

UNIVERSIDADE ESTADUAL PAULISTA
“JULIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL

**ESTUDO DE ASSOCIAÇÃO GENÔMICA E ANÁLISE DE
ENRIQUECIMENTO GENÉTICO DA CRIOTOLERÂNCIA
ESPERMÁTICA EM BOVINOS DA RAÇA NELORE**

Adriana Santana do Carmo

Médica Veterinária

JABOTICABAL – SÃO PAULO - BRASIL

Março de 2012

UNIVERSIDADE ESTADUAL PAULISTA
“JULIO DE MESQUITA FILHO”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL

**ESTUDO DE ASSOCIAÇÃO GENÔMICA E ANÁLISE DE
ENRIQUECIMENTO GENÉTICO DA CRIOTOLERÂNCIA
ESPERMÁTICA EM BOVINOS DA RAÇA NELORE**

Adriana Santana do Carmo

Orientador: Prof. Dr. José Fernando Garcia

Co-orientadores: Prof. Dr. Marc Roger Jean Marie Henry

Prof. Dr. Marcos Vinícius Gualberto Barbosa da Silva

Profa. Dra. Vera Fernanda Martins Houssepian de Lima

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

JABOTICABAL – SÃO PAULO - BRASIL

Março de 2012

C287e Carmo, Adriana Santana do
Estudo de associação genômica e análise de enriquecimento genético da criotolerância espermática em bovinos da raça Nelore / Adriana Santana do Carmo. -- Jaboticabal, 2012
xviii, 124 f. : il. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2012
Orientador: José Fernando Garcia
Banca examinadora: José Bento Sterman Ferraz, Heidge Fukumasu, Roberto Carvalheiro, Joaquim Mansano Garcia
Bibliografia

1. Criotolerância espermática. 2. Genômica funcional. 3. GWAS. 4. Nelore. 5. Serpinas. 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: ESTUDO DE ASSOCIAÇÃO GENÔMICA E ANÁLISE DE ENRIQUECIMENTO GENÉTICO DA CRIOTOLERÂNCIA ESPERMÁTICA EM BOVINOS DA RAÇA NELORE

AUTORA: ADRIANA SANTANA DO CARMO

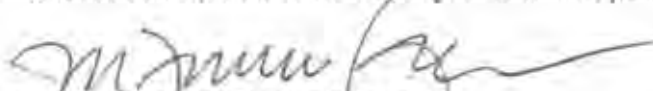
ORIENTADOR: Prof. Dr. JOSE FERNANDO GARCIA

CO-ORIENTADOR: Prof. Dr. MARC ROGER JEAN MARIE HENRY

CO-ORIENTADOR: Prof. Dr. MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA

CO-ORIENTADORA: Profa. Dra. VERA FERNANDA MARTINS HOSSEPIAN DE LIMA

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



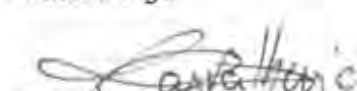
Prof. Dr. JOSÉ BENTO STERMAN FERRAZ

Departamento de Ciências Básicas / Faculdade de Zootecnia e Engenharia de Alimentos de Pirassununga



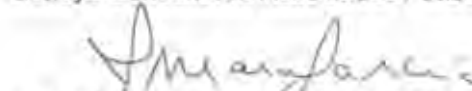
Prof. Dr. HEIDGE FUKUMASU

Departamento de Ciências Básicas / Faculdade de Zootecnia e Engenharia de Alimentos de Pirassununga



Prof. Dr. ROBERTO CARVALHEIRO

Gensys Consultores Associados / Jaboticabal/SP



Prof. Dr. JOAQUIM MANSANO GARCIA

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

DADOS CURRICULARES DA AUTORA

Adriana Santana do Carmo nasceu na cidade de Ponte Nova – MG, em 03 de junho de 1983. Em fevereiro de 2001 ingressou no curso de Medicina Veterinária na Escola de Veterinária da Universidade Federal de Minas Gerais – UFMG. Em julho de 2003 ingressou no Programa Institucional de Bolsas de Iniciação Científica (PIBIC – CNPq) orientada pelo Prof. Dr. Marc Roger Marie Henry permanecendo até dezembro de 2005. Graduiu-se em 06 de janeiro de 2006 ingressando no programa de mestrado em Ciência Animal (área de concentração Reprodução Animal) da UFMG com o mesmo orientador. Em março de 2008 iniciou como aluna do programa de doutorado em Medicina Veterinária (área de reprodução animal) da Universidade Estadual “Júlio de Mesquita Filho” campus de Jaboticabal (UNESP – FCAV) tendo como orientadora a Prof. Dr^a Vera Fernanda Martins Hossepian de Lima. Em julho de 2009 solicitou alteração na orientação, tendo como novo orientador o Prof. Adj. José Fernando Garcia, da UNESP-FMVA, campus de Araçatuba. Em fevereiro de 2010 foi aprovada como bolsista do Programa de Estágio no exterior (PDDE-CAPES), permanecendo seis meses no Bovine Functional Genomics Laboratory (BFGL) do Agricultural Research Services (ARS) do United States Department of Agriculture sob a supervisão do Dr. Tad S. Sonstegard.

“O correr da vida embrulha tudo.
A vida é assim: esquentada e esfria,
aperta e daí afrouxa,
sossega e depois desinquieta.
O que ela quer da gente é coragem.”

(Guimarães Rosa)

Dedico

Esse trabalho e todas as minhas conquistas aos meus pais, Ciriaca e Deoclécio.

Obrigado pelo amor e apoio incondicional nos
muitos caminhos que trilhei para chegar até aqui!

Sem vocês esse sonho não seria possível!

Amo vocês!!!

AGRADECIMENTOS

À Deus pela vida, pelo simples fato de existir e por sempre iluminar e proteger os meus caminhos.

Aos meus pais, Ciriaca e Deoclécio, por tudo que me proporcionaram, pelo amor, pela família maravilhosa que construíram e principalmente por serem exemplos de pessoas fortes e corajosas. Obrigado por sempre estarem presentes em minha vida mesmo que geograficamente distantes. Amo vocês!!

Aos meus irmãos Luciana, Bruno e Pablo, por serem meus melhores amigos e por estarem comigo na alegria e na tristeza. Obrigado por estarem sempre ao meu lado!

Ao Prof. Dr. José Fernando Garcia pela oportunidade, confiança, amizade e todos os ensinamentos preciosos. Obrigado por ter me recebido de portas abertas em Araçatuba e por me deixar fazer parte da sua equipe! Você me ajudou a amadurecer como pessoa e profissional, me ensinando a enxergar além. Serei eternamente grata por tudo!

A Prof^a. Dra. Vera Fernanda Martins Hossepian de Lima pela oportunidade e amizade. Obrigada por ter cuidado tão bem de mim no período em que estive sob sua orientação. Você me deu a oportunidade de conhecer e aprender com pessoas maravilhosas. Muito obrigada por tudo!

Ao Dr. Tad Sonstegard, agradeço imensamente por ter me recebido tão bem em seu laboratório. Sou muito grata por tudo que aprendi com você e sua equipe durante o período de doutorado sanduíche.

Agradeço ao Yuri, meu irmãozinho de orientação, por toda a sua generosidade e paciência. Sem sua ajuda essa tese não teria sido a mesma. Obrigada pelo apoio, pela amizade e pela companhia nos longos fins de semana de trabalho.

Agradeço imensamente à Carol por ter sido mais que uma amiga. Você é uma pessoa muito especial, uma verdadeira irmã de coração. Obrigada pelo apoio, carinho e amizade. Não poderia ter dividido essa aventura com outra pessoa que não fosse você!

Agradeço às Doutoradas Cáriss Maroni Nunes, Luciana de Almeida Regitano e Magda Benavides por serem sempre exemplos de mulheres fortes, profissionais éticas e brilhantes. Obrigada pelo carinho com que sempre me tratam! É uma honra ter convivido e aprendido com vocês.

Serei eternamente grata aos amigos do LBBMA, Fúlvia, Silvana, Fernanda, Diego, Samuel, Pedro, Dani, Ludmilla e Márcinha, pela forma com que me acolheram. Obrigada pela amizade, pelos momentos de excelente convívio, pelas palavras de apoio e por tudo o mais que me ensinaram.

Agradeço à Francine e ao Rodrigo pela amizade, confiança e principalmente pela compreensão. Obrigada por me deixar fazer parte da sua equipe!

Agradeço aos amigos da empresa Deoxi Biotecnologia, Andréa, Érica, Valquíria, Altair e Ismar pela amizade e paciência. Não teria conseguido sem o apoio e a torcida de vocês!

Obrigada aos Professores José Bento Ferraz, Heidge Fukumasu e Roberto Carneiro pelas preciosas sugestões para a elaboração do presente trabalho. Agradeço imensamente por terem me contemplado com seu tempo e experiência!

Agradeço imensamente a empresa Alta Genetics por ter cedido todos os dados fenotípicos e amostras biológicas utilizadas nessa tese. Obrigada pela confiança e colaboração!

Às amigas de toda uma vida, *keridas* do meu coração (Letícia, Buia, Mari, Nera, Bel, Cica), obrigada por sempre estarem na torcida, enviando muita energia positiva para mim! Nossa amizade é um dos meus bens mais preciosos!

Agradeço às bruxas de Jaboticabal (Aline, Ana Paula, Letícia, Fernanda, Verônica e Jaqueline), amigas que fiz para a vida inteira! Tenho muito orgulho de ter sido parte da equipe de vocês! Obrigada por tudo que me ensinaram!

Agradeço a Capes, Fapesp e CNPq pela concessão dos auxílios e da bolsa de doutorado. Sem esses recursos esse trabalho não teria se concretizado.

Agradeço imensamente a todos que participaram de alguma forma desse doutorado!

SUMÁRIO

LISTA DE FIGURAS	xi
LISTA DE TABELAS	xv
LISTA DE ABREVIATURAS.....	xvi
RESUMO	xviii
ABSTRACT	xix
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	2
2.1. Panorama da pecuária bovina no Brasil e no mundo	2
2.2. A Criobiologia dos espermatozóides	5
2.3. Estudos de Associação Genômica (GWAS)	10
2.4. Análise de enriquecimento funcional	30
2.5. A aplicação da genômica na reprodução	33
3. HIPÓTESES E OBJETIVO	35
3.1. Hipóteses	35
3.2. Objetivos gerais	36
3.3. Objetivos Específicos	36
4. MATERIAL E MÉTODOS	36
4.1. Delineamento Experimental: Motilidade Pós-descongelação e Motilidade Pós- teste de Termo Resistência.....	36
4.2. Extração DNA genômico	38
4.3. Análise em micro-arranjo de polimorfismo de sítio único (SNP) de alta densidade	38
4.4. Análise de GWAS	39

4.5. Controle de qualidade dos genótipos.....	42
4.6. Verificação da presença de estratificação na população	44
4.7. Teste de Associação Genômica.....	44
4.8. Análise de Enriquecimento Genético	45
5. RESULTADOS E DISCUSSÃO	46
8. REFERÊNCIAS BIBLIOGRÁFICAS	79
ANEXO 1	97
ANEXO 2	101
ANEXO 3	103

LISTA DE FIGURAS

Figura 1. Evolução do número de doses de sêmen congelado comercializadas no Brasil entre os anos de 1991 a 2010 (ASBIA, 2010).	3
Figura 2. Proporção de doses de sêmen comercializadas no Brasil no ano de 2010 das principais raças de corte (ASBIA, 2010).	4
Figura 3. Relação teórica proposta por AMMAN & HAMERTSED (1993) para explicar a relação entre o número de espermatozoides na dose inseminante e a fertilidade. A inclinação e o ponto máximo da curva dependem das características qualitativas do sêmen. Adaptado de WATSON (2000).	8
Figura 4. Apresentação esquemática dos tipos mais comuns de variações genéticas. Adaptado de FRAZER (2009).	13
Figura 5. Apresentação esquemática da utilização direta (A) e indireta (B) do SNP para elucidar os genes envolvidos com a expressão fenotípica (estrela vermelha representa a mutação causal; estrela cinza representa SNP; linha cheia/tracejada representa a cadeia de DNA; retângulo cinza representa o exon gênico). Adaptado de HISCHHORN e DALY, 2005.	15
Figura 6. Agrupamento da intensidade de fluorescência normalizada em diferentes classes de genótipos representados pelas diferentes cores (AA= vermelho, AB=púrpura e BB=azul) (A); Falha no agrupamento da intensidade de fluorescência normalizada, resultando em <i>No Call</i> onde SNP representados por pontos da cor preta falharam na designação do genótipo (B). Adaptado de ILLUMINA (2010).	17
Figura 7. Detecção do efeito de partidas utilizando análise de componentes principais. Os escores dos componentes principais foram agrupados e representados em Box-plots. Adaptado de Anderson, 2011.	21
Figura 8. Análise de componentes principais (PCA) de uma população estratificada com a formação de três grupos bem definidos no gráfico (A); Inflação generalizada facilmente visualizada com o auxílio do gráfico quantil-quantil (B) (cedido por José Fernando Garcia, 2012).	23

Figura 9. Possíveis cenários e métodos estatísticos para a realização do teste de associação genômica levando em consideração os níveis de estratificação e parentesco da população (proposto por AULCHENKO, 2011).	28
Figura 10. Análise descritiva do fenótipo EBV (Experimento 1) da característica diferença entre a motilidade do sêmen fresco e descongelado (DMD) pelo Histograma, <i>Box plot</i> e <i>Q-Q plot</i>	48
Figura 11. Análise descritiva do fenótipo EBV (Experimento 1) da característica diferença entre a motilidade do sêmen fresco e pós teste de termo resistência (DMT) pelo Histograma, <i>Box plot</i> e <i>Q-Q plot</i>	49
Figura 12. Análise descritiva das médias das medidas fenotípicas (Experimento 2) da característica diferença entre a motilidade do sêmen fresco e descongelado (DMD) pelo Histograma, <i>Box plot</i> e <i>Q-Q plot</i>	50
Figura 14. Análise descritiva das médias das medidas fenotípicas (Experimento 2) da característica diferença entre a motilidade do sêmen fresco e pós teste de termo resistência (DMT) pelo Histograma, <i>Box plot</i> e <i>Q-Q plot</i>	51
Figura 15. Distribuição da frequência alélica menor (MAF) dos SNP antes do controle de qualidade. Linhas verticais abaixo do eixo X representam os SNP individualmente.	55
Figura 16. Distribuição da heterozigosidade dos SNP antes do controle de qualidade. Linhas verticais abaixo do eixo X representam os SNP individualmente.	56
Figura 17. Distribuição dos SNP por cromossomos autossômicos antes e depois do controle de qualidade.	57
Figura 18. Análise de Componentes Principais (PCA) e <i>Heatmap</i> dos fenótipos do Experimento 1 (EBV). Animais representados por cores diferentes de acordo com a magnitude do fenótipo característica diferença entre a motilidade do sêmen fresco (DMD - A) e pós teste de termo resistência (DMT -B).	60
Figura 19. Análise de Componentes Principais (PCA) e <i>Heatmap</i> dos fenótipos do Experimento 2 (médias das medidas fenotípicas. Animais representados por cores diferentes de acordo com a magnitude do fenótipo. Média das medidas fenotípicas de DMD (A) e DMT (B)	61

- Figura 20.** Gráficos Q-Q dos valores de significância (p) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD no Experimento 1 (EBV). Valores de $\lambda = 1,42$ e $1,27$ respectivamente.62
- Figura 21.** Gráficos Q-Q *plot* dos valores de significância (p) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMT no Experimento 1 (EBV). Valores de $\lambda = 1,36$ e $1,27$ respectivamente A e B.63
- Figura 22.** Gráficos Q-Q *plot* dos valores de significância (p) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD no Experimento 2 (Média das medidas fenotípicas). Valores de $\lambda = 1,07$ e $1,14$ respectivamente A e B.64
- Figura 23.** Gráficos Q-Q *plot* dos valores de significância (p) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMT no Experimento 2 (Média das medidas fenotípicas). Valores de $\lambda = 1,07$ e $1,14$ respectivamente A e B.65
- Figura 24.** *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD para o Experimento 1 (EBV). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destacada SNP significativos em ambos os métodos.....67
- Figura 25.** *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT(B) para a característica DMT para o Experimento 1 (EBV). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destacada SNP significativos em ambos os métodos.....68
- Figura 26.** *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação

realizado pelos métodos GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD para o Experimento 2(médias das medidas fenotípicas). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destaca SNP significativos em ambos os métodos.....69

Figura 27. *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT (B) para a característica DMT para o Experimento 2(médias das medidas fenotípicas). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destaca SNP significativos em ambos os métodos.....70

Figura 28. Representação dos cromossomos que apresentaram SNP significativos após o teste de associação realizado com ambos os métodos estatísticos (GRAMMAR e EIGENSTRAT) nos fenótipos (A) DMD no Experimento 1 (EBV), (B) DMT no Experimento 1 (EBV), (C) DMD no Experimento 2 (médias das medidas fenotípicas) e (D) DMT no Experimento 2 (médias das medidas fenotípicas).71

Figura 29. Modelo proposto pelo presente trabalho para as interações entre grupos funcionais encontrados na AEF: linhas contínuas pretas representam grupos de genes específicos; linhas tracejadas pretas representam vias metabólicas. Linha tracejada vermelha representa o possível resultado da interação entre os grupos funcionais arbitrado no presente trabalho. Linha tracejada verde representa o ponto inicial e final do modelo arbitrado no presente trabalho. Via metabólica (1) Baseado nos resultados descritos por SI & OKUNO (1995) e TASH et al. (1996); Via metabólica (2) -Baseado nos resultados descritos por VISCONTI & KOPF; e Via metabólica (3) - Baseado nos resultados descritos por GEIGER et al. (1996) e LU et al. (2011).75

Figura 30. Representação da via metabólica da capacitação espermática proposta por VISCONTI & KOPF (1998). Retângulos coloridos correspondem a grupos funcionais encontrados no presente trabalho: (1) canais e proteínas carreadoras; (2) via de sinalização de cálcio; (3) gene de receptores de cátions; (4) regulação e biossíntese de cAMP; (5) inibidores de protease – SERPINAS; (6) regulação da resposta ao estímulo

externo e (7) regulação positiva dos processo metabólicos envolvendo macromoléculas.	76
--	----

LISTA DE TABELAS

Tabela 1. Sistema BADGE de classificação da reprodutibilidade em estudos de associação genômica (MANLY, 2005).	29
Tabela 2. Distribuição cumulativa do número de SNP que não atingiram os critérios de EHW ($p > 10^{-5}$) para diferentes valores de p .	52
Tabela 3. Distribuição da proporção do <i>call rate</i> (CR) dos SNP.	52
Tabela 4. Número de SNP removidos nas diferentes etapas do controle de qualidade.	53
Tabela 5. Número de genes que se apresentaram associados com as características avaliadas (DMD e DMT) nos diferentes métodos estatísticos (GRAMMAR e EIGENSTRAT).	71
Tabela 6. Índice de Enriquecimento (IE) dos principais grupos funcionais encontrados na AEF com o uso do programa DAVID em ordem decrescente de valor, originados a partir das diferentes combinações de fenótipos e métodos estatísticos para o GWAS. Em destaque os grupos funcionais não utilizados para a proposição do modelo por não apresentar relação clara e direta com dados de literatura.	77

LISTA DE ABREVIATURAS

AEF	- Análise de Enriquecimento Funcional
BADGE	- <i>Better Associations for Disease and Genes</i>
CNV	- Variação do número de cópias
CQ	- Controle de Qualidade
DEP	- Diferença Esperada na Progenie
DMD	- Diferença de motilidade entre sêmen fresco e após descongelação
DMT	- Diferença de motilidade ente sêmen fresco e após TTR
DST	- Distância Genética Padrão de Nei
DC-VC	- Doença comum - variante comum
EBV	- Valor genético estimado
EHW	- Equilíbrio de Hardy-Weinberg
GC	- <i>Genomic Control</i>
GCScore	- <i>Gene Call Score</i>
GEPBV	- Estimação genômica do valor genético predito
GLM	- General Linear Model
GRAMMAR	- Genome- wide Rapid Association using Mixed Model and Regression
GS	- Seleção Genômica
GSEA	- Análise de Enriquecimento Funcional de Genes
GWAS	- Estudo de Associação Genômica
IA	- Inseminação Artificial
IATF	- Inseminação Artificial em Tempo Fixo
IBS	- Idêntico por Estado
IE	- Índice de enriquecimento funcional
INDELS	- Inserção/ Deleção
LD	- Desequilíbrio de Ligação
LG	- Lista de genes
MAF	- Menor Frequência Alélica

MEA	- Análise de Enriquecimento Funcional Modular
MG	- <i>Measured Genotype</i>
MSD	- Motilidade do sêmen descongelado
MSF	- Motilidade do sêmen fresco
MST	- Motilidade do sêmen após o TTR
OLT	- Limite de tolerância osmótico
PC	- Componentes Principais
PCA	- Análise de Componentes Principais
SEA	- Análise de Enriquecimento Funcional Singular
SNP	- Polimorfismo de Nucleotídeo em Sítio Único
TTR	- Teste de Termo Resistência
VE	- Variantes Estruturais
VM	- Vias metabólicas
WGEAS	- <i>Whole Genome Enabled Animal Selection</i>

ESTUDO DE ASSOCIAÇÃO GENÔMICA E ANÁLISE DE ENRIQUECIMENTO GENÉTICO DA CRIOTOLERÂNCIA ESPERMÁTICA EM BOVINOS DA RAÇA NELORE

RESUMO - Estudo de Associação Genômica (GWAS - *Genome Wide Association Study*) foi conduzido para identificar regiões cromossômicas relacionadas à capacidade de criotolerância espermática em bovinos da raça Nelore, empregando os fenótipos: diferença entre a motilidade do sêmen fresco e descongelado (DMD) e diferença entre a motilidade do sêmen fresco e pós-teste de termo-resistência (DMT), com o intuito de compreender melhor os fenômenos que regem tal característica. O perfil de SNP (SNP - *Single Nucleotide Polymorphism*) de cento e vinte e cinco touros foi determinado com o auxílio de micro-arranjo de alta densidade de SNP (BovineHD Illumina®). As estimativas dos valores genéticos dos touros (EBV - *Estimated Breeding Value*) (Experimento 1) e as médias das medidas fenotípicas (Experimento 2) foram calculadas para as características em questão. As metodologias estatísticas GRAMMAR e EIGENSTRAT foram utilizadas para a realização do teste de associação fenótipo / genótipo após procedimento de controle de qualidade dos dados. A análise de enriquecimento funcional foi realizada com auxílio do software DAVID. Essa análise apontou a família de genes *Serpina* (inibidores de proteases), reguladores de cAMP e componentes do citoesqueleto celular como o grupo de genes que apresentam maior importância funcional para as características avaliadas. Todas as metodologias testadas demonstraram-se capazes de identificar regiões genômicas associadas aos fenótipos DMD e DMT, destacando-se os cromossomos 1, 16 e 19.

Palavras-Chave: Criotolerância Espermática, Genômica Funcional, GWAS, Nelore, Serpinas, SNP

GENOME WIDE ASSOCIATION STUDY AND GENETIC ENRICHMENT ANALYSIS OF SPERMATIC CRYOTOLERANCE IN NELLORE CATTLE

ABSTRACT – Genome Wide Association Study (GWAS) was performed in order to identify chromosomal regions related to sperm cryotolerance capacity in Nellore cattle for the phenotypes: difference between fresh and post-thaw semen motility (DMD) and difference between fresh and post thermal resistance test semen motility (DMT), aiming to better understand the phenomena governing this characteristic. SNP (Single Nucleotide Polymorphism) profile of one hundred and twenty five Nellore bulls were determined using high density SNP chip (BovineHD Illumina®). Estimated Breeding Values - EBV (Experiment 1) and phenotypic measures means (Experiment 2) were calculated for the evaluated traits. The statistical methodologies GRAMMAR and EIGENSTRAT were used to test the phenotype / genotype association after data quality control. Functional enrichment analysis was performed using DAVID software. These analyses pointed out to Serpine (protease inhibitor), cAMP metabolism control and cytoskeletal components gene families as the most important functional gene groups for the evaluated traits. All the methods tested shown to be able to identify genomic regions associated with phenotypes DMD and DMT, especially chromosomes 1, 16 and 19.

KEY WORD: Functional Genomics, GWAS, Nellore, Serpin, Spermatic Cryotolerance, SNP

I. INTRODUÇÃO

O aumento da competitividade tem levado à busca por maior eficiência na cadeia produtiva bovina como um todo. Neste contexto, a utilização de sêmen congelado na Inseminação Artificial (IA) é de extrema importância já que permite que touros de alto valor genético sejam utilizados em um grande número de rebanhos (BARBAS; MASCARENHAS, 2009).

A qualidade do sêmen criopreservado é determinada por suas características morfofuncionais. Para isso, avalia-se a qualidade do espermatozoide em relação a essas características, classificando-os de acordo com critérios mínimos pré-determinados que indiquem se o sêmen apresenta qualidade adequada para ser utilizado na IA (WATSON, 2000).

A capacidade espermática de tolerar o processo de criopreservação, apresentando boa qualidade após a descongelação, pode ser influenciada por diversos fatores. Essa característica determina se um animal poderá ter seu sêmen congelado e utilizado amplamente em diversos rebanhos ou se o reprodutor somente poderá ser utilizado em um rebanho, pela monta natural.

A complexidade biológica desse fenômeno, a existência de poucas informações sobre as bases fisiológicas da criotolerância, a dificuldade de obtenção dos dados em grande escala e sua expressão tardia (quando o animal já é praticamente adulto), constituem entraves que dificultam a implantação de critérios para a seleção de animais criotolerantes (DE ROUEN; FRANKE, 1989). Essas particularidades sugerem que a introdução de ferramentas genômicas poderia trazer inúmeros benefícios ao processo de seleção desses animais.

Os Estudos de Associação Genômica (GWAS - *Genome Wide Association Study*) aliados às novas tecnologias de prospecção automatizada de SNP, permitem identificar variações no DNA que estão correlacionados com uma dada característica, apontando para a localização dos genes envolvidos com a expressão do fenótipo de interesse e auxiliando na compreensão dos mecanismos biológicos e fisiológicos desses genes (KU et al., 2010).

O presente trabalho baseia-se na premissa de que a variação na resistência a criopreservação do sêmen entre indivíduos é determinada geneticamente. Avanços na compreensão dos mecanismos de determinação genética dessa característica podem ajudar a explicar a existência de animais mais ou menos criotolerantes. A identificação das vias metabólicas essenciais envolvidas no processo de criopreservação, com consequente conhecimento dos genes e proteínas nelas envolvidas, poderá auxiliar o desenvolvimento de procedimentos de criopreservação mais eficientes, além de trazer luz a particularidades de espécies cujos gametas são altamente sensíveis a esse processo, o que em última instância deve permitir o desenho de estratégias para seleção de animais criotolerantes.

II. REVISÃO DE LITERATURA

2.1. Panorama da pecuária bovina no Brasil e no mundo

No âmbito mundial a pecuária contribui com 40% da produção agrícola total, sendo responsável pela subsistência e segurança alimentar de mais de um bilhão de pessoas. Essa atividade tem sido crucial para a estabilização da economia de países menos desenvolvidos, além de contribuir para a melhoria dos padrões alimentares e redução da pobreza e da fome no mundo (FAO, 2009).

Para se entender a dimensão dessa atividade no Brasil basta dizer que ela é responsável por gerar mais de 220 bilhões de reais por ano, sendo a bovinocultura de corte a principal cadeia produtiva desse setor, com faturamento de mais de R\$ 50 bilhões/ano e geração de cerca de 7,5 milhões de empregos (CEPEA, 2010).

Levantamentos realizados pela Associação Brasileira das Indústrias de Exportação de Carne (ABIEC, 2010) demonstraram que o Brasil atualmente é responsável por 26% do total de carne bovina comercializada no mundo, sendo o maior exportador e o segundo maior produtor de carne, com 1,7 milhões e 9,3 milhões de toneladas equivalente-carcaça respectivamente, produzidas no ano de 2010.

A cadeia agroindustrial da pecuária de corte bovina beneficia vários setores da economia, como a cadeia de produtos e serviços, além de exercer impacto positivo no

aspecto social agregando aproximadamente um milhão de produtores de gado de corte, cerca de 1.000 estabelecimentos industriais de carnes, derivados e serviços de armazenagem, 50 mil pontos de distribuição (BANKUTI; AZEVEDO, 2003; TORRES JR et al., 2005) e mais de 180 empresas de biotecnologia no setor (SEBRAE, 2007).

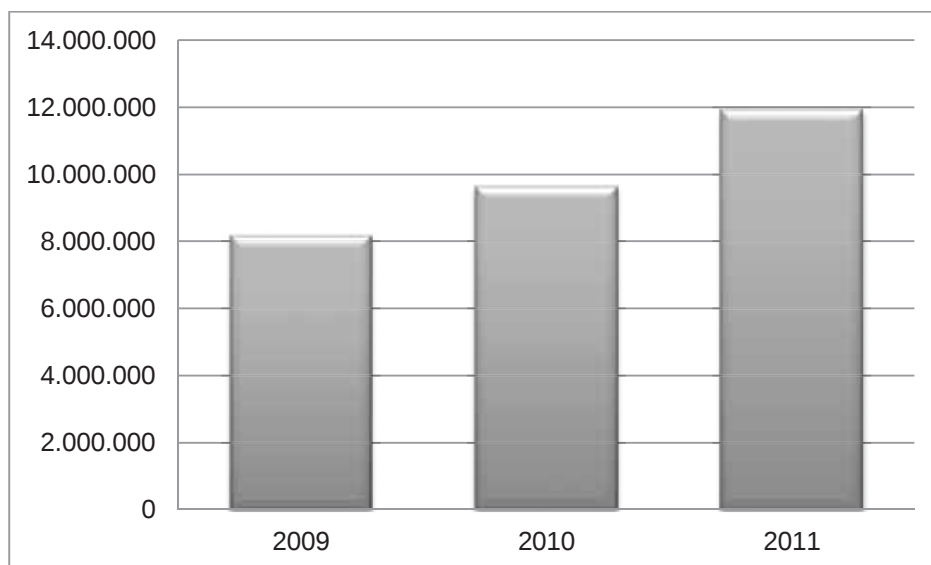


Figura 1. Evolução do número de doses de sêmen congelado comercializadas no Brasil entre os anos de 2009 a 2011 (ASBIA, 2011).

A indústria nacional de comercialização de sêmen também desempenha papel importante no que tange a valorização da pecuária nacional, já que a inseminação artificial foi uma das maiores responsáveis pela expansão rápida e os incrementos significativos no mérito genético dos rebanhos. Como demonstrado pela Figura 1, a cada ano as centrais de inseminação artificial tem aumentado o número de doses comercializadas com o intuito de atender a demanda crescente do mercado. Segundo Associação Brasileira de Inseminação Artificial (ASBIA, 2010), na última década esse mercado se expandiu em aproximadamente 200%. Apenas no ano de 2011, foram vendidas 12 milhões de doses de sêmen, o que representa um aumento de 23,54% em relação ao ano anterior (ASBIA, 2011).

Parte do aquecimento do mercado de sêmen nos últimos anos se deve a Inseminação Artificial em Tempo Fixo (IATF), que viabilizou em termos econômicos e

de manejo, o uso da inseminação artificial em grandes rebanhos, principalmente em regiões onde há problemas com infraestrutura e mão de obra.

No Brasil, 45% das vendas de sêmen congelado são para o mercado leiteiro e 55% para o de corte. Neste último, dentre as mais de 15 raças com sêmen disponível para a venda, o sêmen da raça Nelore se destaca, representando quase 50% do total de vendas da subespécie zebuína (Figura 2).

Dentre os fatores que consagraram a criação dessa raça no país podemos citar sua rusticidade e adaptabilidade, o que permite que animais Nelore sejam criados em diferentes condições de manejo com um bom rendimento de carcaça (PESSUTI; MEZZADRI, 2004).

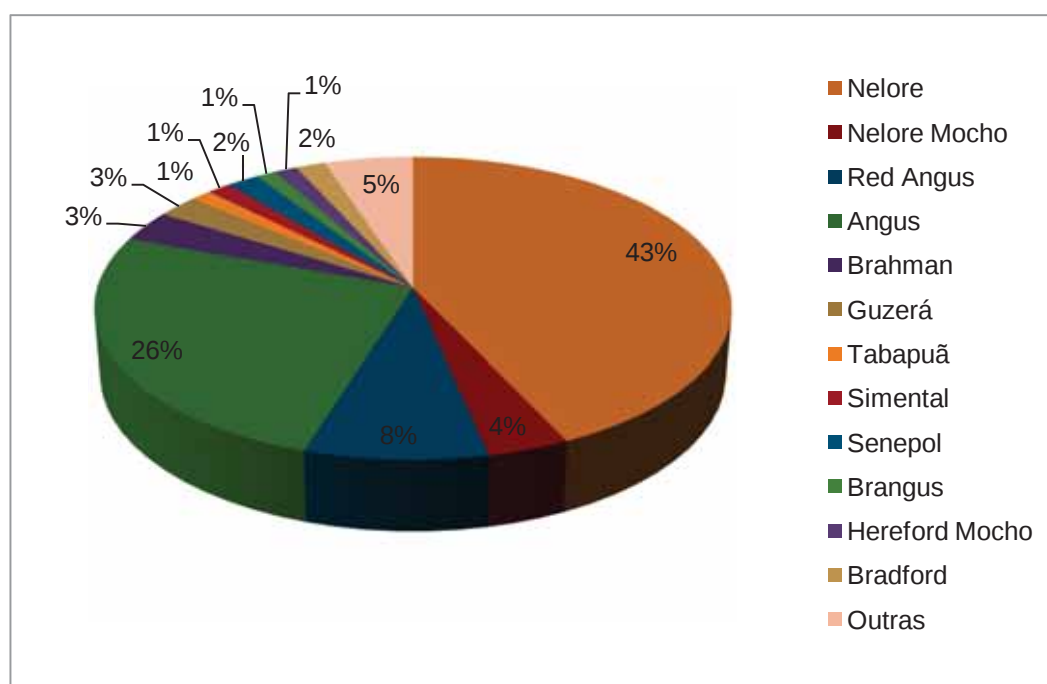


Figura 2. Proporção de doses de sêmen comercializadas no Brasil no ano de 2011 das principais raças de corte (ASBIA, 2011).

Apesar do mercado de sêmen no Brasil estar em franca expansão, estima-se que das 90 milhões de fêmeas em idade reprodutiva no país, apenas 9% sejam inseminadas. Esses dados demonstram o enorme potencial de crescimento que esse

setor ainda apresenta e o aumento da demanda para a comercialização de um número maior de doses de sêmen (ASBIA, 2011).

2.2. A Criobiologia dos espermatozoides

A utilização de sêmen congelado na IA é essencial para o melhoramento dos rebanhos nacionais já que permite que touros de alto valor genético sejam utilizados em diferentes regiões geográficas. Touros com alta capacidade reprodutiva produzem em média cerca de 60.000 doses de sêmen congelado por ano, que podem ser utilizados em diferentes rebanhos intensificando os ganhos pelo melhoramento genético e reduzindo o potencial de transmissão de doenças (CURRY, 2000; BARBAS; MASCARENHAS, 2009).

De forma geral, os protocolos de criopreservação de sêmen incluem a redução da temperatura, a desidratação celular, a congelação e a descongelação (MEDEIROS, 2002). Segundo HALLAP et al. (2006), o processo de criopreservação apresenta várias etapas potencialmente danosas à célula espermática, reduzindo a porcentagem de espermatozoides férteis em até 50%. Os danos à célula espermática podem ser ultraestruturais, físicos, químicos e/ou funcionais, e podem ser causados por fenômenos como o choque pelo frio, o “efeito solução” e a formação de cristais de gelo intracelular (CURRY, 2000).

O choque pelo frio ocorre principalmente quando a temperatura é reduzida de 30°C para 0°C durante o processo de criopreservação (WATSON, 1981). A redução de temperatura causa a desorganização da porção lipídica da membrana plasmática causando fraturas irreversíveis. Durante a congelação ocorre a inativação da enzima aminofosfolípídeo transferase, responsável por manter a polarização dos fosfolípidos concentrados na camada interna da membrana. Com a inativação da enzima, decorrente da redução progressiva da temperatura, as membranas passam do estado fluido para o de gel, e as cadeias de ácidos graxos, que antes eram distribuídas aleatoriamente, se paralelizam formando uma estrutura mais rígida, portanto menos fluida, consequentemente mais susceptível a rupturas e fusões (HAMMERSTED, 1990).

A desorganização da membrana também propicia o agrupamento das proteínas integrais com consequente perda da função de canais iônicos específicos, como por exemplo, os canais de cálcio. Segundo ROBERTSON & WATSON (1986), a congelação aumenta a permeabilidade dos canais de cálcio, aumentando o influxo desse íon para o meio intracelular e contribuindo para a capacitação precoce da célula espermática, diminuindo suas chances como gameta fecundante.

De forma geral, a membrana plasmática possui citoesqueleto celular a ela ancorado, sendo que qualquer alteração dessa estrutura pode repercutir negativamente na função da célula. Esse fato foi observado por SPUNGIN (1995), que verificou o efeito da redução da temperatura sobre a despolimerização dos filamentos de F-actina do citoesqueleto, o que permitia a aproximação e consequente fusão das membranas plasmática e acrossomal externa, em um processo similar à reação acrossômica.

Muitos autores têm descrito a evidente redução da motilidade espermática após a descongelação (WATSON, 2000; ESTEVES et al., 2007; BARBAS; MASCARENHAS, 2009). Segundo WATSON (1995), esse fato justifica-se pela ocorrência de alterações no transporte ativo e na permeabilidade das membranas na região da cauda, além da redução da disponibilidade de energia celular e/ou possíveis danos ao axonema. O mesmo autor relata que o processo de criopreservação propicia a formação de cristais de gelo intra-mitocôndriais que causam danos físicos a essa organela, prejudicando processos como a fosforilação oxidativa, o transporte de prótons e a produção de ATP.

Alterações nas membranas e nas diferentes organelas das células espermáticas ocorrem simultaneamente nas diferentes fases da criopreservação. Uma característica patognomônica da ocorrência de danos devido ao choque pelo frio é a alteração do padrão de motilidade espermática. Células com a viabilidade comprometida apresentam movimentos em círculo ou retrógrado, além de dobras nas peças intermediárias (WATSON, 2000).

A adição e remoção dos agentes crioprotetores ao meio de congelação causam um estresse osmótico transitório na célula denominado “efeito solução” (MAZUR et al., 1972). Quando uma suspensão celular é resfriada, a água pura do meio externo se solidifica, tornando os solutos extracelulares progressivamente mais concentrados. O

ambiente extracelular hiperosmótico induz a saída brusca da água do meio intracelular por osmose, levando a célula a uma desidratação intensa. O limite de tolerância osmótica (OLT), representado pela extensão da retração ou expansão da célula em resposta às alterações de osmolaridade, varia entre as distintas espécies, sendo importante para o estabelecimento da quantidade e tempo ótimos para a adição e a remoção do crioprotetor sem que ocorram danos à célula (AGCA; CRISTER, 2002).

As taxas de resfriamento utilizadas no protocolo de congelação são de extrema importância para evitar a formação de cristais de gelo intracelular que podem mecanicamente danificar organelas. A velocidade em que o resfriamento ocorre, deve permitir que a água deixe a célula por osmose antes que os cristais de gelo sejam formados. Para minimizar esses danos, utiliza-se taxa de resfriamento que variam entre 15 a 60°C/minuto (WATSON, 2000).

Apesar dos extensos danos celulares, o sêmen congelado bovino tem sido utilizado há algumas décadas com taxas de concepção equivalentes às aquelas obtidas através da monta natural. Segundo WATSON (2000), o sucesso na utilização de sêmen congelado nessa espécie deve-se ao uso de elevada concentração de espermatozoides na dose inseminante. A espécie bovina é especialmente beneficiada com a congelação do sêmen, pois necessita de relativamente poucos espermatozoides para apresentar bons índices de fertilidade.

AMMAN & HAMMERSTED (1993), propuseram modelo teórico (Figura 3) para explicar a relação entre a fertilidade e a dose inseminante no qual a inseminação deve ser realizada com uma quantidade de células espermáticas suficiente para atingir o ponto mais alto da curva garantindo assim fertilidade satisfatória.

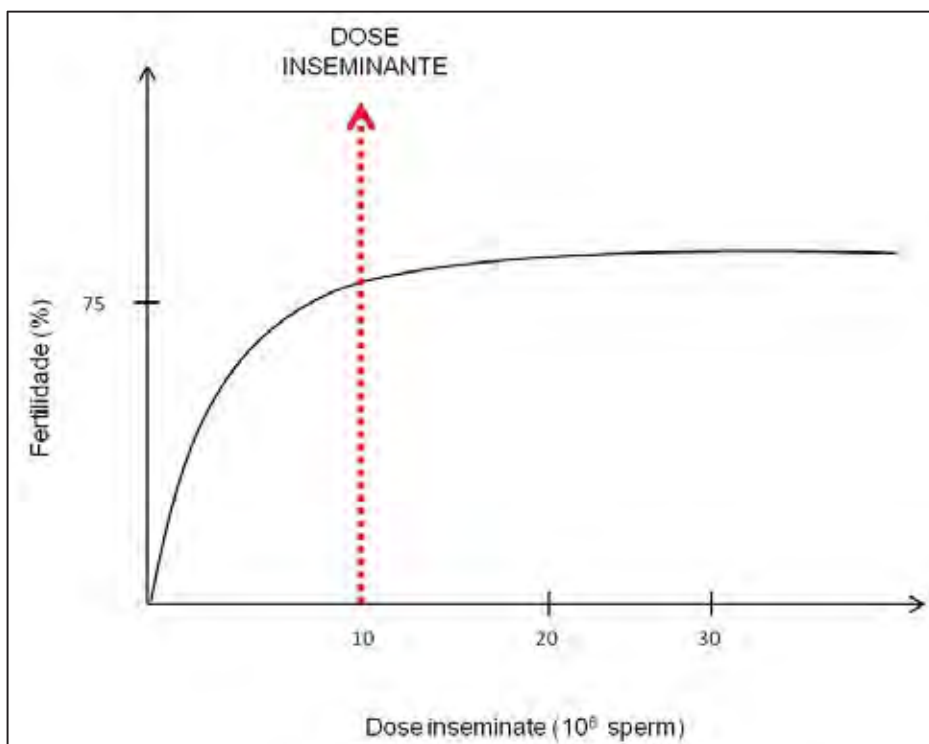


Figura 3. Relação teórica proposta por AMMAN & HAMERTSED (1993) para explicar a relação entre o número de espermatozoides na dose inseminante e a fertilidade. A inclinação e o ponto máximo da curva dependem das características qualitativas do sêmen. Adaptado de WATSON (2000).

Nas últimas décadas, muitas pesquisas têm sido realizadas com o intuito de revelar quais são os principais fatores que afetam a criotolerância da célula espermática (BELL et al., 1993; GAO et al., 1993; CURRY et al., 1994; CURRY et al., 1995; GILMORE et al., 1996). Os mecanismos responsáveis por essa característica ainda não foram elucidados, mas evidências demonstram a importante influência de fatores como espécie, protocolo de criopreservação e perfil genético (HOLT, 2000).

A utilização da motilidade como parâmetro de avaliação da criotolerância espermática é controversa (JANUSKAUSKAS et al., 1995), já que espermatozoides que apresentem boa motilidade após descongelação, podem ser incapazes de fertilizar um oócito caso apresentem danos acrossomais (STROM et al., 1997). Entretanto, esse parâmetro é um indicativo da proporção de espermatozoides vivos após esse processo, além de ser a manifestação da integridade estrutural e funcional dessa célula.

HANDELSMAN (2002) relata que o estudo desse parâmetro exige atenção especial já que o mesmo não apresenta distribuição normal devido ao modo como é mensurado (escala de 1 a 100% com acréscimos/decréscimos de 5%). Segundo esse autor, mesmo depois de transformados os valores da motilidade podem não apresentar normalização na distribuição.

Particularidades espécies-específicas tais como a forma, o volume, o tamanho e a composição das organelas podem influenciar na sensibilidade das células à congelamento. Estudos comparativos entre a forma da cabeça dos espermatozoides de diferentes espécies demonstraram correlação negativa entre o tamanho da cabeça do espermatozoide e a criotolerância (GAO et al., 1993; BARBAS e MASCARENHAS, 2009).

A variação individual na criotolerância do sêmen é reconhecida em muitas espécies domésticas, incluindo os bovinos (PARKINSON; WHITFIELD, 1987; KOIVISTO et al., 2008). Touros podem ser classificados como mais ou menos criotolerantes dependendo das características de seus espermatozoides. SHAMSUDIN & LARSSON (1993), relataram que características da estrutura da membrana espermática, relacionadas a variações entre linhagens de uma mesma raça, podem conferir ao espermatozoide resistência a congelamento.

O mesmo foi observado por TADA et al. (1993), que relataram maior criotolerância em camundongos de determinadas linhagens. Resultados semelhantes foram relatados por THURSTON et al. (2002) na espécie suína, que identificaram diferenças específicas no DNA de animais classificados como mais ou menos criotolerantes.

A criotolerância pode variar entre touros distintos e entre ejaculados do mesmo touro, sendo que entre 5 a 15% dos ejaculados são usualmente descartados após o congelamento pelas centrais de inseminação artificial (DEFOIN et al., 2008).

A identificação de touros com alta criotolerância traria inúmeros benefícios para as centrais de inseminação, tais como a) redução do descarte de palhetas após a congelamento devido à baixa qualidade espermática; b) redução dos custos com manutenção de touros na fase preparatória para coletas c) otimização do número de

espermatozoides por palheta, maximizando o uso de touros com sêmen de alta qualidade e d) personalização de protocolos de congelamento de sêmen de touros com baixa tolerância a congelação, mas com alto valor econômico (AMANN; KATZ, 2004).

2.3. Estudos de Associação Genômica (GWAS)

Apesar dos resultados positivos provenientes da aplicação de programas de melhoramento no rebanho nacional ao longo dos últimos 20 anos, características de baixa herdabilidade e de difícil mensuração, ou que não podem ser diretamente mensuradas, apresentam baixa resposta aos métodos de seleção tradicional.

A capacidade de compreender a manifestação de características complexas e estimar os efeitos de segmentos cromossômicos na expressão dessas características permite aumentar a eficiência de seleção pela utilização destas informações na predição do valor genético (FRAZER, 2009).

Nesse sentido, a aplicação de conceitos e ferramentas de genética molecular pode contribuir para suplantando algumas destas limitações.

Graças ao Projeto Genoma Bovino, concluído em 2009, e consequente disponibilidade de micro-arranjos de alta densidade de SNP e de vasta aplicabilidade (Bovine 50K – painel contendo mais de 54.000 SNP – Illumina Inc. - USA) (THE BOVINE HAPMAP CONSORTIUM, 2009; THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009), a primeira utilização do conceito *Whole Genome Enabled Animal Selection* (WGEAS) (ou Seleção Genômica - GS) ocorreu em bovinos de leite na América do Norte no final da primeira década do século XXI.

O resultado prático da aplicação da estratégia originalmente proposta por MEUWISSEN et al. (2001) foi a determinação de DEPs Genômicas (GEPBV), que passaram a ser publicadas oficialmente pelas associações de criadores de raças leiteiras em países como Estados Unidos, Canadá, Austrália, Nova Zelândia e Holanda (HAYES et al., 2009).

A aplicação prática da GS em rebanhos bovinos de leite tem mostrado ser possível a predição do mérito genético de animais jovens, antes mesmo de introduzi-los

no programa de melhoramento, direcionando a escolha dos melhores reprodutores (WEIGEL et al., 2010).

Além desta já consagrada aplicação da genômica em bovinos de leite, baseada no aumento da acurácia de programas de melhoramento através da determinação de DEPs Genômicas (HAYES et al., 2009), outra importante perspectiva de aplicação dos conhecimentos genômicos, consiste na identificação e mapeamento dos genes que interferem na expressão de características de importância econômica, visando melhorar a compreensão do controle genético de características complexas, ou seja, aquelas sob controle poligênico, e em última instância, propor novas estratégias para a seleção dos melhores indivíduos de uma população.

Os Estudos de Associação Genômica (GWAS - *Genome Wide Association Study*) tem caráter altamente inovador, pois permitem a identificação de associações estatísticas entre centenas de locos genômicos e características complexas, aumentando o entendimento sobre importantes vias moleculares e biológicas (SCOTT et al., 2007).

O GWAS foi um conceito originalmente utilizado pela comunidade científica envolvida com a genética epidemiológica humana, e tem por finalidade a identificação de regiões do genoma relacionadas com fenótipos específicos para posterior investigação de suas funções biológicas (NATIONAL HUMAN GENOME RESEARCH INSTITUTE, 2011).

A abordagem GWAS representa uma grande evolução em relação à metodologia clássica de análise de genes candidatos, uma vez que não é necessário conhecimento prévio profundo dos mecanismos fisiológicos envolvidos na manifestação da característica investigada (TAYLOR et al., 1998; WANG et al., 2005).

Três fatores principais favoreceram a popularização dessa metodologia. O primeiro deles foi o modelo descrito por LANDER (1996), denominado de doença comum, variante comum (DC-VC). O modelo DC-VC baseia-se no fato de que o perfil genético das doenças comuns é determinado por variações genéticas comuns na população (MAF – menor frequência alélica inferior a 0,05). Sendo assim, a base

genética das doenças pode ser descoberta pela identificação de variantes com diferentes frequências alélicas em classes fenotípicas distintas (SEBASTIANI, 2009).

Para que essa abordagem se tornasse viável era necessário o acesso às variantes genômicas através de tecnologias de genotipagem massal. Esse problema foi resolvido com o desenvolvimento de micro-arranjos para a detecção de milhares de variações simultaneamente. Esses micro-arranjos permitem a varredura do genoma de uma população, a um baixo custo e de maneira muito rápida, comparado ao que vinha sendo feito com o uso de outras tecnologias. A partir do desenvolvimento dessas ferramentas, foi possível a disponibilização de catálogos que permitiram o acesso a diversas informações sobre as variantes genômicas nas diferentes espécies (INTERNATIONAL HAPMAP CONSORTIUM, 2005, BOVINE INTERNATIONAL CONSORTIUM, 2009).

De forma geral, a maioria das variações genéticas são teoricamente neutras não representando contribuição para a variação fenotípica. As demais variações podem ser classificadas quanto à sua forma de apresentação, podendo ser denominadas como polimorfismos estruturais, quando as diferenças entre indivíduos ocorrem em mais de um par de bases, ou de sítio único, quando as diferenças ocorrem em apenas um par de bases (Figura 4) (FRAZER, 2009).

Dentre as variações estruturais (VE) mais comuns podemos citar as inserções-deleções (INDELS), as inversões de sequência, as substituições em bloco e as variações no número de cópias (CNV) (REDON et al., 2006; COOPER et al., 2008; KORN et al., 2008; BICKHART et al., 2012).

Os polimorfismos de nucleotídeo em sítio único (SNP) são a forma mais comum de variação no genoma. Os SNP são em sua maioria dialélicos, sendo o alelo menos frequente (MAF) presente em pelo menos 1% da população testada (BROOKES, 1999). Apesar da alta recorrência na observação de mutações em um único nucleotídeo ao longo dos genomas (em média uma a cada 1.000 pares de base), muitas dessas variações não são SNP propriamente ditos e sim variantes raras, pois não estão presentes em mais de 1% da população (LI; SADLER, 1991; WANG et al., 1998 e 2005).



Figura 4. Apresentação esquemática dos tipos mais comuns de variações genéticas. Adaptado de FRAZER (2009).

Estudo realizado por GABRIEL (2002), na espécie humana demonstrou as primeiras evidências de que, com exceção de algumas regiões onde a recombinação ocorre com alta frequência (*hotspots*), o genoma é caracterizado por uma estrutura de blocos que se apresentam altamente correlacionados com os SNP pelo princípio de desequilíbrio de ligação (LD). O LD é a associação não aleatória de alelos em dois ou mais locos no genoma, formando blocos de haplótipos que segregam juntos na população. Segundo SEBASTIANI et al. (2003), apenas alguns SNP, denominados de *tagSNP*, seriam suficientes para representar a variabilidade de um bloco haplotípico (GABRIEL, 2002).

Na espécie bovina, a primeira grande iniciativa de identificação de SNP foi realizada pelo THE BOVINE HAPMAP CONSORTIUM em 2009. Esse trabalho interrogou 37.470 SNP em 497 animais de 19 raças distintas demonstrando diferenças na MAF de SNP das subespécies taurina e zebuína. Raças taurinas apresentaram SNP com MAFs médias superiores (30% SNP com MAF >0.3) às das raças zebuínas (19% SNP com MAF >0.3) e maior LD a longas distâncias no genoma.

Os SNP podem ser originados por diversos processos naturais ou induzidos, entre eles a ocorrência natural de mutações e a sua seleção, a recombinação homóloga

e, em casos mais raros, os agentes mutagênicos externos os principais causadores desse tipo de polimorfismo (MOUNTAIN et al., 1992). Apesar de serem gerados continuamente no DNA dos indivíduos, a maioria dos polimorfismos de sítio único foram originados após a especiação (diferenciação das espécies), e antes da formação de diferentes populações. Sendo assim, indivíduos da mesma espécie apresentam 85% de SNP comuns e 15% de SNP específicos para cada população, o que demonstra a estabilidade desse tipo de variação (BARBUJANI et al., 1997).

Com relação à distribuição dos SNP ao longo do genoma, de acordo com o THE BOVINE HAPMAP CONSORTIUM (2009), a maioria dos SNP descritos na espécie bovina se localiza em regiões intergênicas (63,74%), sendo a minoria localizada nas regiões intrônicas (34,9%) e exônicas (1,35%) do genoma.

Os SNP podem estar associados ao fenótipo diretamente, quando o SNP é a mutação causal que altera a regulação e/ou expressão do gene, ou indiretamente, quando o SNP está ligado à mutação causal por desequilíbrio de ligação (LD) (Figura 5).

Segundo PEARSON & MANOLIO (2011), os estudos de associação representam novos desafios na interpretação de dados, além de utilizar novas terminologias e modelos estatísticos. De forma geral, essa abordagem requer alguns elementos essenciais para ser realizada com sucesso (CANTOR, 2010).

O primeiro desafio consiste na necessidade de utilização de amostras provenientes de uma população representativa do fenótipo de interesse. O tamanho amostral deve ser suficientemente grande para detectar até mesmo pequenas associações com os SNP variantes, já que polimorfismos que explicam até 1% do fenótipo podem ser efetivos na elucidação de suas bases biológicas (CANTOR, 2010; GOLDSTEIN, 2009). A utilização de um grande número de amostras torna essa metodologia bastante dispendiosa, sendo que a realização do GWAS em múltiplas etapas pode minimizar os custos de estudos dessa natureza. Sendo assim, pode ser realizada uma etapa de triagem onde se analisa número pequeno de animais utilizando micro-arranjos de SNP de alta densidade. A etapa seguinte consiste em interrogar apenas os SNP que demonstraram forte associação com os fenótipos em um número

maior de animais (FEUGAND et al., 2009). A utilização de múltiplas etapas beneficia principalmente os estudos envolvendo características complexas cuja mensuração dos fenótipos é onerosa, difícil e demorada (FRAZER, 2009).

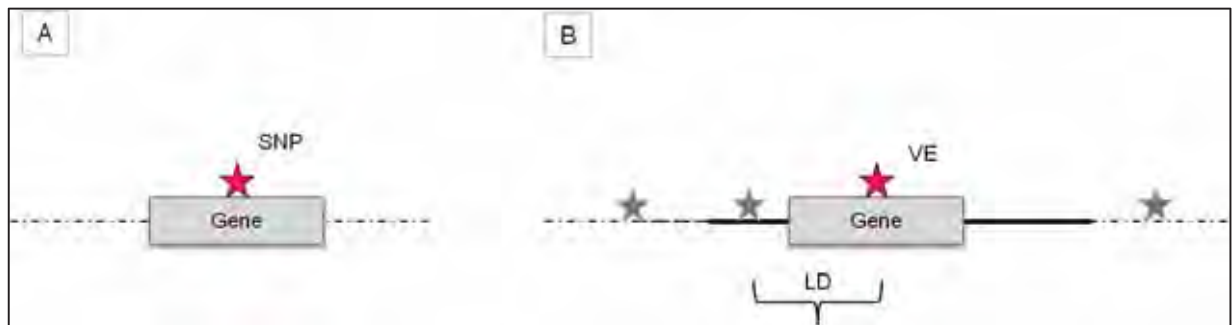


Figura 5. Apresentação esquemática da utilização direta (A) e indireta (B) do SNP para elucidar os genes envolvidos com a expressão fenotípica (estrela vermelha representa a mutação causal; estrela cinza representa SNP; linha cheia/tracejada representa a cadeia de DNA; retângulo cinza representa o exon gênico). Adaptado de HISCHHORN e DALY, 2005.

Um dos elementos mais crítico no GWAS é a utilização de métodos estatísticos suficientemente poderosos que permitam detectar associações sem nenhum tipo de viés, descartando aquelas consideradas falsas (CHAN, 2009; CANTOR, 2010). Diversos autores proveram listas extensas de fatores e métodos a serem considerados na análise (SACCONI et al., 2008; CHAN, 2009; CANTOR, 2010).

O emprego de métodos estatísticos apropriados é de particular importância, pois em algumas estratégias para o GWAS realiza-se a comparação entre centenas de milhares de SNP, um a um, o que aumenta as taxas de falsos positivos ou erro do tipo I (CHAN, 2009). Um processo eficiente de controle de qualidade (CQ) dos SNP, testes estatísticos robustos, análise de haplótipos e limites de significância elevados são alternativas normalmente utilizadas para diminuir as associações espúrias (EASTON et al., 2007; SLADEK et al., 2007; THE WELLCOME TRUST CASE CONTROL CONSORTIUM, 2007; BOLORMAA et al., 2010).

Devido ao grande número de SNP e amostras utilizadas no GWAS, pequenos erros no genótipo podem resultar em níveis inaceitáveis de erros do tipo I (falso-positivo) e II (falso negativo). A realização de um procedimento de CQ dos SNP e das

amostras dever ser cuidadosamente realizado como etapa precedente ao teste de associação (MANOLIO, 2007; ZIEGLER, 2008).

A aplicação de diferentes critérios no CQ visa à remoção de SNP e/ou amostras que introduziriam vieses nos resultados sem que ocorra perda no poder de detecção de associações verdadeiras. Para se entender a magnitude dos impactos negativos de erros do tipo I em estudos de associação, basta considerar uma situação hipotética onde um estudo realizado com 770.000 SNP e nível de significância de 0,001 apresentaria pelo menos 770 SNP espuriamente associados ao fenótipo devido a esse tipo de erro.

Segundo FREIMER et al. (2005), não existem protocolos ou limiares de controle de qualidade rígidos para o GWAS. Embora alguns estudos e/ou bancos de dados de genótipos / fenótipos exijam que os dados atinjam determinados critérios para a inclusão no estudo, esse limiares são bastante flexíveis dependendo das estratégias de análise adotadas pelo pesquisador. A escolha dos parâmetros de CQ é tipicamente baseada em atributos que objetivam refletir a utilidade e a integridade do genótipo de um dado SNP. De forma geral, esse procedimento é dividido em CQ dos genótipos de cada SNP (a) e do conjunto de genótipos de SNP de uma dada amostra (b).

2.3.1. Controle de qualidade dos genótipos de cada SNP

2.3.1.1. Call Rate e Gene Call Score (GCScore)

A designação do genótipo dos SNP é determinada pela intensidade de fluorescência emitida pela reação físico-química que ocorre no sítio de hibridização específico para cada SNP, que é captada por diferentes tipos de plataformas de detecção. Para tanto, é necessário minimizar diferenças nas intensidades da fluorescência ocasionadas por fatores não biológicos. Esse processo é denominado de normalização do sinal, e consiste na transformação dos níveis de fluorescência em escalas coordenadas de força e contraste, que são interpretados como r e θ respectivamente, nas coordenadas polares dos sinais alélicos (XIAO et al., 2007).

Intensidades normalizadas geralmente resultam na formação de três possíveis grupos que representam os genótipos hipotéticos AA, AB e BB (Figura 7).

Apesar dos algoritmos responsáveis pela designação do genótipo serem eficientes, uma pequena fração de SNP não será agrupada na posição esperada, resultando em erros na designação dos genótipos e até mesmo a não designação dos mesmos (*No Call*) (TEO et al., 2007).

SNP com sinais de fluorescência agrupados terão seus genótipos designados a um determinado valor de confiança (*GC Score*) que dependerá da distância do SNP do centro (*cluster center*) do grupo no qual ele foi designado. Como demonstrado pela Figura 6, o *GC Score* decresce à medida que o SNP se afasta do centro do grupo até um limite mínimo para detecção (McCARTHY et al., 2008).

Grandes quantidades de *No Call* e baixo *GC Score* podem acarretar na redução do poder de associação devido à diminuição do número e da confiabilidade das informações que serão utilizadas na análise. Além disso, podem acarretar vieses nos resultados, alterando as estimativas das frequências alélicas das amostras analisadas (FU et al., 2009).

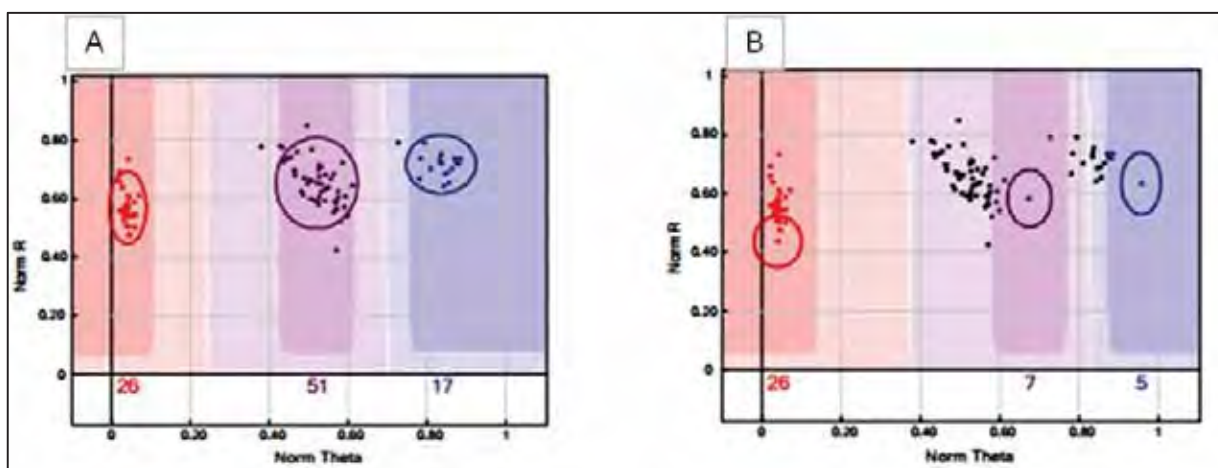


Figura 6. Agrupamento da intensidade de fluorescência normalizada em diferentes classes de genótipos representados pelas diferentes cores (AA= vermelho, AB=púrpura e BB=azul) (A); Falha no agrupamento da intensidade de fluorescência normalizada, resultando em *No Call* onde SNP representados por pontos da cor preta falharam na designação do genótipo (B). Adaptado de ILLUMINA (2010).

2.3.1.2. Equilíbrio de Hardy-Weinberg

O equilíbrio de Hardy-Weinberg (EHW) preconiza que em populações grandes, homogêneas, com acasalamento ao acaso e onde as duas cópias de um mesmo gene são herdadas de forma independente, as frequências alélicas se mantêm constantes por muitas gerações. SNP que desviam do EHW podem indicar endogamia, estratificação na população, seleção ou associação do fenótipo-genótipo (BALDING, 2006). Erros durante o processo de determinação do perfil de SNP também podem ocasionar desvios de EHW, o que tem levado esse parâmetro a ser bastante utilizado no controle de qualidade dos SNP (McCARTHY et al., 2008).

2.3.1.3. Determinação da Menor Frequência Alélica (MAF)

A menor frequência alélica (MAF) refere-se à frequência do alelo SNP menos comum em uma dada população. Na maioria dos GWAS descritos na literatura, observa-se a remoção dos SNP que apresentam MAF abaixo de 2% com o intuito de aumentar o poder de detecção da associação genótipo-fenótipo. Como os chips comerciais apresentam SNP com diversas MAFs, esse procedimento muitas vezes acarreta na remoção de muitos SNP da análise (GORLOV et al., 2008). O descarte dos SNP com MAF abaixo de um limite pré-estabelecido se justifica pela premissa de que esses genótipos podem ser produto de baixa qualidade da determinação do perfil de SNP e/ou erros na designação dos mesmos (TABANGIN et al., 2009).

2.3.2. Controle de qualidade do conjunto de genótipos de SNP de uma dada amostra

2.3.2.1. Verificação de possíveis erros humanos

O poder do estudo de associação pode ser consideravelmente reduzido se o genótipo for designado a fenótipos errados. Amostras duplicadas ou com identificação trocada devem ser identificadas e corrigidas antes da realização do teste de associação.

Um dos métodos utilizados para verificar a presença de amostras duplicadas é o cálculo da distância genética (DST – *Nei's standard genetic distance*) entre as mesmas. Esse cálculo é baseado no número de alelos de SNP que os indivíduos compartilham, segundo a fórmula geral:

$$DST = 1 - \frac{IBS2 + 0,5 IBS1}{N}$$

Onde $IBS2$ (IBS – *identical by state*) é quando os dois possíveis alelos de um locus específico são compartilhados por dois indivíduos, $IBS 1$ é quando apenas um dos dois possíveis alelos de um loco específico é compartilhado por dois e N é o número de indivíduos do estudo (NEI & ROYCHOUDHURY, 1972). A análise de distância genética é realizada em uma matriz ($N \times N$), onde amostras duplicadas teoricamente compartilham dois alelos em todos os SNP ($IBS 2$) e apresentam DST igual a zero (ou seja não existe distancia genética entre duas amostras iguais, o que caracterizaria erro humano).

A troca de amostras pode ser verificada pela identificação do sexo dos indivíduos analisados. Este procedimento é particularmente pertinente quando o sexo é uma variável levada em consideração no estudo ou quando os animais com perfil de SNP analisado são todos do mesmo sexo. Para tanto se deve avaliar a heterozigosidade observada no cromossomo X de todas as amostras. Essa metodologia baseia-se no fato de que em mamíferos os machos são heterogaméticos, apresentando apenas um cromossomo X, portanto apresentam baixa proporção de heterozigosidade nesse cromossomo (LAURIE et al., 2010).

A heterozigosidade observada também permite identificar se duas amostras foram colocadas erroneamente uma sobre a outra durante a montagem das placas para

as análises do perfil de SNP. Caso isso ocorra, o DNA proveniente de dois animais será analisado como o de um único animal, sendo a proporção de heterozigosidade muito superior à média dos demais animais do estudo (LAURIE et al., 2010).

2.3.2.2. Amostras com baixa qualidade de DNA

Quando o GWAS é realizado em um grupo amostral com centenas de amostras, é inevitável a variação na qualidade e na concentração das amostras de DNA processadas. A baixa qualidade e/ou concentração de DNA pode levar a falhas na detecção de SNP, com redução na emissão de fluorescência e aumento do número de genótipos não designados ou *No Call* (FU et al., 2009).

A proporção de *No Call* interfere diretamente no *call rate* das amostras, que é um dos principais parâmetros de qualidade do procedimento de análise de SNP calculado pela divisão do número de SNP com genótipos designados (AA + AB + BB) pelo número de SNP analisados no estudo. Amostras com baixo *call rate* devem ser removidas, com o intuito de minimizar qualquer viés na interpretação dos resultados.

2.3.2.3. Efeitos na comparação de análises realizadas independentemente

Mesmo quando o DNA apresenta alta qualidade e concentração, alguns fatores confundidores podem ser introduzidos no estudo devido a variações no procedimento analítico. Em estudos realizados com um grande número de amostras, os DNAs individuais são agrupados em diversas partidas para posterior processamento. Essas partidas são analisadas independentemente, sendo que a utilização de informações provenientes de diferentes partidas pode ocasionar em alterações na frequência alélica dos SNP dificultando a interpretação dos resultados (LAMBERT & BLACK, 2012).

A análise de componentes principais (PCA – *Principal Component Analysis*) é um método estatístico multivariado que produz componentes não correlacionados a partir de uma matriz de dados com variáveis potencialmente correlacionadas. Nesse tipo de análise, a maior variação dentro do banco de dados é representada pelo primeiro

componente principal (PC), a segunda maior variação é explicada pelo segundo PC, seguido de outros componentes que podem explicar a variação em magnitude decrescente (ANDERSON et al., 2010).

Essa abordagem permite identificar partidas que apresentem variação fora do padrão, como apresentado na Figura 7 que ilustra estudo hipotético realizado com informações genóticas provenientes de 16 partidas de análise de SNP idênticos, porém realizadas em momentos distintos, no qual dois PC são suficientes para demonstrar os efeitos introduzidos pelas partidas 11, 12, 14 e 16. Nas partidas 11, 12 e 14, apenas alguns indivíduos são responsáveis pelo efeito. Já na partida 16, todas as amostras processadas desviam-se das demais. Em estudos onde o efeito entre partidas está presente deve-se remover os indivíduos divergentes e designar novamente os genótipos.

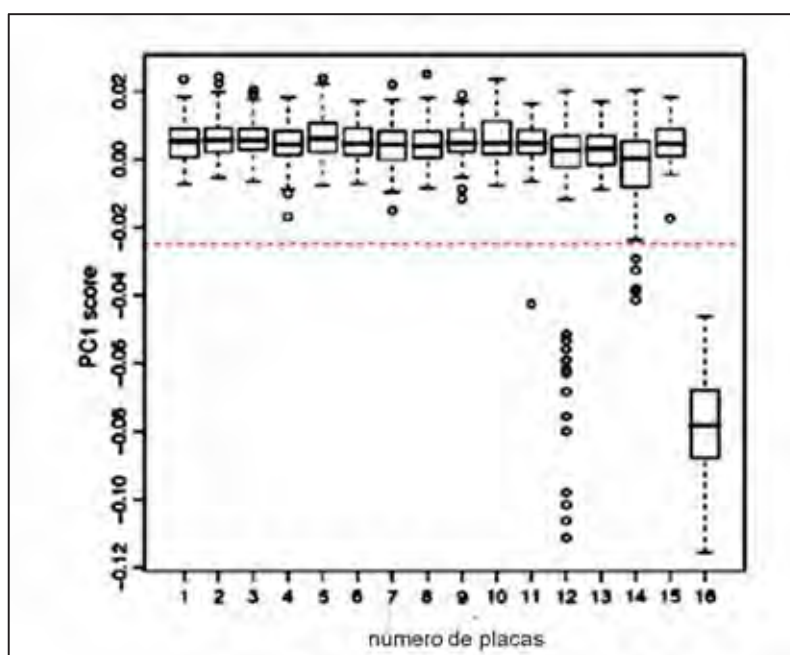


Figura 7. Detecção do efeito de partidas utilizando análise de componentes principais. Os escores dos componentes principais foram agrupados e representados em Box-plots. Adaptado de Anderson, 2011.

2.3.2.4. Estratificação da população

O efeito da estratificação da população nos estudos de associação tem sido amplamente discutido na literatura, e vem sendo citado como um dos principais fatores confundidores na interpretação dos resultados desse tipo de estudo (PRICE et al., 2006; McCARTHY et al., 2008; LAURIE et al., 2010). A estratificação da população ocorre geralmente quando as amostras são provenientes de populações ancestrais distintas. A diferença alélica entre as populações ancestrais pode acarretar uma associação significativa entre alelos e fenótipos, entretanto sem efeito causal direto ou indireto com o fenótipo (CARDON; PALMER, 2003).

Mesmo um pequeno grau de estratificação da população pode afetar análise do tipo GWAS devido aos grandes tamanhos amostrais necessários para detectar SNP envolvidos na expressão de características complexas (MARCHINI et al., 2004).

Para a verificação da estratificação da população pode ser utilizada a análise de componentes principais, sendo suficiente a utilização de apenas dois PCs para identificar grupos originados em diferentes populações ancestrais. Os escores dos PCs podem ser utilizados como covariáveis nos modelos de regressão no teste de associação, com o intuito de ajustar os dados para a ancestralidade da população (PRICE et al., 2006).

Uma das consequências mais importantes da presença de estratificação é a inflação do número de SNP associados ao fenótipo. Isso pode ser facilmente visualizado com a utilização de um gráfico quantil-quantil (*Q-Q plot*). Esse gráfico mostra a ocorrência de divergência entre os valores de significância observados e aqueles esperados na hipótese nula do teste estatístico χ^2 . Segundo PATTERSON et al. (2006), desvios da diagonal do gráfico (hipótese nula) geralmente sugerem associação verdadeira dos SNP com o fenótipo, exceto em caso que o desvio ocorre de forma generalizada, indicando inflação do teste potencialmente devido a estratificação conforme apresentado na Figura 8.

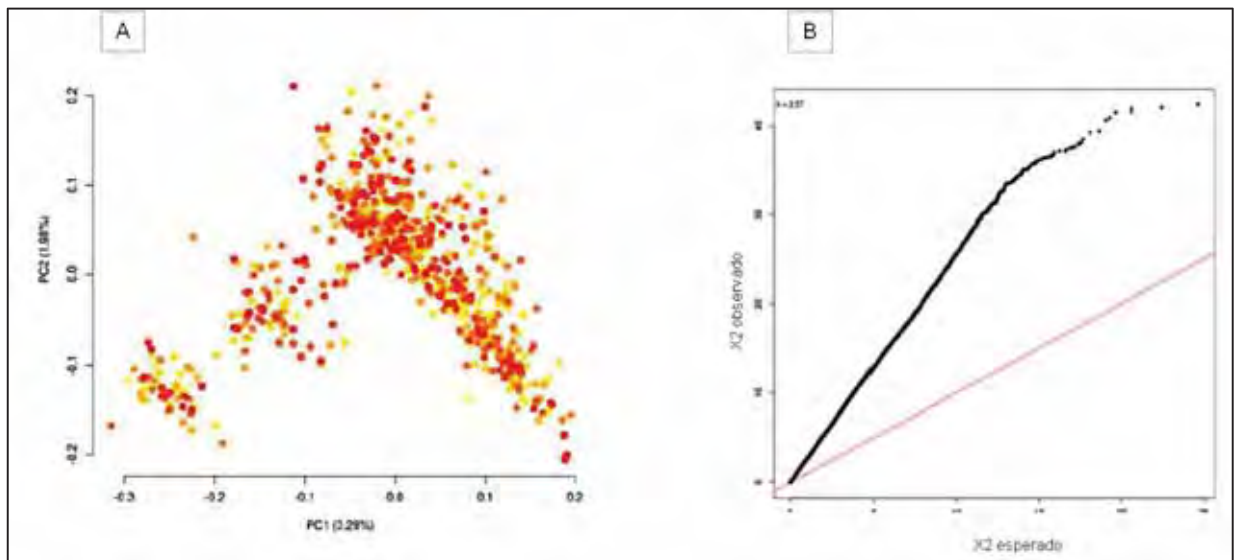


Figura 8. Análise de componentes principais (PCA) de uma população estratificada com a formação de três grupos bem definidos no gráfico (A); Inflação generalizada facilmente visualizada com o auxílio do gráfico quantil-quantil (B) (cedido por José Fernando Garcia, 2012).

2.3.3. Metodologias estatísticas utilizadas nos testes de associação

Além dos procedimentos de controle de qualidade, cuidados especiais devem ser tomados na escolha dos métodos estatísticos mais adequados para testar a existência de associação entre o fenótipo e o genótipo.

Apesar de muitos GWAS terem sido realizados na espécie humana, o que propiciou progressos consideráveis no desenvolvimento de métodos estatísticos específicos, até a presente data relativamente poucos estudos dessa natureza foram realizados em bovinos (BARENDSE et al., 2007; COLE et al., 2009; HAYES et al., 2009; BOLORMAA et al., 2010; PRYCE et al., 2010). A utilização do GWAS em diferentes espécies e fenótipos implica na busca por metodologias alternativas de análise devido à natureza das características avaliadas e às restrições quanto ao tamanho amostral (SCUTERI et al., 2007; WEEDON et al., 2008).

A maioria dos estudos realizados em humanos é do tipo caso-controle, onde a característica avaliada é binária (McCARTHY et al., 2008). Nesses casos, realiza-se a comparação entre os genótipos dos indivíduos acometidos (caso) e não acometidos

(controle) por determinada condição. O mesmo não acontece com as características de produção tipicamente estudadas em animais como os bovinos, onde os fenótipos de interesse econômico são do tipo contínuo ou ordinal, em que a variação dos mesmos entre os indivíduos de uma população pode não ser tão óbvia (BOLORMAA et al., 2010).

Uma das maneiras mais simples de analisar estudos de associação do tipo caso-controle é testar a hipótese nula (não associação) entre linhas e colunas de uma matriz 2 x 3, que contém a contagem das três classes de genótipos (AA, BB, AB), entre os casos e os controles. Para tanto, podem ser utilizados testes estatísticos como o teste de Pearson e o exato de Fischer (BALDING et al., 2008).

Em estudos com características quantitativas, o efeito do SNP sobre o fenótipo pode ser analisado como o contraste das médias de diferentes classes de genótipos sobre o fenótipo em um modelo linear. Apesar de rápida, essa abordagem ignora a covariância entre os indivíduos aparentados aumentando a taxa de falsos positivos. Para resolver esse problema, BOERWINKLE et al. (1986), propuseram a utilização de um modelo misto que considera a relação de parentesco entre os indivíduos como um componente aleatório, e estima os resíduos corrigidos para as relações de parentesco e para os efeitos fixos como mostrado a seguir:

(1)

$$Y_i = \mu + \sum_j \beta_j C_{ji} + G_i + e_i$$

Onde Y_i é o fenótipo do i -ésimo indivíduo, μ é a média da população, β_j é o efeito do j -ésimo efeito fixo ou covariável, C_{ji} é o efeito do j -ésimo efeito fixo ou covariável no i -ésimo indivíduo, G_i e e_i são a matriz de efeitos aleatórios e os resíduos do i -ésimo indivíduo, respectivamente. Nesse modelo assume-se que os efeitos aleatórios seguem uma distribuição normal multivariada com média 0. A variância do modelo poligênico é definida como $\Phi\sigma_G^2$, onde Φ é a matriz de parentesco e σ_G^2 é a variância genética aditiva do poligene.

Os resíduos são calculados de acordo com a equação:

(2)

$$Y_i^* = Y - \left(\hat{\mu} + \sum_j \hat{\beta}_j C_j + \hat{G}_i \right) = e_i$$

Onde $\hat{\beta}_j$ é o valor estimado do j-ésimo efeito fixo e $\hat{G}_i$ is té o valor estimado da contribuição do poligene (valor genético).

Essa metodologia, denominada de *Measured Genotype* (MG) utiliza os resíduos estimados na equação acima como a característica quantitativa (variável dependente) para o teste de associação com os genótipos em uma regressão linear simples como a mostrada a seguir:

(3)

$$\hat{e}_i = \mu + k g_i + e_i$$

Onde $\hat{e}_i$ é o vetor de resíduos da equação (2), μ é a média, g_i é o vetor de genótipos do i-ésimo indivíduo, k é o efeito do marcador e e_i é vetor de resíduos aleatórios dos i-ésimos indivíduos.

Esse método é bastante eficiente em estudos que testam poucos SNP, em populações não estratificadas e pouco aparentadas (BOERWINKLE, 1986). Caso o estudo não apresente essas características, métodos estatísticos específicos podem ajustar os dados para a existência de estratificação ou de amostras aparentadas com o intuito de reduzir os erros no teste de associação.

Em estudos de associação genômica espera-se que apenas uma pequena proporção de SNP sejam verdadeiramente associados ao fenótipo, sendo a grande maioria deles caracterizados pela hipótese nula da distribuição estatística χ^2 . Levando essa premissa em consideração, DEVLIN & ROEDER (1999) propuseram uma metodologia denominada de *Genomic Control* (GC). Esse método baseia-se no fato de que em grandes populações é possível estimar um fator de inflação λ para inferir a estratificação presente.

Segundo os autores, o fator de inflação pode ser estimado a partir da regressão do fenótipo pelo número de locos analisados. Para cada equação de regressão estima-se o teste estatístico T^2 , segundo a equação:

(4)

$$T_i^2 = \beta_i^2 \frac{\beta_i^2}{\text{Var}(\beta_i)}$$

Onde β_i é o efeito do i-ésimo SNP sobre o fenótipo e $\text{Var}(\beta_i)$ é a variância do efeito estimado. O fator de inflação então é calculado pela seguinte equação:

(5)

$$\lambda = \text{mediana} \frac{(T_1^2, T_2^2, \dots, T_n^2)}{0,456}$$

Onde 0,456 é o valor da mediana da distribuição χ^2 com a variância não central igual a ϕ . Nesse método, valores de λ maiores do que 1 indicam estratificação da população. Sendo assim, os valores dos testes estatísticos são corrigidos pelo λ com o intuito de minimizar os erros do tipo I.

AMIN et al. (1999), propuseram a metodologia GRAMMAR (*Genome-wide Rapid Association using Mixed Model and Regression*), que combina as metodologias MG e GC. Nesse método o teste de associação pode ser realizado da mesma maneira que no MG, onde os resíduos ambientais estimados a partir do modelo poligênico aditivo serão utilizados como fenótipos no teste de associação, sendo corrigido pelo fator de inflação λ (GC) quando há estratificação.

Métodos de associação estruturados também são comumente utilizados como opção para estudos onde a estratificação está presente. Essa abordagem baseia-se em alocar indivíduos a sub-populações e testar a associação nas diferentes sub-populações separadamente (PRITCHARD et al. 2000).

Outro método popularmente utilizado para a correção da estratificação é o EIGENSTRAT. Essa metodologia consiste de três etapas distintas, sendo a primeira

delas a utilização de PCA dos genótipos para inferir a variação genética. A partir dos escores dos PCs é realizada a correção dos genótipos computando os resíduos da regressão linear. Dessa forma a associação é realizada utilizando genótipos corrigidos previamente para a estratificação (PRICE et al., 2006).

De acordo com AULCHENKO (2011), o método estatístico mais adequado para realizar o teste de associação depende dos níveis de parentesco e estratificação da população (Figura 9). De acordo com esse autor, a aplicabilidade dos diferentes métodos pode ser distinta dependendo do tipo de cenário:

1) Grupo amostral pouco aparentado e com ancestralidade comum, determinando baixos valores para o fator de inflação λ . Nesse caso, a aplicação do método GC será suficiente para controlar o viés residual.

2) Grupo amostral pouco aparentado e proveniente de populações distintas. Nesse caso, a aplicação do método de associação estruturada é mais recomendada.

3) Grupo amostral muito aparentado. Nesse caso, deve ser considerado como método de escolha o modelo misto para a realização do teste de associação.

4) Grupo amostral com níveis moderados de parentesco e estratificação. Nesse caso deve-se utilizar o EIGENSTRAT ou métodos alternativos que ajustem os dados pelos escores dos PCs.

Os métodos de associação podem se basear ou não em informações de parentesco acuradas e podem ser afetados negativamente quando essas informações estão incompletas ou erradas. Para sanar esse problema, AMIN et al. (2007), recomendam que informações provenientes do pedigree genômico sejam consideradas devido a sua maior confiabilidade.

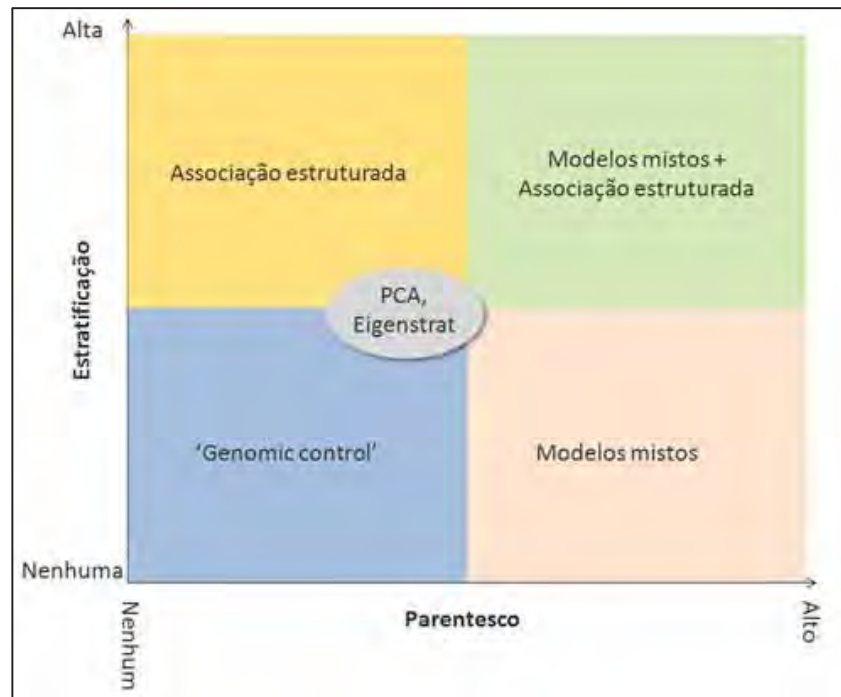


Figura 9. Possíveis cenários e métodos estatísticos para a realização do teste de associação genômica levando em consideração os níveis de estratificação e parentesco da população (proposto por AULCHENKO, 2011).

Além de escolher um teste de associação adequado para as características do estudo proposto, outra medida importante para reduzir os erros do tipo I é a correção dos valores de significância para as comparações múltiplas. A correção mais utilizada tradicionalmente, principalmente em estudos com humanos, é a correção de Bonferroni. Esse ajuste foi proposto por BENJAMIN & HOCHBERG (1995), e consiste em corrigir os valores de significância para o número de comparações independentes realizadas pelo teste de associação (valor de significância (α) / n). Um dos principais problemas dessa correção é o aumento do erro do tipo II tornando esse teste excessivamente conservativo. Além disso, esse teste impõe níveis de significância excessivamente altos, o que dificulta a identificação de associação em estudos nos quais o SNP tem efeito moderado sobre a característica (ASTON; CARREL, 2009).

A reprodutibilidade de um estudo pode ser favorecida pela utilização de um valor de significância rigoroso ou pela combinação desse valor com outros critérios. Com o

intuito de oferecer alternativas para indicar a reprodutibilidade de um estudo, sem que o ajuste de Bonferroni seja utilizado, MANLY (2005) propôs o sistema BADGE (*Better Associations for Disease and Genes*) de classificação. Esse sistema contempla uma grande variação nos valores de significância dividindo os mesmos em classes de acordo com a confiabilidade da associação representada no *Manhattan Plot*. Sendo assim, a última classe apresenta o valor de significância menos rigoroso, e a primeira classe o mais rigoroso, de tal modo que a grandeza das classes descreva a força da associação genótipo/fenótipo. Os valores de significância e a denominação de cada classe desse sistema estão representados na Tabela 1.

Tabela 1. Sistema BADGE de classificação da reprodutibilidade em estudos de associação genômica (MANLY, 2005).

1° Classe	$p < 2 \times 10^{-7}$	Sugere reprodutibilidade sob pressuposição conservadora para associação sem evidências prévias.
2° Classe	$p < 5 \times 10^{-6}$	Sugere reprodutibilidade sob pressuposição conservadora para características previamente mapeadas (~100 cM) ou outras fracas evidências de associação. Sugere reprodutibilidade sob pressuposição não conservadora para associação sem evidências prévias.
3° Classe	$p < 1 \times 10^{-4}$	Sugere reprodutibilidade sob pressuposição conservadora para características previamente mapeadas em pequenas regiões (~5 cM). Sugere reprodutibilidade sob pressuposição não conservadora para associação com evidências prévias fracas.
4° Classe	$p < 2 \times 10^{-3}$	Sugere reprodutibilidade sob pressuposição conservadora para características previamente mapeadas em pequenas regiões.
5° Classe	$p < 5 \times 10^{-2}$	Confirma associação de primeira e segunda classe em amostras de população independente.

2.4. Análise de enriquecimento funcional

Os desafios para a geração de conhecimentos úteis e aplicáveis a partir do emprego de abordagens analíticas no âmbito, as quais invariavelmente geram enormes quantidades de informação, não se assentam apenas nas etapas de processamento e análise de dados, mas principalmente naquelas que envolvem sua interpretação. Nesse contexto, novas técnicas de mineração de dados, capazes de extrair informações biológicas consistentes, são necessárias para interpretar o significado biológico de um dado experimento. Sendo assim, pode-se dizer que um dos maiores desafios da biologia computacional atual é o de prover métodos automáticos robustos que permitam inferir a funcionalidade de uma longa lista de genes derivada de estudos de associação GWAS (FONTANILLO et al., 2011).

Os conhecimentos gerados nos GWAS podem ser posteriormente utilizados em áreas como transcriptômica, epigenômica, proteômica e mapeamento metabólico. A combinação dinâmica das informações provenientes desses estudos permite integrar as diferentes partes do todo, gerando um conceito denominado de Biologia de Sistemas (CHUANG, 2010).

Aspecto essencial dessa abordagem é o de traduzir as assinaturas moleculares em informações que possam auxiliar o entendimento de mecanismos biológicos. Nos últimos anos, diversos métodos e ferramentas foram desenvolvidos para interpretar longas listas de genes ou proteínas utilizando informações disponíveis nos bancos de dados biológicos (FONTANILLO et al., 2011).

Basicamente, essa abordagem objetiva avaliar a frequência dos termos funcionais da lista de genes, aplicando testes estatísticos para determinar aqueles termos significantemente representados ou enriquecidos (WEBBER, 2011).

Informações curadas provenientes de diferentes fontes, tais como Gene Ontology (GO) (THE GENE ONTOLOGY CONSORTIUM, 2000) e KEGG (KANEHISA; GOTO, 2000), são comumente utilizadas nesse contexto. Essas fontes provêm mapas e

informações que agrupam padrões comuns de diferentes interações entre os componentes do sistema (KANEHISA; GOTO, 2000).

Diversas ferramentas foram desenvolvidas para a aplicação do enriquecimento funcional, apesar de cada aplicação introduzir suas próprias fontes de variação tais como: diferentes testes estatísticos, fontes de termos e organismos com informação disponível, de forma geral todos realizam o mesmo tipo de análise produzindo poucas diferenças nos resultados.

De acordo com FONTANILLO et al. (2011), a análise de enriquecimento funcional (AEF) pode ser classificada em três categorias distintas: análise de enriquecimento singular (SEA), análise de enriquecimento de genes (GSEA) e análise de enriquecimento modular (MEA).

Uma das primeiras abordagens da AEF foi estabelecer limites de significância para cada gene com o intuito de selecionar os mais importantes para a análise de enriquecimento. Essa metodologia é denominada de análise de enriquecimento singular (SEA) e pode apresentar algumas limitações tais como: a) nenhum gene atingir os limites de significância necessários para serem incluídos na análise devido a seus efeitos modestos na característica; b) resultados podem não apresentar termos biológicos comuns entre os genes incluídos na análise, dificultando a interpretação dos resultados e c) resultados de diferentes estudos sobre a mesma característica podem apresentar poucos termos funcionais comuns.

Com o intuito de superar essas limitações, SUBRAMANIAN & TAMAYO (2005) propuseram o GSEA. Esse método baseia-se na comparação entre uma lista de genes (LG) ordenados de acordo com sua correlação com o fenótipo e os genes descritos em uma via metabólica específica (VM). Para tanto, são realizados cálculos do índice de enriquecimento, níveis de significância e níveis de significância corrigidos para testes múltiplos.

O índice de enriquecimento (IE) reflete em que proporção a LG está representada em uma dada VM. O índice é calculado da seguinte forma: quando um componente da LG está numa VM, o índice é acrescido de um determinado valor. Caso contrário, o índice é decrescido desse valor. A magnitude do valor que será acrescido

ou decrescido do índice a cada comparação depende da correlação do gene com o fenótipo e do software utilizado.

O valor de significância nominal do IE é estimado pelo teste de permutação dos genes. Basicamente o que ocorre é o cálculo de uma nova estimativa de IE a cada permutação, gerando distribuição nula para os índices. O valor de significância nominal do IE é calculado com base na distribuição nula gerada pelas permutações e corrigido posteriormente para comparações múltiplas.

É importante lembrar que esse método foi primariamente desenvolvido para dados provenientes de estudos de expressão gênica, onde a LG é determinada por genes com expressão diferencial em determinadas classes de fenótipos. Apesar disso, SUBRAMANIAN & TAMAYO (2005) afirmam que a utilização dessa metodologia, com LG provenientes de estudos de associação genômica tem sido bastante promissora para a formulação de hipóteses biológicas.

As abordagens SEA e GSEA avaliam cada gene independentemente ignorando as relações entre eles. A análise de enriquecimento modular (MEA) permite a integração de informações provenientes de diferentes fontes e identifica combinações de termos enriquecidos para uma lista de genes ou proteínas. Essa abordagem representa um avanço nas análises de enriquecimento, pois permite a integração de anotações heterogêneas descobrindo combinações significativas entre elas.

Atualmente, existem mais 70 de tipos de softwares capazes de realizar AEF (HUANG et al., 2009a). Dentre os mais populares podemos destacar o software DAVID (HUANG et al., 2009b), que permite a integração de diversos bancos de dados simultaneamente e o uso algoritmos avançados de análise de enriquecimento modular.

Esse software permite realizar a análise funcional seguindo conceito gênico próprio (DAVID *Gene Concept*) para agrupar genes e proteínas espécies-específicas provenientes de diversos bancos de dados públicos como o NCBI, PIR e Uniprot/SwissProt, considerando valores IE maiores do que 2,00 como de relevância para investigações mais detalhadas. Esse método agrupa dezenas de milhares de identificadores provenientes de mais de 65.000 espécies em 1,5 milhões de proteínas e genes únicos.

2.5. A aplicação da genômica na reprodução

Atualmente existem mais de setecentos genes descritos como associados à reprodução/fertilidade em diferentes espécies, possibilitando sua utilização como candidatos para o estudo da biologia das características reprodutivas (ASTON; CARREL, 2009).

No que tange a fertilidade em bovinos, FEUGAND et al. (2009) realizaram GWAS multi-estágio em touros da raça Holandesa sendo que na primeira fase foram analisados vinte touros de alta e baixa fertilidade, evidenciando quatro SNP significativamente associados à característica ($p \leq 10^{-4}$), e na segunda fase apenas esses SNP foram interrogados em uma população de 200 animais confirmando a associação significativa de apenas dois deles (rs29024867 e rs41257187; $p < 0.05$). Os resultados confirmaram a importância do gene da Integrina Beta 5 (IGB5) na fertilidade de bovinos, embora não tenha sido definido um papel específico para esse gene.

Estudo realizado por PRYCE et al. (2010), demonstrou que setenta e três SNP estavam associados à fertilidade em setecentos e oitenta vacas da raça Holandesa. Os SNP em questão estavam posicionados principalmente nos cromossomos BTA6, BTA14, BTA20, BTA29, e embora o resultado não tenha se repetido na população de validação, os autores observaram um haplótipo fortemente associado a essa característica no cromossomo 18 em duas populações de validação distintas. Resultados semelhantes foram encontrados em estudos realizados por KUHN et al., (2003), ASHWELL et al., (2004), HOLMBERG & ANDERSSON-EKLUND (2006), e COLE et al., (2009).

Como causas mais frequentes de infertilidade no macho, destacam-se fatores que sofrem influência genética tais como a oligozoospermia, azoospermia e baixa qualidade seminal. Sendo assim, alguns grupos de pesquisa têm realizado GWAS com o intuito de descobrir SNP associados a esses fenótipos como preditores indiretos da fertilidade (ASTON; CARREL, 2009 e 2010; XI-HUANG, et al., 2011).

ASTON & CARREL (2009) em estudo piloto em humanos, não observaram SNP significativamente associados a essas características após a correção de Bonferroni. Mesmo assim esses autores consideraram como importantes candidatos para estudos posteriores vinte SNP que apresentaram significância sem a aplicação de correções para as comparações múltiplas ($p \leq 10^{-5}$), caracterizando-se como um exemplo da importância de estudos prévios com GWAS para direcionar estudos futuros, mesmo quando as associações detectadas inicialmente sejam são consideradas marginais. Esses mesmos autores realizaram segunda fase do trabalho (ASTON; CARREL, 2010) interrogando oitenta e quatro SNP selecionados na primeira fase. Dessa vez, apenas quatorze SNP foram significativamente associados ($p < 0,05$) às características sem a aplicação de correção para as comparações múltiplas.

FORTES et al. (2010), realizaram um GWAS para as características idade a puberdade, intervalo de anestro e primeira ovulação pós-parto em animais das raças Tropical Composta e Brahman, e observaram que os SNP significativamente associados apresentavam-se em sua maioria nas regiões cromossômicas BTA 14 e BTA 5 nas raças Brahman e Tropical Composta, respectivamente.

Em um dos estudos de associação mais completos já realizados em bovinos, COLE et al., (2011) analisaram trinta e uma características produtivas, incluindo cinco características reprodutivas de importância econômica. A taxa de prenhez foi avaliada como uma medida indireta de fertilidade da vaca e se apresentou geneticamente correlacionada com fenótipos produtivos, como a duração da vida produtiva e o escore de células somáticas.

Ao analisar essas três características simultaneamente, esses autores observaram dois grupamentos de SNP significativamente associados. O primeiro deles localizado em uma região de 15,7 Mb do BTA 7, correspondendo a 1.166 genes, destacando-se o gene INSR (gene receptor de insulina). O segundo grupamento localizado em uma região de 106 Mb do BTAX, próximo ao gene LOC 520057 (gene da proteína inibidora de fosfatase do tipo I). Adicionalmente foram localizadas regiões associadas especificamente a taxa de prenhez no cromossomo BTA1 e BTA3, destacando-se o gene ATP1B4 (gene ATPase transportadora de Na^+/K^+ beta 4

polipeptídeo) e GRIA3 (gene receptor de Glutamato 3) no BTAX no fenótipo taxa de prenhez.

Ao analisar o efeito do touro na facilidade de parto e na taxa de natimortos, através da análise de suas filhas, o estudo demonstrou que SNP associados a essas características encontram-se agrupados em uma região de 15,82 MB no cromossomo BTA18, correspondendo a 1.322 genes, com destaque para os genes PGLYRP1-IGF1 e LOC787057. A característica efeito da vaca na taxa de natimorto apresentou-se associada a uma região genômica específica de 211,67Kb no BTA 23, próximo ao gene MOCS1-LRFN2 e a genes CD82 do cromossomo BTA15. Segundo esses autores, o gene MOCS em questão já foi relatado na literatura em estudos de associação de natimortos em humanos.

A utilização da abordagem GWAS para identificar SNP associados a diferentes características reprodutivas, a grande variedade de métodos de análise e alta abrangência demonstrada por seus resultados apontam para os desafios impostos pela aplicação dessa metodologia na atualidade. O foco do presente trabalho foi de identificar regiões cromossômicas associadas a característica qualitativas do sêmen através do emprego de GWAS em bovinos da raça Nelore, com o intuito de apresentar novas perspectivas para o entendimento das vias metabólicas e dos processos fisiológicos responsáveis por seu controle.

3. HIPÓTESES E OBJETIVO

3.1. Hipóteses

A metodologia GWAS, aliada à análise de enriquecimento funcional, é capaz de identificar regiões cromossômicas em bovinos da raça Nelore associadas aos fenótipos: diferença na motilidade entre o sêmen fresco e descongelado (DMD) e diferença na motilidade sêmen fresco e após teste de termo resistência (DMT).

3.2. Objetivos gerais

Identificar e caracterizar as regiões cromossômicas associadas aos fenótipos DMD e DMT, levando a predição de mapas metabólicos que permitam a identificação de mecanismos reguladores da criotolerância espermática dos bovinos e seus efeitos na manifestação fenotípica.

3.3. Objetivos Específicos

1. Comparar a utilização do valor genômico estimado (EBV) e das médias das medidas fenotípicas para os fenótipos DMD e DMT em estudo de associação genômica (GWAS).

2. Identificar regiões genômicas potencialmente associadas à manifestação dos fenótipos DMD e DMT.

3. Propor possível (eis) mecanismo (s) de interação dos genes localizados nas regiões cromossômicas identificadas com relação a manifestação dos fenótipos DMD e DMT.

4. MATERIAL E MÉTODOS

4.1. Delineamento Experimental: Motilidade Pós-descongelação e Motilidade Pós-teste de Termo Resistência

Para avaliação dos fenótipos foram coletados cerca de 50 ejaculados de 125 animais da raça Nelore, que compõe o quadro de doadores de sêmen da Empresa Alta Genetics do Brasil Ltda., em Uberaba – MG, entre 2005 e 2010.

Antes integrarem o grupo experimental, todos os animais foram submetidos a avaliação clínica geral e do sistema reprodutivo de acordo com procedimento

preconizado pelo Manual do Colégio Brasileiro de Reprodução Animal para exame andrológico e avaliação de sêmen (1998) com intuito de garantir que os animais encontravam-se saudáveis. Cada um dos touros teve os ejaculados colhidos com auxílio de vagina artificial de acordo com procedimentos de rotina. Todos os ejaculados foram submetidos ao mesmo protocolo de criopreservação.

A avaliação espermática consistiu da observação da motilidade e vigor espermático, realizada imediatamente após a coleta do sêmen, após a descongelação e após o teste de termo resistência (TTR). O quadro com a análise descritiva das médias e desvios padrão dos fenótipos de cada touro encontra-se no Anexo 1.

A avaliação convencional da motilidade e do vigor foi baseada na análise visual da porcentagem de espermatozoides móveis em (escala de 0 a 100%) e do escore da energia das células em movimento (em escala de 1 a 5), respectivamente. Para a realização dessas avaliações, uma gota de sêmen foi colocada entre lâmina e lamínula (pré-aquecidas a 37°C) sob microscópio óptico em aumento 100X.

A longevidade do sêmen e sua resistência à temperatura foram avaliadas através do TTR lento. O teste foi realizado incubando o sêmen descongelado a 38°C por 3 horas (VIANNA et al., 2009). Foram realizadas avaliações de motilidade e vigor após o fim do período de incubação.

Os fenótipos considerados para a realização do presente estudo foram a diferença de motilidade entre o sêmen fresco (MSF) e o descongelado (MSD), denominada DMD e a diferença de motilidade entre o sêmen fresco (MSF) e o sêmen pós TTR (MST), denominada de DMT.

Para análise da criotolerância espermática foram consideradas as diferenças entre motilidades avaliadas em dois tempos distintos, ao invés de seus valores absolutos, com o intuito de eliminar potenciais confundidores na avaliação dessa característica. Sendo assim, um animal hipotético A que apresentou MSF = 80% e MSD = 60% ($80-60 = 20$) será igualmente criotolerante ao animal B que apresenta MSF = 60% e MSD = 40% ($60-40 = 20$), apesar do touro A apresentar maior MSF e MSD do que o touro B.

Foram descartadas as avaliações de qualidade seminal que estavam acima ou abaixo de quatro desvios padrão da média. Essa medida foi tomada para evitar que avaliações não representativas do animal, devido à ocorrência de situações pontuais, fossem incluídas na análise.

4.2. Extração DNA genômico

A extração do DNA genômico foi realizada a partir do sêmen dos animais, utilizando protocolo baseado na extração com fenol:clorofórmio:álcool isoamílico (25:24:1), segundo rotina utilizada no laboratório de Bioquímica e Biologia Molecular Animal (LBBMA-UNESP, Campus Araçatuba).

Após a extração do DNA, sua concentração foi determinada (em ng/mL), assim como seu grau de pureza (relação entre as densidades óticas a 260 e 280nm) utilizando-se espectrofotômetro (Nanodrop - Thermo Fisher Scientific, Waltham, Massachusetts, EUA).

4.3. Análise em micro-arranjo de polimorfismo de sítio único (SNP) de alta densidade

O perfil de SNP das amostras foi determinado pela plataforma Infinium com o ensaio HDBovine Illumina[®], que contém aproximadamente 777.000 SNP, dos quais 538.517 são polimórficos para a subespécie *Bos primigenius indicus*. Estudos de validação prévios demonstraram que a plataforma em questão apresenta *call rate* médio de 99,6% e reprodutibilidade de 100%, sendo adequada para estudos com animais dessa subespécie (Illumina, 2010).

O protocolo de análise do perfil de SNP foi realizado como o preconizado pelo fabricante e está disponível em detalhes no endereço eletrônico www.illumina.com. Em suma, o protocolo em questão consistiu de seis etapas principais distintas:

Amplificação: o DNA das amostras foi desnaturado, neutralizado e submetido a um processo de amplificação isotérmico. Essa etapa objetivou aumentar a

concentração de DNA centenas de vezes para que isso não fosse um fator limitante no processo, reduzindo a qualidade dos resultados.

Fragmentação: o produto de amplificação foi fragmentado por um processo enzimático do tipo *end-point*, isto é, os sítios de fragmentação são determinados por atributos específicos da sequência de DNA, o que evita a ocorrência de fragmentação excessiva.

Hibridização: o DNA amplificado e fragmentado se anela (hibridiza) à sequência de 50 bases que está covalentemente ligada à nano partícula *bead*, sendo que cada *bead* corresponde a um locus específico do DNA.

Extensão de base-única: ocorre a extensão de uma única base nucleotídica da sequência, que irá incorporar os corantes que permitirão a determinando os genótipos.

Coloração: coloração da lâmina (micro-arranjo). Diversos corantes foram incorporados aos *beads* sendo responsáveis pela emissão de fluorescência.

Visualização das Imagens: Um conjunto de laser foi utilizado para excitar os fluoróforos que estavam ligados aos *beads* na lâmina, permitindo que o sinal de fluorescência fosse captado e analisado.

4.4. Análise de GWAS

4.4.1. Análise dos fenótipos

De forma geral, os estudos de associação são realizados com características mensuradas uma única vez, o que difere do presente estudo onde o mesmo touro teve diversas amostras de sêmen avaliadas em distintas ocasiões. As duas abordagens mais empregadas para analisar características quantitativas longitudinais em animais são: a) utilização do valor genético do animal para a característica avaliada (EBV) e b) utilização das médias das medidas fenotípicas para cada indivíduo. Sendo assim, o presente estudo analisou os fenótipos seguindo as duas abordagens possíveis em dois experimentos distintos:

a) Experimento 1: Estimativa do valor genético (EBV) para DMD e DMT em cada touro foi considerada como fenótipo.

b) Experimento 2: Média das medidas fenotípicas para DMD e DMT em cada touro foi considerada como fenótipo.

A utilização do EBV como fenótipo nos estudos de associação em bovinos já é bastante consagrada (PRYCE et al., 2010 e COLE et al., 2011). No presente estudo, o pequeno número de indivíduos do grupo amostral impactou negativamente o cálculo do EBV dos animais, impossibilitando que estimativas de herdabilidade fossem calculadas.

Para suplantar essa dificuldade, foi utilizada a estimativa obtida por CHANDLER et al. (1985), para a motilidade pós descongelação (DMD). O mesmo procedimento não pôde ser utilizado para o cálculo de herdabilidade da motilidade pós TTR (DMT), já que essas estimativas não foram encontradas na literatura.

Sendo assim, a herdabilidade arbitrada para o cálculo da EBV de ambos os fenótipos foi baseada naquela previamente descrita para DMD ($h^2 = 0,20$), dada a alta correlação entre DMD e DMT. Apesar da herdabilidade utilizada não ter sido calculada da forma convencional, optamos por esse procedimento por entender que a mesma aproxima-se mais do real do que a herdabilidade estimada com base no banco de dados do estudo ($h^2 = 0,05$). Tal decisão baseou-se em trabalho de CHANDLER et al. (2000) que relataram como complicador para o cálculo da herdabilidade da motilidade o fato de que essa característica é mensurada em uma escala de 0 a 100%, em incrementos/decrêscimos de 5%, o que seria responsável pela variância heterogênea da característica, afetando os cálculos de herdabilidade.

Apesar do cálculo do EBV ter sido realizado com um modelo que considera o ambiente onde as medidas foram mensuradas permanentemente, foram utilizados critérios para exclusão das estimativas menos acuradas com o intuito de assegurar a confiabilidade das mesmas. Foram removidos antes do teste de associação, os animais que apresentaram acurácia $< 0,5$.

Baseados em RASMUSSEN-TORVIK et al. (2010) que relataram cuidados a serem tomados para a utilização da média dos valores mensurados como característica fenotípica, quando os valores apresentam relativa estabilidade entre as repetições, utilizou-se a média como alternativa para a detecção de sinais de associação que não seriam passíveis de detecção em estudos com tamanho amostral reduzido. Sendo assim, a utilização da média dos valores mensurados por touro, após a remoção das avaliações marginais com valores com ± 4 desvios padrão da média, foi utilizada.

a) Experimento 1

A análise uni-característica foi conduzida com o intuito de estimar os valores genéticos para DMD e DMT, levando em consideração o desempenho do próprio indivíduo além dos possíveis efeitos de ambiente que poderiam influenciar as características em questão.

Foi utilizado o procedimento GLM (*Generalized Linear Model*) do pacote estatístico SAS para verificar os efeitos ambientais que afetavam significativamente a característica avaliada, considerando todo o banco de dados fenotípico ($n=148$) touros. O resultado detalhado das análises está apresentado no Anexo 2.

A equação que melhor ajustou os fenótipos ($r^2 = 0,18$ e $0,23$ para DMD e DMT, respectivamente) considerou o ano e estação em que foi realizada a coleta como efeitos fixo, a idade do touro na data da coleta (linear e quadrático) como covariável, e o indivíduo avaliado como efeito aleatório. Apesar do r^2 ter sido baixo, não haviam mais informações disponíveis para a realização de melhorias no ajuste do modelo.

A estimativa dos parâmetros genéticos foi realizada segundo a equação proposta por HENDERSON (1953):

(6)

$$Y = Xf + Za + Wc + e$$

No qual, Y é vetor de observações da variável estudada, f é vetor de efeitos fixos, a é vetor dos valores genéticos aditivos dos animais avaliados, c é vetor de efeito permanente de ambiente, e é o vetor de efeitos residuais, e X , Z e W são as matrizes de incidência dos efeitos fixos, aleatórios e efeito permanente, respectivamente. Os padrões genéticos do modelo foram estimados utilizando-se o software MTDFREML (BOLDMAN et al., 1995).

Os EBVs dos touros foram avaliados quanto a sua normalidade (teste Shapiro-Wilk) e a presença de marginais com a utilização da biblioteca GenABEL (AULCHENKO et al., 2007) do pacote estatístico R.

b) Experimento 2

As médias das medidas fenotípicas como característica fenotípica para cada touro foram calculadas e avaliadas quanto a sua normalidade (teste Shapiro-Wilk), heterogeneidade de variância (teste de Levene) e presença de marginais com a utilização da biblioteca GenABEL (AULCHENKO et al., 2007) do pacote estatístico R.

4.5. Controle de qualidade dos genótipos

Os procedimentos de controle de qualidade objetivaram a manutenção da integridade e a utilidade dos SNP que foram testados. Para tanto, os SNP foram avaliados e excluídos seguindo diversos critérios em diferentes etapas.

O procedimento de controle de qualidade dos genótipos foi realizado através do uso de rotinas personalizadas nas linguagens Perl e da biblioteca GenABEL (AULCHENKO et al., 2007) do pacote estatístico R, sendo dividido em duas categorias: controle de qualidade dos genótipos de cada SNP e controle de qualidade do conjunto de genótipos de SNP de uma dada amostra.

4.5.1. Controle de qualidade dos genótipos de cada SNP

1) SNP mapeados no cromossomo 0 – Devido a falhas na montagem do genoma bovino (versão UMD 3.1) alguns SNP contidos no ensaio BovineHD Illumina® não foram posicionados adequadamente no genoma, tendo sua localização remetida a um cromossomo fictício denominado de 0. Devido a ausência de coordenadas genômicas precisas para esses SNP, os mesmos foram excluídos da amostra.

2) SNP localizados no DNA mitocondrial – Esse critério se justificou devido a baixa heterozigosidade e consequente baixa ou nula informatividade do DNA mitocondrial para o estudo proposto.

3) SNP localizados nos cromossomos sexuais X e Y – Esse critério se justificou devido as limitações que os algoritmos atualmente disponíveis apresentam para determinação desses genótipos (*Genotype Calling*). Uma das razões para esse fato é que a versão atual da montagem do genoma apresenta falhas nesses cromossomos.

4) SNP com a mesma coordenada genômica – Durante o desenvolvimento do ensaio BovineHD Illumina® observou-se que alguns SNP com designações diferentes possuíam a mesma coordenada genômica. Uma das razões para esse fato é que houve redundância de alguns SNP durante o processo de fabricação do SNP chip BovineHD Illumina®. Por essa razão todos os SNP que se encontravam nessa situação foram retirados da análise.

5) SNP com baixo GCScore – SNP com GCScore inferior a 0,7 foram retirados da análise. Esse critério foi utilizado para assegurar a qualidade dos genótipos, já que SNP que apresentam GCScores < 0,7 representam níveis de confiabilidade decrescentes (de intermediário a baixa).

6) SNP altamente correlacionados – Determinados ensaios de determinação do perfil de SNP podem apresentar grupos de SNP altamente correlacionados que podem ter uma grande influência sobre alguns componentes principais durante a avaliação da estratificação pelo PCA. LAURIE et al. (2010), recomendam que esses sejam retirados do banco de dados para permitir que a estratificação da população seja avaliada sem viés, portanto foram removidos SNP que apresentaram $r^2 \geq 0,995$ entre si.

7) SNP com baixa MAF – Foram removidos SNP com $MAF < 0,02$. Esse limite é preconizado por diversos autores e objetiva retirar SNP cujo genótipo foi designado erroneamente durante sua determinação.

8) SNP com baixo *call rate* – Foram removidos SNP com *call rate* $< 0,98$ objetivando eliminar SNP com baixa qualidade de genotipagem.

9) SNP que desviam do EHW - Foram removidos SNP com EHW $p < 10^{-5}$ objetivando eliminar genótipos designados erroneamente durante sua determinação.

4.5.2. Controle de qualidade do conjunto de genótipos de SNP das amostras

1) Identificação de possíveis erros na determinação do perfil de SNP - Foi realizada a verificação do sexo das amostras utilizando a proporção de heteozigosidade do cromossomo X. Além disso, foi analisada a existência de amostras duplicadas no banco de dados, utilizando o coeficiente de parentesco genômico IBS ($IBS > 0.95$ = amostras duplicadas) entre os pares de amostras.

2) Remoção de indivíduos com baixo *call rate* - Foram removidos indivíduos com *call rate* $< 0,90$ objetivando remover aqueles em que o procedimento de determinação do perfil de SNP apresentou qualidade duvidosa.

4.6. Verificação da presença de estratificação na população

A aplicação de PCA utilizando os SNP e os indivíduos que atenderam aos critérios estabelecidos para controle de qualidade com base no parentesco genômico foi conduzida. A utilização do PCA objetivou avaliar a relação entre os indivíduos, com o intuito de evidenciar uma possível subestrutura na população.

4.7. Teste de Associação Genômica

O teste de associação foi escolhido de acordo com o comportamento de estratificação e parentesco das amostras com base no diagrama proposto por AULCHENKO et al. (2007).

A confiabilidade dos resultados foi verificada através do cálculo do fator de inflação λ , no qual valores de λ superiores a 1 são indicativos de inflação generalizada do teste. A distribuição dos valores estatísticos foi ajustada para a inflação do teste quando o valor de λ foi maior ou igual a um. Como controle complementar, os valores de p não ajustados foram inspecionados por meio de um gráfico de quantil-quantil (Q-Q).

Os testes de associação foram realizados utilizando-se os métodos GRAMMAR e EIGENSTRAT. No Experimento 1, o método GRAMMAR foi equivalente ao método *Genomic Control*, já que os ajustes foram realizados para a estimativa do EBV, sem adição de covariáveis extras no modelo poligênico do teste. No Experimento 2 foram incluídos no modelo poligênico: estação (primavera, verão, outono e inverno), ano da coleta do sêmen e o número de avaliações por touro como efeitos fixos, além da idade do animal no momento da coleta como covariável. Para corrigir o valor de p para múltiplos testes, o ajuste de Bonferroni foi aplicado.

Finalmente, os resultados foram apresentados na forma de *Manhattan Plots* com o co-logaritmo dos valores de p corrigidos para λ . Foram representados nesse gráfico os limiares de significância ajustados para múltiplos testes (Bonferroni) e as classificações propostas pelo sistema BADGE para facilitar a interpretação dos resultados. Estas análises foram realizadas utilizando a biblioteca GenABEL do pacote estatístico R.

4.8. Análise de Enriquecimento Genético

Após a realização do teste de associação foram considerados significativos os SNP com valores de significância maiores do que 0,01%. Os genes que participaram da AEF foram selecionados levando em consideração sua associação direta e indireta com o SNP.

Para a identificação dos genes associados diretamente com os SNP (SNP intra-gênicos), foram utilizadas as coordenadas genômicas dos mesmos. Rotinas do software Plink foram utilizadas para identificar os genes associados indiretamente com os SNP,

isto é, ligados por LD ao SNP. Foi estimado o LD entre todos os SNP posicionados a até 1 Mb dos SNP com valores de significância maiores do que 0,01%, sendo selecionados para a AEF apenas os genes com r^2 superior a 0,8.

A AEF foi realizada pelo software DAVID (versão 6.7) aplicando-se a análise de agrupamento funcional com rigor médio. Os IE e os valores de p nominal e ajustado foram calculados para cada termo do grupo. Apenas os grupos com IE > 2 foram considerados de importância funcional.

Para o cálculo do valor de significância do agrupamento funcional foi utilizado o teste exato de Fisher modificado (*EASE Score*). Esse teste determina se existe diferença estatística na proporção de termos em cada grupo. O software utiliza estatística *Kappa* para verificar se os termos pertencem a mesma categoria funcional.

5. RESULTADOS E DISCUSSÃO

A análise descritiva das características avaliadas nos dois experimentos (Histograma, *Box-plot* e *Q-Q plot* - Figuras 10 a13) demonstrou que os fenótipos não apresentaram distribuição normal (teste de Shapiro-Wilk, $p < 0,05$).

Tentativas de normalização da distribuição dos dados fenotípicos foram realizadas através de transformações logarítmicas e *Box-cox* (informações complementares no anexo 3). Apesar de transformadas, as medidas fenotípicas não se apresentaram normalizadas (teste de Shapiro-Wilk, $p < 0,05$).

Uma das hipóteses para explicar esse fato, refere-se ao número limitado de animais avaliados, o que pode ter interferido na distribuição dos fenótipos. O melhor valor estimado do fator de correção λ na transformação *Box-cox* (DMD= 0,86 DMT = 1) indicou que apesar de não ter atingido a normalidade segundo o teste aplicado, a distribuição dos dados ficou próxima da normal ($\lambda = 1$).

De acordo com HANDELSMAN (2002), as características qualitativas do sêmen apresentam particularidades que dificultam a ocorrência desse tipo de distribuição. Segundo esse autor, tais características podem não atingir a normalidade mesmo após transformação de seus valores por diferentes métodos. Dada a proximidade da

normalização, verificada pela forma de sino dos histogramas apresentados nas Figuras 10 e 11, justificou-se a realização dos testes de associação genômica com fenótipos não transformados.

Como a distribuição encontrava-se próxima a normal, aliado ao fato do tamanho amostral limitado a 125 animais, optou-se por mantê-los na análise com intuito de preservar o maior número de amostras possível. Além disso, consideramos que a permanência de indivíduos com fenótipos poderia ser interessante nas etapas de associação uma vez que poderia explicar as causas das variações fenotípicas.

Após a definição dos ajustes dos fenótipos e animais foram realizados os procedimentos de análise descritiva dos genótipos. Primeiramente foram removidos do banco de dados de SNP, todos aqueles relacionados ao DNA mitocondrial, aos cromossomos sexuais e ao cromossomo 0 contidos no painel inicial de 786.799 SNP. Esse procedimento resultou em 735.238 SNP úteis para a submissão aos controles de qualidade subsequente. As análises descritivas dos parâmetros de qualidade dos SNP selecionados compreenderam a MAF (Figura 14), heterozigosidade (Figura 15), EHW (Tabela 2) e *call rate* (Tabela 3). Além da heterozigosidade média e o *call rate* dos indivíduos.

A análise descritiva dos genótipos demonstrou que a distribuição das MAF foi semelhante ao discutido anteriormente pelo THE BOVINE HAPMAP CONSORTIUM (2009) nas raças zebuínas. Segundo esses autores, 19% dos SNP apresentaram MAF superiores a 0,3. No presente estudo, apenas 10% dos SNP apresentaram MAF superiores a esse limite. Essa diferença pode ser justificada pelo fato de no estudo realizado pelo THE BOVINE HAPMAP CONSORTIUM (2009) a MAF média foi contabilizada considerando MAF de três raças zebuínas distintas (Brahman, Nelore e Gir) e não apenas da raça Nelore como no presente estudo.

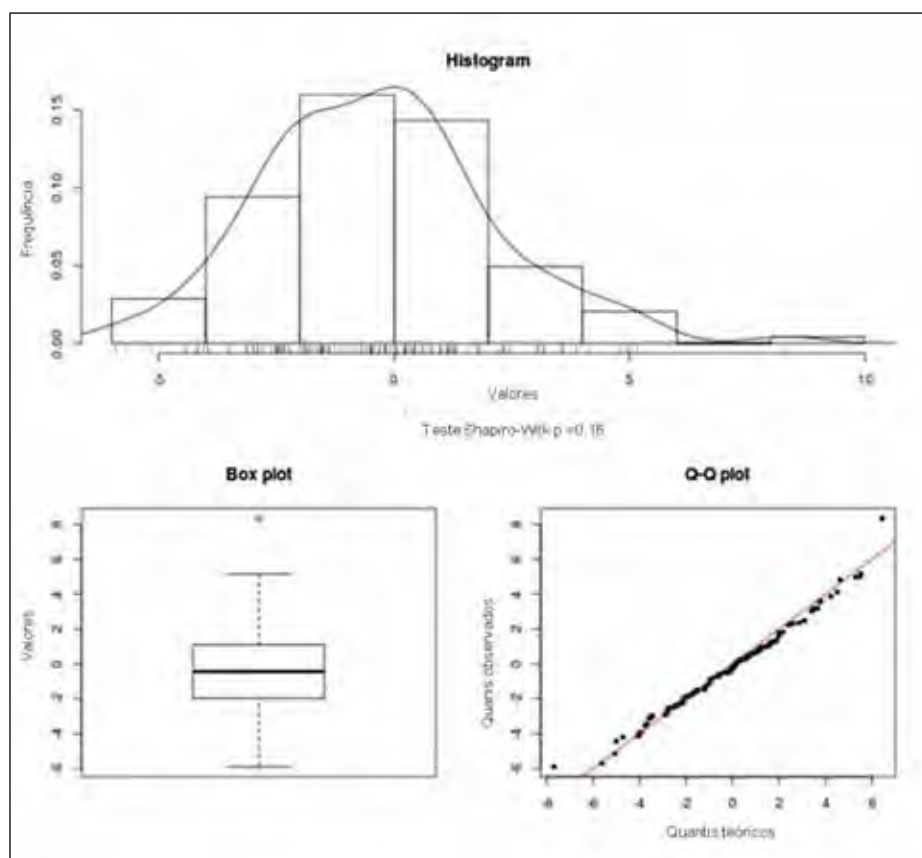


Figura 10. Análise descritiva do fenótipo EBV (Experimento 1) da característica diferença entre a motilidade do sêmen fresco e descongelado (DMD) pelo Histograma, *Box plot* e *Q-Q plot*.

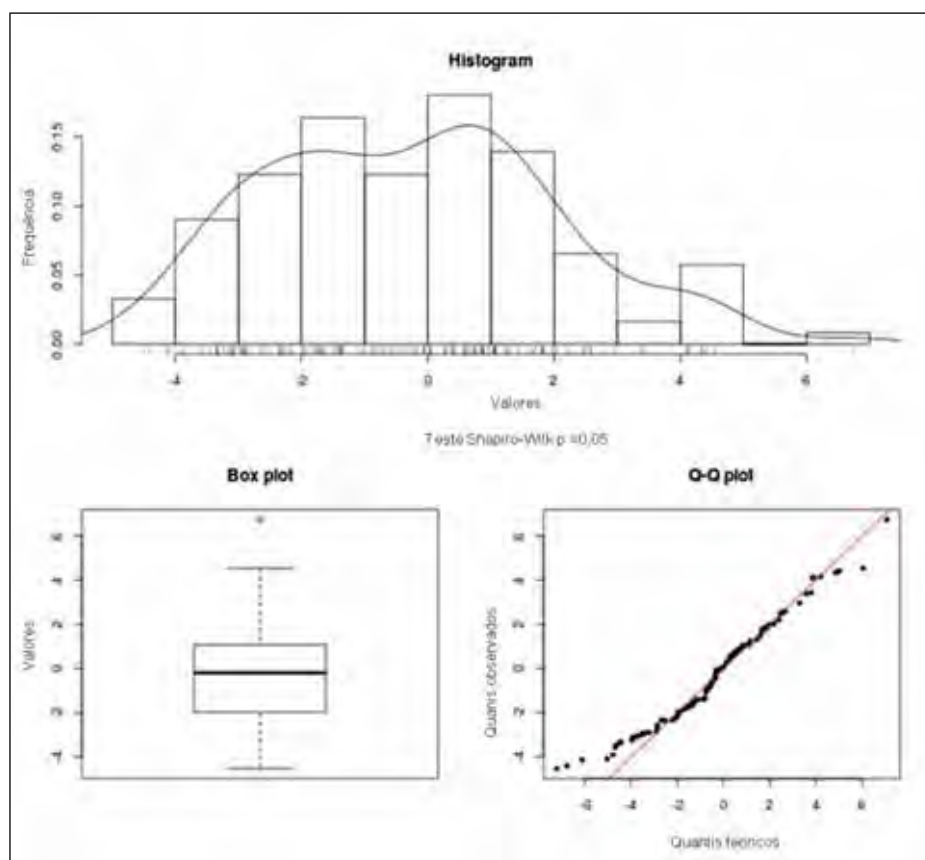


Figura 11. Análise descritiva do fenótipo EBV (Experimento 1) da característica diferença entre a motilidade do sêmen fresco e pós teste de termo resistência (DMT) pelo Histograma, *Box plot* e *Q-Q plot*.

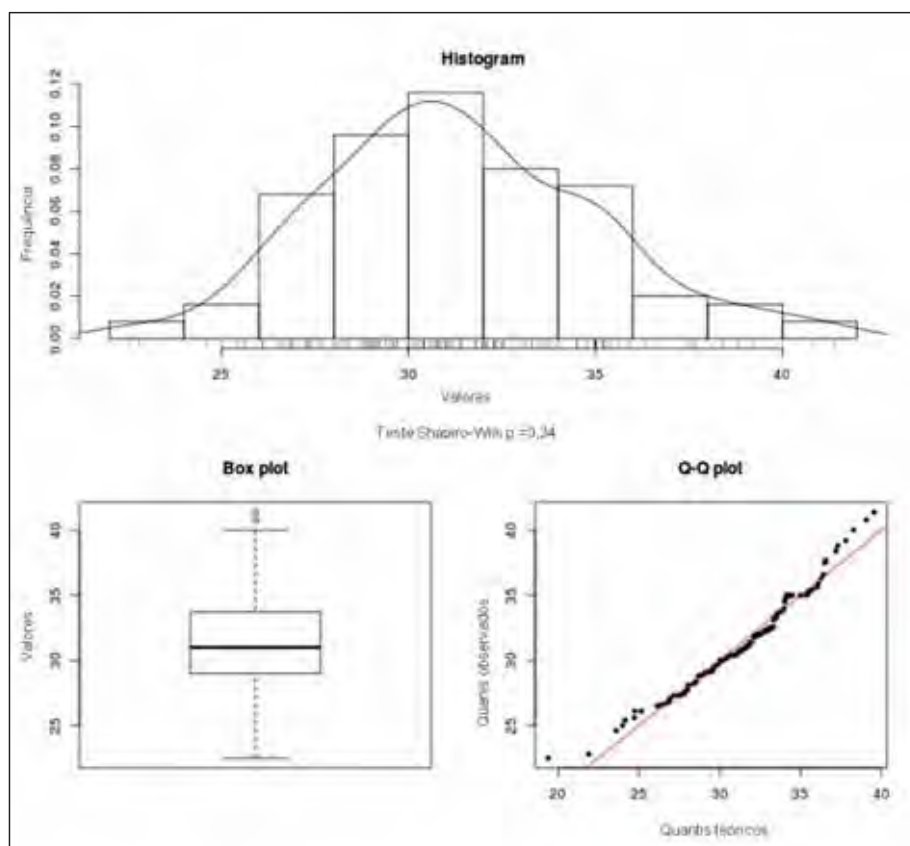


Figura 12. Análise descritiva das médias das medidas fenotípicas (Experimento 2) da característica diferença entre a motilidade do sêmen fresco e descongelado (DMD) pelo Histograma, *Box plot* e *Q-Q plot*.

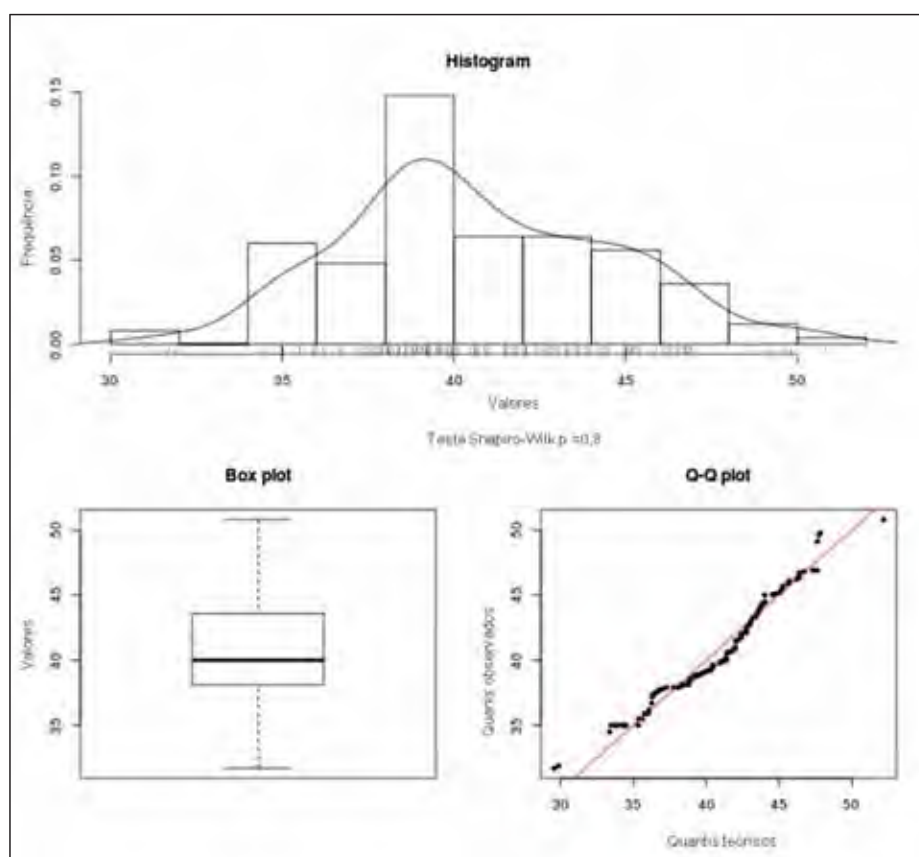


Figura 13. Análise descritiva das médias das medidas fenotípicas (Experimento 2) da característica diferença entre a motilidade do sêmen fresco e pós teste de termo resistência (DMT) pelo Histograma, *Box plot* e *Q-Q plot*.

Tabela 2. Distribuição cumulativa do número de SNP que não atingiram os critérios de EHW ($p > 10^{-5}$) para diferentes valores de p.

	$p \leq 10^{-5}$	$p \leq 10^{-4}$	$p \leq 10^{-3}$	$p \leq 10^{-2}$	$p \leq 5 \times 10^{-2}$	Todos p
Nº SNP	3.370	4.522	6.374	12.433	29.928	735.238
% SNP	0,05	0,06	0,09	1,7	4,1	100

Tabela 3. Distribuição da proporção do *call rate* (CR) dos SNP.

	$CR \leq 0,9$	$0,9 < CR \leq 0,95$	$0,95 < CR \leq 0,99$	$CR > 0,99$
Nº SNP	51.960	16.978	33.888	632.412
%SNP	7,1	2,3	4,6	0,86

A análise de heterozigosidade observada demonstrou que existe número de SNP informativos em quantidade suficiente para serem utilizadas em estudos com animais zebuínos. Esses resultados reforçam os estudos de validação realizados pela empresa ILLUMINA (2010) que demonstraram ser possível o uso do micro-arranjo BovineHD Illumina® com sucesso em estudos dessa subespécie. Quanto a avaliação do comportamento do conjunto de SNP com relação a cada amostra individual, a heterozigosidade média no grupo do estudo foi de $0,21 \pm 0,01$. Da totalidade de amostras avaliadas, 63,1% apresentaram *call rate* entre 90-95% e 36,9% entre 95-99%.

Por não atenderem aos critérios adotados no CQ conforme apresentado na seção Material e métodos (seção 4.5), 42,34% dos SNP foram excluídos das análises, restando 311.346 SNP do painel inicial de 735.238 (Figura 16). A Tabela 4 apresenta o

número de SNP removidos em cada uma das etapas do CQ. Todos os 125 amostras atenderam os critérios de qualidade pré-estabelecidos.

Segundo LAURIE et al. (2010), o limite de exclusão de SNP pela MAF, quando a plataforma Infinium é utilizada para determinação do perfil de SNP das amostras, deve ser de 0,01 devido a alta desempenho do ensaio. Para reduzir possíveis impactos negativos de eventuais erros, critério de exclusão de SNP pela MAF, adotado Np presente estudo ($MAF < 0,02$), foi utilizado com sucesso por diversos autores em trabalhos anteriores (WANG et al., 2005; ANDERSON et al., 2010). Apesar de grande quantidade de SNP ter sido excluída (208.352) devido a esse critério, cabe ressaltar que cerca de 25.000 desses SNP excluídos apresentaram MAF igual a zero, sendo portanto não informativos para o estudo.

Tabela 4. Número de SNP removidos nas diferentes etapas do controle de qualidade.

Critério	Nº de SNP removidos
SNP com mesma coordenada genômica	54
SNP correlacionados ($r^2 \geq 0,995$)	142.891
$MAF < 0,02$	208.352
<i>Call rate</i> $< 0,98$	73.315
$EHW > 10^{-5}$	3.553

Não foram utilizados critérios para exclusão de SNP baseados na heterozigosidade observada no presente estudo. A distribuição de heterozigosidade dos SNP em geral, assim como a média da heterozigosidade das amostras ($0,21 \pm 0,01$)

demonstraram a qualidade das análises para determinação dos perfis de SNP, descartando possíveis erros ou contaminação. Segundo LAURIE et al. (2001) para a plataforma BovineHD Illumina® devem ser removidos SNP dos cromossomos autossômicos e pseudo autossômicos com heterozigosidade superior a 0.3, portanto acima dos valores encontrados no presente estudo ($0,21 \pm 0,01$).

Critérios para a exclusão de SNP que desviam do EHW vem sendo discutidos por diversos autores. Os valores de significância aceitos para desvios de EHW controversamente variam entre 10^{-3} a 10^{-7} (SAMANI et al., 2007; ZIEGLER, 2008; LAURIE et al, 2010). De maneira geral a recomendação é que sejam excluídos SNP com desvios de pelo menos 10^{-4} . Segundo ZIEGLER (2008), essa recomendação deve-se ao fato de que outros processos, além de erros na análise do perfil de SNP tais como a seleção, podem causar desvio de EHW. A distribuição de EHW apresentada na Tabela 2 demonstra que menos de 1% dos SNP foram eliminados com o critério de $EHW < 10^{-5}$ utilizado no presente estudo.

Limites de *call rate* acima de 97% são normalmente relatados para exclusão de amostras (ANDERSON et al 2010).Entretanto, ZIEGLER (2008) afirma que valores acima de 90% já seriam suficientes para garantir a qualidade do estudo. Como o tamanho amostral do presente estudo é limitado, adotamos o critério menos rigoroso para a remoção das mesmas, sendo excluídas amostras que apresentaram *call rate* abaixo de 90%.

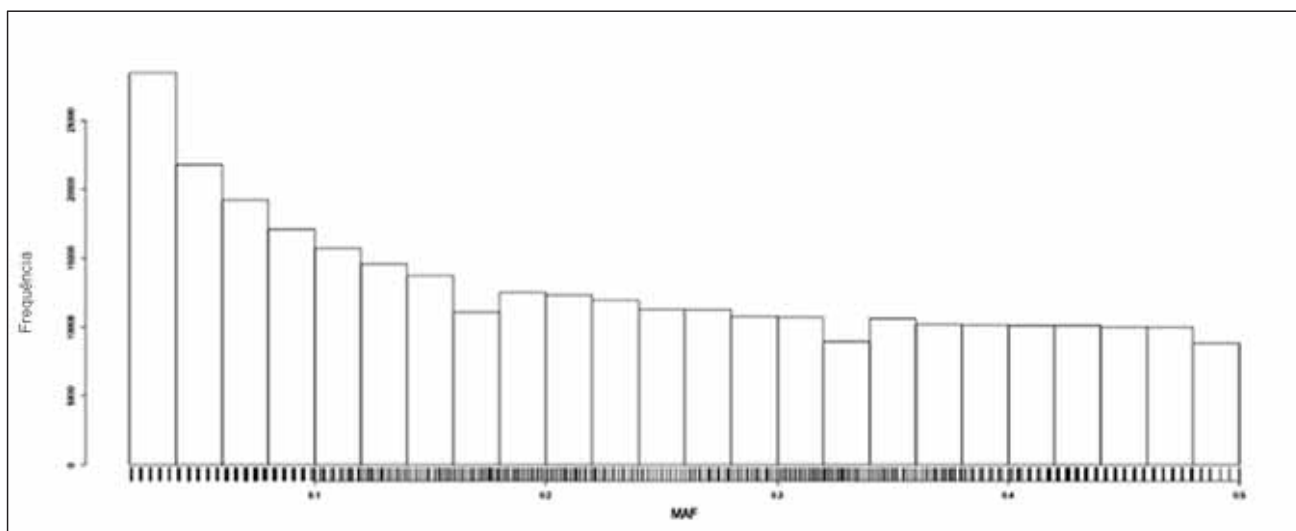


Figura 14. Distribuição da frequência alélica menor (MAF) dos SNP antes do controle de qualidade. Linhas verticais abaixo do eixo X representam os SNP individualmente.

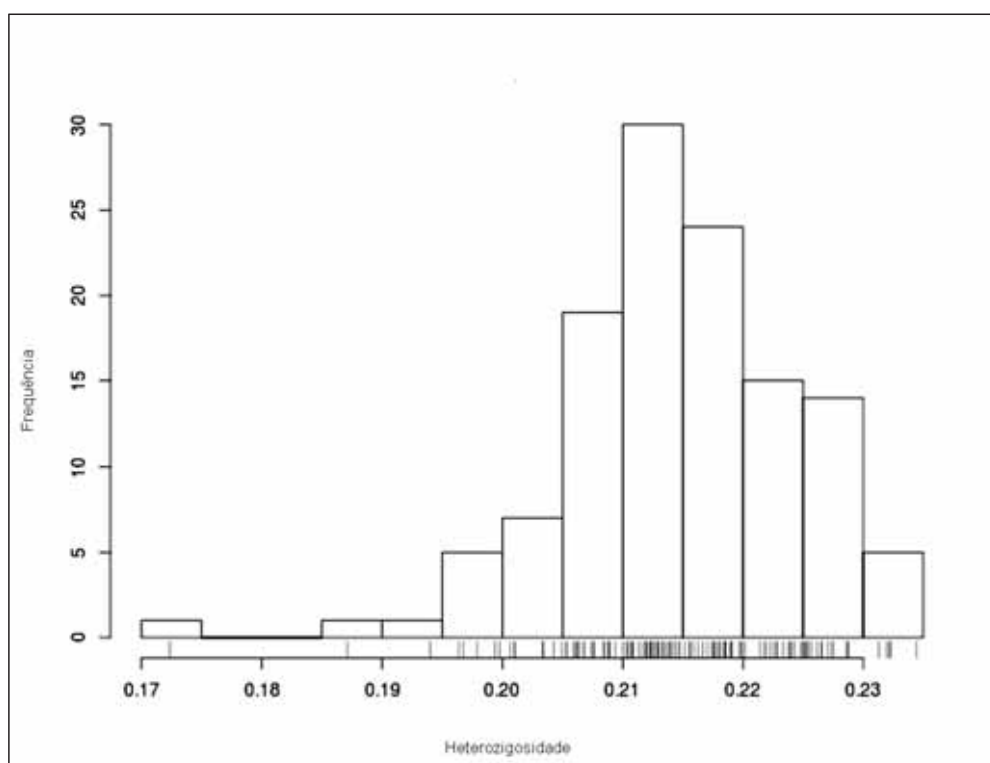


Figura 15. Distribuição da heterozigosidade dos SNP antes do controle de qualidade. Linhas verticais abaixo do eixo X representam os SNP individualmente.

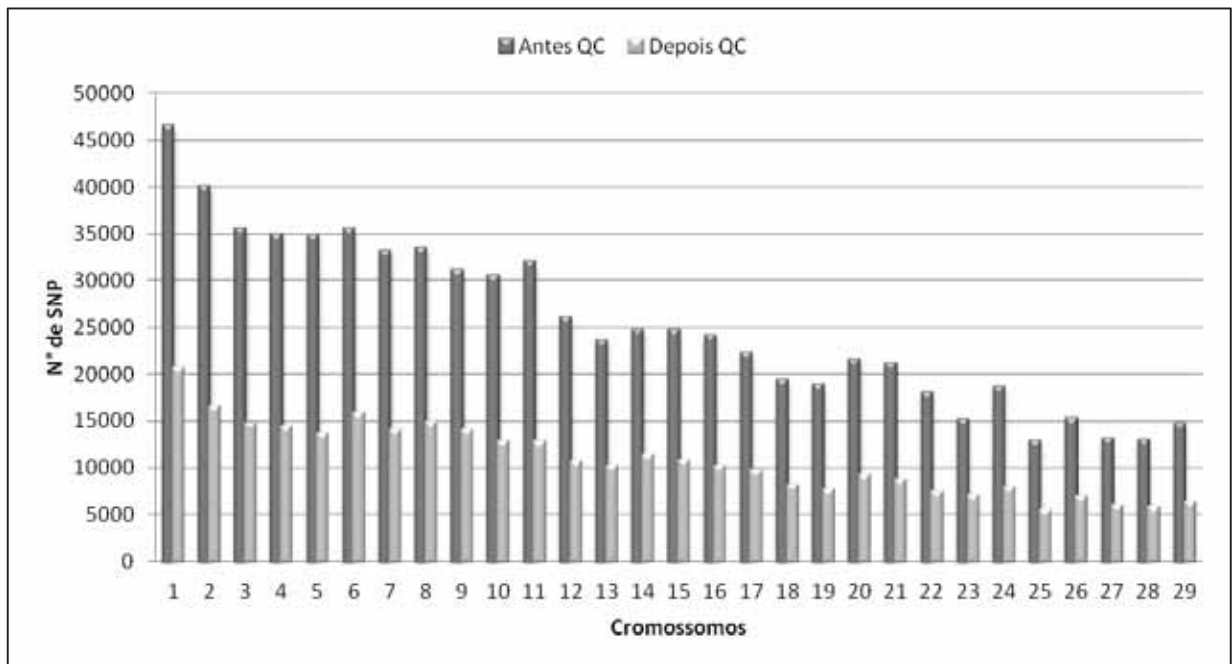


Figura 16. Distribuição dos SNP por cromossomos autossômicos antes e depois do controle de qualidade.

O número de SNP removidos após o CQ depende do rigor dos filtros aplicados. Segundo ANDERSON et al. (2010), estudos realizados com pequenos grupos amostrais devem utilizar critérios mais conservadores de exclusão para eliminar os possíveis vieses decorrentes do pequeno número de indivíduos na análise.

No Experimento 1, três animais foram excluídos antes do teste de associação devido a baixa acurácia do EBV estimado (acurácia < 0,50). Esses animais também foram retirados do Experimento 2 para que o banco de dados analisado fosse o mesmo em ambos os experimentos, sendo assim, o grupo amostral finalmente foi constituído por 122 animais.

A análise de componentes principais (PCA) baseada no parentesco genômico (Figuras 17 e 18) demonstrou que as amostras em estudo pertencem a apenas um grupo principal, não apresentando subestrutura evidente.

Já a análise de estratificação da população pelo cálculo do fator de inflação λ (DEVLIN & ROEDER, 2008) demonstrou em ambos os experimentos presença de inflação após o QC ($\lambda > 1$) nos fenótipos avaliados (Figura 19 a 22).

O PCA é um método amplamente utilizado com o objetivo de identificar a presença de sub-populações distintas em grupo de amostras (THE BOVINE HAPMAP CONSORTIUM, 2009; PRICE et al., 2010; HAO et al., 2010).

No presente estudo, a maioria dos touros avaliados é proveniente do mesmo programa de melhoramento, o que sugere que os mesmos possuem a mesma ancestralidade, o que ajuda a justificar a formação de apenas um grupo (dados não apresentados).

Ainda nessa análise, os indivíduos foram classificados de acordo com a magnitude de seu fenótipo (*Heatmap*), permitindo a identificação de classes fenotípicas distintas. No Experimento 1, de forma geral, os animais mais criotolerantes apresentaram-se agrupados à direita do bloco principal (cor vermelha) enquanto os menos criotolerantes (cor amarela) encontraram-se à esquerda do mesmo. O mesmo comportamento se repetiu no experimento dois, mas de forma menos evidente, estando os animais criotolerantes na parte superior do bloco.

A subdivisão entre classes fenotípicas evidenciadas pelo *Heatmap* reforçou a hipótese de que componentes genéticos são determinantes para a expressão de tolerância a criopreservação. A análise do *Heatmap* do Experimento 2 demonstrou menor variação fenotípica do que o do Experimento 1 (Figuras 17 e 18). Como touros em regime de coleta nas centrais de inseminação teoricamente apresentam pouca variação nos parâmetros qualitativos do sêmen, devido ao efeito de seleção da fase preparatória para coletas, provavelmente a utilização das médias das características por animal (Experimento 2) representem melhor o fenótipo avaliado do que os EBVs (Experimento 1).

Os diferentes grupos fenotípicos evidenciados nessa análise mostraram que o modo de ajuste dos fenótipos é importante para a interpretação dos resultados. Mesmo tendo como base para a análise de componentes principais os mesmos genótipos, o EBV e as médias das medidas fenotípicas das mesmas características se comportaram como fenótipos diferentes. Isso pode ter ocorrido devido a baixa capacidade de ajuste do modelo utilizado para estimar o EBV e o pequeno tamanho amostral utilizado para calcular essa estimativa, acarretando possível viés na análise. Apesar de o grupo

amostral ser pequeno, estudos de associação genômica realizados para características de difícil mensuração, como as do presente estudo, foram realizados com sucesso por FEUGAND et al. (2009) com tamanhos amostrais similares.

Apesar da estratificação da população não ter sido evidenciada (Figuras 17 e 18), os *Q-Q plots* demonstraram a presença de algum confundidor. Esse fato foi confirmado pelo valor do fator de inflação λ , superior a 1 em todas as características e testes avaliados. Estudos prévios demonstram que valores de λ superiores a esse limite são indicativos de inflação generalizada do teste, demandando a realização da correção dos valores de significância para a redução dos erros do tipo I (PRICE et al., 2006).

Os *Q-Q plots* demonstraram que o método EIGENSTRAT foi mais eficiente em corrigir a inflação presente no grupo de amostras do Experimento 1 do que o método GRAMMAR. No Experimento 2 ocorreu o contrário (GRAMMAR mais eficiente que EIGENSTRAT). Tal fato poder ser evidenciado pela redução dos valores de λ nos testes realizados, o que indica a menor ocorrência de associações espúrias (Figuras 19 a 22).

No Experimento 2, o método GRAMMAR permitiu que covariáveis adicionais fossem utilizadas para ajustar as médias das medidas fenotípicas, o que não ocorre no método EIGENSTRAT que utiliza apenas os escores dos PCs para tal ajuste. A maior capacidade de ajuste do modelo poligênico utilizado no método GRAMMAR permitiu a maior redução dos valores do fator de inflação λ .

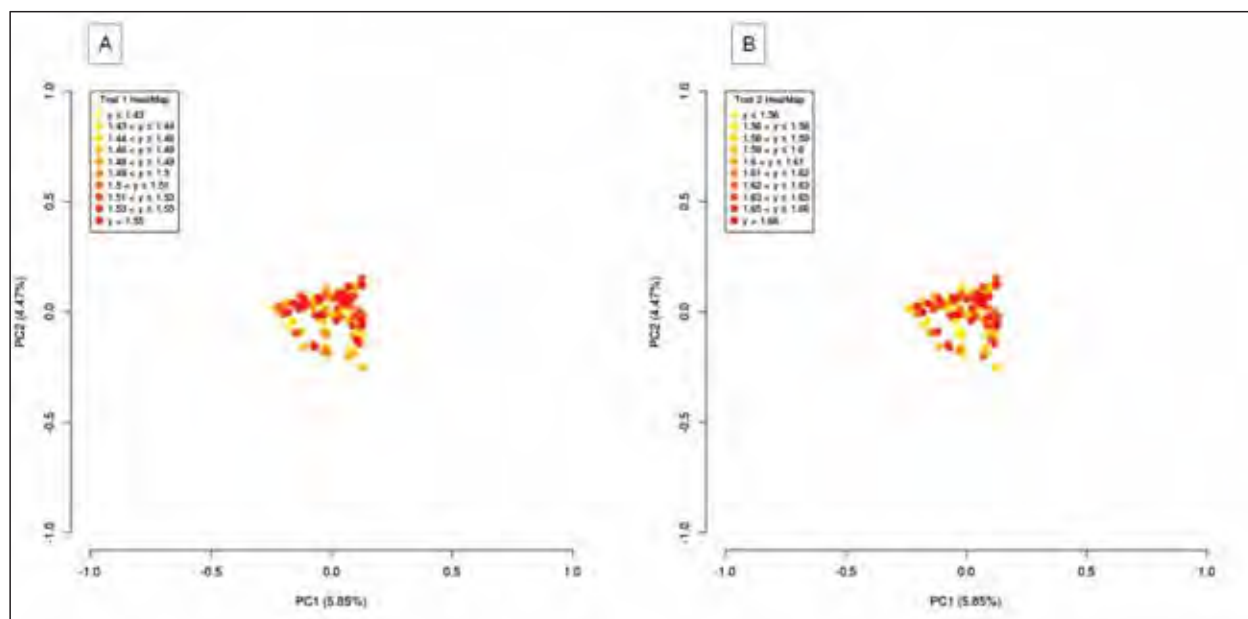


Figura 17. Análise de Componentes Principais (PCA) e *Heatmap* dos fenótipos do Experimento 1 (EBV). Animais representados por cores diferentes de acordo com a magnitude do fenótipo característica diferença entre a motilidade do sêmen fresco (DMD - A) e pós teste de termo resistência (DMT -B).

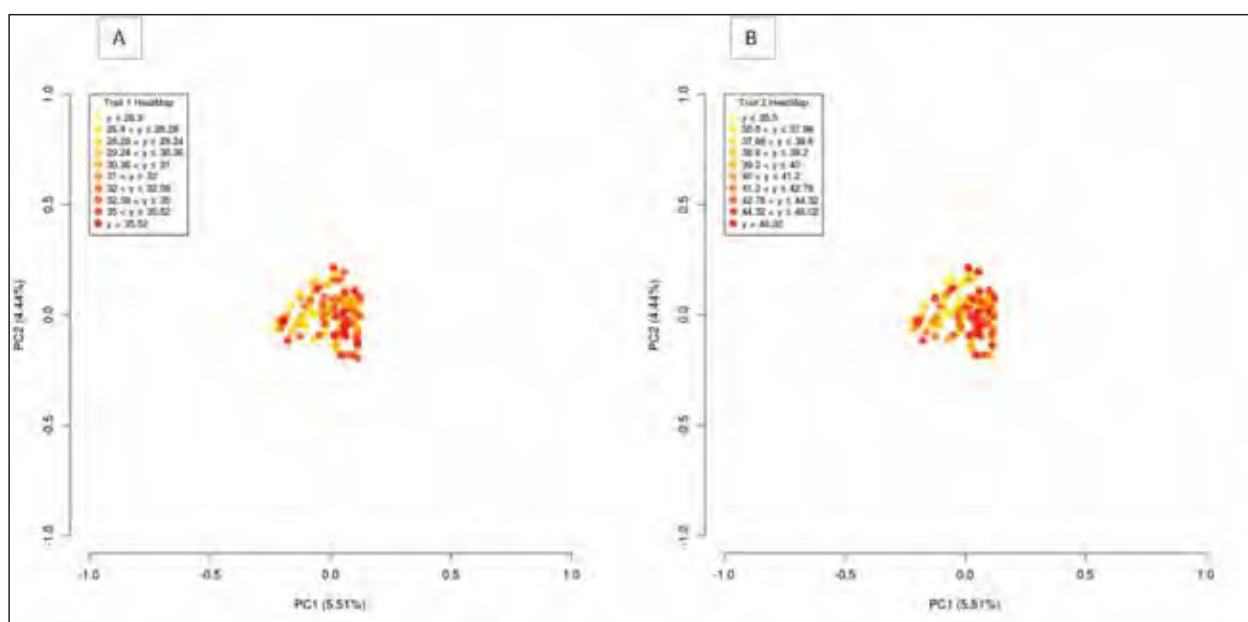


Figura 18. Análise de Componentes Principais (PCA) e *Heatmap* dos fenótipos do Experimento 2 (médias das medidas fenotípicas). Animais representados por cores diferentes de acordo com a magnitude do fenótipo. Média das medidas fenotípicas de DMD (A) e DMT (B)

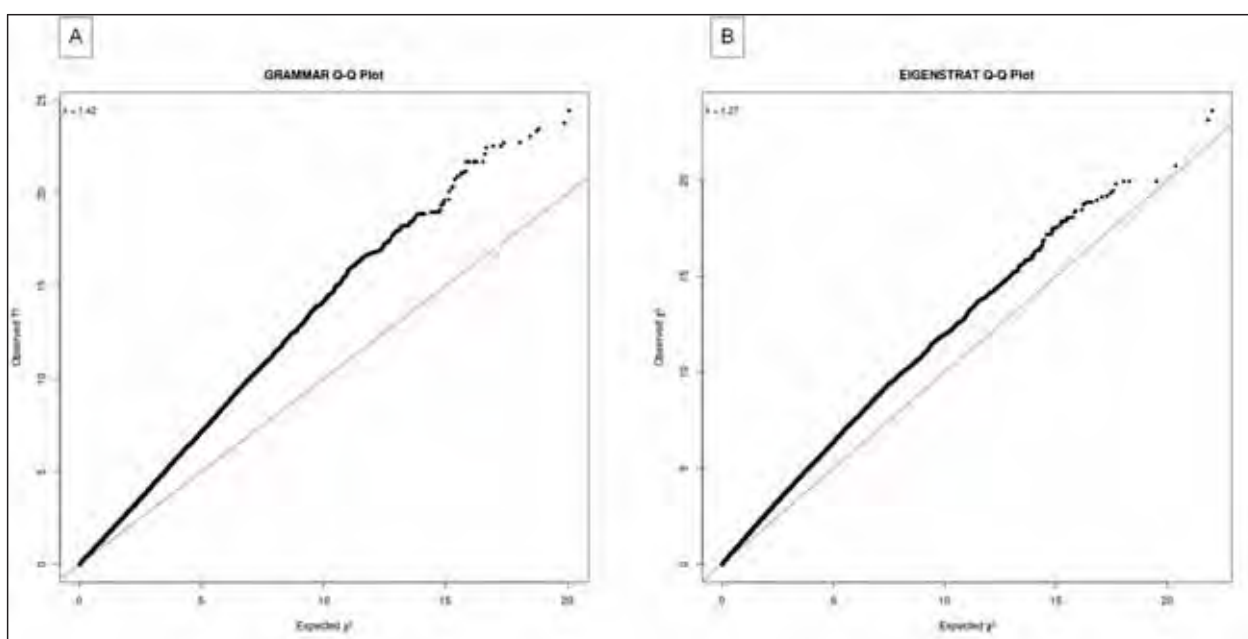


Figura 19. Gráficos Q-Q dos valores de significância (p) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD no Experimento 1 (EBV). Valores de $\lambda = 1,42$ e $1,27$ respectivamente.

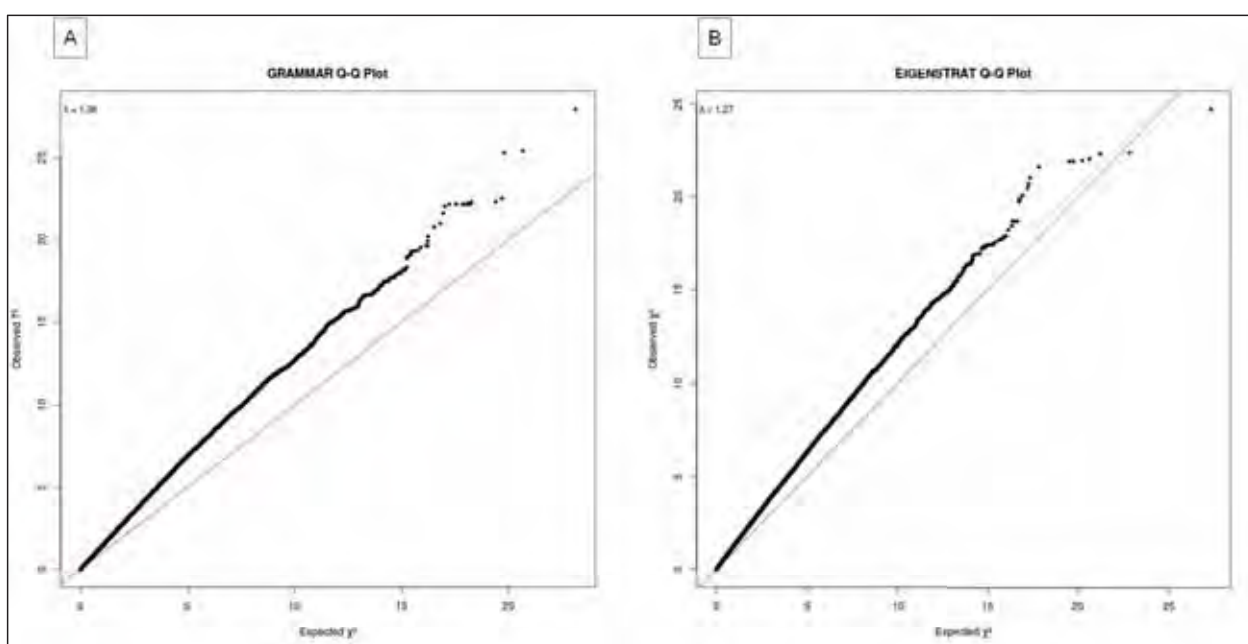


Figura 20. Gráficos *Q-Q plot* dos valores de significância (p) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMT no Experimento 1 (EBV). Valores de $\lambda = 1,36$ e $1,27$ respectivamente A e B.

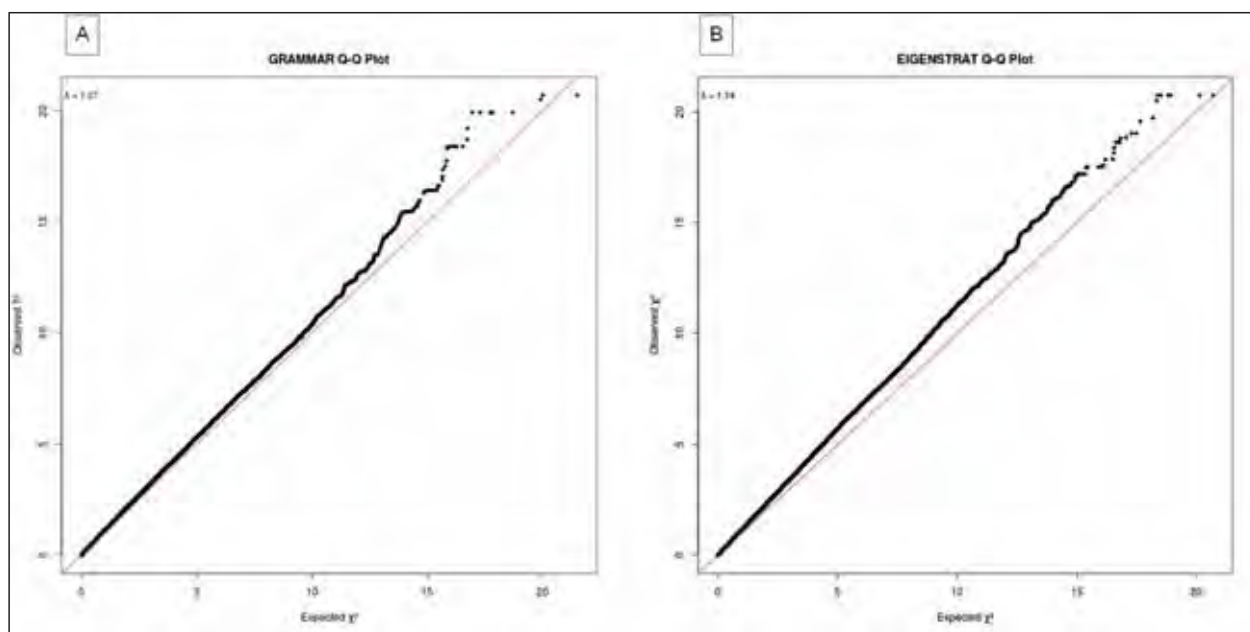


Figura 21. Gráficos *Q-Q plot* dos valores de significância (p) não ajustados dos testes de associação realizados com a Metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD no Experimento 2 (Média das medidas fenotípicas). Valores de $\lambda = 1,07$ e $1,14$ respectivamente A e B.

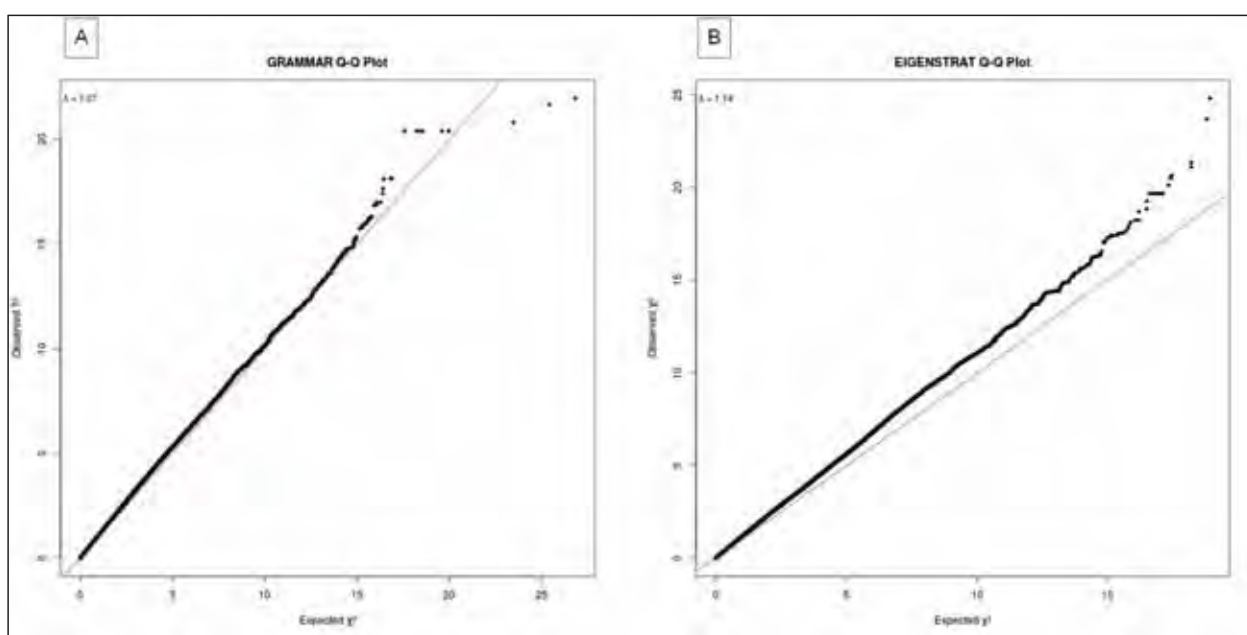


Figura 22. Gráficos *Q-Q plot* dos valores de significância (*p*) não ajustados dos testes de associação realizados com a metodologia GRAMMAR (A) e EIGENSTRAT (B) para a característica DMT no Experimento 2 (Média das medidas fenotípicas). Valores de $\lambda = 1,07$ e $1,14$ respectivamente A e B.

As Figuras 23 a 26 representam os gráficos do tipo *Manhattan Plot* referentes aos testes de associação em ambos os experimentos. Por sua análise pode ser notada grande semelhança entre dos diferentes resultados nos testes (GRAMMAR e EIGENSTRAT) realizados com os mesmos fenótipos (DMD e DMT).

De forma geral, os gráficos não apresentam picos altos evidentes, apontando para regiões genômicas significativamente associadas com o fenótipo, mas sim picos de magnitude média a baixa, além de difusos por diversos cromossomos indicando que muitos SNP tem efeito intermediário a baixo sobre a característica. De acordo com o sistema BADGE, os picos de maior significância se posicionaram na 3ª classe para a característica DMD e na 4ª classe para DMT.

A Figura 27 apresenta a análise comparativa dos cromossomos que apresentaram SNP mais significativos comum nas metodologias utilizadas para a realização do teste de associação indicando haver relação com o fenótipo entre as distintas combinações fenótipos (EBV) e métodos estatísticos (GRAMMAR e EIGENSTRAT) utilizados.

Os *Manhattan Plot* mostraram-se semelhantes quando o mesmo fenótipo foi testado pelos métodos GRAMMAR e EIGENSTRAT. Esse resultado era esperado já que se trata da mesma característica sob análise. Os fenótipos DMD e DMT também apresentaram padrão de picos semelhantes, principalmente no Experimento 2. Esse resultado também era esperado devido à alta correlação entre os dois fenótipos ($r^2=0,74$).

Para a identificação de genes indiretamente associados à característica de interesse (em LD com os SNP significativos), foi utilizada a estratégia de determinar o LD entre um dado SNP e os outros posicionados a uma determinada distância utilizando o software Plink. Esse procedimento foi utilizado por diversos autores para determinar o padrão de LD de regiões cromossômicas específicas (BUSH et al., 2009; BOHMANOVA et al., 2010). Entretanto, como ainda não existem padrões de LD relatados na raça Nelore, consideramos apenas os genes com $r^2 > 0,8$ do SNP significativo para a análise de enriquecimento funcional.

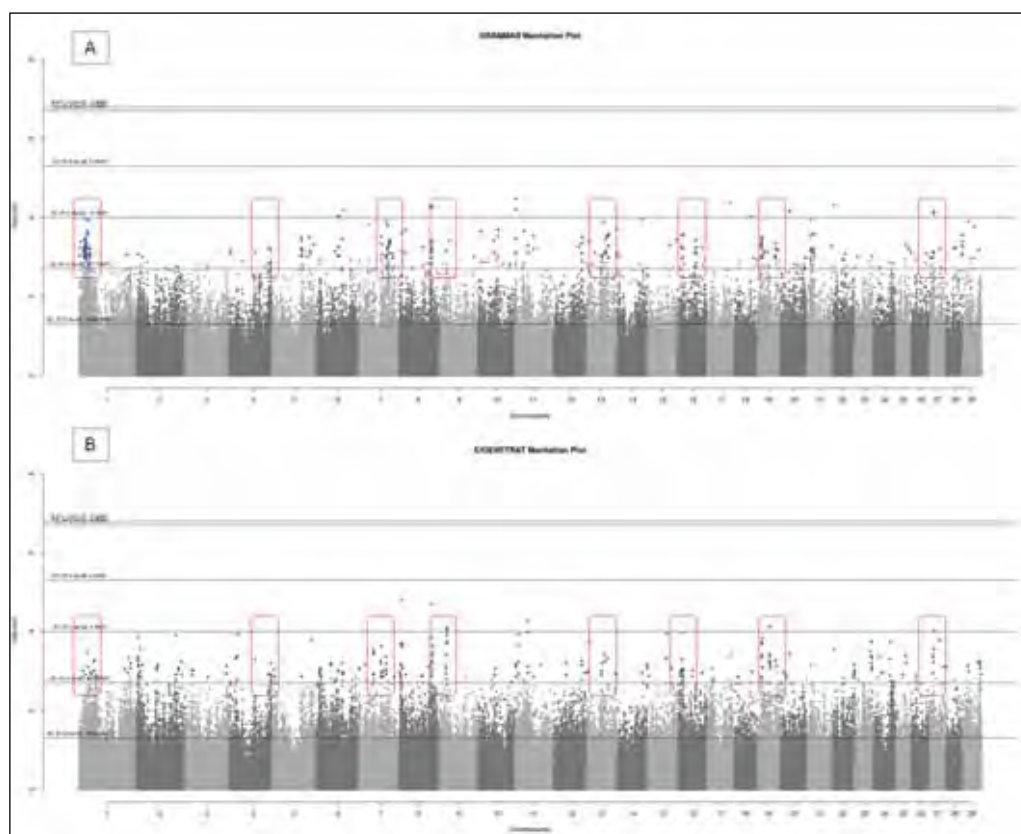


Figura 23. *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD para o Experimento 1 (EBV). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destaca SNP significativos em ambos os métodos.

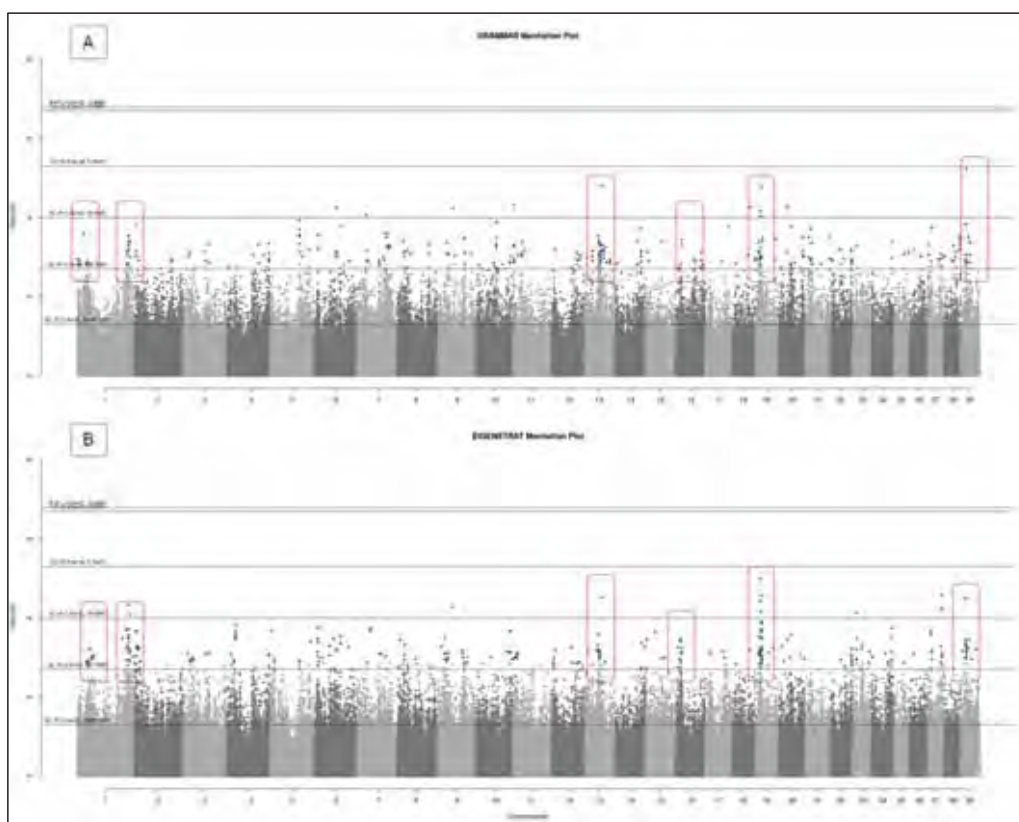


Figura 24. *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT(B) para a característica DMT para o Experimento 1 (EBV). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destacada SNP significativos em ambos os métodos.

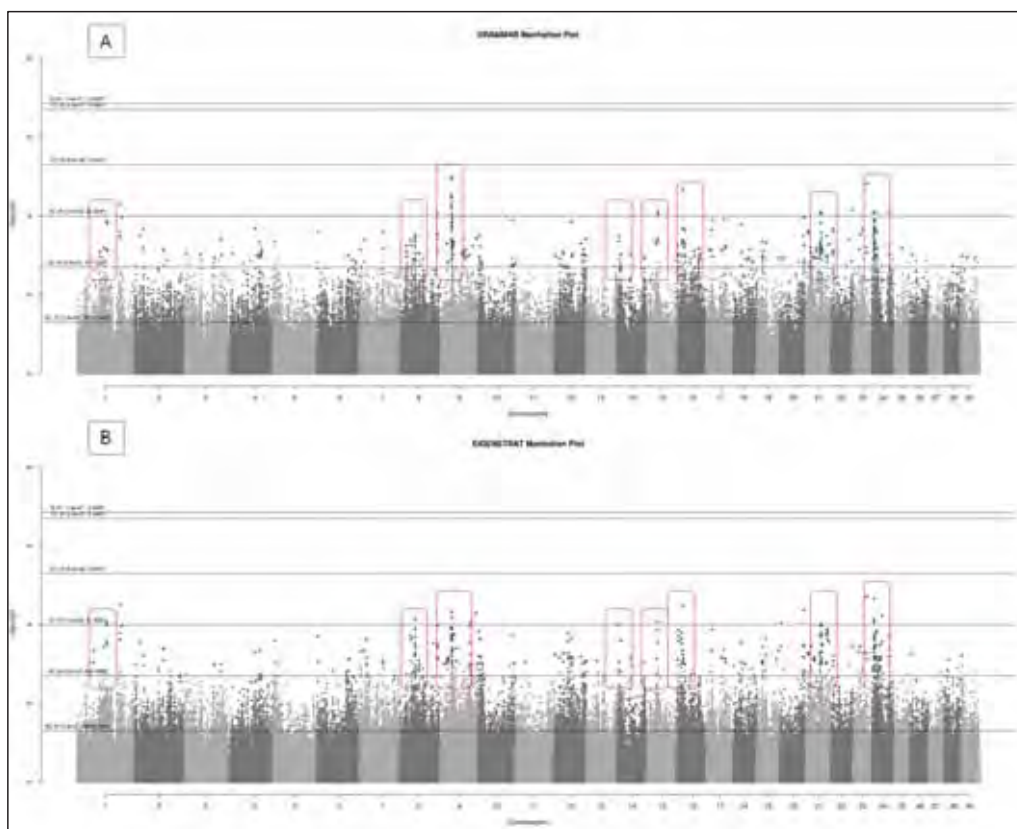


Figura 25. *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT (B) para a característica DMD para o Experimento 2(médias das medidas fenotípicas). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destaca SNP significativos em ambos os métodos.

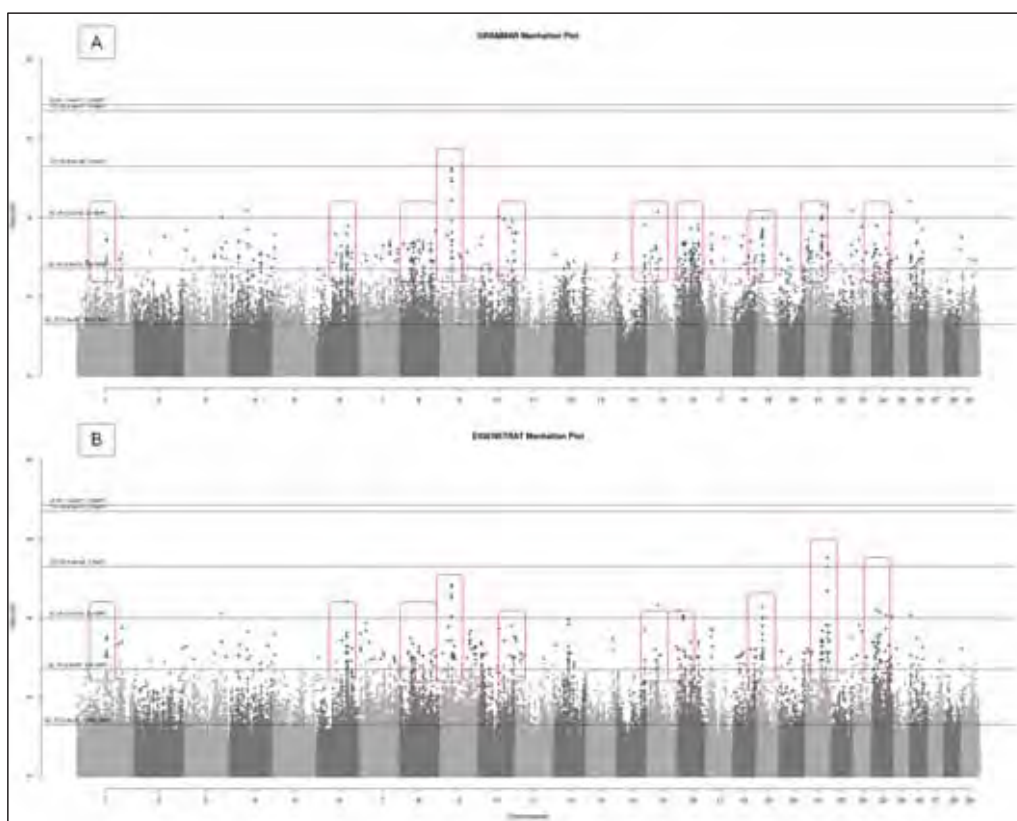


Figura 26. *Manhattan Plot* dos valores de significância (p) ajustados pelo Método de Bonferroni e classificados pelo sistema BADGE (MANLY, 2005) do teste de associação realizado pelos métodos GRAMMAR (A) e EIGENSTRAT (B) para a característica DMT para o Experimento 2 (médias das medidas fenotípicas). SNP com valores de significância 0,01 mais altos estão representados de azul. Retângulo vermelho destaca SNP significativos em ambos os métodos.

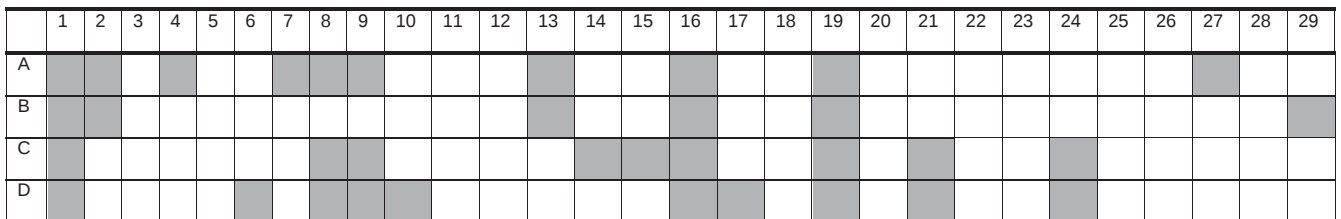


Figura 27. Representação dos cromossomos que apresentaram SNP significativos após o teste de associação realizado com ambos os métodos estatísticos (GRAMMAR e EIGENSTRAT) nos fenótipos (A) DMD no Experimento 1 (EBV), (B) DMT no Experimento 1 (EBV), (C) DMD no Experimento 2 (médias das medidas fenotípicas) e (D) DMT no Experimento 2 (médias das medidas fenotípicas).

Tabela 5. Número de genes que se apresentaram associados com as características avaliadas (DMD e DMT) nos diferentes métodos estatísticos (GRAMMAR e EIGENSTRAT).

Característica	Teste de associação	Nº de genes
Experimento 1		
DMD	GRAMMAR	196
	EIGENSTRAT	209
DMT	GRAMMAR	241
	EIGENSTRAT	383
Experimento 2		
DMD	GRAMMAR	266
	EIGENSTRAT	210
DMT	GRAMMAR	350
	EIGENSTRAT	302

Com base nos resultados das análises do GWAS, a análise de enriquecimento funcional (AEF) foi conduzida com o apoio do software DAVID (5). Apenas a análise de enriquecimento funcional para a característica DMD no Experimento 1 (análise de EBV) não apresentou grupos gênicos funcionais com $IE > 2$. Tal observação pode ter sido causada pelo menor número de genes selecionados para a AEF nessa característica (Tabela 5). Apesar disso, os grupos funcionais com menor IE para essa característica, se assemelharam aqueles observados para os demais fenótipos.

O grupo funcional que mais se destacou, não apenas pelo IE, mas por estar presente nos dois métodos de associação utilizados, foi o de inibidores de proteases. Dentre os genes desse grupo se destacaram os genes da superfamília de Serpinas.

As Serpinas compõem a superfamília de proteínas que possuem em comum a presença de um domínio único que determina sua estrutura e função. A maioria dos membros dessa família são inibidores de proteases com um mecanismo de ação único e atuação em diferentes tecidos (GETTINS, 2002).

Segundo LEE et al. (2011), inibidores de protease do tipo serina (família das serpinas) estão presentes nas células germinativas dos testículos de camundongos e envolvidos na regulação da espermatogênese e inibição da apoptose. Segundo esses autores, camundongos com mutações nesse gene apresentam fertilidade reduzida, diminuição da concentração espermática e espermatogênese anormal.

Esses inibidores de protease também foram identificados por LU et al. (2011) na mucosa do epitélio da vesícula seminal, epidídimo e vasos deferentes. Nos testículos, foram localizados nas espermatogônias, espermatócitos, espermátides, células de Leydig, espermatozoide, além do acrossoma como uma proteína intrínseca da superfície dessa região.

Ainda segundo esses autores, esse inibidor de protease é detectado predominantemente nos espermatozoides não capacitados, sendo perdida antes que esse processo se inicie. Além disso, o mesmo estudo demonstrou evidências de que essa protease pode inibir a albumina sérica bovina indutora de capacitação impedindo que o espermatozoide se una ao oócito.

Apesar de não terem sido realizados estudos de expressão diferencial dos inibidores de protease do tipo Serpina entre touros bons e maus congeladores no presente estudo, a menção a trabalhos prévios sobre a influência das Serpinas sobre a propriedade capacitante do espermatozoide sugere o aprofundamento dos estudos das funções dessa proteína. seria extremamente importante para prevenir que o espermatozoide se capacite precocemente com o processo de criopreservação.

Outro membro da superfamília de Serpinas, os inibidores de protease tipo serina – heparina dependente, já foram descritas como importantes inibidores da penetração do espermatozoide em oócitos humanos livres de zona pelúcida. De acordo com ESPAÑA et al. (2007), essa substância está presente no líquido seminal em uma concentração de 2,2 a 3,7 micromolar. Aproximadamente duas horas após a ejaculação essa protease perde sua atividade devido a sua interação com outras proteases. Uma concentração de apenas 0,04 micromolar inibe 50% da capacidade de penetração espermática, reduzindo a fertilidade. Essas informações também auxiliam na qualificação desse gene como alvo de estudos futuros.

Além das já citadas, outras serpinas desempenham importantes funções na reprodução dos machos tais como aquelas relacionadas com a inibição da proteína C, que é responsável por inibir a reação acrossômica devido a inibição da protease acrossina (GEIGER et al., 1996), e Serpinas $\alpha 1$ – anti-tripsina, que estão relacionadas a motilidade progressiva do espermatozoide humano. DING et al. (2007), afirmam que essa última tem papel importante na maturação do espermatozoide pelo epidídimo. Os resultados dos trabalhos desses autores sugerem que esses genes devem ser incluídos em estudos mais profundos sobre os fenótipos DMD e DMT, já que os mesmos demonstram ser fundamentais para a manifestação desses fenótipos. Ambos os genes relatados foram selecionados pela AEF com IE > 2 no presente estudo.

As Serpinas também podem atuar indiretamente na reprodução já que são responsáveis pela secreção da forma ativa de neurotransmissores envolvidos com a regulação do eixo hipotálamo-hipófise, tais como o Neuropeptídeo Y e os opióide endógenos (HWANG et al., 2009). O controle da liberação de GnRH pelos Neuropeptídeos Y é de fundamental importância para a produção de espermatozoide

funcionais e capazes de tolerar processos como a criopreservação, tornando esse gene alvo para estudos subsequentes.

Além das Serpinas, outros grupos funcionais que mereceram destaque são os genes envolvidos na dinâmica dos componentes do citoesqueleto celular e da ativação da cAMP. É de conhecimento geral que após a descongelação do sêmen, a célula espermática passa do estado inerte para um estado ativo, sendo obrigada a reestabelecer suas funções e sua motilidade. A ativação do cAMP induz a ativação flagelar na maioria dos mamíferos através da fosforilação oxidativa de proteínas relacionadas à motilidade (TASH et al., 1996). Estudos realizados por SI & OKUNO (1995) demonstraram que a o cAMP ativa a motilidade espermática pela regulação do deslizamento entre os microtúbulos, evidenciando a alteração nas propriedades deslizantes dos microtubulos espermáticos pela ativação de uma proteína cAMP dependente.

Com base em relatos encontrados na literatura, o presente trabalho propõe um modelo para a interação dos principais grupos funcionais / genes observados na AEF (Figura 28).

Interessantemente, o modelo proposto a partir da AEF se assemelha ao descrito por VISCONTI & KOPF (1998). Esses autores descreveram via metabólica detalhada com os principais fatores envolvidos na capacitação espermática. Dentre os fatores citados, muitos foram encontrados no presente estudo como grupos funcionais importantes para a manifestação das características DMD e DMT, auxiliando a compreensão da interação entre eles, já que a capacitação precoce da célula espermática é uma das consequências negativas da criopreservação (Figura 29).

Além disso, ao se listar todos os principais grupos funcionais indicados pela AEF, utilizando distintos fenótipos e modelos estatísticos, observou-se que a vasta maior parte deles encaixava-se no modelo (Tabela 6).

A similaridade entre os resultados do presente estudo e os descritos na literatura indica a consistência e interação dos grupos funcionais observados e apontam para novas perspectivas de pesquisa científica nesse campo através da exploração mais profunda dos grupos funcionais apontados nesse trabalho.

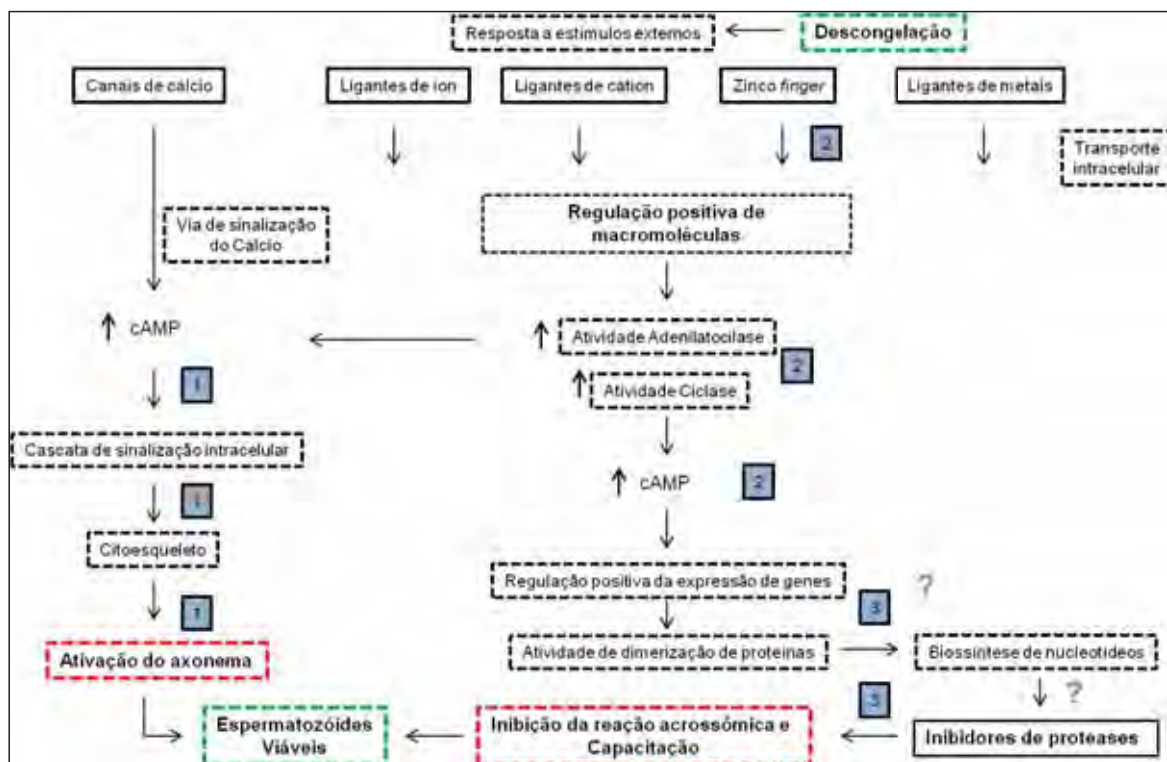


Figura 28. Modelo proposto pelo presente trabalho para as interações entre grupos funcionais encontrados na AEF: linhas contínuas pretas representam grupos de genes específicos; linhas tracejadas pretas representam vias metabólicas. Linha tracejada vermelha representa o possível resultado da interação entre os grupos funcionais arbitrado no presente trabalho. Linha tracejada verde representa o ponto inicial e final do modelo arbitrado no presente trabalho. Via metabólica (1) Baseado nos resultados descritos por SI & OKUNO (1995) e TASH et al. (1996); Via metabólica (2) -Baseado nos resultados descritos por VISCONTI & KOPF; e Via metabólica (3) - Baseado nos resultados descritos por GEIGER et al. (1996) e LU et al. (2011).

Tabela 6. Índice de Enriquecimento (IE) dos principais grupos funcionais encontrados na AEF com o uso do programa DAVID em ordem decrescente de valor, originados a partir das diferentes combinações de fenótipos e métodos estatísticos para o GWAS. Em destaque os grupos funcionais não utilizados para a proposição do modelo por não apresentar relação clara e direta com dados de literatura.

Grupo Funcional	Nº de vezes que aparece na AEF	Índice de Enriquecimento
Regulação da resposta ao estímulo externo	2	3,11
Regulação da resposta ao estímulo externo	2	2,6
Inibidores de protease	5	2,41
Inibidores de protease	5	2,23
Inibidores de protease	5	2,17
Canais de cálcio e proteínas carreadoras	2	2,01
Atividade de dimerização de proteínas	1	1,72
Regulação e biossíntese de cAMP	2	1,67
Inibidores de protease	5	1,57
Inibidores de protease	5	1,46
Canais de cálcio e proteínas carreadoras	2	1,25
Regulação positiva dos processo metabólicos envolvendo macromoléculas	1	1,23
Transporte intracelular	1	1,21
Via de sinalização do cálcio	1	1,19
Regulação e biossíntese de cAMP	2	1,19
Componentes do citoesqueleto	1	0,9

7. CONCLUSÕES

1. Tanto o valor genômico estimado (EBV) quanto as médias das medidas fenotípicas para os fenótipos DMD e DMT utilizados em teste de associação genômica (GWAS), mostraram-se capazes de identificar regiões genômicas associadas aos fenótipos em questão.

2. Regiões cromossômicas foram identificadas como associadas aos fenótipos DMD e DMT, sendo que se destacaram regiões específicas nos cromossomos 1, 16 e 19 em todos os tipos de fenótipos ou métodos utilizados para o GWAS.

3. Adicionalmente, regiões cromossômicas foram identificadas como associadas aos fenótipos de interesse de forma distinta entre os Experimentos 1 (2 e 13) e Experimento 2 (8, 9, 21 e 24).

4. Os grupos funcionais que mais se destacaram na análise de enriquecimento funcional (AEF) foram as famílias de inibidores de proteases Serpina, reguladores de cAMP celular e componentes do citoesqueleto, permitindo a proposição de modelo de interação entre os grupos funcionais em linha com o conhecimento prévio existente na literatura.

8. REFERÊNCIAS BIBLIOGRÁFICAS¹

ABIEC. Associação Brasileira das Indústrias Exportadoras de Carnes (2010). Rebanho Bovino Brasileiro. Disponível em: www.abiec.com.br/3_rebanho.asp. Acesso em: 03 de Out de 2011.

AGCA, Y.; GILMORE, J.; BYERS, M. et al. Osmotic characteristics of mouse spermatozoa in the presence of extenders and sugars, **Biol. Reprod.** v. 67, p. 1493–1501, 2002.

AMANN, R. P.; KATZ, D. F. Reflections on CASA after 25 years. **Journal of Andrology**. v. 25, p. 317–325, 2004.

AMIN, N.; VAN DUIJN, C.; AULCHENKO, Y. A genomic background based method for association analysis in related individuals. **PloS one**, v.2, n. 12, p. 1274, 2007.

AMMAN, R. P.; HAMMERSTEDT, R. H. In vitro evaluation of sperm quality: an opinion. **Journal of Andrology**. v. 14, p. 397–406, 1993.

ANDERSON, C. A.; PETTERSON F. M.; CLARKE, G. M. et al. Data quality control in genetic case-control association studies. **Nature Protocols**, v.5, n.9, p. 1564–1573, 2010.

ANDERSON, C. A. Data Quality Control. Analysis of Complex Disease Association Studies, p.95. Elsevier Inc. 2011.

¹ 1 Normalização documental para produção científica da UNESP: normas para apresentação de referências segundo a NBR 6023: 2002 da ABNT. São Paulo, 2003. disponível em :<http://www.biblioteca.unesp.br/pages/normalização.pdf>.

ASBIA. Associação Brasileira de Inseminação Artificial (2010). Disponível em: www.asbia.org.br. Acesso em: 04 de Out de 2011.

ASBIA. Associação Brasileira de Inseminação Artificial (2011). Disponível em: www.asbia.org.br. Acesso em: 27 de Março de 2012.

ASHWELL, M. S.; HEYEN, D. W.; SONSTEGARD, T. S.; VAN TASSELL, C. P. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. **Journal of Dairy Science**. v. 87, p. 468–475, 2004.

ASTON, K. I.; CARREL, D. T. Genome-Wide Study of Single-Nucleotide Polymorphisms Associated With Azoospermia and Severe Oligozoospermia. **Journal of Andrology**. v. 30, n. 6, 2009.

ASTON, K. I.; KRAUSZ, C.; LAFACE, I.; RUIZ-CASTANÉ, E.; CARREL, D. T. Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. **Human Reproduction**. v.25, n. 6, p. 1383–1397, 2010.

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

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

BALDING, D.J. A tutorial on statistical methods for population association studies. **Nat Rev Genet**. v. 7, n.10 781-91, 2006.

BANKUTI, F. I.; AZEVEDO, P. F. Na clandestinidade: o mercado informal de carne bovina (2004). Disponível em: www.fearp.usp.br. Acesso em: 03 de Out de 2011.

BARBAS, J. P.; MASCARENHAS, E. R. D. Cryopreservation of domestic animal sperm cells. **Cell Tissue Bank**. v.10, p.49–62, 2009.

BARBUJANI, G.; MAGAGNI, A.; MINCH, E.; An apportionments of human DNA diversity. **Proc. Natl. Acad. Sci**. v. 94, p. 4516–4519, 1997

BARENDSE, W. The effect of measurement error of phenotypes on genome wide association studies. **BMC Genomics**. v. 12, p. 232, 2011.

BELL, M.; WANG, R.; HELLSTROM, W. J.; SIKKA, S. K.; Effect of cryoprotective additives and cryopreservation protocol on sperm membrane lipid peroxidation and recovery of motile human sperm. **Journal of Andrology**. v. 14, p. 472–478, 1993.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: A Partical and Powerful approach for multiple testing. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 57, n 1, p. 289-300, 1995.

BICKHART, D. M.; HOU, Y., SCHROEDER, S. S. et al. Copy number variation of individual cattle genomes using next-generation sequencing. **Genome Research**, In press.

BOERWINKLE, E.; TURNER, S. T.; WEINSHILBOUM, R. et al. Analysis of the distribution of erythrocyte sodium lithium countertransport in a sample representative of the general population. **Genet Epidemiol**. v. 3, p.365-378, 1986.

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

BOLDMAN, K.G.; KRIESE, L.A.; VAN VLECK, L.D. et al. A manual for use for MTDFREML. A set of programs to obtain estimates of variance and covariance

[DRAFT]. Lincoln: Department of Agriculture / Agricultural Research Service, 1995. 120p.

BOLORMAA, S. PRYCE, J. E.; HAYES, B. J.; GODDARD, M. E. Multivariate analysis of a genome-wide association study in dairy cattle. **Journal of Dairy Science**. v. 93, p. 3818–3833, 2010.

BROOKES, A. J. The essence of *SNP*. **Gene**. v.234, p. 177–186, 1999.

BUSH, W. S.; CHEN, G.; TORSTENSON, E. S. et al. LD-Spline: Mapping SNPs on genotyping platforms to genomic regions using patterns of linkage disequilibrium. **BioData Mining**. v. 2, n.7, 2009.

CANTOR, R. M.; LANGE, K. SINSHEIMER, J, S. Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. **The American Journal of Human Genetics**. v. 86, p. 6–22, 2010

CARDON, L. R, PALMER, L. J. Population stratification and spurious allelic association. **Lancet**, v. 361, p. 598, 2003.

CEPEA. Centro de Estudos Avançados de Economia Aplicada. Disponível em: www.cepea.esalq.usp.br. Acesso em 04 de Out de 2011.

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, p. 149–156, 2009.

CHANDLER, J. E.; ADKINSON, R. W.; et al. Environmental and Genetic Sources of Variation for Seminal Quality in Mature Holstein Bulls' **J Dairy Sci**, v. 68, p.1270-1279, 1985.

CHUANG, H.; HOFREE, M.; IDEKER, T. A decade of Systems Biology. **Reviews in Advance**. v. 20, p. 15, 2010.

COLE, J. B.; VANRADEN, P. M.; O'CONNELL, J. R.; VAN TASSELL, C. P. et al., 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. et al., Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics**. v. 12, p.408, 2011.

COOPER, G. M.; ZERR, T.; KIDD, J. M.; EICHLER, E. E.; NICKERSON, D. A. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. **Nature Genetics**. v. 40, p. 1199–1203, 2008.

CURRY, M. R.; MILLAR, J. D.; WATSON, P. F. Calculated optimal cooling rates for ram and Human sperm cryopreservation fail to conform with empirical observations. **Biology of Reproduction**, v. 51, p. 1014–1021, 1994.

CURRY, M. R.; REDDING, B. J.; WATSON, P. F. Determination of water permeability coefficient and its activation energy for rabbit spermatozoa. **Cryobiology**. v. 82, p. 175–181, 1995.

CURRY, M. R. Cryopreservation of semen from domestic livestock. **Reviews of Reproduction**. v. 5, p. 46–52, 2000.

DEFOIN, L.; GRANADOS, A.; DONNAY, I. Analysing motility parameters on fresh Bull semen could help to predicty resistance to freezing: a preliminary study. **Reproduction of Domestic Animals**. v. 43, p. 606-611, 2008.

DE ROUEN, S. M., FRANKE D. E. Effects of sire breed, breed type and age and weight at breeding on calving rate and date in beef heifers first exposed at three ages. **Journal of Animal Science**. v. 67, p. 1128-1137, 1989.

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

DING, Z.; QU, F.; GUO, W. Identification of sperm forward motility-related proteins in human seminal plasma. **Molecular Reproduction and Development**. v.74, n. 9, p. 1124-1131, 2007.

EASTON, D. F.; POOLEY, K. A.; DUNNING, A. M. et al., Genome-wide association study identifies novel breast cancer susceptibility loci. **Nature**. v. 447, p.1087-93, 2007.

ESPAÑA, F.; NAVARRO, S.; MEDINA, P. et al.The role of protein C inhibitor in human reproduction. **Semin. Thromb. Hemost.** v.33, n.1, p.41-45, 2007.

ESTEVEES, S. C.; SPAINE, D. M.; CEDENHO, A.P. Effects of pentoxifylline treatment before freezing on motility, viability and acrosome status of poor quality human spermatozoa cryopreserved by the liquid nitrogen vapor method. **Brazilian Journal of Medical and Biological Research**. v. 40, p. 985-992, 2007.

FAO. Food And Agriculture Organization Of The United Nations. The State of food and agriculture – Livestock in Balance, Roma, 2009, 167p.

FEUGAND, J. M.; KAYA, A.; PAGE, G. P.; CHEN, L.; MEHTA, T. et al., Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. **BMC Genomics**. v. 10, p.176, 2009.

FONTANILLO, C.; NOGALES-CADENA, R. et al. Functional Analysis beyond Enrichment: Non-Redundant Reciprocal Linkage of Genes and Biological Terms. **PloS one**. v. 6, n.9, p.242, 2011.

FORTES, M. R. S.; REVERTER, A.; ZHANGA, Y.; COLLIS, E.; NAGARAJ, S. H.; JONSSONA, N.; PRAYAGAA, K. C.; BARRIS, W.; HAWKEN, R. J. Association weight matrix for the genetic dissection of puberty in beef cattle. **PNAS**. v. 107, n.31, 13642–13647. 2010.

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.

FREIMER, N. B.; SABATTI, C. Guidelines for association studies in Human Molecular Genetics, **Hum. Mol. Genet.**, v.14, p. 2481-2483, 2005.

FU, W.; WANG, Y.; WANG, Y, et al. Missing call bias in high-throughput genotyping. **BMC Genomics**.v. 10, p.106, 2009.

GABRIEL, S. B.; The Structure of Haplotype Blocks in the Human Genome. **Science**. v. 296, p. 2225, 2002

GAO, G. Y.; ASHWORTH, E.; WATSON, P. F.; KLEINHANS, F. W.; MAZUR, P.; CRITSER, J. K.; Hyperosmotic tolerance of human spermatozoa: separate effects of glycerol, sodium chloride and sucrose on spermolysis. **Biology of Reproduction**. v. 49, p. 112–123, 1993.

GEIGER, M.; ZECHMEISTER-MACHHART, M.; UHRIN, P. et al. Protein C inhibitor (PCI). **Immunopharmacology**. v. 32, n.1-3, p.53-56, 1996.

GETTINS, P. G. W. Serpin Structure, Mechanism, and Function. **Chem. Rev.**v. 102, p. 4751-4803, 2002.

GILMORE, J. A.; DU, J.; TAO, J.; PETER, A. T.; CRITSER, J. K. Osmotic properties of boar spermatozoa and their relevance to cryopreservation. **Journal of Reproduction and Fertility**. v. 107, p. 87–95, 1996.

GOLDSTEIN, D. B. Common genetic variation and human traits. **The New England Journal of Medicine**, v. 360, p.1696–1698, 2009.

GORLOV, I. P.; GORLOVA, O. Y.; SUNYAEV, S. R. et al. Shifting paradigm of association studies: value of rare singlenucleotide polymorphisms. **Am. J. Hum. Genet.**, v. 82, p.100–112, 2008.

HALLAP, T., NAGY, S., JAAKMA, U., JOHANNISSON, A., RODRIGUEZ-MARTINEZ, H. Usefulness of a triple fluorochrome combination Merocyanine 540 / Yo-Pro 1 / Hoechst 33342 in assessing membrane stability of viable frozen-thawed spermatozoa from Estonian Holstein AI bulls. **Theriogenology**, v. 65, p. 1122–1136, 2006.

HAMMERSTEDT, R. H.; GRAHAM, J. K.; NOLAM, J. P. Cryopreservation of mammalian sperm: What we ask them to survive? **Journal of Andrology**, v.11, p.73-88, 1990.

HANDELSMAN, D. J. Optimal Power transformations for Analysis of Sperm Concentration and Other Semen Variables. **Journal of Andrology**. v.23, p. 629–634, 2002.

HAO, K.; CHUDIN, E.; GREENAWALT, D. et al. Magnitude of Stratification in Human Populations and Impacts on Genome Wide Association Studies. **PLoS ONE**, v. 5, n. 1, p. 8695, 2010.

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. Estimation of variance and variance components. **Biometrics**, v. 9, p. 226-252, 1953.

HIRSCHHORN, J. N.; DALY, M. J. Genome-wide association studies for common diseases and complex traits. **Nature Reviews**. v. 6, p. 95-198, 2005.

HOLMBERG, M.; ANDERSSON-EKLUND, L. Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. **Journal of Dairy Science**. v. 89, p. 3664–3671, 2006.

HOLT, W. V. Fundamental aspects of sperm cryobiology: the importance of Species and individual differences. **Theriogenology**. v. 53, p. 47-58 , 2000.

HUANG, D. W.; SHERMAN, B. T.; LEMPICK, I R. A. et al. DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. **Nucleic Acids Res.**, v.35, n.1, 2007.

HUANG, D. W.; SHERMAN, B. T.; LEMPICK, I R. A. .Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Res.**, v.37, n.1, 2009a.

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

HWANG, S.; BUNDEY, R.; TONEFF, T.; HOOK, V. Endopin Serpin Protease Inhibitors Localize with Neuropeptides in Secretory Vesicles and Neuroendocrine Tissues. **Neuroendocrinology**. v. 89, n. 2, p. 210–216, 2009.

ILLUMINA. Product literature. Disponível em: <<http://www.illumina.com/support/literature.ilmn>>. Acesso em: 24 de Ago. 2011.

JANUSKAUSKAS, A., RODRIGUEZ-MARTINEZ, H. Assessment of sperm viability by measurements of ATP, membrane integrity and motility in frozen thawed bull semen. **Acta. Vet. Scan.**, v. 36, p. 571 – 574, 1995.

KANEHISA, M.; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Res.**, v. 28, p. 27-30, 2000.

KOIVISTO, M. B., COSTA, M. T. A., PERRI, S. H. V., VICENTE, W. R. R. The effect of season on semen characteristics and freezability in *Bos indicus* and *Bos Taurus* bulls in the southeastern region of Brazil. **Reproduction of Domestic Animals**. v. 44, p. 587–592, 2009.

KORN, J. M. et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. **Nature Genetics**. v. 40, p. 1253–1260, 2008.

KU, C. S.; LOY, E. Y. The pursuit of genome-wide association studies: where are we now? **Journal of Human Genetics** v. 55, p.195–206, 2010.

KUHN, C.; BENNEWITZ, J.; REINSCH, N.; XU, N.; THOMSEN, H.; LOOFT, C. et al., Quantitative trait loci mapping of functional traits in the German Holstein cattle population. **Journal of Dairy Science**. v. 86, p. 360–368, 2003.

LAMBERT, C. G.; BLACK, L. G. Learning from our GWAS mistakes: from experimental design to scientific method. **Biostatistics**. v. 13, n. 2, p. 195-203, 2012.

LAURIE, C. C.; DOHENY, K. F.; MIREL, D. B. et al. Quality Control and Quality Assurance in Genotypic Data for Genome-Wide Association Studies. **Genetics of Epidemiology**. v.34, p. 591–602, 2010

LEE, B.; PARK, I.; JIN, S. et al. Impaired spermatogenesis and fertility in mice carrying a mutation in the Spink2 gene expressed predominantly in testes. *J. Biol. Chem.*, v. 286, n. 33, p. 29108 – 291017, 2011

LI, W.; SADLER, L. A.; Low nucleotide diversity in man. **Genetics**. v. 54, p. 129, 513–523, 1991.

LU, C. H.; LEE, R. K.; HWU, Y. M. et al. SERPINE2, a serine protease inhibitor extensively expressed in adult male mouse reproductive tissues, may serve as a murine sperm decapacitation factor. **Biology of Reproduction**. v.84, n.3, p. 514-525, 2010

MANLY, K. F. Reliability of statistical associations between genes and disease. **Immunogenetics**. p. 57, p. 549–558, 2005.

MANOLIO, T. A. Genomewide Association Studies and Assessment of the Risk of Disease. **The new england journal of medicine**. v. 363, p. 166-76, 2010.

MARCHINI, J. et al The effects of human population structure on large genetic association studies. **Nat. Genet.** v. 36, p. 512–517, 2004.

MAZUR, P.; LEIBO, S. P.; CHU, E.H. Y. A two-factor hypothesis of freezing injury. **Exp Cell Res**; v. 71, p. 345-355, 1972

McCARTHY, M. I.; ABECASIS, G. R.; CARDON, L. R.; GOLDSTEIN, D. B.; LITTLE, J.; IOANNIDIS, J. P. A.; HIRSCHHORN, J. N. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. **Nature Reviews Genetics**. v. 9, p. 356–369, 2008.

MANUAL para exame andrológico e avaliação de sêmen animal. 2. ed. Belo Horizonte: CBRA., 1998. 49p.

MEDEIROS, C. M. O.; FORELL, F.; PLIEIRA, A. T. D.; RODRIGUES, J. L. Current status of sperm cryopreservation. Why isn't it better? **Theriogenology**. v. 57, p. 327-344, 2002.

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).

MOUNTAIN, J. L.; LIN, A. A.; BOWCOCK, A. M. et al. Evolution of modern humans: evidence from nuclear DNA polymorphisms. **Philos. Trans. Roy. Soc. London B: Biol. Sci.** v. 337, p. 159–165, 1992.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE. A Catalog of Published Genome Wide Association Studies. Disponível em: <http://www.genome.gov/GWASudies/>. Acesso em: 23 de Ago de 2011.

NEI, M.; ROYCHOUDHURY, A. K. Gene differences between caucasioan, negro and Japanese population. **Science**, v. 177, p. 434-436, 1972.

PARKINSON, T. J.; WHITFIELD, C. H.; Optimisation of freezing conditions for bovine spermatozoa. **Theriogenology**. v. 27, p. 781–797, 1987.

PATTERSON, N.; PRICE, A. L.; REICH, D. Population structure and eigenanalysis. **PLoS Genet**, v.2, p.190, 2006.

PEARSON, T. A.; MANOLIO, T. A. How to Interpret a Genome-wide Association Study. **Journal os American Medical Association**. v. 299, n. 11, p. 1335-1344, 2008.

PESSUTI, O.; MEZZADRI, F. P. Atualidade e perspectiva da pecuária paranaense. In: Simpósio Internacional de Reprodução Animal Aplicada, 1., 2004, Londrina. Anais., Londrina, 2004. p. 21-27, 2004.

PRICE, A. L.; PATTERSON, N. J.; PLENGE, M. J. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, v. 38, n. 8, p.8, p.904, 2006.

PRICE, A. L.; ZAITLEN, N. A.; REICH, D. et al. New approaches to population stratification in genome-wide association studies. **Nat Rev Genet**.. v. 11, n.7, p.: 459–463, 2010.

PRITCHARD, J. K; STEPHENS, M. et al. Inference of population structure using multilocus genotype data. **Genetics**, v.155, p. 945–59, 2000.

PRYCE, J. E.; BOLORMAA, S.; CHAMBERLAIN, P. J.; SAVIN, K. et al., A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. **Journal of Dairy Science**. v. 93, p. 3331–3345, 2010.

RASMUSSEN-TORVIK, L. J.; ALONSO, A. Impact of repeated measures and sample selection on genomewide association studies of fasting glucose. **Genet Epidemiol.** v. 34, n. 7, p. 665–673, 2010.

REDON, R. et al. Global variation in copy number in the human genome. **Nature.** v. 444, p. 444–454, 2006.

ROBERTSON, L.; WATSON, P.F. Calcium transport in diluted or cooled ram semen. **Journal of Reproduction and Fertility.** v.77, 1986
177–185.

SACCONE, S. F.; SACCONE, N. L.; SWAN, G. E.; MADDEN, P. A.; GOATE, A. M. et al., Systematic biological prioritization after a genome-wide association study: an application to nicotine dependence. **Bioinformatics.** v. 24, p.1805-1811, 2008.

SAMANI, N. J.; ERDMANN, J.; HALL, A.S. et al. Genomewide association analysis of coronary artery disease. **N Engl J Med.**, v. 357, p. 443–453, 2007.

SCOTT, L. J.; MOHLKE, K. L.; BONNYCASTLE, L. L.; WILLER, C. J.; LI, Y.; DUREN, W. L. et al., A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. **Science.** v. 316, p. 1341–1345, 2007.

SCUTERI, A.; SANNA, S.; CHEN, W. M. et al.: Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. **PLoS Genetics.** v. 3, p.115, 2007.

SEBASTIANI, P.; LAZARUS, R.; WEISS, S.T. et al. Minimal haplotype tagging. **Proc Natl Acad Sci.**, v. 100, p. 9900–9905, 2003

SEBASTIANI, P.;TIMOFEEV, N.; DWORKIS, D.A. Genome-wide association studies and the genetic dissection of complex traits. **American Journal of Hematology**. v. 84, p. 504–515, 2009.

SEBRAE. Serviço Brasileiro de Apoio às Micro e Pequenas Empresas. Estudo de Empresas de Biotecnologia do Brasil. Fundação Biomina. 54p. 2007.

SHAMSUDDIN, M.; LARSSON, B. *In vitro* development of bovine embryos after fertilization using semen from different donors. **Reproduction of Domestic Animals**. v.28, n.2, p.77-84, 1993.

SI, Y.; OKUNO, M. Activation of Mammalian Sperm Motility by Regulation of Microtubule Sliding Via Cyclic Adenosine 5'-Monophosphate-Dependent Phosphorylation. **Biology of reproduction**, v. 53, p. 1081-1087, 1995.

SLADEK, R.; ROCHELEAU, G.; RUNG, J.; DINA, C.; SHEN, L. et al., A genome-wide association study identifies novel risk loci for type 2 diabetes. **Nature**. v. 445, p. 881–885, 2007.

SPUNGIN, B.; MARGALIT, I.; BREITBART, H. Sperm exocytosis reconstructed in a cell free system. Evidence for the involvement of phospholipase C and actin filaments in membrane fusion. **Journal of Cell Science**. v. 108, p. 2525–2535, 1995.

STRÖM, B.; ROTA, A.; LINDE-FORSBERG, C. In vitro characteristics of canine spermatozoa subjected to two methods of cryopreservation. **Theriogenology**. v.48, p.247-256, 1997.

SUBRAMANIAN, A.; TAMAYO, P.; MOOTHA, V. K. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. **PNAS**. v.102, n.43, p. 15545–15550, 2005.

TABANGIN, M. E.; WOO, J. G.; MARTIN, L. J. The effect of minor allele frequency on the likelihood of obtaining false positives. **BMC Proceedings**, v. 3 (Suppl 7):S41, 2009.

TADA, N.; SATO, M.; YAMANOI, J.; MIZOROGI, T. et al., Cryopreservation of mouse spermatozoa in the presence of raffinose and glycerol. **Journal of Reproduction and Fertility**. v. 89, p.511—516, 1990.

TASH, J. S.; HIDAKA, H.; MEANS, A. R. Axokinin phosphorylation by cAMP-dependent protein kinase is sufficient for activation of sperm flagellar motility. **J Cell Biol** v. 103, p. 649-655. 1986.

TAYLOR, J. F.; COUTINHO, L. L.; HERRING, K. L.; GALLAGHER JUNIOR, D. S. et al., Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle. **Animal Genetics**. v. 29, p. 194–201, 1998.

TEO, Y. Y.; INOUE, M.; SMALL, K. S. et al. A genotype calling algorithm for the Illumina BeadArray platform. **Bioinformatics**, v. 23, p. 2741, 2007.]

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

THE GENE ONTOLOGY CONSORTIUM. Gene ontology: tool for the unification of biology. **Nature Genetetics**, v. 25, n.1, p. 25-29, 2000.

THE WELLCOME TRUST CASE CONTROL CONSORTIUM. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. **Nature**. v. 447, p.661–83, 2007.

THURSTON, L. M.; SIGGINS, K.; MILEHAM, A. J.; WATSON, P. F.; HOLT, W. V. Identification of Amplified Restriction Fragment Length Polymorphism Markers Linked to Genes Controlling Boar Sperm Viability Following Cryopreservation. **Biology of Reproduction**. v.66, p.545–554, 2002.

TORRES JR, A. M.; TITO ROSA, F. R.; TONINI, M. G. O. Como andam os negócios. **Agroanalysis**. v. 25. n. 6. p. 40-42, 2005.

VIANNA, F. P., PAPA, F. O., ZAHN, F. S., MELO, C. M., DELL'AQUA Jr., J. A. Thermoresistance sperm tests are not predictive of potential fertility for cryopreserved bull semen. **Animal Reproduction Science**, v. 113, p. 279-282, 2009.

VISCONTI, P. E.; KOPF, G. S. Regulation of Protein Phosphorylation during Sperm Capacitation. **Biology of Reproduction**. v.59, p. 1-6, 1998.

XIAO, Y.; SEGAL, M. R.; YANG, Y. H. ET AL. A multi-array multi-SNP genotyping algorithm for Affymetrix SNP microarrays. **Bioinformatics**. v.23, p.1459–1467, 2007.

WANG, D. G.; FAN, J. B.; SIAO, C. J.; BERNO, A.; YOUNG, P. et al., Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. **Science**. v. 280, p. 1077–1082, 1998.

WANG, W. Y. S.; 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.

WATSON, P.F. The effects of cold shock on sperm cell membranes. In: Morris, G.J., Clarke, A. _Eds., Effects of Low Temperatures on Biological Membranes. Academic Press, London, 1981, pp. 189–218.

WATSON, P.F. Recent developments and concepts in the cryopreservation of spermatozoa and the assessment of their post-thawing function. *Reproduction, Fertility and Development*. v. 7, p. 871–891, 1995.

WATSON, P. F. The causes of reduced fertility with cryopreserved semen. ***Animal Reproduction Science***. v. 60–61, p. 481-492, 2000.

WEBBER, C. Functional Enrichment Analysis with Structural Variants: Pitfalls and Strategies. ***Cytogenet Genome Res.***, v. 135, p. 277–285, 2011.

WEEDON, M. N.; LETTRE, G.; FREATHY, R. M. et al., A common variant of HMGA2 is associated with adult and childhood height in the general population. ***Nature Genetics***. v. 39: p. 1245– 1250, 2007.

WEIGEL, K. A.; DE LOS CAMPOS, G.; VÁZQUEZ, A. I.; VAN TASSELL, C. P.; ROSA, G. J. M.; GIANOLA, D.; O'CONNELL, J. R.; VANRADEN, P. M.; WIGGANS, G. R. Genomic selection and its effects on dairy cattle in breeding programs. In: *Proceedings of 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, Germany, p. 342., 2010.

XI-HUANG, J.; SCOTT, M. B.; PU, X. Y. et al., Association between single-nucleotide polymorphisms of *DNMT3L* and infertility with azoospermia in Chinese men. ***Reproductive Biomedicine Online***. In Press.

ZIEGLER, A.; KÖNING , I. R.; THOMPSON, J. R. Biostatistical Aspects of Genome-Wide Association Studies. ***Biometrical Journal***. v. 50, p. 8–28, 2008.

ANEXO 1

Tabela 1. Análise descritiva das médias e desvios padrão (DP) das medidas fenotípicas DMD e DMT.

Id touro	Média DMD	DP DMD	Média DMT	DP DMT
1	36,3	8,5	46,2	7,3
2	21,4	4,8	31,4	6,9
3	30,0	7,1	38,6	7,5
4	30,0	6,3	39,3	5,6
5	35,6	4,9	45,5	5,0
6	28,0	8,0	40,0	4,0
7	32,1	7,0	38,6	6,9
8	31,3	7,9	39,4	6,2
9	29,8	10,0	38,5	8,5
10	37,1	8,1	46,4	6,3
11	25,8	5,8	34,2	2,0
12	27,5	5,9	39,8	5,0
13	40,0	9,0	46,7	7,9
14	24,2	6,6	32,5	9,4
15	27,4	7,5	38,2	5,9
16	29,2	4,9	39,2	5,8
17	30,0	7,9	40,0	5,0
18	28,0	5,7	40,0	3,5
19	32,3	8,5	40,0	7,3
20	32,4	7,1	39,8	7,1
21	29,0	6,5	40,0	7,9
22	32,3	9,8	41,7	8,4
23	26,8	7,1	42,1	7,6
24	32,3	7,2	44,3	4,9
25	32,5	6,1	45,0	6,3
26	38,4	10,1	49,8	9,5
27	26,5	9,1	38,4	7,7
28	26,1	9,4	35,0	10,6
29	33,5	13,1	44,5	8,6
30	35,9	11,4	45,9	10,4
31	29,6	10,8	42,1	8,4

Id touro	Média DMD	DP DMD	Média DMT	DP DMD
32	27,4	8,4	38,9	7,7
33	32,0	5,4	43,9	7,1
34	30,4	9,3	39,9	8,7
35	30,8	12,4	36,7	6,1
36	41,4	12,5	50,8	6,9
37	33,3	8,0	44,2	8,5
38	35,3	7,4	41,5	7,9
39	30,2	6,8	39,3	8,1
40	35,0	4,6	41,5	4,3
41	35,6	7,5	46,9	7,9
42	30,1	8,7	40,6	8,7
43	33,1	7,1	40,7	6,6
44	37,6	14,0	46,7	10,9
45	27,3	4,8	38,1	5,2
46	32,9	10,2	45,5	7,7
47	30,5	8,6	41,0	10,4
48	32,1	7,7	39,8	6,6
49	31,1	7,3	38,4	7,7
50	29,7	9,3	43,9	6,4
51	27,7	6,7	38,8	6,3
52	28,5	8,2	42,8	7,6
53	30,4	9,9	41,8	9,1
54	29,1	7,9	38,1	7,4
55	32,5	9,5	44,5	8,6
56	31,0	10,0	46,1	8,9
57	39,5	11,7	51,8	9,8
58	33,6	10,5	45,5	9,4
59	30,5	8,8	41,4	7,9
60	30,2	7,6	38,9	7,4
61	31,8	8,3	39,9	8,7
62	30,6	7,4	37,6	8,0
63	29,1	8,8	39,1	7,6
64	34,5	9,1	46,8	6,4
65	32,0	11,6	41,0	9,5
66	29,7	8,7	35,5	10,5

Id touro	Média DMD	DP DMD	Média DMT	DP DMD
67	22,5	10,0	38,4	10,2
68	27,7	11,0	40,6	9,5
69	27,6	8,1	38,6	6,5
70	33,8	8,8	42,6	8,5
71	35,2	10,6	45,3	13,5
72	37,5	14,1	49,1	10,5
73	44,4	12,8	50,5	12,5
74	26,0	7,9	35,0	10,1
75	30,4	8,2	42,0	7,2
76	28,2	8,0	37,8	6,7
77	32,0	6,2	40,5	7,6
78	30,7	7,8	38,8	6,9
79	39,2	10,0	44,4	8,1
80	35,0	8,0	40,6	7,4
81	28,1	6,2	37,9	5,9
82	29,0	9,8	39,0	9,5
83	30,4	6,3	38,2	5,6
84	28,3	7,5	37,7	7,2
85	28,9	9,5	38,1	8,2
86	27,4	7,0	34,5	10,4
87	28,8	11,8	37,4	11,3
88	27,3	8,6	38,0	7,3
89	31,1	8,0	40,1	7,7
90	32,0	6,2	41,2	6,3
91	20,0	11,4	33,9	8,7
92	44,3	6,7	53,6	4,8
93	37,1	9,4	46,5	6,0
94	31,4	7,2	40,9	8,0
95	25,0	7,7	35,1	7,2
96	26,3	9,4	40,9	8,3
97	28,3	6,2	37,5	6,2
98	32,6	8,7	43,8	8,2
99	30,7	8,7	38,4	7,8
100	29,1	6,8	38,9	6,7
101	31,8	8,8	42,9	7,1

Id touro	Média DMD	DP DMD	Média DMT	DP DMD
102	34,5	6,2	43,0	6,7
103	31,0	10,6	40,6	9,0
104	26,3	10,9	34,1	10,3
105	33,2	7,0	42,5	7,4
106	29,4	6,0	35,9	4,6
107	32,0	10,2	42,4	8,7
108	25,7	9,7	32,7	8,7
109	38,8	9,4	46,9	7,8
110	35,2	8,9	45,2	7,9
111	40,8	7,3	49,6	7,5
112	28,9	6,3	35,9	4,2
113	26,1	8,3	40,0	7,8
114	33,7	9,0	45,0	10,3
115	25,6	6,9	33,9	6,1
116	34,0	2,2	43,0	4,5
117	36,6	12,6	45,1	9,2
118	34,8	9,6	45,1	9,0
119	34,5	7,6	47,5	9,8
120	32,2	9,7	44,1	8,7
121	25,0	10,0	31,9	9,1
122	32,8	13,8	43,5	10,7
123	30,0	5,8	38,6	5,7
124	32,1	9,9	43,6	9,4
125	31,1	9,0	39,2	7,6

Tabela 2. Análise descritiva das características DMD e DMT

	Média	Variância	C.V	Mínimo	Máximo
DMD	31,2816	18,2055	13,64	20	44,4
DMT	41,112	18,6465	10,5034	31,4	53,6

C.V = coeficiente de variação

ANEXO 2

Tabela 1. Estatística do Procedimento GLM (*Generalized Linear Model*) para avaliar a capacidade de ajuste do modelo utilizado para estimar o EBV (Experimento 1) da característica DMD (SAS, 2007).

	GL	S.Q.	Q.M.	Valor de F	Pr > F
Modelo	166	116176,4	699,8	8.95	< 0,0001
Erro	6595	515964,7	78,2		
Correção Total	6761	632141,1			

* G.L = graus de liberdade; S.Q = Soma dos quadrados; Q.M = Quadrados mínimos

Tabela 1. Estatística do Procedimento GLM (*General Linear Model*) para avaliar o efeito das covariáveis no EBV (Experimento 1). da característica DMD (SAS, 2007).

	GL	Tipo III SS	Q.M.	Valor de F	Pr > F
Touros	147	89549,4	609,1	7,79	< 0,0001
Ano* estação do ano	17	21670,5	1274,7	16,29	< 0,0001
Idade	1	2293,4	2293,4	29,31	< 0,0001
Idade ²	1	1132,1	1132,1	14,47	0,0001

* G.L = graus de liberdade; S.Q = Soma dos quadrados; Q.M = Quadrados mínimos

$$R^2 = 0,18$$

Coeficiente de variação = 28,34

Tabela 3. Estatística do Procedimento GLM (*Generalized Linear Model*) para avaliar a capacidade de ajuste do modelo utilizado para estimar o EBV (Experimento 1) da característica DMT (SAS, 2007).

	GL	S.Q.	Q.M.	Valor de F	Pr > F
Modelo	166	127753,8	769,6	11,92	<0,0001
Erro	6595	425611,1	64,53		
Correção Total	6761	553365,0			

* G.L = graus de liberdade; S.Q = Soma dos quadrados; Q.M = Quadrados mínimos

Tabela 4. Estatística do Procedimento GLM (*General Linear Model*) para avaliar o efeito das covariáveis no EBV (Experimento 1). da característica DMT (SAS, 2007).

	GL	Tipo I SS	Q.M.	Valor de F	Pr > F
Touros	147	69638,3	473,7	7,34	< 0,0001
Ano* estação do ano	17	20786,8	1222,7	18,9	< 0,0001
Idade	1	5611,2	5611,2	86,95	0,0005
Idade ²	1	774,2	473,7	12,0	0,0001

* G.L = graus de liberdade; S.Q = Soma dos quadrados; Q.M = Quadrados mínimos

$$R^2 = 0,23$$

Coeficiente de variação = 19,42

ANEXO 3

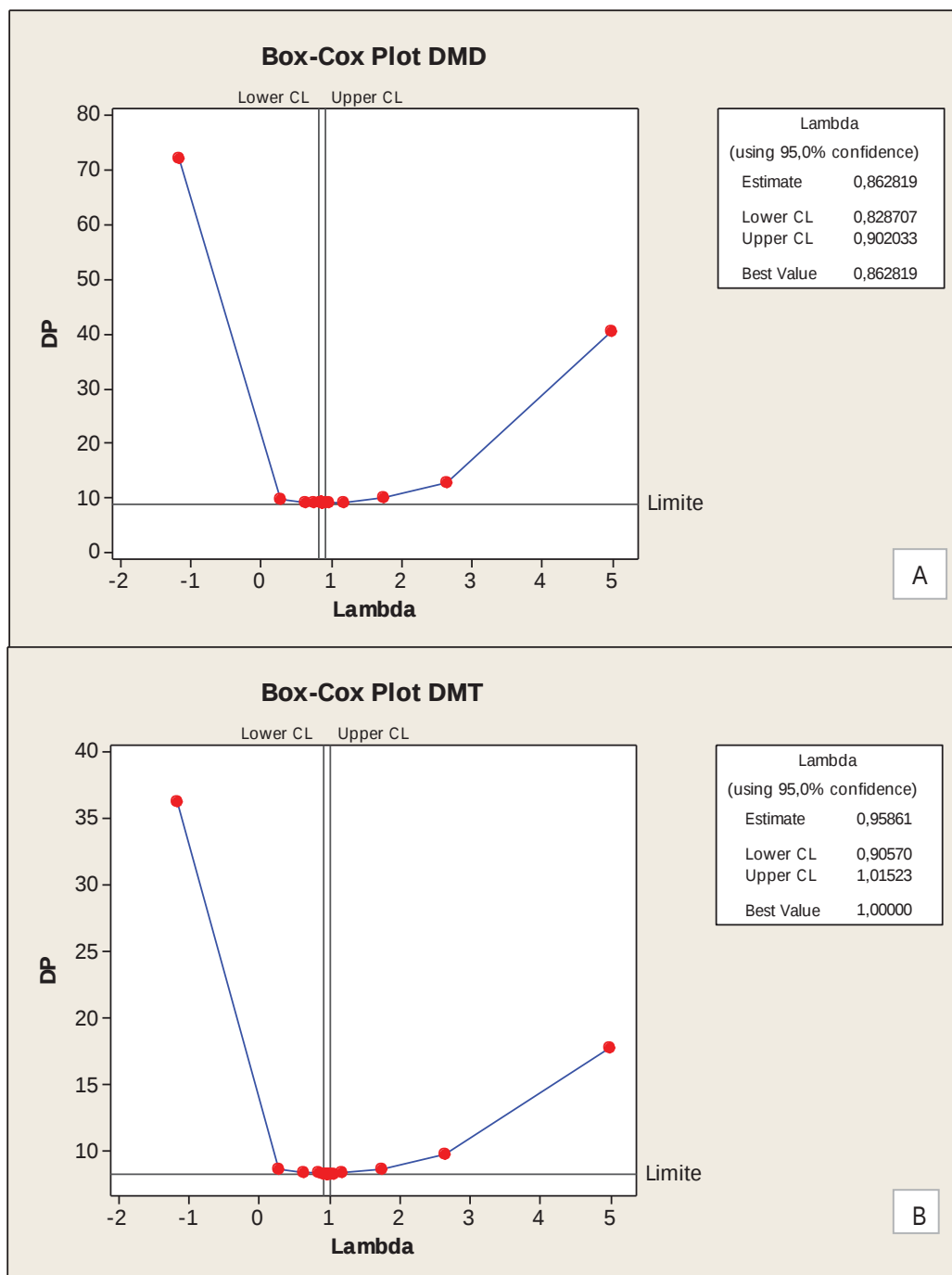


Figura 1. Estimativa do fator de correção ótimo de lambda para as características DMD e DMT.