porcine ITIH SSC13 cluster and its effect on litter size in an Iberian × Meishan F2 population. **Animal Reproduction Science**, v.128, p.85-92, 2011.

BALDING, D. J. A tutorial on statistical methods for population association studies. **Nature Reviews Genetics**, v.7, p.781-791, 2006.

BARUSELLI, P.S.; GIMENES, L.U.; SALES, J.N.S. Fisiologia reprodutiva de fêmeas taurinas e zebuínas. **Revista Brasileira de Reprodução Animal**, v.31, n.2, p.205-211, 2007.

BELLENGUEZ, C.; STRANGE, A.; FREEMAN, C. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. **Bioinformatics Applications Note**, v.28, n.1, p.134-135, 2012.

BENJAMIN, Y.; HOCHBERG, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society Series B**, v.57, p.289-300, 1995.

BHATTACHARYA, S.; BUNICK, C.G.; CHAZIN, W.J. Target selectivity in EF-hand calcium binding proteins. **Biochimica et Biophysica Acta**, v.1742, p.69-79, 2004.

BLASCHEK M.; KAYA A.; ZWALD, N.; MEMILI, E.; KIRKPATRICK, B.W. A whole-genome association analysis of noncompensatory fertility in Holstein bulls. **Journal of Dairy Science**, v.9, 4695-4699, 2011.

BOLIGON, A.A.; RORATO, P.R.N.; ALBUQUERQUE, L.G. Correlações genéticas entre medidas de perímetro escrotal e características produtivas e reprodutivas de fêmeas da raça Nelore. **Revista Brasileira de Zootecnia**, v.36, p.565-571, 2007.

BOLIGON, A.A.; ALBUQUERQUE, L.G. Correlações genéticas entre escores visuais e características reprodutivas em bovinos Nelore usando inferência bayesiana. **Pesquisa Agropecuária Brasileira**, v.45, p.1412-1418, 2010.

BONI, R.; GUALTIERI, R.; TALEVI, R.; TOSTI, E. Calcium and other ion dynamics during gamete maturation and fertilization. **Theriogenology**, v.68, S156-S164, 2007.

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

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

CattleQTLdb. 2013. Release 19. <http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>. Acesso em: 10 de Jan. 2013.

CAVALCANTE, F.A.; MARTINS FILHO, R.; CAMPELLO, C.C.; LÔBO, R.N.B.; MARTINS, G.A. Período de serviço em rebanho Nelore na Amazônia Oriental. **Revista Brasileira de Zootecnia**, v.30, n.5, p.1456-1459, 2001.

CHAN, E.K.F.; HAWKEN, R.; REVERTER, A. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. **Animal Genetics**, v.40, n.2, p.149-56, 2009.

CHAZIN, W.J. Relating form and function of EF-hand calcium binding proteins. **Accounts of Chemical Research**, v.44, p.171-179, 2011.

CHENOWETH, P.J. Aspects of reproduction in female *Bos indicus* cattle: a review. **Australian Veterinary Journal**, v.71, n.12, p.422-426, 1994.

CLARKE, H.G.; HOPE, S.A.; BYERS, S.; RODGERS, R.J. Formation of ovarian follicular fluid may be due to the osmotic potential of large glycosaminoglycans and proteoglycans. **Reproduction**, v.132, p.119-131, 2006.

CLOP, A.; VIDAL, O.; AMILLS, M. Copy number variation in the genomes of domestic animals. **Animal Genetics**, v.43, p.501-517, 2012.

COLE, J.B.; VANRADEN, P.M.; O'CONNELL, J.R.; VAN TASSELL, C.P.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; WIGGANS, G.R. Distribution and location of genetic effects for dairy traits. **Journal of Dairy Science**, v.92, p.2931-2946, 2009.

COLE, J.B.; WIGGANS, G.R.; MA, L.; SONSTEGARD, T.S.; LAWLOR, T.J.; CROOKER, B.A.; VAN TASSELL, C.P.; YANG, J.; WANG, S.; MATUKUMALLI, L.K.; DA, Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics**, v.12, 408, 2011.

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

COUTINHO, L.H.; ROSÁRIO, M.F. Biotecnologia Animal. **Estudos Avançados**, v.24, p.123-147, 2010.

CUSHMAN, R.A.; ALLAN, M.F.; KUEHN, L.A. Characterization of biological types of cattle: Indicator traits of fertility in beef cattle. **Revista Brasileira de Zootecnia**, v.37, p.116-121, 2008.

DALY, M.J.; RIOUX, J.D.; SCHAFFNER, S.F.; HUDSON, T.J.; LANDER, E.S. High-resolution haplotype structure in the human genome. **Nature Genetics**, v.29, n.2, p.229-232, 2001.

DAY, M.L.; IMAKAWA, K.; WOLFE, P.L.; KITTOCK, R.J.; KINDER, J.E. Endocrine mechanism of puberty in heifers. Role of hypothalamo-pituitary estradiol receptors in the negative feedback of estradiol on luteinizing hormone secretion. **Biology of Reproduction**, v.37, p.1054-1065, 1987.

DENG, H.W.; CHEN, W.M. Re: Biased tests of association: comparison of allele frequencies when departing from Hardy-Weinberg proportions. **American Journal of Epidemiology**, v.151, p.335-357, 2000.

DEVLIN, B.; ROEDER, K. Genomic control for association studies. **Biometrics**, v.55, p.997-1004, 1999.

DIAS, L.T.; EL FARO, L.; ALBUQUERQUE, L.G. 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.

DOBSON, H.; KAMONPATANA, M. A review of female cattle reproduction with special reference to a comparison between buffaloes, cows and zebu. **Journal of Reproduction and Fertility**, v.77, p.1-36, 1986.

DONATO, J.JR.; FRAZAO, R.; ELIAS, C.F. The PI3K signaling pathway mediates the biological effects of leptin. **Arquivos Brasileiros de Endocrinologia e Metabologia**, v.54, n.7, p.591-602, 2010.

ELER, J.P. Estratégias para antecipar a puberdade em novilhas de corte. In: IV SIMPÓSIO DE PRODUÇÃO DE GADO DE CORTE, 2004, Viçosa. **Anais...Viçosa**, 2004. p. 449-458.

FAN, B.; DU, Z.Q.; GORBACH, D.M.; ROTHCHILD, M.F. Development and application of high-density SNP arrays in genomic studies of domestic animals. **Asian-Australian Journal of Animal Science**, v.23, p.833-847, 2010.

FINLAY, E.K.; BERRY, D.P.; WICKHAM, B.; GORMLEY, E.P.; BRADLEY, D.G. A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. **PLoS One**, v.7, e30545, 2012.

FLICEK, P.; AHMED, I.; AMODE, M.R.; BARRELL, D.; BEAL, K.; BRENT, S.; CARVALHO-SILVA, D.; CLAPHAM, P.; COATES, G.; FAIRLEY, S. et al. Ensembl 2013. **Nucleic Acids Research**, v.41, 2013.

FORTES, M.R.; REVERTER, A.; ZHANG, Y.; COLLIS, E.; NAGARAJ, S.H.; JONSSON, N.N.; PRAYAGA, K.C.; BARRIS, W.; HAWKEN, R.J. Association weight matrix for the genetic dissection of puberty in beef cattle. **Proceedings of the National Academy of Sciences USA**, v.107, n.31, p.13642-13647, 2010.

FORTES, M.R.S.; REVERTER, A.; NAGARAJ, S.H. A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. **Journal Animal Science**, v.89, p.1669-1683, 2011.

FORTES, M.R.S.; LEHNERT, S.A.; BOLORMAA, S.; REICH, C.; FORDYCE, G.; CORBET, N.J.; WHAN, V.; HAWKEN, R.J.; REVERTER, A. Finding genes for economically important traits: Brahman cattle puberty. **Animal Production Science**, v.52, p.143-150, 2012a.

FORTES, M.R.S.; LI, Y.; COLLIS, E.; ZHANG, Y.; HAWKEN, R.J. The IGF1 pathway genes and their association with age of puberty in cattle. **Animal Genetics**, v.44, p.91-95, 2012b.

FRAZER, K. A.; MURRAY, S. S.; SCHORK, N. J.; TOPOL, E.J. Human genetic variation and its contribution to complex traits. **Nature Reviews**, v.10, p.241-251, 2009.

FROMMLET, F.; RUHALTINGER, F.; TWAROG, B.; BOGDAN, M. Modified versions of Bayesian information criterion for genome-wide association studies. **Computational Statistics and Data Analysis**, v.56, p.1038-1051, 2012.

GABRIEL, S.B.; SCHAFFNER, S.F.; NGUYEN, H.; MOORE, J.M.; ROY, J.; BLUMENSTIEL, B.; HIGGINS, J.; DEFELICE, M.; LOCHNER, A.; FAGGART, M.; LIU-CORDERO, S.N.; ROTIMI, C.; ADEYEMO, A.; COOPER, E.; WARD, R.; LANDER, E.S.; DALY, M.J.; ALTSHULER, D. The structure of haplotype blocks in the human genome. **Science**, v.296, p.2225-2229, 2002.

GODDARD, M.E.; HAYES, B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews Genetics**, v.10, p.381-391, 2009.

GODDARD, K. A.; ZIEGLER, A.; WELLEK, S. Adapting the logical basis of tests for Hardy-Weinberg Equilibrium to the real needs of association studies in human and medical genetics. **Genetic Epidemiology**, v.33, n.7, p.569-580, 2009.

GUO, T.; LIU, X.F.; DING, X.B.; YANG, F.F.; NIE, Y.W.; AN, Y.J.; GUO, H. Fat-1 transgenic cattle as a model to study the function of ω -3 fatty acids. **Lipids in Health and Disease**, v.10, p.244, 2011.

HAWKEN, R.J.; ZHANG, Y.D.; FORTES, M.R.S.; COLLIS, E.; BARRIS, W.C. Genome-wide association studies of female reproduction in tropically adapted beef cattle. **Journal Animal Science**, v.90, n.5, p.1398-1410, 2011.

HAYES, J. B.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M.E. Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science**, v.92, p.433-443, 2009.

HENDERSON, C.R. **Applications of linear models in animal breeding**. 1984. Guelph, Ontario, Canada: University of Guelph.

HIENDLEDER, S.; BAUERSACHS, S.; BOULESTEIX, A.; BLUM, H.; ARNOLD, G.J.; FROHLICH, T.; WOLF, E. Functional genomics: tools for improving farm animal health and welfare. **Revue scientifique technique**, v.24, p.354-77, 2005.

HIRSCHHORN, J.N.; DALY, M.J. Genome-wide association studies for common diseases and complex traits. **Nature Reviews Genetics**, v.6, p.95-108, 2005.

HUANG, D.W.; SHERMAN, B.T.; LEMPICKI, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. **Nature Protocols**, v.4, n.1, p.44-57, 2009a.

HUANG, D.W.; SHERMAN, B.T.; LEMPICKI, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Research**, v.37, n.1, p.1-13, 2009b.

ILLUMINA Inc. Product literature. Disponível em <<http://www.illumina.com/support/literature.ilmn>>. Acesso em: 10 de Set. 2011.

IRELAND, J.J.; ZIELAK-STECIWKO, A.E.; JIMENEZ-KRASSEL, F.; FOLGER, J.; BETTEGOWDA, A.; SCHEETZ, D.; WALSH, S.; MOSSA, F.; KNIGHT, P.G.; SMITH, G.W. et al. Variation in the ovarian reserve is linked to alterations in intrafollicular estradiol production and ovarian biomarkers of follicular differentiation and oocyte quality in cattle. **Biology Reproduction**, v.80, p.954-964, 2009.

IRELAND, J.J.; SMITH, G.W.; SCHEETZ, D.; JIMENEZKRASSEL, F.; FOLGER, J.K.; IRELAND, J.L.; MOSSA, F.; LONERGAN, P.; EVANS, A.C. Does size matter in females? An overview of the impact of the high variation in the ovarian reserve on ovarian function and fertility, utility of anti-Mullerian hormone as a diagnostic marker for fertility and causes of variation in the ovarian reserve in cattle. **Reproduction, Fertility and Development**, v.23, 1-14, 2011.

JOHNSTON, D.J.; BARWICK, S.A.; CORBET, N.J. Genetics of heifer puberty in two tropical beef genotypes in northern Australia and associations with heifer- and steer-production traits. **Animal Production Science**, v.49, p.399-412, 2009.

JOHNSTON, D.J.; TIER, B.; GRASER, H.U. Beef cattle breeding in Australia with genomics: opportunities and needs. **Animal Production Science**, v.52, p.100-106, 2012.

KALIVE, M.; FAUST, J.J.; KOENEMAN, B.A.; CAPCO, D.G. Involvement of the PKC family in regulation of early development. **Molecular Reproduction and Development**, v.77, p.95-104, 2010.

KEMPER, K.E.; GODDARD, M.E. Understanding and predicting complex traits: knowledge from cattle. **Human Molecular Genetics**, 2012. Disponível em: < doi: 10.1093/hmg/dds332 >.

KHATKAR, M.S.; NICHOLAS, F.W.; COLLINS, A.R.; ZENGER, K.R.; CAVANAGH, J.A.; BARRIS, W.; SCHNABEL, R.D.; TAYLOR, J.F.; RAADSMA, H.W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics**, v.9, 187, 2008.

KIM, E.S.; KIRKPATRICK, B.W. Linkage disequilibrium in the North American Holstein population. **Animal Genetics**, v.40, p.279-288, 2009.

KINSELLA, R.J.; KÄHÄRI, A.; HAIDER, S.; ZAMORA, J.; PROCTOR, G.; SPUDICH, G.; ALMEIDA-KING, J.; STAINES, D.; DERWENT, P.; KERHORNOU, A.; KERSEY, P.; FLICEK, P. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford), 2011. Disponível em: < <http://www.ncbi.nlm.nih.gov/pmc/articles/pmc3170168/> >. Acesso em: 05 de Jan. 2013.

KOZIOL, M.J.; GARRETT, N.; GURDON, J.B. Tpt1 activates transcription of oct4 and nanog in transplanted somatic nuclei. **Current Biology**, v.17, n.9, p.801-807, 2007.

LAUREANO, M.M.M.; OTAVIANO, A.R.; LIMA, A.F.L.; COSTA, R.B.; SALMAN, A.K.D.; SENA, J.A.D.; TONHATI, H.; ALBUQUERQUE, L.G. Characterization and polymorphism screening of IGF-I and prolactin genes in Nelore heifers. **Italian Journal Animal Science**, v.8, p.277-283, 2009.

LAURIE, C.C.; DOHENY, K.F.; MIREL, D.B.; PUGH, E.H.; BIERUT, L.J.; BHANGALE, T.; BOEHM, F.; CAPORASO, N.E.; CORNELIS, M.C.; EDENBERG, H.J. et al. Quality control and quality assurance in genotypic data for genome-wide association studies. **Genetic Epidemiology**, v.34, n.6, p.591-602, 2010.

LI, S.; CHEN, X.; DING, Y.; LIU, X.; WANG, Y.; HE, J. Expression of translationally controlled tumor protein (TCTP) in the uterus of mice of early pregnancy and its possible significance during embryo implantation. **Human Reproduction**, v.26, p.2972-2980, 2011.

LITONJUA, A.A.; CELEDO, J.C. **Gene Association Studies**. Encyclopedia of Respiratory Medicine. Oxford: Academic Press, London, 2006, 240-247p.

LIU, G.; BICKHART, D. Copy number variation in the cattle genome. **Functional & Integrative Genomics**, p.1-16, 2012.

LUNA-NEVAREZ, P.; BAILEY, D.W.; BAILEY, C.C. VANLEEUEWEN, D.M.; ENNS, R.M.; SILVER, G.A.; DEATLEY, K.L.; THOMAS, M.G. Growth characteristics, reproductive performance, and evaluation of their associative relationships in Brangus cattle managed in a Chihuahuan desert production system. **Journal Animal Science**, v.88, p.1891-1904, 2010.

MALTECCA, C.; GRAY, K.A.; WEIGEL, K.A.; CASSADY, J.P.; ASHWELL, M. A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. **Animal Genetics**, v.42, p.585-591, 2011.

MAQUIVAR, M.G. **Nutritional regulation of precocious puberty in heifers**. 2011. 182p. Tesis - The Ohio State University, 2011.

MATUKUMALLI, K.L.; LAWLEY, C.T.; SCHNABEL, R.D.; TAYLOR, J.F.; ALLAN, M.F.; HEATON, M.P.; O'CONNELL, J.; MOORE, S.S.; SMITH, T.P.; SONSTEGARD, T.S.; VAN TASSELL, C.P. Development and characterization of a high density SNP genotyping assay for cattle. **PLoS One**, v.4, n.4, e5350, 2009.

MEIRELLES, S.L.; ESPASANDIN, A.C.; MATTAR, M.; QUEIROZ, S.A. Genetic and environmental effects on sexual precocity traits in Nellore cattle. **Revista Brasileira de Zootecnia**, v.38, n.8, p.1488-1493, 2009.

MERICO, D.; ISSERLIN, R.; STUEKER, O.; EMILI, A.; BADER, G.D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. **PLoS One**, v.5, n.11, e13984, 2010.

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

MILAZZOTTO, M.P.; RAHAL, P.; NICHI, M.; MIRANDA-NETO, T.; TEIXEIRA, L.A.; FERRAZ, J.B.S.; ELER, J.P.; CAMPAGNARI, F.; GARCIA, J.F. 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, p.274-279, 2008.

MONSALVES, F.M. **Valor econômico e impacto da seleção para precocidade reprodutiva de fêmeas na raça Nelore**. 2008. 52p. Dissertação (Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinárias – UNESP, 2008.

MOORHEAD, M.; HARDENBOL, P.; SIDDIQUI, F. FALKOWSKI, M.; BRUCKNER, C.; IRELAND, JAMES.; JONES, H.B.; JAIN, M.; WILLIS, T.D.; FAHAM, M. Optimal genotype determination in highly multiplexed SNP data. **European Journal of Human Genetics**, v.14, p.207-15, 2006.

MOSKVINA, V.; SCHMIDT, K.M. On multiple-testing correction in genome-wide association studies. **Genetic Epidemiology**, v.32, p.567-573, 2008.

MUCARI, T.B.; ALENCAR, M.M.; BARBOSA, P.F.; BARBOSA, R.T. Análise genética do período de gestação em animais de um rebanho Canchim: estimação de parâmetros genéticos e escolha entre modelos animais alternativos. **Revista Brasileira de Zootecnia**, v.40, n.6, p.1211-1216, 2011.

NOGUEIRA, G.P. Puberty in South American *Bos indicus* (Zebu) cattle. **Animal Reproduction Science**, v.82-83, p.361-372, 2004.

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

PARE G. Genome-wide association studies—data generation, storage, interpretation, and bioinformatics. **Journal of Cardiovascular Translational Research**, v.3, p.183-188, 2010.

PAUSCH H.; FLISIKOWSKI, K.; JUNG, S.; EMMERLING, R.; EDEL, C.; GÖTZ, K.U.; FRIES, R. Genome-Wide Association Study Identifies Two Major Loci Affecting Calving Ease and Growth-Related Traits in Cattle. **Genetics**, v.187, p.289-297, 2011.

PAUSCH, H.; WANG, X.; JUNG, S.; KROGMEIER, D.; EDEL, C.; EMMERLING, R.; GOTZ, K.-U. FRIES, R. Identification of QTL for UV-protective eye area pigmentation in cattle by progeny phenotyping and genome-wide association analysis. **PLoS One**, v.7, e36346, 2012.

PENÑAGARICANO, F.; WEIGEL, K.A.; KHATIB, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. **Animal Genetics**, v.43, s.1, 65-71, 2012.

PEREIRA, E.; ELER, J. P.; FERRAZ, J. B.S. Correlação genética entre perímetro escrotal e algumas características reprodutivas na raça Nelore. **Revista Brasileira de Zootecnia**, v.29, n.6, p.1676-1683, 2000.

PEREIRA, E.; ELER, J.P.; COSTA, F.A.A.; FERRAZ, J.B.S. Análise genética da idade ao primeiro parto e do perímetro escrotal em bovinos da raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.53, n.1, p.116-121, 2001.

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.

PETERS, S.O.; KIZILKAYA, K.; GARRICK, D.J.; FERNANDO, R.L.; REECY, J.M.; WEABER, R.L., SILVER, G.A.; THOMAS, M.G. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. **Journal of Animal Science**, 2012. Disponível em: <doi: 10.2527/jas.2012-5580>.

PONGPANICH, M.; SULLIVAN, P.F.; TZENG, J.Y. A quality control algorithm for filtering SNP in genome-wide association studies. **Bioinformatics**, v.26, p.1731-1737, 2010.

PRASAD, V.; OKUNADE, G.W.; MILLER, M.L.; SHULL, G.E. Phenotypes of SERCA and PMCA knockout mice. **Biochemical and Biophysical Research Communications**, v.322, n.4, p.1192-1203, 2004.

PRICE, A.L.; PATTERSON, N.J.; PLENGE, R.M.; WEINBLATT, M.E.; SHADICK, N.A.; REICH, D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, v.38, n8, p.904-909, 2006.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.R.; BENDER, D.; MALLER, J.; SKLAR, P.; BAKKER, P.I.W.; DALY, M.J.; SHAM, P.C. PLINK: a toolset for whole-genome association and population-based linkage analysis. **American Journal of Human Genetics**, v.81, p.559-574, 2007.

RAMAKRISHNAPPA, N.; RAJAMAHENDRAN, R.; LIN, Y.; LEUNG, P.C.K. GnRH in non-hypothalamic reproductive tissues. **Animal Reproduction Science**, v.88, p.95-113, 2005.

RANDEL, R.D. Seasonal effects on female reproductive functions in the bovine (Indian breeds). **Theriogenology**, v.21, p.170-185, 1984.

RAWLINGS, N.C.; EVANS, A.C.O.; HONARAMOOZ, A.; BARTLEWSKIA, P.M. Antral follicle growth and endocrine changes in prepubertal cattle, sheep and goats. **Animal Reproduction Science**, v.78, p.259-270, 2003.

REDDY, P.; LIU, L.; ADHIKARI, D.; JAGARLAMUDI, K.; RAJAREDDY, S.; SHEN, Y.; DU, C.; TANG, W.; HAMALAINEN, T.; PENG, S.L.; LAN, Z.J.; COONEY, A.J.; HUHTANIEMI, I.; LIU, K. Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool. **Science**, v.319, p.611-613, 2008.

REDING, S.C.; STEP NOSKI, A.L.; CLONINGER, E.W.; OATLEY, J.M. THY1 is a conserved marker of undifferentiated spermatogonia in the pre-pubertal bull testis. **Reproduction**, v.139, p.893-903, 2010.

RIGON, C.; ANDRISANI, A.; FORZAN, M.; D'ANTONA, D.; BRUSON, A.; COSMI, E.; AMBROSINI, G.; TIBONI, G.M.; CLEMENTI, M. Association study of AMH and AMHRII polymorphisms with unexplained infertility. **Journal of Fertility and Sterility**, v.94, p.1244-1248, 2010.

RINCON, G.; WEBER, K.L.; Van EENENNAAM, A.L. Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. **Journal Dairy Science**, v.94, p.6116-6121, 2011.

ROCHA, J.C.M.C.; TONHATI, H.; ALENCAR, M.M.; LÔBO, R.B. Componentes de variância para o período de gestação em bovinos de corte. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.57, n.6, p.784-791, 2005.

ROSSE, C. LINCH, M.; KERMORGANT, S.; CAMERON, A.J.; BOECKELER, K.; PARKER, P.J. PKC and the control of localized signal dynamics. **Nature Review Molecular Cellular Biology**, v.11, p.103-112, 2010.

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

SAMBROOK, J.; FRISTCH, E.F.; MANIATIS, T. **Molecular cloning. A laboratory manual**. Cold Spring Harbour Laboratory Press: New York, 1989. E3-E4.

SANTANA Jr, M.L.; ELER, J.P.; FERRAZ, J.B.S.; MATTOS, E.C. Genetic relationship between growth and reproductive traits in Nellore cattle. **Animal**, v.6, n.4, p.565-570, 2012.

SCHULMAN, N.F.; SAHANA, G.; ISO-TOURU, T.; MCKAY, S.D.; SCHNABEL, R.D.; LUND, M.S.; TAYLOR, J.F.; VIRTÄ, J.; VILKKI, J.H. Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. **Animal Genetics**, v.42, p.263-269, 2011.

SCHWARZENBACHER, H.; DOLEZAL, M.; FLISIKOWSKI, K.; SEEFRIED, F.; WURMSER, C.; SCHLÖTTERER, C.; FRIES, R. Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. **BMC Genomics**, v.13, 48, 2012.

SEMMELMANN, C.E.N., LOBATO, J.F.P., ROCHA, M.G. Efeito de sistemas de alimentação no ganho de peso e desempenho reprodutivo de novilhas Nelore acasaladas aos 17/18 meses. **Revista Brasileira de Zootecnia**, v.30, n.3, p.835-843, 2001.

SEN, A.; CHOUDHARY, E.; INSKEEP, E.K.; FLORES, J.A. Effects of selective protein kinase C isozymes in prostaglandin F 2alpha-induced Ca²⁺ signaling and luteinizing hormone induced progesterone accumulation in the mid-phase bovine corpus luteum. **Biology Reproduction**, v.72, p.976-84, 2005.

SHIOTSUKI, L.; SILVA, J.A.V.; ALBUQUERQUE, L.G. Associação genética da prenhez aos 16 meses com o peso a desmama e o ganho de peso em animais da raça Nelore. **Revista Brasileira de Zootecnia**, v.38, n.7, p.1211-1217, 2009.

SILVA, J.A.V.; OLIVEIRA, H.N.; ELER, J.P.; FERRAZ, J.B.S. Parâmetros genéticos de prenhez em novilhas aos 14 meses e de perímetro escrotal de machos nelore usando inferência bayesiana. In: SIMPÓSIO NACIONAL DE MELHORAMENTO ANIMAL, IV, 2002. **Anais...** SBMA, 2002.

SILVA, J.A.V.; DIAS, L.T.; ALBUQUERQUE, L.G. Estudo Genético da Precocidade Sexual de Novilhas em um Rebanho Nelore. **Revista Brasileira de Zootecnia**, v.34, n.5, p.1568-1572, 2005.

SMITH, J.G.; NEWTON-CHEH, C. Genome-wide association study in humans. **Methods in Molecular Biology**, v.573, p.231-258, 2009.

SNELLING, W.M.; CUSHMAN, R.A.; FORTES, M.R.S. How SNP chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. **Journal Animal Science**, v.90, n.4, p.1152-1165, 2011.

SODHI, M.; SCHOOK, LB. **Genomics Research: Livestock Production**. Encyclopedia of Biotechnology in Agriculture and Food, 2011, p.311-315.

SPENCER, T.E.; SANDRA, O. WOLF, E. Genes involved in conceptus-endometrial interactions in ruminants: insights from reductionism and thoughts on holistic approaches. **Reproduction**, v.135, p.165-179, 2008.

STEVENS, E.L.; HECKENBERG, G.; ROBERSON, E.D.O. Inference of relationships in population data using Identity-by-Descent and Identity-by-State. **PLoS Genetics**, v.7, e1002287, 2011.

SU, G.; BRØNDUM, R. F.; MA, P.; GULDBRANDTSEN, B.; AAMAND, G.P.; LUND, M.S. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. **Journal Dairy Science**, v.95, p.4657-4665, 2012.

SUN, Y.V. Integration of biological networks and pathways with genetic association studies. **Human Genetics**, v.131, p.1677-1686, 2012.

TABANGIN, M.E.; WOO, J.G.; MARTIN, L.J. The effect of minor allele frequency on likelihood of obtaining false positives. **BMC Proceedings**, v.3, S41, 2009.

TANAKA, A.L.R. **Eficiência reprodutiva de fêmeas Nelore**. 2010. 116p. Tese (Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinárias – UNESP, 2010.

TANI, T.; SHIMADA, H.; KATO, Y.; TSUNODA, Y. Bovine oocytes with the potential to reprogram somatic cell nuclei have a unique 23-kDa protein, phosphorylated transcriptionally controlled tumor protein (TCTP). **Cloning Stem Cells**, v.9, n.2, p.267-280, 2007.

TEO, Y.Y.; INOUE, M.; SMALL, K.S.; GWILLIAM, R.; DELOUKAS, P.; KWIATKOWSKI, D.P.; CLARK, T.G. A genotype calling algorithm for the Illumina BeadArray platform. **Bioinformatics**, v.23, p.2741-2746, 2007.

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

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

THE INTERNATIONAL HAPMAP CONSORTIUM. The international hapmap project. **Nature**, v.426, p.789-796, 2003.

THE WELLCOME TRUST CASE CONTROL CONSORTIUM. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. **Nature**, v.447, p.661-678, 2007.

Van EENENNAAM, A.L.; Drake, D.J. Where in the beef-cattle supply chain might DNA tests generate value? **Animal Production Science**, v.52, p.185-196, 2012.

VANRADEN, P.M.; VAN TASSELL, C.P.; WIGGANS, G.R.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; SCHENKEL, F.S. Invited review: reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v.92, p.16-24, 2009.

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

VISSER, J.A.; DE JONG, F.H.; LAVEN, J.S.E.; THEMMEN, A.P.N. Anti-Mullerian hormone: a new marker for ovarian function. **Reproduction**, v.131, p.1-9, 2006.

WANG, Q.S.; ZHAO, H.B.; PAN, Y.C. SNPknow: a web server for functional annotation of cattle SNP markers. **Canadian Journal of Animal Science**, v.91, p.247-253, 2011.

WANG, W.Y.; BARRATT, B.J.; CLAYTON, D.G.; TODD, J.A. Genome-wide association studies: theoretical and practical concerns. **Nature Reviews Genetics**, v.6, p.109-118, 2005.

WEINER, M.P.; HUDSON, T.J. Introduction to SNP: discovery of markers for disease. **Biotechniques**, v.10, s.4-7, p.12-13, 2002.

WEIR, B.S.; ANDERSON, A.D.; HEPLER, A.B. Genetic relatedness analysis: Modern data and new challenges. **Nature Reviews Genetics**, v.7, p.771-780, 2006.

WEIR, B.S.; LAURIE, C.C. Characterizing allelic association in the genome era. **Genetics Research**, v.92, n.5-6, p.461-470, 2010.

WILLIAMS, G.L.; AMSTALDEN, M. Understanding postpartum anestrus and puberty in the beef female. In: APPLIED REPRODUCTIVE STRATEGIES IN BEEF CATTLE, 2010, San Antonio. **Proceedings...**San Antonio, USA, 2010. p.55-71.

WILTBANK, J.N.; KASSON, C.W.; INGALLS, J.E. Puberty in cross-bred and straightbred beef heifers on two levels of feed. **Journal Animal Science**, v.29, p.602-605, 1969.

WITTKE-THOMPSON, J.K.; PLUZHNIKOV, A.; COX, N.J. Rational inferences about departures from Hardy-Weinberg equilibrium. **American Journal of Human Genetics**, v.76, p.967-986, 2005.

YAN, T.; HOU, B.; YANG, Y. Correcting for cryptic relatedness by a regression-based genomic control method. **BMC Genetics**, v.10, 78, 2009.

ZEGGINI, E.; RAYNER, W.; MORRIS, A.P. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. **Nature Genetics**, v.37, p.1320-1322, 2005.

ZHANG, H.; WANG, Z.; WANG, S.; LI, H. Progress of genome wide association study in domestic animals. **Journal of Animal Science and Biotechnology**, v.3, n.1, p.26, 2012a.

ZHANG, J.; OTOLEDO, S.M.; PANDEY, B.N.; GUO, G.; PAIN, D.; LI, H.; AZZAMA, E.I. Role of the translationally controlled tumor protein in DNA damage sensing and repair. **Proceedings of the National Academy of Sciences USA**, v.109, E926-E933, 2012b.

ZHANG Z.; ERSOZ, E.; LAI, C.Q.; TODHUNTER, R.J.; TIWARI, H.K.; GORE, M.A.; BRADBURY, P.J.; YU, J.; ARNETT, D.K.; ORDOVAS, J.M.; BUCKLER, E.S. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**, v.42, p.355-360, 2010.

ZHENG, W .; LIU, K. The emerging role of maternal phosphatidylinositol 3kinase (PI3K) signaling in manipulating mammalian preimplantation embryogenesis. **Cell Cycle**, v.10, p.178-179, 2011.

ZHENG, W.; NAGARAJU, G.; LIU, Z.; LIU, K. Functional roles of the phosphatidylinositol 3-kinases (PI3Ks) signaling in the mammalian ovary. **Molecular and Cellular Endocrinology**, v.356, p.24-30, 2012.

ZIEGLER, A.; KÖNIG, I.R. **A Statistical Approach to Genetic Epidemiology. Concepts and Applications**. 2 ed. Wiley-VCH, Weinheim, 2006. p.497.

ZIEGLER, A.; KÖNIG, I.R.; THOMPSON, J.R. Biostatistical aspects of genome-wide association studies. **Biometrical Journal**, v.50, p.1-21, 2008.

ZHOU X.; STEPHENS, M. Genome-wide efficient mixed-model analysis for association studies. **Nature Genetics**, v.44, p.821-824, 2012.

10. Apêndice

10.1. Tabelas Suplementares

Tab. suppl. 1. Número de SNP removidos de cada experimento de acordo com o critério de controle de qualidade (QC) aplicado.

<i>Critério de QC</i>	<i>Experimento 1</i>	<i>Experimento 2</i>		<i>Experimento 3</i>	
	<i>(IPP)</i>	<i>EBV1 (IPP)</i>	<i>EBV2 (PG)</i>	<i>EBV1 (IPP)</i>	<i>EBV2 (PG)</i>
"Call rate" > 98%	97.344	96.854	100.469	97.524	95.535
MAF > 2%	188.762	195.018	193.786	189.852	188.617
Desvio de HWE ($p > 10^{-5}$)	2.509	8.260	11.691	13.980	14.751

Tab. supl. 2. Distribuição do número de marcadores por cromossomo antes e após o controle de qualidade.

BTA	Cromossomo em pb ^A	N° de SNP no painel	N° de SNP após QC			
			Experimento 1 ^B		Experimento 2	
			EBV1 ^C	EBV2 ^D	EBV1 ^E	EBV2 ^F
1	158.337,067	46.891	29.768	28.374	29.671	29.823
2	137.070,424	40.356	25.663	23.162	25.202	25.467
3	121.430,405	35.985	22.989	21.176	23.108	23.199
4	120.829,699	35.357	21.806	20.078	21.763	21.889
5	121.191,424	35.251	21.040	19.067	20.876	20.960
6	119.458,736	35.821	24.081	21.897	23.415	23.521
7	112.638,659	33.552	20.864	19.725	21.094	21.233
8	113.384,836	33.834	22.821	20.672	22.553	22.611
9	105.708,250	31.351	20.436	19.228	20.415	20.594
10	104.305,016	30.868	18.671	17.354	18.697	18.760
11	107.310,763	32.252	19.506	17.867	19.869	19.929
12	91.163,125	26.636	15.588	14.432	15.557	15.670
13	84.240,350	23.839	14.515	13.484	15.026	15.048
14	84.648,390	25.009	17.061	15.593	17.149	17.198
15	85.296,676	25.188	15.141	13.977	15.494	15.586
16	81.724,687	24.428	15.405	13.933	15.364	15.467
17	75.158,596	22.456	14.354	13.344	14.505	14.582
18	66.004,023	19.583	12.056	11.189	12.565	12.610
19	64.057,457	19.091	11.251	9.724	11.382	11.458
20	72.042,655	21.653	13.219	12.158	13.310	13.398
21	71.599,096	21.395	13.204	12.243	13.621	13.691
22	61.435,874	18.150	10.973	10.046	11.120	11.201
23	52.530,062	15.420	9.816	8.896	9.730	9.790
24	62.714,930	18.762	11.656	10.691	11.966	11.968
25	42.904,170	13.026	8.052	7.129	8.011	8.053
26	51.681,464	15.355	10.686	9.265	10.047	10.032
27	45.407,902	13.304	8.132	7.472	8.353	8.394
28	46.312,546	13.179	8.396	7.489	8.356	8.422
29	51.505,224	14.917	8.943	8.105	8.741	8.791
X	148.823,899	40.253	-	-	-	-
Y	-	1.423	-	-	-	-
Mit	~16.338	346	-	-	-	-
0 ^G	9.828,056	1.867	-	-	-	-
Total	-2.640 10e-6	786.798	466.093	427.770	466.960	469.345

^AUMD 3.1; ^Bn=96; ^Cn=299; ^Dn=510; ^En=1.076; ^Fn=1.221; ^Gmarcadores não posicionados em nenhum cromossomo.

Tab. supl. 3. Densidade de marcadores do BovineHD BeadChip antes e após o controle de qualidade.

<i>N inicial</i>	<i>Média (kb)</i>	<i>Mediana (kb)</i>	<i>Experimento</i>	<i>Subgrupo</i>	<i>N pós-QC</i>	<i>Média (kb)</i>	<i>Mediana (kb)</i>
			1	-	466.093	5,41	3,52
			2	1	427.770	5,86	3,75
786.798	3,40	2,69		2	426.171	5,88	3,76
			3	1	466.960	5,37	3,52
				2	469.345	5,34	3,51

Tab. supl. 4. Estatísticas descritivas dos efeitos de substituição alélica estimados a partir do conjunto de marcadores testados em cada análise do tipo GWAS.

Parâmetro	Idade ao Primeiro Parto			Período Gestacional		
	EXP1	EXP2	EXP3	EXP2	EXP3	EXP3
Média	0,015	-0,012	-0,010	0,010	-0,001	-0,001
Mediana	0,010	-0,007	-0,006	0,008	-0,001	-0,001
Variância	0,304	0,249	0,066	0,038	0,009	0,009
Desvio-padrão	0,551	0,499	0,258	0,196	0,096	0,096
Min.	-4,13	-4,18	-2,35	-1,57	-0,68	-0,68
Q1.	-0,29	-0,28	-0,15	-0,10	-0,05	-0,05
Q3.	0,31	0,26	0,13	0,12	0,05	0,05
Máx.	4,43	3,64	1,90	1,54	0,69	0,69
Curtose	2,90	2,79	2,83	2,51	2,54	2,54
Assimetria	0,10	-0,10	-0,10	0,08	-0,01	-0,01

* estimados em dias, exceto EXP1 estimado em meses.

Tab. supl. 5. Distribuição dos SNP “top” por cromossomo selecionados para as análises funcionais.

<i>N SNP “top” /BTA</i>	<i>Idade ao Primeiro Parto</i>			<i>Período Gestacional</i>		
	<i>EXP1</i>	<i>EXP2</i>	<i>EXP3</i>	<i>EXP2</i>	<i>EXP3</i>	
1	2	2	1	2	12	
2	-	1	4	5	1	
3	2	1	5	-	2	
4	1	-	1	1	11	
5	-	9	1	3	5	
6	1	3	2	3	2	
7	2	6	7	1	-	
8	-	8	1	4	1	
9	4	7	2	1	1	
10	37	1	3	1	3	
11	7	3	7	1	-	
12	5	2	2	1	1	
13	1	4	-	-	1	
14	-	-	14	9	1	
15	-	8	2	-	2	
16	2	4	-	1	2	
17	3	-	3	-	2	
18	3	-	-	-	1	
19	8	2	-	5	3	
20	-	1	1	-	3	
21	-	-	-	-	-	
22	7	-	-	-	1	
23	3	-	4	-	-	
24	4	5	-	-	3	
25	5	-	6	-	-	
26	-	46	1	-	-	
27	-	-	3	-	-	
28	-	1	-	-	4	
29	-	-	2	-	2	

Tab. supl. 6. Sumário dos conjuntos de dados obtidos das análises do tipo GWAS e empregados nas etapas de análise funcional.

<i>Exp.</i>	<i>Subgr.</i>	<i>N</i>	<i>SNP após QC</i>	<i>SNP "top"</i>	<i>SNP intra-gênicos</i>	<i>Genes mapeados</i>	<i>Genes com termos anotados *</i>
1	-	96	466.093	96	38	91	53 (58%)
2	1	297	427.770	114	73	104	63 (61%)
	2	508	426.171	38	10	37	13 (35%)
3	1	1068	466.960	72	29	72	50 (69%)
	2	1221	469.345	64	35	50	30 (60%)

* número e respectivo aproveitamento de genes com informação disponível para mineração de termos funcionais.

Tab. supl. 7. Posicionamento dos SNP "top" utilizados nas análises funcionais com relação à estrutura do gene candidato.

Classificação	Idade ao Primeiro Parto			Período Gestacional	
	EXP1	EXP2	EXP3	EXP2	EXP3
<i>Intrônico</i>	29	64	22	5	33
<i>Intergênico</i>	58	41	43	28	29
<i>Upstream^A</i>	4	2	6	3	2
<i>Downstream^B</i>	5	4	-	2	-
<i>Sinônimo</i>	-	2	-	-	-
<i>Não-sinônimo</i>	-	1	1	-	-
<i>Total</i>	96	114	72	38	64

^AVariante da sequência 5' do gene;^BVariante da sequência 3' do gene;

Tab. supl. 8. Marcadores SNP localizados em exons de genes candidatos identificados para EBV1 (IPP)

<i>Exp.</i>	<i>Nome</i>	<i>BTA</i>	<i>Posição</i>	<i>Consequência para o transcrito</i>	<i>Variantes alélicas</i>	<i>Ensembl ID</i>
2	rs109403435	26	22.653,239	sinônimo	T/C	ENSBTAG0000000005780
2	rs135360696	19	25.185,813	sinônimo	A/G	ENSBTAG000000014806
2	rs137388765	15	83.605,457	não-sinônimo	T/G	ENSBTAG000000002450
3	rs42490238	9	24.127,805	não-sinônimo	C/T	ENSBTAG0000000003919

10.2. Figuras Suplementares

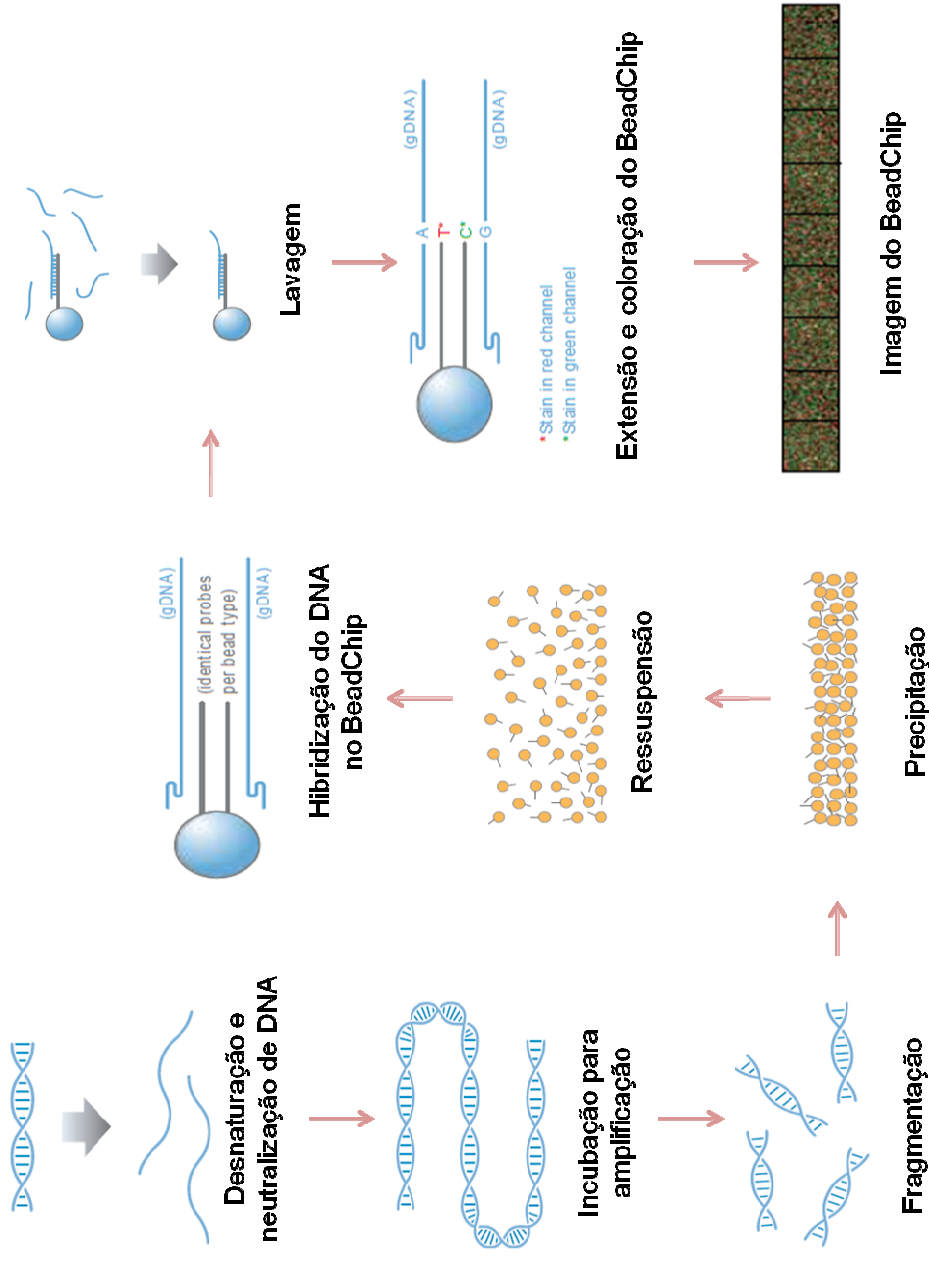


Fig. suppl. 1. Ilustração do ensaio de genotipagem Infinium HD (Adaptado de Illumina, 2011).

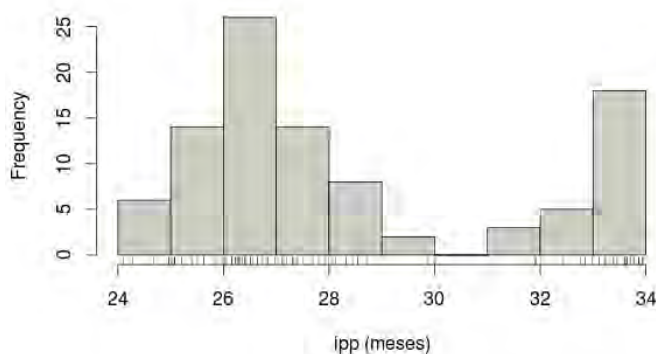
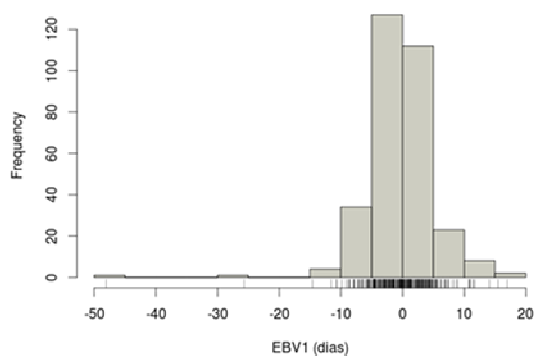
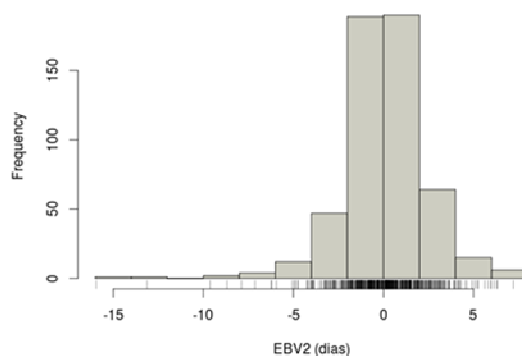


Fig. supl. 2. Experimento1. Distribuição dos dados de IPP (meses). Linhas verticais abaixo do eixo x representam os animais individualmente.

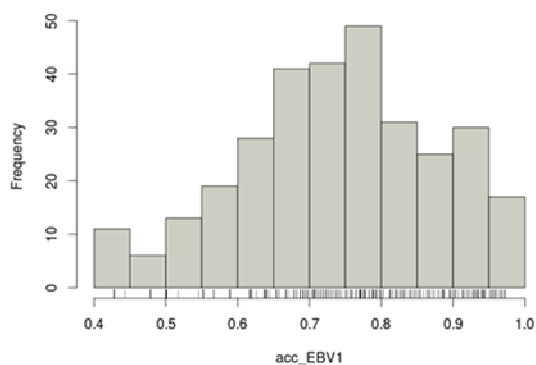
A



C



B



D

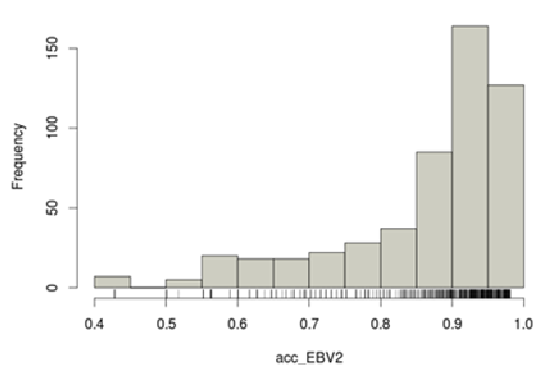


Fig. supl. 3. Experimento 2. Distribuição dos EBVs (dias) e suas respectivas acurácias. EBV1 (IPP) (figuras A e B) e EBV2 (PG) (figuras C e D), respectivamente. Linhas verticais abaixo do eixo x representam os animais individualmente.

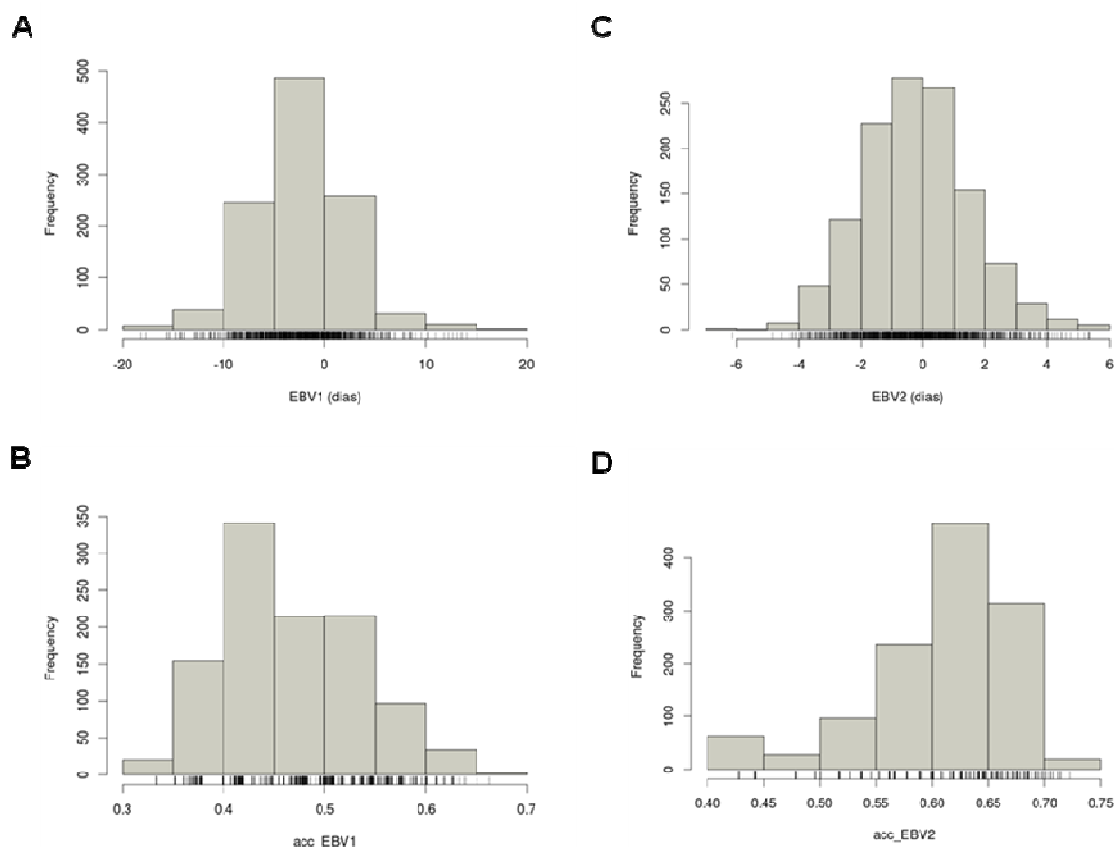


Fig. supl. 4. Experimento 3. Distribuição dos EBVs (dias) e suas respectivas acurácias. EBV1 (IPP) (figuras A e B) e EBV2 (PG) (figuras C e D), respectivamente. Linhas verticais abaixo do eixo x representam os animais individualmente.

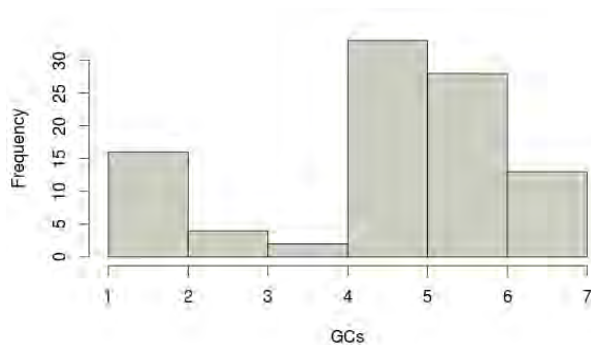


Fig. supl. 5. Experimento1. Distribuição dos grupos de contemporâneas (GCs) para os dados de IPP.

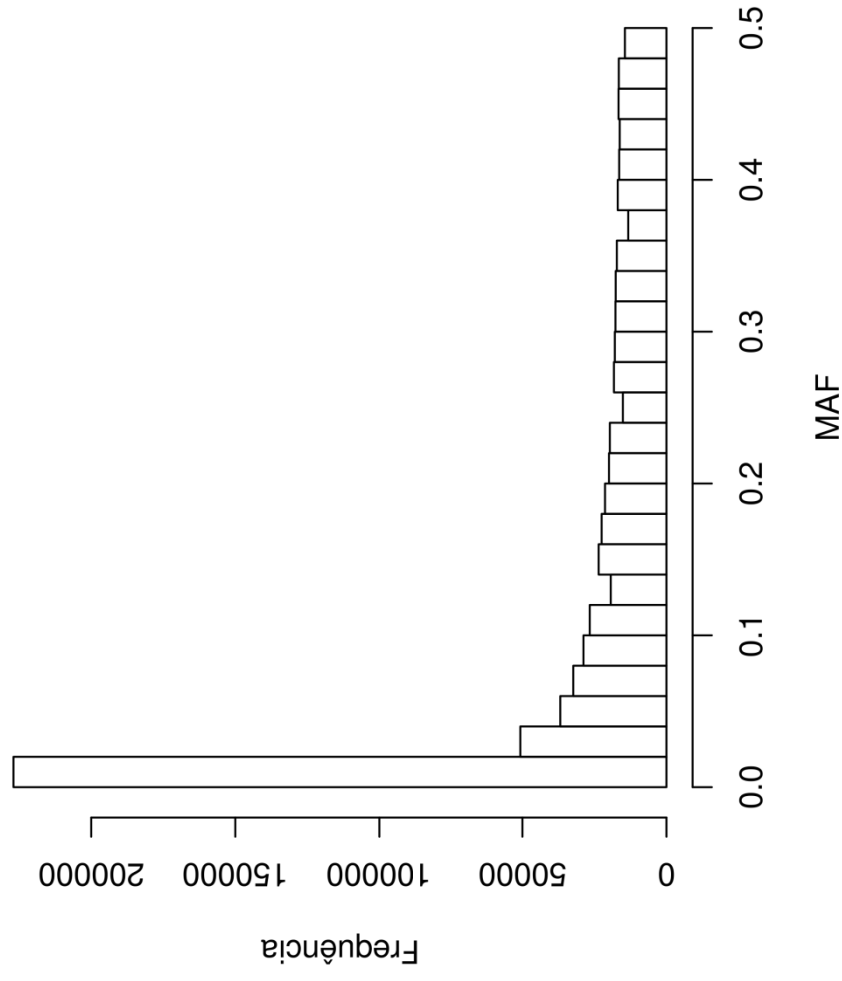


Fig. supl. 6. Experimento 1. Distribuição da Frequência do Menor Alelo (MAF) dos marcadores SNP do ensaio BovineHD antes do controle de qualidade.

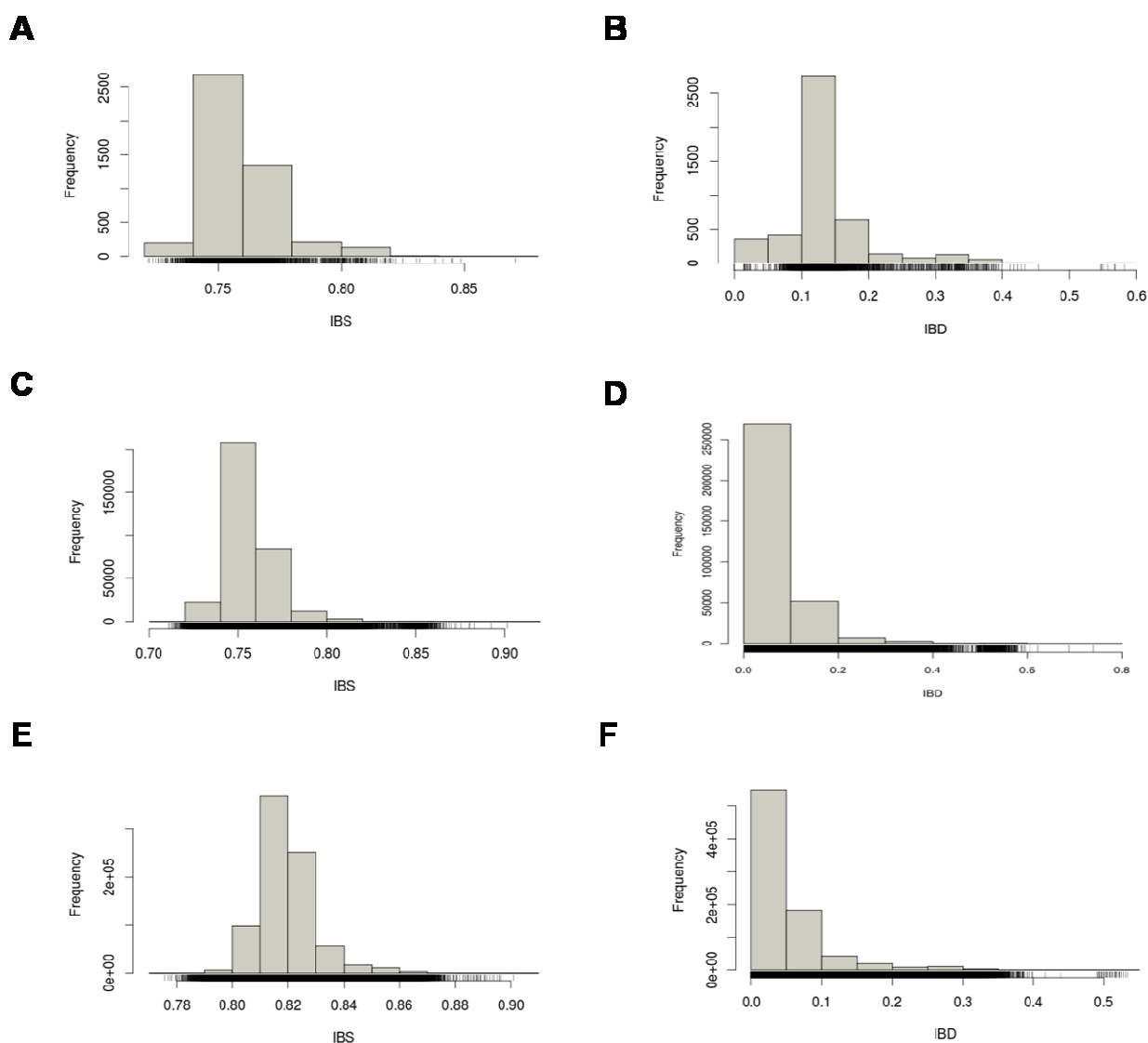


Fig. suppl. 7. Distribuição dos valores de IBS (à esquerda) e das proporções estimadas de IBD (à direita), figuras A-B (Experimento 1), C-D (Experimento 2) e E-F (Experimento 3), respectivamente. Linhas verticais abaixo do eixo x representam as comparações aos pares.

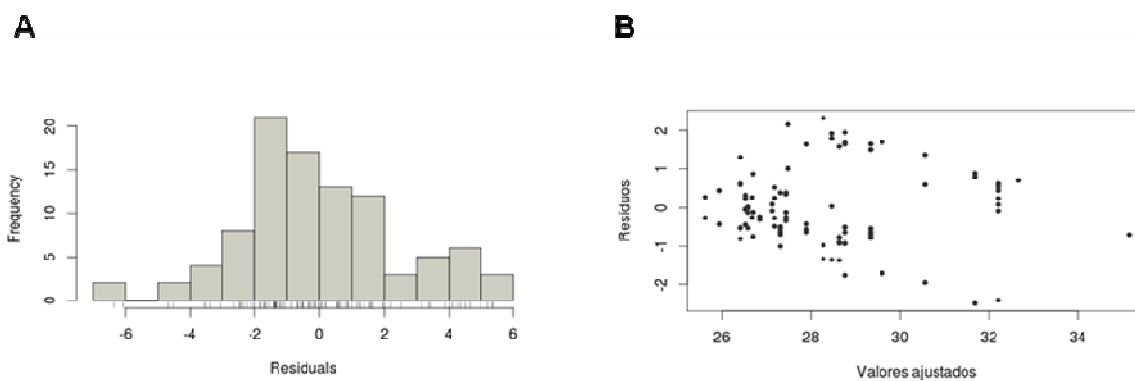


Fig. supl. 8. Experimento 1. Distribuição (A) e dispersão (B) dos resíduos testados na análise de associação. Linhas verticais abaixo do eixo x representam os valores individualmente.

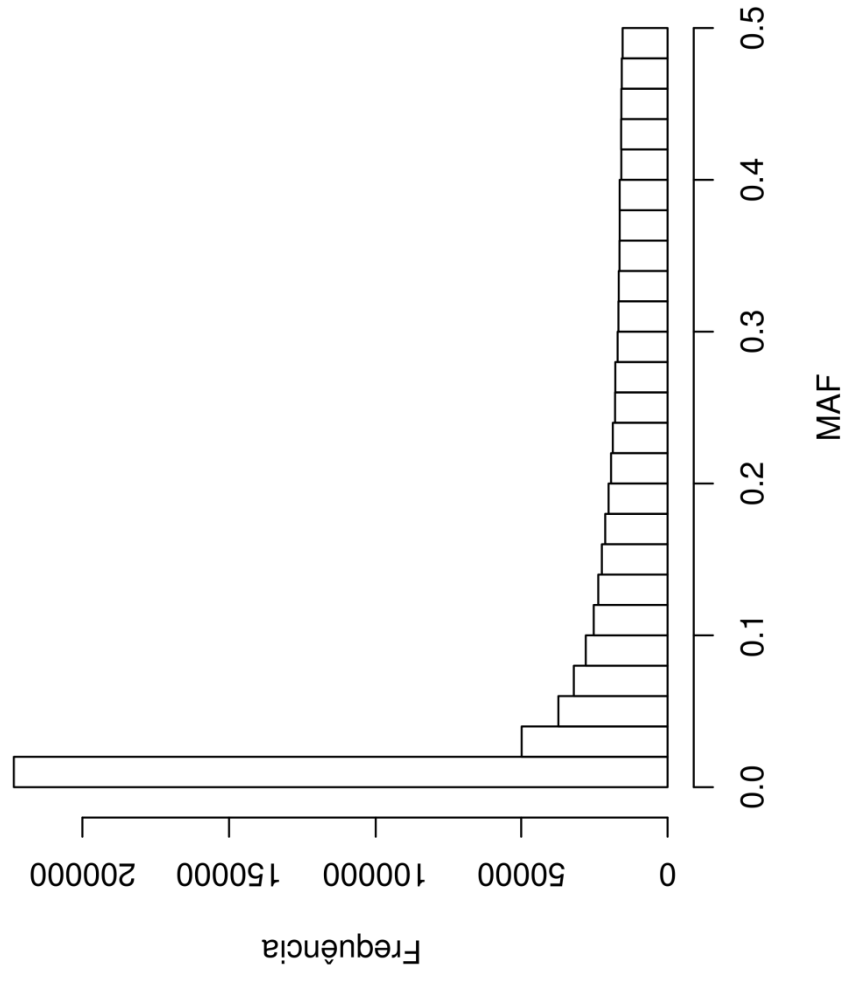


Fig. supl. 9. Experimento 2. Distribuição da Frequência do Menor Alelo (MAF) dos marcadores SNP do ensaio BovineHD antes do controle de qualidade.

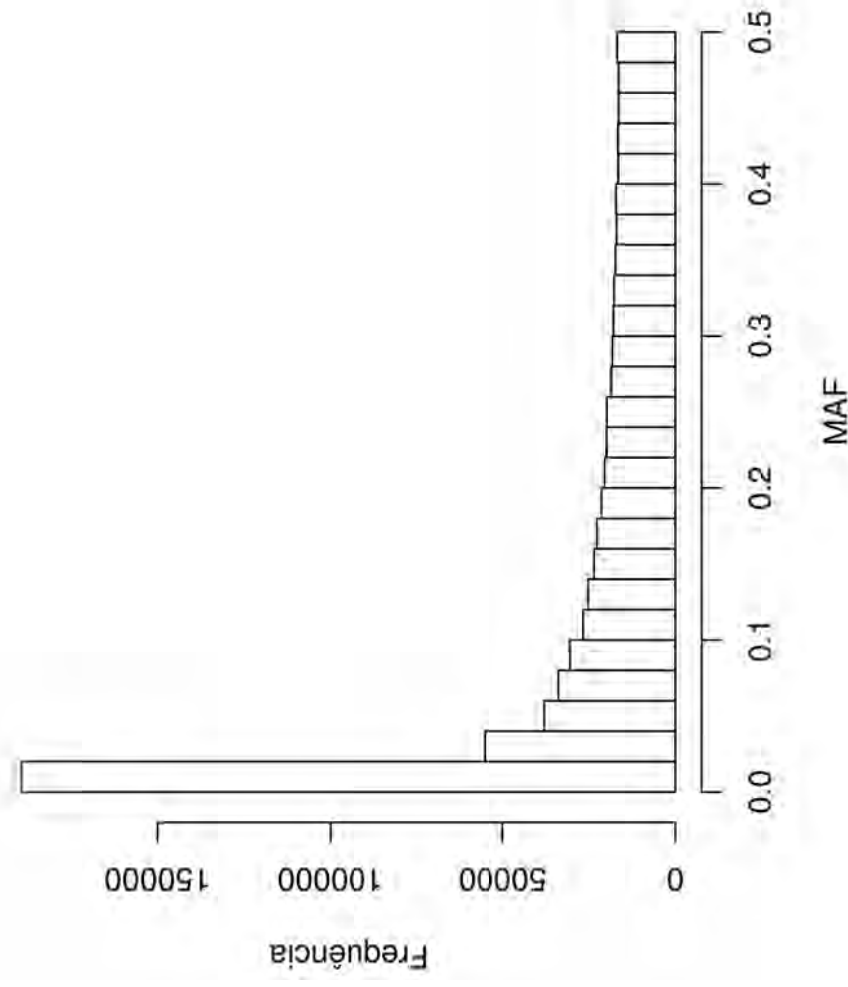


Fig. supl. 10. Experimento 3. Distribuição da Frequência do Menor Alelo (MAF) dos marcadores SNP do ensaio BovineHD antes do controle de qualidade.

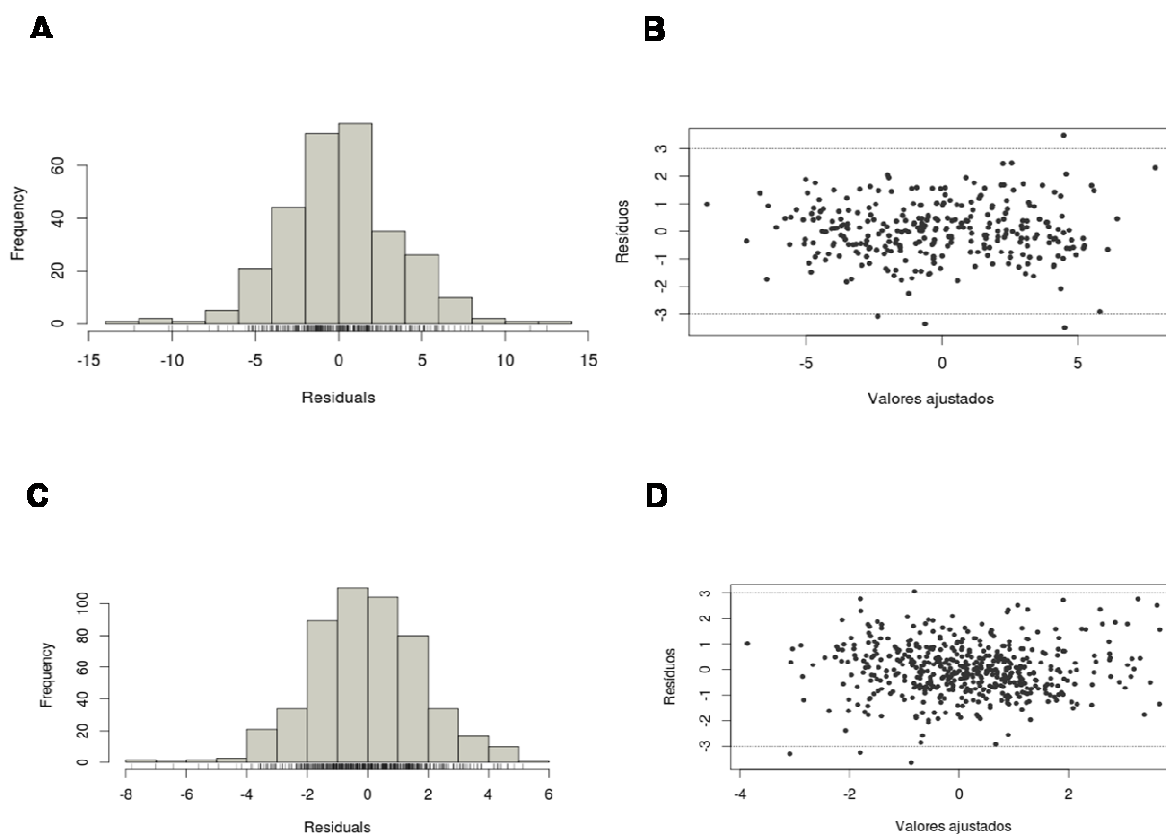


Fig. supl. 11. Experimento 2. Distribuição (A; C) e dispersão (B, D) dos resíduos testados na análise de associação dos subgrupos EBV1 (IPP) e EBV2 (PG), respectivamente. Linhas verticais abaixo do eixo x representam os valores individualmente.

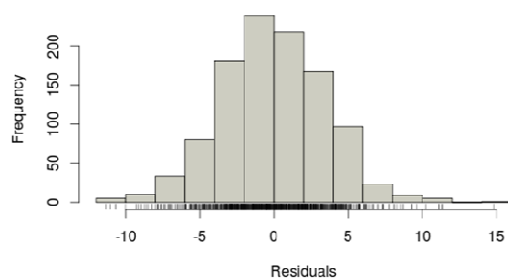
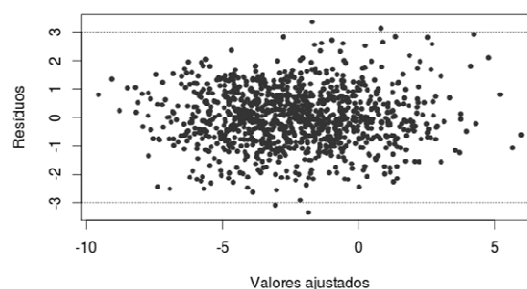
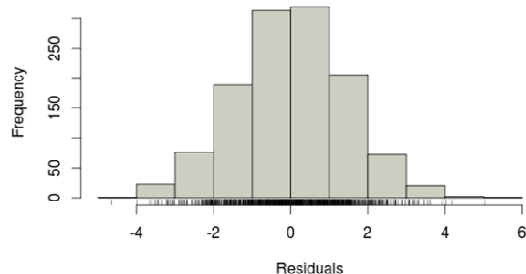
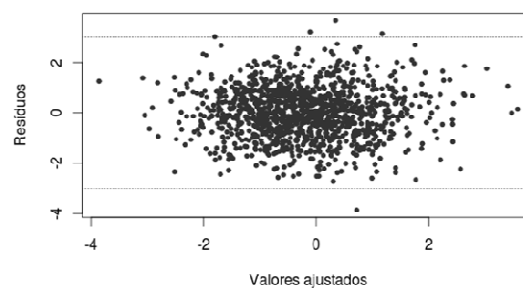
A**B****C****D**

Fig. supl. 12. Experimento 3. Distribuição (A; C) e dispersão (B, D) dos resíduos testados na análise de associação dos subgrupos EBV1 (IPP) e EBV2 (PG), respectivamente. Linhas verticais abaixo do eixo x representam os valores individualmente.