UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS

GERSON ANTÔNIO DE OLIVEIRA JÚNIOR

**Associação genômica ampla (GWAS) aplicada a
características reprodutivas de novilhas da raça Nelore**

Pirassununga
2017

GERSON ANTÔNIO DE OLIVEIRA JÚNIOR

**Associação genômica ampla (GWAS) aplicada a
características reprodutivas de novilhas da raça Nelore**

Tese apresentada à Faculdade de Zootecnia e Engenharia
de Alimentos da Universidade de São Paulo, como parte
dos requisitos para obtenção do título de Doutor em
Ciências do programa de pós-graduação em Biociência
Animal

Área de concentração: Biociência Animal

Orientador: Dr. José Bento Sterman Ferraz

Coorientador: Dr. Marcos Vinícius Gualberto Barbosa da Silva

Pirassununga
2017

Ficha catalográfica elaborada pelo
Serviço de Biblioteca e Informação, FZEA/USP,
com os dados fornecidos pelo(a) autor(a)

dG382a de Oliveira Júnior, Gerson Antônio
Associação genômica ampla (GWAS) aplicada a
características reprodutivas de novilhas da raça
Nelore / Gerson Antônio de Oliveira Júnior ;
orientador José Bento Sterman Ferraz ; coorientador
Marcos Vinícius Gualberto Barbosa da Silva. --
Pirassununga, 2017.
81 f.

Tese (Doutorado - Programa de Pós-Graduação em
Biociência Animal) -- Faculdade de Zootecnia e
Engenharia de Alimentos, Universidade de São Paulo.

1. Melhoramento Genético Animal. 2. Genômica . 3.
Marcador Molecular. 4. Nelore. 5. Reprodução. I.
Sterman Ferraz, José Bento, orient. II. Gualberto
Barbosa da Silva, Marcos Vinícius, coorient. III.
Título.

FOLHA DE AVALIAÇÃO

Nome: OLIVEIRA JÚNIOR, Gerson Antônio de

Título: Associação genômica ampla (GWAS) aplicada a características reprodutivas de novilhas da raça Nelore

Tese apresentada ao Programa de Pós-Graduação em Biociência Animal da Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo para a obtenção do Título de Doutor em Ciências.

Data: ____/____/____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Prof. Dr. _____

Instituição: _____ Julgamento: _____

Dedico este trabalho a minha família e a todos que participaram de forma positiva dessa minha jornada.

Sem vocês não seria possível.

Agradecimentos

Primeiramente agradeço a Deus por ter me dado a oportunidade de ter a família que tenho. Minha mãe Maria Zélia Borges de Oliveira, meu pai Gerson Antônio de Oliveira Baesse e meu irmão Régis Borges de Oliveira. Sem eles nada disso faria sentido em minha vida. Sei o quanto se esforçaram e se abdicaram para que pudesse estudar e chegar ao título de Doutor. Serei eternamente grato.

À todos que me deram força e incentivos nos momentos difíceis e também celebraram os momentos alegres. Tive muita sorte de estar perto de pessoas ótimas que foram fundamentais em meu crescimento pessoal e profissional.

Minha companheira, Aline Fernanda de Souza, que sempre esteve ao meu lado. Sem ela tudo seria bem mais difícil.

Ao Dr. José Bento Sterman Ferraz, que aceitou ser meu orientador e me deu a oportunidade de trabalhar nesse projeto, colaborando muito com meu crescimento pessoal e profissional.

Ao Dr. Dorian J. Garrick e todos os membros da *Iowa State University* pela oportunidade de trabalhar com eles durante meu estágio BEPE. Sem dúvida é uma das pessoas mais fantásticas que já conheci.

Ao Dr. Haja N. Kadarmideen e todo seu grupo de pesquisa da *University of Copenhagen* por todo suporte acadêmico que recebi e pela parceria no projeto *Genomic Improvement of Fertilization Traits* (GIFT).

Meu grande amigo Dr. Ricardo Ventura, que sempre foi aberto a discussões e valiosos conselhos. Obrigado pela paciência e dedicação que teve comigo.

Ao Dr. Joanir Pereira Eler pela amizade e por todas discussões acadêmicas.

Ao meu coorientador Dr. Marcos Vinicius G. B Silva pela oportunidades oferecidas, conselhos e discussões e com certeza colaboraram com meu aprendizado.

Aos professores Dr. César Gonçalves de Lima e Dra. Rachel Santos Bueno Carvalho por todos os seus ensinamentos.

À Elisângela Chicaroni Mattos de Oliveira que sempre foi muito paciente e prestativa.

À Vera Lucia (Dona Vera) por sua amizade e seus deliciosos petiscos.

À todos os alunos do Grupo de Melhoramento Genético (Barbara, Bruna, Bruno, Gerardo, Lenin, Lucas, Isabela, Laís, Miguel, Minos, Pâmela, Yuri) que sempre estiveram comigo em todos os momentos.

À Fundação de Amparo a Pesquisa (FAPESP), pela concessão da bolsa de estudo no país e de estágio no exterior e por todo recurso disponibilizado na minha formação (Processos: 2013/12097-9 / 2014/26354-6).

À Faculdade de Zootecnia e Engenharia de Alimentos, pela oportunidade de realização do curso de doutorado.

Quem se arrisca a andar por ares nunca antes respirados ou pensar fora da curva tem grandes chances de encontrar pedras no caminho. No entanto, ninguém é digno de contribuir para a ciência se não usar suas dores e insônias nesse processo. Não há céu sem tempestade. Risos e lágrimas, sucessos e fracassos, aplausos e vaias fazem parte do currículo de cada ser humano, em especial daqueles que são apaixonados por produzir novas ideias.

(autor desconhecido)

A formalização do fim é apenas a reinvenção de um novo começo.

Resumo

OLIVEIRA JR, G. A. Associação genômica ampla (GWAS) aplicada a características reprodutivas de novilhas da raça Nelore. 2017. 81 f. Tese (Doutorado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2017.

Considerando a importância econômica da precocidade sexual no sistema de produção de bovinos de corte, o objetivo deste estudo foi identificar regiões genômicas associadas a características reprodutivas tais como a prenhez precoce e número de folículos antrais em novilhas da raça Nelore, identificando processos biológicos envolvidos para um melhor entendimento da variabilidade fenotípica destas características. O banco de dados foi composto por 2.283 novilhas, filhas de 78 touros, provenientes de três fazendas comerciais do Brasil. Destas, 1.267 foram genotipadas com aproximadamente 74 K SNPs. Após análises de controle de qualidade, permaneceram informações de 1.255 animais e 64.800 marcadores. A informação genotípica de 42 touros também estava disponível e touros não genotipados com mais de 5 progênies genotipadas foram imputados para a mesma densidade de marcador das novilhas. Análises de associação genômica ampla (GWAS) foram utilizadas considerando uma metodologia Bayesiana (BayesB) para identificar potenciais janelas genômicas (~ 1 Mb) responsáveis pela explicação de $\geq 1\%$ da variância genética das características. Essas regiões foram utilizadas para a pesquisa de genes bem como nas análises de enriquecimento utilizando os termos MeSH. Ainda, blocos haplotípicos dessas regiões genômicas foram ajustados como efeitos classificatórios no modelo BayesB, com o objetivo de estimar a influência de cada alelo dos touros heterozigotos sobre os fenótipos. Os coeficientes de herdabilidade preditos foram $0,28 \pm 0,07$ e $0,49 \pm 0,09$ para PP e NF, respectivamente, com correlação genética de $-0,21 \pm 0,29$. Os resultados do GWAS apontaram janelas genômicas significativas nos cromossomos 5, 14 e 18 para PP e 2, 8, 11, 14, 15, 16 e 22 para NF. As análises de enriquecimento revelaram termos significativos ($p < 0,05$) associados com PP: *Munc-18 Proteins*, *Fucose* e *Hemoglobins*, e com NF: *Cathepsin B*, *Receptors-Neuropeptide* e *Palmitic Acid*. Um total de 15 alelos haplotípicos tiveram efeitos diferentes ($p < 0,10$) de sua cópia alternativa para PP e 16 para NF. As análises genômicas contribuíram para uma melhor compreensão da estrutura genética das características reprodutivas PP e NF, podendo colaborar com estratégias de seleção para melhoria dos índices reprodutivos de bovinos da raça Nelore.

Palavras-chave: genômica. Prenhez. Precocidade sexual. Bovinos de corte.

Abstract

OLIVEIRA JR, G. A. Genome-wide association study (GWAS) for reproductive traits in Nellore heifers. 2017. 81 f. Thesis (Doctoral) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2017.

Considering the economic importance of sexual precocity in beef cattle production system, the objective of this study was to explore genomic regions associated with heifer pregnancy (HP) and total number of antral follicle (NF) in a population of Nellore (*Bos indicus*) heifers. The dataset was composed by 2,283 animals, which were daughters of 78 sires, coming from three different commercial farms in Brazil. From these, 1,267 were genotyped with around 74 K SNP markers. After quality control analyses, the remained dataset was formed by 1,255 animals and 64,800 markers. Genotypic information of 42 sires was also available, and non-genotyped sires with more than five genotyped progenies in the dataset were imputed for the same marker density of the heifers. A Genome-wide Association Study (GWAS) was performed under a BayesB method to identify potential genomic windows (~1 Mb) that explained at least 1% of the total additive genetic variance of the traits. The pointed genomic regions were used in MeSH enrichment analyses as well as for genes searching. Haplotype blocks that span these potential genomic regions were fitted as classificatory effects in the BayesB model, aiming to estimate the effect of each allele of the heterozygous sires for these regions. The estimated heritabilities coefficients were 0.28 ± 0.07 and 0.49 ± 0.09 for HP and NF, respectively, with genomic correlation of -0.21 ± 0.29 . The GWAS results pointed to significant genomic windows on chromosomes 5, 14 and 18 for HP and 2, 8, 11, 14, 15, 16 and 22 for NF. The MeSH enrichment analyses revealed significant ($p < 0.05$) terms associated with HP: *Munc-18 Proteins*, *Fucose* and *Hemoglobins*, and with NF: *Cathepsin B*, *Receptors-Neuropeptide* and *Palmitic Acid*. The pointed regions by GWAS harbored important genes that can help explain the genetic basis of the considered traits. A total of 15 haplotype alleles obtained better results ($p < 0.10$) than their alternative copy for HP and 16 for NF. The genomic analyses contributed to a better understanding of the genetic control of the reproductive traits HP and NF and can be useful to selection strategies in improving reproductive indices in Nellore cattle.

Keywords: Genomics. Pregnancy. Sexual precocity. Beef cattle.

Lista de Tabelas

Tabela 1 – Valores médios ($\pm$ desvio padrão) para idade (meses) e peso (Kg) mensurados no início da estação de monta, taxa de prenhez (PP, %) e número de folículos antrais (NF) das novilhas genotipadas (após CQ) por fazenda.....	29
Tabela 2 – Estatística descritiva dos coeficientes de herdabilidade (h^2) e correlação genética (r) para prenhez precoce (PP) e o número de folículos antrais (NF) das novilhas Nelore.....	35
Tabela 3 – Descrição das janelas genômicas que explicaram pelos menos 0,5% da variância genética total para a característica de prenhez precoce em novilhas da raça Nelore.	39
Tabela 4 – Descrição das janelas genômicas que explicaram pelos menos 0,5% da variância genética total para a característica de número de folículos antrais em novilhas da raça Nelore.	40
Tabela 5 – Termos MeSH relacionados com prenhez precoce em novilhas Nelore com seus respectivos <i>p-values</i>	42
Tabela 6 – Termos MeSH relacionados com número de folículos antrais em novilhas Nelore com seus respectivos <i>p-values</i>	44
Tabela 7 – Blocos Haplotípicos considerando as regiões do estudo de GWAS que explicaram mais de 1% da variância genética total da característica de prenhez precoce em novilhas da raça Nelore.	45
Tabela 8 – Blocos Haplotípicos considerando as regiões do estudo de GWAS que explicaram mais de 1% da variância genética total da característica número de folículos antrais.....	48

Lista de Figuras

Figura 1 – Protocolo de inseminação artificial em tempo fixo (IATF) usado para as novilhas Nelore. Dia 0: Implante de orelha a base de Norgestomet, plus 2,0 mg de benzoato de estradiol (EB). Dia 8: implante de estradiol cipionato (EC), 0,265 mg de cloprostenol sódico (PGF) e 300 IU de gonadotrofina coriônica equina (eCG). Dia 10: Inseminação em tempo fixo (IATF).....	28
Figura 2 – Valores médios ($\pm$ desvio padrão) de desequilíbrio de ligação (r^2) entre marcadores SNP adjacentes por cromossomos na população de novilhas Nelore estudada.....	36
Figura 3 – Média ($\pm$ desvio padrão) do desequilíbrio de ligação (r^2) entre marcadores SNP por cromossomos na população de novilhas Nelore estudada.	36
Figura 4 – Comportamento do desequilíbrio de ligação médio (r^2) entre pares de marcadores em função da distância.....	37
Figura 5 – Manhattan <i>plot</i> do estudo de associação genômico amplo (GWAS) para a característica prenhez precoce em novilhas da raça Nelore, considerando limite de significância (linha vermelha) em 1% da proporção da variância genética.	38
Figura 6 – Manhattan <i>plot</i> do estudo de associação genômico amplo (GWAS) para a característica de número total de folículos antrais em novilhas Nelore, considerando limite de significância (linha vermelha) em 1% da proporção da variância genética.	38
Figura 7 – Distribuição da diferença dos efeitos estimados <i>a posteriori</i> entre dois alelos haplotípicos do cromossomo 5 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de prenhez de novilhas.....	46
Figura 8 – Distribuição da diferença dos efeitos estimados <i>a posteriori</i> entre dois alelos haplotípicos do cromossomo 14 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de prenhez de novilhas.....	47
Figura 9 – Distribuição da diferença dos efeitos estimados <i>a posteriori</i> entre dois alelos haplotípicos do cromossomo 18 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de prenhez de novilhas.....	47
Figura 10 – Distribuição da diferença dos efeitos estimados <i>a posteriori</i> entre dois alelos haplotípicos do cromossomo 22 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de número de folículos antrais.	49
Figura 11 – Distribuição da diferença dos efeitos estimados <i>a posteriori</i> entre dois alelos haplotípicos do cromossomo 11 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de número de folículos antrais.	49
Figura 12 – Distribuição da diferença dos efeitos estimados <i>a posteriori</i> entre dois alelos haplotípicos do cromossomo 8 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de número de folículos antrais.	50

Figura 13 – Cluster considerando os 2 primeiros autovetores que mais explicaram a variância dos valores da matriz G (*Eigenvector*) dos touros que apresentaram blocos haplotípicos com efeito significativo ($p < 0,10$) para prenhez precoce (PP), número de folículos antrais (NF) ou ambas características.....64

Sumário

1. Introdução	14
2. Revisão de Literatura	16
2.1. Características Reprodutivas	16
2.2 Estudo de Associação Genômica Ampla	17
2.2.1 Controle de Qualidade	19
2.3 Desequilíbrio de Ligação (LD)	20
2.4 Seleção Genômica	21
2.5 Imputação	22
2.6 Aplicação da genômica na reprodução.....	23
3. Hipótese	26
4. Material e Métodos	27
4.1 Dados Fenotípicos	27
4.1.1 Bando de Dados	27
4.1.2 Características.....	27
4.2 Dados genotípicos	28
4.3 Estimação dos Componentes de variância	29
4.4 Estudo de Associação Genômica Ampla (GWAS)	30
4.5 Busca gênica e enriquecimento funcional.....	32
4.6 Imputação de touros não-genotipados.....	32
4.7 Análise de Haplótipos	33
4.8 Sequenciamento	34
5. Resultados	35
5.1 Dados Fenotípicos	35
5.2 Estudo de Associação Genômica Ampla	35
5.3 Análises de enriquecimento funcional utilizando os termos MeSH.....	42
5.4 Imputação dos animais não-genotipados.....	45
5.5 Estudo dos Haplótipos.....	45
6. Discussão	51
6.1 Estudos de Associação Genômica Ampla.....	52
6.2 Genes Associados.....	53
6.2.1 Prenhez em Novilhas	53
6.2.2 Número de Folículos Antrais.....	57
6.3 Análises de enriquecimento funcional utilizando os termos MeSH.....	61
6.4 Imputação de touros não genotipados e teste de parentesco	62
6.5 Estudo de Haplótipos	63
7. Conclusão.....	65
8. Referências Bibliográfica.....	66
APÊNDICE A.....	80

1. Introdução

Nas próximas décadas, a demanda por proteína para consumo humano aumentará significativamente, enquanto que a disponibilidade de terras produtivas é provável que diminua ou, otimistamente, continue a mesma. Portanto, uma melhor compreensão das características de interesse econômico ligadas à produção animal seria essencial para atingir a taxa exigida de expansão da produtividade (SIMIANER, 2016).

A precocidade sexual é uma importante característica, determinando a eficiência da cadeia de produção de carne, especialmente para animais zebuínos (*Bos indicus*), os quais podem demandar mais de três anos para atingir a puberdade (ABEYGUNAWARDENA; DEMATAWEWA, 2004). Aumentar a taxa de concepção de novilhas jovens resulta em diluição dos custos de reposição do rebanho, maior oferta de animais para venda ou reprodução, maior produtividade acumulada por animal e, conseqüentemente, maior rentabilidade para os produtores (AKANNO et al., 2015; ELER et al., 2014; GUARINI et al., 2014).

A fertilidade nas fêmeas pode ser mensurada por diferentes maneiras, incluindo a determinação de prenhez em novilhas expostas a reprodução. Essa característica tem como vantagem a mensuração direta nos animais, o qual é um processo simples e de baixo custo, permitindo sua utilização em larga escala. Alguns estudos em raças taurinas (*Bos taurus*) (AKANNO et al., 2015; MOKHTARI et al., 2015), bem como em zebus e bovinos compostos (PORTO-NETO et al., 2015; SANTANA JR et al., 2015), têm usado essa característica como medida de precocidade sexual, uma vez que, animais que não atingem a puberdade na época desejável, iniciam sua vida reprodutiva tardiamente, ocasionando aumento do custo de produção aos produtores.

Os avanços nas tecnologias genômicas permitem hoje a genotipagem de centenas de milhares de marcadores de DNA espalhados por todo o genoma, em um processo rápido e acessível (KHATKAR; RANDHAWA; RAADSMA, 2014). A inclusão dessa informação nas avaliações genéticas permite melhores predições e decisões de seleção, principalmente em características poligênicas, tais como as reprodutivas, reduzindo o intervalo entre gerações, melhorando os ganhos genéticos e afetando positivamente a taxa de ganho genético da população (BERRY; KEARNEY, 2011; MEUWISSEN; HAYES; GODDARD, 2001).

Marcadores genômicos, tais como o polimorfismo de um único nucleotídeo (SNP), podem ser usados em estudos de associação genômica ampla (GWAS), a qual é uma estratégia já bem estabelecida para a identificação de regiões genômica relacionadas com características de interesse em diferentes espécies (HU; PARK; REECY, 2016). As análises de GWAS têm como

pressuposição o desequilíbrio de ligação (LD) entre pelo menos um marcador e uma mutação de efeito maior ou locus de características quantitativas (QTL), responsáveis pela variação fenotípica observada. Blocos genômicos com baixa taxa de recombinação são conhecidos por haplótipos, os quais podem apresentar maior nível de LD com regiões QTL quando comparados com marcadores pontuais, indicando alelos com importantes efeitos no desempenho dos animais.

Considerando que tecnologias reprodutivas, tais como a aspiração folicular, fertilização in vitro (FIV) e a transferência de embriões, são massivamente adotadas em rebanhos zebuínos, o número de folículos antrais (NF) tornou-se uma relevante característica, definindo o desempenho dos animais para essas tecnologias (BARUSELLI et al., 2015). Essas tecnologias podem auxiliar os programas de melhoramento animal ao permitir o maior uso de fêmeas de alto valor genético e, conseqüentemente, um aumento nos ganhos genéticos da população. A característica de NF já foi associada com diferentes características reprodutivas, tais como funcionalidade ovariana (IRELAND et al., 2009), resposta a superovulação (SINGH et al., 2004), longevidade (JIMENEZ-KRASSEL et al., 2015) e fertilidade (MOSSA et al., 2012). A circulação do hormônio anti-Mülleriano foi positivamente associada com NF em camundongos (DURLINGER; VISSER; THEMME, 2002) e mulheres (FANCHIN et al., 2003). Contudo, estudos mostraram que animais zebuínos produzem mais oócitos, bem como mais blastocistos que os animais taurinos, apesar dos primeiros apresentarem piores índices reprodutivos (GIMENES et al., 2015; SALES et al., 2015).

De acordo com Kappes et al. (2000), fatores genéticos que afetam a reprodução ainda não estão bem elucidados e a compreensão mais detalhada desses fatores poderia levar a uma melhoria substancial na seleção genética dessas características. Com isso, a identificação de regiões genômicas associadas a essas características poderia ampliar a compreensão dos efeitos genéticos relacionados na expressão fenotípica, servindo como base para estudos de expressão genica e colaborando com a eficiência reprodutiva em bovinos.

O objetivo do trabalho foi detectar regiões genômicas associadas com as características reprodutivas de prenhez precoce e número de folículos antrais em novilhas da raça Nelore, bem como a identificação de genes candidatos envolvidos em processos biológicos associados à expressão desses fenótipos. Ainda, identificar touros heterozigotos para essas regiões genômicas foram avaliados com o intuito de eleger possíveis alelos haplotípicos com efeito significativo sobre as características avaliadas.

2. Revisão de Literatura

2.1. Características Reprodutivas

Características reprodutivas são de extrema importância dentro do sistema de produção agropecuário por influenciar a idade ao primeiro parto, o custo de manutenção de fêmeas vazias e a quantidade de animais que serão destinados ao mercado (DAY; NOGUEIRA, 2013; ELER et al., 2002). Ainda, fêmeas que não pariram em uma dada estação de monta são, normalmente, eliminadas do rebanho, gerando custos de reposição desses animais (ELER et al., 2008). De Vries (2006) estimou um prejuízo de aproximadamente US\$550 a cada prenhez perdida em animais da raça Holandesa, atribuindo um ganho de US\$278 para cada nova prenhez no rebanho.

A prenhez em novilhas está relacionada com a maturidade sexual dos animais. Quando observada como característica binária, atribui-se 1 para animais prenhes e 0 para vazios, podendo ser definida como a probabilidade de prenhez aos 14 meses de idade (PP14) de novilhas exposto à estação de monta (ELER et al., 2002). Neste caso, para implementação do processo de seleção, os animais entram em estação de monta entre 12 a 15 meses de idade e, consequentemente, tendo o primeiro parto aproximadamente aos 24 meses de idade.

A eficiência reprodutiva está relacionada com o início da vida produtiva das fêmeas, impactando o intervalo entre gerações, a intensidade de seleção e, consequentemente, a taxa de ganho genético do rebanho (FORTES et al., 2012a; VAN MELIS et al., 2010). A seleção de novilhas sexualmente mais precoces é uma prática adotada por alguns programas de melhoramento genético animal. Porém, os índices de seleção, quase que exclusivamente, consideram apenas características de crescimento. Segundo Cammack, Thomas e Enns (2009), o progresso genético das características reprodutivas foi modesto usando os métodos quantitativos convencionais, por se tratar de características mensuradas em apenas um sexo e por, em grande parte, serem selecionadas apenas de forma indireta.

Day e Nogueira (2013) destacam a existência de variação genética para início da puberdade de novilhas Nelore, oferecendo uma oportunidade de seleção genética. Os valores dos coeficientes de herdabilidade preditos para características relacionadas com puberdade na literatura são variados, influenciados pela população, raça, efeitos ambientais e número amostral. A revisão feita por Cammack, Thomas e Enns (2009) demonstrou valores variando de 0,05 a 0,60 para diferentes características reprodutivas em fêmeas. Trabalhos considerando a característica de PP14 em Nelore relataram coeficientes de herdabilidade próximos a 0,50 (ELER et al., 2002; SANTANA JR et al., 2013; SHIOTSUKI et al., 2009).

Existem também medidas indiretas para a precocidade sexual em novilhas, com destaque para o perímetro escrotal (PE), o qual vem sendo amplamente utilizado como indicador de características reprodutivas nas fêmeas (SANTANA JR. et al., 2011). Os autores reportaram correlação genética de -0,30 entre PP14 e PE, sendo ligeiramente inferior às descritas por Eler et al. (2006) e Van Melis et al. (2010), que estimaram correlação de -0,27 e -0,29, respectivamente. Ambos concluem que dada à baixa correlação genética observada entre PE e PP14, seria mais efetivo selecionar a característica na própria fêmea, no caso a probabilidade de prenhez de novilhas.

2.2 Estudo de Associação Genômica Ampla

O genoma bovino (*Bos taurus*) é composto por mais de 2670 milhões de pares de base (UMD 3.1.1), sendo aproximadamente 94% distribuídos nos 29 cromossomos autossômicos, 5% no cromossomo sexual X e o restante no cromossomo Y e DNA mitocondrial (GEER et al., 2009). Mutações genômicas podem ser introduzidas a cada nova geração, sendo o tipo mais abundante no genoma os polimorfismos de um único nucleotídeo (SNP¹). Sua frequência pode ser afetada por pressão seletiva, e também induzidos por recombinações homólogas e fatores exógenos (MOUNTAIN et al., 1992). Estas se baseiam na troca de apenas uma das bases nitrogenadas (A, C, T, G) da cadeia da molécula de DNA, como ilustrado a seguir:

Indivíduo 1: CAT**C**CAGC

Indivíduo 2: CAT**T**CAGC

Os SNPs são em sua maioria dialélicos, sendo que para uma variação ser considerada um marcador molecular do tipo SNP é necessário que o alelo de menor frequência (MAF²) esteja presente em pelo menos 1% da população avaliada (BROOKES, 1999). Atualmente estão descritos aproximadamente 100 milhões deles no genoma bovino (GEER et al., 2009), os quais têm sido estudados com mais frequência em relação a outros marcadores genéticos devido a baixa taxa de mutação e a facilidade de genotipagem. Acredita-se que aproximadamente 85% dos SNPs são comuns dentro de uma mesma espécie e 15% específicos para cada população, demonstrando a estabilidade dessas variações genômicas e sugerindo que marcadores significativos para uma raça ou população podem não serem em outras (HAWKEN et al., 2012).

A partir do início do século XXI, os avanços biotecnológicos permitiram a genotipagem de centenas de milhares de SNPs espalhados por todo genoma, possibilitando estudos tais como

¹ Single-Nucleotide Polymorphism

² Minor allele frequency

os Estudos de Associação Genômica Ampla (GWAS³) e a Seleção Genômica (GS⁴) a serem realizados em diferentes espécies animais e vegetais (HABIER; FERNANDO; DEKKERS, 2007).

Os estudos de GWAS tem caráter inovador, permitindo a identificação de associações entre milhares de loci genômicos e características complexas, aumentando o entendimento sobre a base genética destas. A metodologia foi inicialmente desenvolvida para estudos epidemiológicos em humanos (MCCARTHY et al., 2008), entretanto vem sendo aplicada em diferentes espécies animais e vegetais (BARBAN et al., 2016; SANTANA et al., 2014; WOLFE et al., 2016). Os estudos buscam associar regiões do genoma com os fenótipos de interesse, a fim de identificar possíveis regiões de maior efeito sobre a característica e posterior investigação de funções biológicas, visando aumentar a compreensão da influência genética sobre a expressão fenotípica (YANG et al., 2013).

Uma das vantagens do GWAS seria a não exigência de conhecimento prévio profundo dos processos fisiológicos e bioquímicos envolvidos na manifestação dos fenótipos estudados (TAYLOR et al., 1998; WANG et al., 2005). Como desafios envolvendo esses estudos estariam a adequação de modelos e metodologias estatísticas para a estimação de um grande número de efeitos a partir de um limitado número de observações, o tamanho amostral, a colinearidade advinda do desequilíbrio de ligação entre os marcadores e a interpretação dos resultados (CANTOR; LANGE; SINSHEIMER, 2010; PEARSON; MANOLIO, 2008; RESENDE et al., 2008). O tamanho amostral deveria ser suficiente para identificar pequenas associações genômicas, sendo que marcadores que explicam 1% da variação fenotípica já podem ser importantes no entendimento biológico da característica (CANTOR; LANGE; SINSHEIMER, 2010; GOLDSTEIN, 2009).

As metodologias Bayesianas aplicadas nos estudos genômicos apresentam vantagens em relação a métodos clássicos, predizendo os efeitos de todos os loci de forma simultânea (ex. BayesA), assumindo variâncias específicas para cada locus (ex. BayesA, BayesB), ou seja, assumem que os marcadores podem explicar diferentes proporções da variação genética da característica analisada. Os efeitos preditos dos marcadores são então o efeito de um determinado locus ajustado para todos os outros loci incluídos no modelo (GARRICK; FERNANDO, 2013).

³ Genome Wide Association Study

⁴ Genomic Selection

A metodologia Bayesiana combina informações *a priori* dos parâmetros a serem preditos com as informações do banco de dados disponível, gerando uma distribuição *a posteriori* para o cada parâmetro do modelo, a qual é usada nas inferências (SORENSEN; GIANOLA, 2002). É desejável que as distribuições *a posteriori* sofram mínima influência das informações *a priori*, sendo para isto necessário um banco de dados informativo (GARRICK; FERNANDO, 2013).

Diferentes sequências de uma mesma região genômica podem ser mais frequentes em indivíduos com determinado padrão fenotípico, sugerindo possível associação estatística entre a característica analisada e o locus (QTL⁵) (RESENDE et al., 2012). Gabriel et al. (2002) demonstrou que, com exceção a determinadas regiões com alta taxa de recombinação (*hotspots*), o genoma é composto por regiões altamente correlacionadas de SNPs unidos pelo princípio do desequilíbrio de ligação. Com isto formam-se blocos genômicos, denominados de haplótipos, que segregam juntos na população. Uma das pressuposições dos estudos de GWAS é que pelo menos um marcador do tipo SNP esteja em LD com os genes ou regiões genômicas relacionados com a característica de interesse (CANTOR; LANGE; SINSHEIMER, 2010).

2.2.1 Controle de Qualidade

Os processos de controle de qualidade (CQ) dos SNPs são importantes para diminuir a frequência de falsos positivos (erro tipo I) e, consequentemente, associações espúrias (BOLORMAA et al., 2011; CHAN; HAWKEN; REVERTER, 2009), sendo precedentes às análises de associação. Os critérios utilizados no CQ são específicos para cada estudo, não havendo um padrão rígido ou protocolo a ser seguido, com limiares flexíveis dependendo dos objetivos dos pesquisadores (FREIMER; SABATTI, 2005). O que se busca é sempre a redução de falsos positivos sem comprometer o poder estatístico.

Os principais critérios considerados no CQ são:

- *GenCall Score* (GCScore)
- *Call rate* amostral
- *Call rate* por SNP
- Menor frequência alélica (MAF)
- Equilíbrio de Hardy-Weinberg
- Remoção de amostras duplicadas

⁵ Quantitative Trait Locus

2.3 Desequilíbrio de Ligação (LD⁶)

O LD pode então ser definido como associação não aleatória entre alelos presentes em diferentes loci do genoma. O LD pode ser influenciado por vários fatores tais como seleção, taxa de recombinação e mutação, endogamia e distância entre os loci, ou seja, quanto mais perto o marcador molecular estiver do QTL, maior a ligação entre eles.

Pode-se estimar LD a partir da frequência alélica, dado por:

$$D = \text{prob}(AB)\text{prob}(ab) - \text{prob}(Ab)\text{prob}(aB)$$

considerando um locus com alelos A e a e outro com alelos B e b . Esta estatística, proposta por Hill (1981) é dependente das frequências dos alelos individuais, não sendo aplicável para em comparações de LD em múltiplos loci. Lewontin (1964) propôs um D' que seria o próprio D padronizado pelo D máximo. Porém, também é de pouca precisão quando estimada a partir de pequenas populações ou com alelos de baixa frequência (MCRAE et al., 2002).

A estatística r^2 (HILL; ROBERTSON, 1968), se tornou a mais popular em estudo de LD, sendo estimativa por:

$$r^2 = \frac{D^2}{[\text{prob}(A)\text{prob}(a)\text{prob}(B)\text{prob}(b)]}$$

Os valores de r^2 variam de zero a um, referindo-se à ausência ou completo LD, respectivamente, sendo valores acima de 0,2 considerados como úteis para estudos genômicos (HAYES et al., 2007). A interpretação genética desse parâmetro seria medir a influência dos marcadores na expressão de QTLs, sendo um indicativo do número de marcadores e fenótipos necessários para estudos genômicos (RESENDE et al., 2012).

A combinação de alelos de loci adjacentes de um cromossomo é denominada por haplótipos. São regiões genômicas normalmente com baixa taxa de mutação, sendo compartilhadas entre parentes de gerações próximas. Mutações pontuais que influenciam a variabilidade de determinado fenótipo podem não ser diretamente detectadas nos estudos de GWAS, mas graças a conservação dos blocos haplotípicos, a região genômica pode ser evidenciada por associação com SNPs em desequilíbrio de ligação (LD) dentro do haplótipo (SCHERER, 2016).

⁶ Linkage disequilibrium

2.4 Seleção Genômica

O principal ponto na melhoria nos índices de produção é a obtenção de valores genéticos acurados. O melhoramento genético tradicional possivelmente continuará sendo crucial na melhoria genética em características reprodutivas e produtivas dos animais domésticos (KADARMIDEEN, 2010). Porém, a inclusão de informações moleculares tem mostrado importante em programas de melhoramento uma vez que complementam os bancos de dados das análises genéticas, ajudando nas decisões de seleção, principalmente de animais jovens (DEKKERS, 2004; DJARI et al., 2013).

Polimorfismos do tipo SNP são considerados fontes de variação de mérito genético, portanto marcadores em LD com algum QTL podem ser utilizados como critério extra para identificação de indivíduos candidatos à seleção, o que aumentaria a acurácia na avaliação genética (RESENDE et al., 2012). As predições baseadas em informações de pedigree usam o parentesco assumido entre os animais, não considerando similaridade entre indivíduos teoricamente não relacionados. A informação genômica corrige esse viés adicionando a informação real de parentesco entre os animais, fazendo com que a matriz de parentesco tradicional (A) seja substituída pela matriz de parentesco genômico (G) (DAETWYLER et al., 2013; VANRADEN, 2008).

Os marcadores moleculares já são comercialmente empregados na GS (GODDARD; HAYES, 2007; MEUWISSEN; HAYES; GODDARD, 2001) em diferentes espécies tais como suínos, aves, ovinos e bovinos (DEKKERS, 2004). Comparado com a seleção tradicional, a inclusão da informação genômica resulta em consideráveis benefícios na resposta a seleção de animais jovens, que ainda não tiveram seus fenótipos coletados, com potencial redução dos custos em até 90% dos programas de melhoramento, reduzindo o intervalo entre gerações, antecipando decisões de seleção e melhor controle do nível endogâmico do rebanho (MEUWISSEN; HAYES; GODDARD, 2001; PRYCE; HAYES; GODDARD, 2012; SCHAEFFER, 2006). García-Ruiz et al. (2016) demonstraram a importância da GS na raça Holandesa nos Estados Unidos, sendo sua importância comparada com a da inseminação artificial na época de sua introdução.

Os valores genômicos são obtidos a partir da soma dos efeitos estimados para cada SNP, ponderando a presença ou não do alelo do marcador. Esse efeito é estimado a partir de uma população em que se tem informações fenotípicas e genotípicas, conhecida por população de treinamento. Os valores estimados formam, então, uma equação de predição, a qual é

posteriormente aplicada em uma população de validação, formada por animais jovens previamente genotipados.

Para verificar a acurácia de predição dos valores genômicos, métodos de validação estão sendo empregados. A metodologia de validação cruzada pode ser realizada de diferentes maneiras, considerando cenários em que todos os indivíduos fazem parte da população de treinamento pelo menos uma vez ou, em alguns casos, em que alguns indivíduos não façam parte da população de referência (DAETWYLER et al., 2013). O trabalho de Saatchi et al. (2011) propôs uma abordagem para validação cruzada que usa cluster *k-means*, sendo este baseado no relacionamento genômico entre os animais para particionar a população em blocos (k), minimizando o relacionamento entre as populações treinamento e validação.

2.5 Imputação

A imputação genômica é uma metodologia estatística que permite a predição do genótipo de loci não genotipados a partir do LD entre marcadores, podendo também utilizar as informações de parentesco entre animais (VENTURA et al., 2014). A técnica tem se tornado comum nos programas de melhoramento de raças leiteiras (WIGGANS et al., 2012), com estudos em diferentes espécies tais como suínos (CLEVELAND; HICKEY, 2013; XIANG et al., 2015), aves (FULTON, 2012; HEIDARITABAR et al., 2015), ovinos (BOLORMAA et al., 2015; VENTURA et al., 2016) e bovinos (BOISON et al., 2015; CARVALHEIRO et al., 2014).

Em estudos de imputação genômica normalmente se usa parte da população genotipada na densidade de SNPs desejável como população de validação (PV). Nesta, os loci que estão ausentes no chip de menor densidade são mascarados, fazendo com que esses animais tenham a mesma quantidade de marcadores que os disponíveis no chip de menor densidade. Os demais animais que permaneceram genotipados com maior densidade de SNPs formam então a população de referência (PR). Os marcadores mascarados na PV serão então preditos, usando as informações da PR (CALUS et al., 2014). A acurácia de imputação é estimada comparando os SNPs imputados com os verdadeiramente genotipados.

No processo de imputação, os haplótipos da PR são distintos em paterno e materno pelo processo conhecido como faseamento. As informações disponíveis da PV (LD, pedigree) são então utilizadas para determinar os genótipos ausentes nos haplotipos destes animais (HICKEY et al., 2013). Esse processo pode ser afetado por inúmeros fatores tais como a densidade e distribuição dos marcadores nos genótipos da PV, a distância genética entre os animais da PR

e da PV, a frequência alélica, o padrão de LD das regiões a serem imputadas (HICKEY et al., 2012).

Carvalho et al. (2014) avaliaram a eficiência de imputação em uma população de animais da raça Nelore. No trabalho, os autores consideraram diferentes densidades de genótipos para PV (7K a 75K), imputando para alta densidade (~ 777K). A imputação a partir de 15K foi considerada acurada, com valores de correlação superiores a 0,97. Para inferir os haplotipos/genótipos, o estudo comparou dois *softwares*, sendo um que considera apenas informações da população (Beagle (BROWNING; BROWNING, 2008)), e outro que pode combinar informações da população com dados de pedigree (FImpute (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014)). Apesar da imputação não ter sido muito afetada pela informação de pedigree, o FImpute resultou em maiores valores de acurácia para diferentes cenários de imputação na população estudada. Chud et al. (2015) demonstraram semelhantes resultados em uma população de animais da raça sintética Canchim, com melhores resultados de imputação quando a PV foi genotipada com pelo menos 50K marcadores e utilizando o software FImpute.

Também é possível a imputação genômica de animais não genotipados a partir da informação genômica de parentes próximos. Boison et al. (2014) avaliaram esse procedimento utilizando dados reais e simulados de uma população de animais da raça Gir (*Bos indicus*). No trabalho, melhores valores de acurácias de imputação para um indivíduo foram conseguidos quando consideradas pelo menos informações de 5 progênies genotipadas. Concluíram que valores de acurácia de imputação acima de 0,90 seriam suficientes para uma adequada estimativa de regiões genômicas idênticas por descendência (IBD⁷). Similarmente, BERRY et al. (2014) também avaliaram a imputação de animais não genotipados, concluindo que imputação dos haplótipos parentais pode ser realizada de forma acurada ($\geq 0,96$) a partir da informação de pelo menos cinco progênies genotipadas.

2.6 Aplicação da genômica na reprodução

Darwash, Lamming e Wooliams (1997) definiram fertilidade de vacas leiteiras como a habilidade do animal em engravidar e se manter prenhe após expostas a estação de monta no momento da ovulação. Para isso, processos biológicos como início da atividade ovariana,

⁷ Identical by descendent

detecção de estro, ovulação, ciclicidade e ausência de doenças reprodutivas devem ser considerados (ROYAL; MANN; FLINT, 2000).

Fortes et al. (2013a) confirmaram a natureza poligênica de característica reprodutivas de machos e fêmeas, reportando importantes regiões genômicas associadas com estas nos 30 cromossomos. No trabalho, os autores destacaram regiões nos cromossomos 1, 5, 14 e 16 para características reprodutivas em fêmeas, tais como taxa de prenhez, taxa de ovulação, prenhez em novilhas, entre outras. Para Burns, Fordyce e Holroyd (2010), a idade à puberdade e idade ao primeiro parto seriam fatores principais de influência na vida reprodutiva das fêmeas.

Dadas as diferenças genéticas entre animais zebuínos e taurinos, diferentes regiões de QTL têm sido reportadas na literatura associadas com características reprodutivas (CANAVEZ et al., 2012; HAWKEN et al., 2012). Fortes et al. (2013a) elaboraram uma tabela com os principais resultados de associação genômica com características reprodutivas em bovinos. Para os autores, a descoberta de regiões genômicas que influenciam características de interesse zootécnico é um processo em andamento, com possíveis aplicações na indústria.

No trabalho de Gaddis, Null e Cole (2016), os autores compararam diferentes abordagens para estudos de GWAS em característica reprodutivas em bovinos da raça Holandesa. Diferentes densidades de genotipagem foram testadas demonstrando que as principais associações foram detectadas independentemente da quantidade de marcadores disponíveis. Contudo, painéis com alta densidade de marcadores resultaram em um maior número de regiões genômicas associadas com os fenótipos testados e, no geral, apresentaram menor erro de predição dos efeitos dos marcadores (menor FDR⁸). Os autores ainda sugeriram que análises multivariadas em estudos de GWAS podem ser uma interessante opção, com maior poder estatístico.

Fortes et al. (2011) estudaram a característica de idade a puberdade, definida como idade à ocorrência do primeiro corpo lúteo (AGECL⁹) e outras 21 características em novilhas de uma raça tropical composta e Brahman (*Bos indicus*) na Austrália, reportando 48% dos genes selecionados em comum entre as duas raças utilizando a metodologia de associação de matriz ponderada (AWM¹⁰), previamente descrito em Fortes et al. (2010). As análises de enriquecimento funcional indicaram importante vias metabólicas associadas como atividades de orientação aniônica do cérebro, adesão celular, sinapse e receptores de glutamato, sugerindo

⁸ False discovery rate

⁹ Age at the first *corpus luteum*

¹⁰ Association weight matrix

influência no hormônio liberador de gonadotrofina (GnRH¹¹). A raça tropical composta, a qual tem pouca influência de animais zebuínos, atingiram a puberdade foram em média três meses mais cedo que animais Brahman.

Analisando os fenótipos de AGECL e idade dos machos ao atingirem 26 cm de perímetro escrotal, Fortes et al. (2012a) observaram 32 SNPs associados com ambas características. Esses marcadores estavam localizados no cromossomo 14, entre 21 e 28 Mb aproximadamente. Os autores sugerem que esse cromossomo seria fundamental para bovinos da raça Brahman, com possíveis efeitos pleiotrópicos com crescimento dos animais. Esse efeito foi posteriormente confirmado, com destaque para o alelo C do SNP rs109231213, localizado próximo ao gene *PLAG1*, com resultados significativos sobre os fenótipos de altura de costela, peso, consumo alimentar e idade à puberdade em machos e fêmeas (FORTES et al., 2013b).

No trabalho de Irano et al. (2016), 10 regiões genômicas foram significativamente associadas com prenhez em um banco de dados composto por 1,770 novilhas Nelore de aproximadamente 16 meses de idade. Os autores reportaram coeficiente de herdabilidade média de 0,30 para a característica e as regiões associadas foram distribuídas nos cromossomos 5, 6, 7, 14, 18, 21 e 27, explicando 8.9% da variância genética total.

Höglund et al. (2015) reportaram 17.388 marcadores do tipo SNP espalhados em 25 cromossomos, associados com um índice de fertilidade de fêmeas de gado taurino. Esse índice incluiu características tais como o número de inseminações por parição, intervalo de anestro, entre outras.

Em um estudo com a raça Brangus (3/8 Brahman + 5/8 Angus), 18 QTLs do cromossomo 15 foram associados com ambos os fenótipos de prenhez em novilhas e concepção no primeiro serviço (PETERS et al., 2013). Os autores reportaram coeficiente de herdabilidade de 0,10 para prenhez em novilhas, com QTL específicos para essa característica presentes nos cromossomos 2, 4, 10, 13 e 20.

Devido à grande variedade de métodos de análise e alta abrangência de resultados, a identificação de regiões genômicas associadas a diferentes características reprodutivas em bovinos Nelore é de extrema relevância. O foco do presente trabalho foi identificar regiões genômicas associadas a características reprodutivas tais como a prenhez precoce e número de folículos antrais em novilhas da raça Nelore, identificando processos biológicos envolvidos para um melhor entendimento da variabilidade fenotípica destas características.

¹¹ Gonadotropin-releasing hormone

3. Hipótese

- Existem regiões genômicas de efeito maior sobre a expressão das características reprodutivas estudadas.

4. Material e Métodos

4.1 Dados Fenotípicos

4.1.1 Bando de Dados

O banco de dados foi formado por 2.283 novilhas da raça Nelore, filhas de 78 touros, com informações de prenhez aos 14 meses e, em aproximadamente 48% desses animais (1.099) foi mensurado o número de folículos antrais. Essas informações foram coletadas em três fazendas comerciais de bovinos de corte localizadas no Estado do Mato Grosso do Sul, sendo armazenadas e analisadas pelo Grupo de Melhoramento Animal e Biotecnologia da Faculdade de Zootecnia e Engenharia de Alimentos de Pirassununga, São Paulo. Os animais tinham idade média de 15,8 meses, sendo criados em condições ambientais semelhantes, com dieta a base de pasto, com água e sal mineral disponíveis *ad libitum*.

Os animais foram alocados em grupos de contemporâneos formados pela combinação das informações de rebanho de origem, grupo de manejo e ano de nascimento do animal. Registros de grupos contemporâneos sem variabilidade (onde, por exemplo, todos os animais apresentavam o mesmo estado de prenhez) e com menos de dez animais não foram considerados nas análises. Além disso, também foram excluídos animais com idade ou peso superiores a 3,5 desvios padrão acima ou abaixo da média geral.

4.1.2 Características

Prenhez precoce (PP)

O diagnóstico de prenhez foi realizado com auxílio de equipamento de ultrassom (Chison 8200VET, Kylumax, com transdutor de 7,5 MHz) ou por palpação retal, 40 dias após o protocolo de inseminação artificial em tempo fixo (Figura 1). A característica foi considerada como categórica, com valor 1 (sucesso) atribuído a novilhas diagnosticadas prenhes e 0 (falha) atribuídas àquelas que não estavam prenhes no momento do diagnóstico.

O uso de animais próximos aos 15 meses de idade é uma prática já utilizada como estratégia de selecionar para precocidade sexual, fazendo parte da rotina de manejo das fazendas (ELER et al., 2002). Em países tropicais tais como o Brasil, a estação de nascimento é normalmente distribuída entre os meses de julho e novembro, com a estação de monta começando em outubro. Diante disso, as novilhas entram na estação de monta entre 12 meses (nascidos em novembro) e 16 meses (nascidos em julho) de idade.

Número de Folículos Antrais (NF)

Os folículos visíveis (≥ 3 mm de diâmetro) pela imagem do ultrassom (transdutor linear transretal de 7,5 MHz, Mindray M5Vet, China) em ambos os ovários foram contados e registrados no dia 4 do protocolo de IATF (Figura 1) para estabelecer o número total de folículos antrais recrutados na manifestação de ondas foliculares.

Figura 1 – Protocolo de inseminação artificial em tempo fixo (IATF) usado para as novilhas Nelore. Dia 0: Implante de orelha a base de Norgestomet, plus 2,0 mg de benzoato de estradiol (EB). Dia 8: implante de estradiol cipionato (EC), 0,265 mg de cloprostenol sódico (PGF) e 300 IU de gonadotrofina coriônica equina (eCG). Dia 10: Inseminação em tempo fixo (IATF).

4.2 Dados Genotípicos

Amostras de pelo da cauda foram coletadas de 1.267 novilhas para posterior extração do DNA nuclear. Este foi usado na genotipagem utilizando o chip *GGP Bos indicus HD*, especialmente desenvolvido para animais zebuínos, com densidade de 74,677 marcadores do tipo SNP. O processo de extração do DNA e genotipagem foram realizados pela empresa GeneSeek®.

Dos 78 touros (pais nas novilhas), 42 foram previamente genotipados com a plataforma *Illumina® BovineHD BeadChip* que contém 777,962 SNPs espalhados por todo o genoma. Esses animais foram incluídos no banco de dados para as análises de imputação e detecção dos blocos haplotípicos.

Análises de controle de qualidade (CQ) foram realizadas utilizando o pacote PREGSF90 (AGUILAR et al., 2013), com intuito de reduzir associações espúrias e, por conseguinte, aumentar a precisão das análises genômicas. As informações de marcadores presentes no

cromossomo Y, mitocondrial ou não atribuídos a cromossomos (não anotados) não foram consideradas nas análises. Os parâmetros de CQ utilizados para remove SNPs pouco informativos foram: alelos de menor frequência (MAF) < 0,02, taxa de genotipagem por SNP < 0,95 e, quando a proporção observada de marcadores heterozigotos diferiu da proporção esperada (Equilíbrio de Hardy-Weinberg) em valores > 0,15. Animais com taxa de genotipagem < 0,90 também foram excluídos do banco de dados. Os marcadores presentes no cromossomo X foram considerados apenas na análise de GWAS. Ao termino dos procedimentos de CQ, permaneceram no banco de dados informações de 64.753 SNPs e 1.255 novilhas. A origem desses animais e suas informações fenotípicas estão demonstradas na Tabela 1.

Tabela 1 – Valores médios ($\pm$ desvio padrão) para idade (meses) e peso (Kg) mensurados no início da estação de monta, taxa de prenhez (PP, %) e número de folículos antrais (NF) das novilhas genotipadas (após CQ) por fazenda.

	Fazenda1	Fazenda2	Fazenda3
N	579	299	377
Idade	16,1 ($\pm$ 1,20)	14,1 ($\pm$ 0,87)	17,2 ($\pm$ 0,52)
Peso	283,2 ($\pm$ 28,92)	260,6 ($\pm$ 19,20)	276,5 ($\pm$ 16,73)
PP	33	35	36
NF	12,07 ($\pm$ 4,68)	10,99 ($\pm$ 3,80)	–

N = número de animais.

4.3 Estimação dos Componentes de variância

As matrizes de covariâncias genética e residual para PP e NF foram preditas utilizando a metodologia de etapa única (*single step*) (LEGARRA; AGUILAR; MISZTAL, 2009), com abordagem bayesiana. O modelo estatístico considerado para a análise bicaracterística foi:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \quad [1]$$

em que y_1 e y_2 são os vetores com as informações fenotípicas das características 1 (PP) e 2 (NF) observadas nos animais genotipados e não genotipados; b_1 e b_2 os vetores de solução dos efeitos classificatórios para as características 1 e 2, respectivamente, incluindo o grupo contemporâneo e a covariável idade da novilha no momento das mensurações; a_1 e a_2 os vetores dos efeitos genéticos aditivos aleatórios para as características 1 e 2, respectivamente; ϵ_1 e ϵ_2 os vetores de soluções dos efeitos residuais para as características 1 e 2, respectivamente; X_1 , X_2 , Z_1 e Z_2 são matrizes de incidência relacionando b_1 , b_2 , a_1 e a_2 com a variável dependente y . As análises foram efetuadas pelo programa THRGIBBS2F90 (MISZTAL et al., 2002), sendo os parâmetros para PP preditos usando um modelo limiar, relacionando o fenótipo observado em escala categórica com uma escala normal contínua subjacente.

A matriz de covariância de a e ϵ assumiu:

$$Var \begin{bmatrix} a_1 \\ a_2 \\ \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} H\sigma_{a1}^2 & 0 & 0 & 0 \\ 0 & H\sigma_{a2}^2 & 0 & 0 \\ 0 & 0 & I\sigma_{\epsilon1}^2 & 0 \\ 0 & 0 & 0 & I\sigma_{\epsilon2}^2 \end{bmatrix} \quad [2]$$

sendo σ_{a1}^2 e σ_{a2}^2 os componentes de variância dos efeitos genético aditivo para as características 1 e 2, respectivamente; $\sigma_{\epsilon1}^2$ e $\sigma_{\epsilon2}^2$ os componentes residuais para as características 1 e 2, respectivamente; I uma matriz identidade e H a matriz de relacionamento que combina informações dos animais genotipados e não-genotipados, considerando a informação de pedigree como descrito em Aguilar et al. (2010), com a inversa estimada por:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \quad [3]$$

onde A^{-1} é a inversa da matriz de relacionamento aditiva (pedigree); A_{22}^{-1} a inversa da matriz de relacionamento aditiva incluindo apenas os animais genotipados e G^{-1} a inversa da matriz de relacionamento genômica estimada de acordo com Vanraden (2008):

$$G = \frac{MM'}{2 \sum p_i(1 - p_i)} \quad [4]$$

em que M é a matriz de incidência dos marcadores cujos elementos na i -ésima coluna são $0-2p_i$, $1-2p_i$ e $2-2p_i$ para os genótipos AA, AB e BB, respectivamente; M' é a transposta da matriz de incidência; p_i é a frequência do alelo B do i -ésimo marcador.

Foram utilizadas um total de 2.200.000 de iterações para as análises, sendo que as primeiras 200.000 foram descartadas (*burn-in*) e o intervalo de amostragem foi de 200. A convergência das cadeias de Markov (MCMC¹²) foi verificada pelo teste de convergência de Geweke e inspeção visual dos resultados com auxílio o pacote *boa* (SMITH, 2007) do software R (R CORE TEAM; R DEVELOPMENT CORE TEAM, 2013).

4.4 Estudo de Associação Genômica Ampla (GWAS)

O desequilíbrio de ligação (LD) entre os marcadores foi estimado utilizando o quadrado da correlação (r^2) (HILL; ROBERTSON, 1968), calculado por:

¹² Markov Chain Monte Carlo

$$r^2 = \frac{D^2}{[prob(A)prob(a)prob(B)prob(b)]} \quad [5]$$

sendo $D = prob(AB)prob(ab) - prob(Ab)prob(aB)$, A , a e B , b os quatro alelos de dois loci. As análises foram realizadas com o software Plink v1.9 (CHANG et al., 2015).

A associações entre os marcadores moleculares e as informações fenotípicas foram realizadas utilizando um modelo misto sob a metodologia BayesB, no software *Gensel* (FERNANDO; GARRICK, 2012). O método BayesB assume uma distribuição t com média nula como informação *a priori* para os efeitos dos SNPs e variância específica para cada alelo sob uma distribuição qui-quadrado invertida escalonada, com v_a graus de liberdade e S_a^2 o parâmetro de escala (FERNANDO; GARRICK, 2013). A ideia é que a maioria dos SNPs possuem pequenos efeitos (aproximadamente zero), exceto por alguns mais próximos à mutação causal que teriam efeitos maiores, dando mais sentido biológico para as análises (MEUWISSEN; HAYES; GODDARD, 2001, 2016).

O modelo estatístico utilizado no GWAS foi:

$$y = Xb + \sum_{j=1}^K a_{ij}\beta_j\delta_j + \varepsilon \quad [6]$$

sendo y , X e b previamente definidos na equação [1]; K o número total de marcadores do tipo SNP; a_{ij} um vetor representando a covariável de SNP no locus j para o animal i , codificados de acordo com o número de alelos B (0,1,2); β_j o efeito de substituição alélica do locus j , assumindo $\beta_j|\pi, \sigma_j^2 \sim \delta_j N(0, \sigma_j^2)$ quando $\delta_j = 1$ e $\beta_j = (1 - \delta_j)N(0, \sigma_{\beta i}^2 = 0)$ quando $\delta_j = 0$; onde δ_j é uma variável indicadora; e $\varepsilon|\sigma_\varepsilon^2 \sim N(0, \sigma_\varepsilon^2)$. A condição *a priori* para δ_j foi:

$$(\delta_j|\pi) \begin{cases} 1; \text{probabilidade}(1 - \pi) \\ 0; \text{probabilidade}(\pi) \end{cases}, \text{ sendo } \pi = 0.999.$$

Para a inferência sobre a distribuição *a posteriori* dos efeitos estimados dos SNPs, uma cadeia de 140.000 iterações foi gerada, utilizando-se o método de MCMC. As primeiras 20.000 iterações foram descartadas (*burn-in*) e as 120.000 seguintes foram utilizadas, com intervalo de amostragem a cada 100 iterações. Em vez de considerar os efeitos dos marcadores individualmente, a proporção da variância explicada por janelas genômicas não sobrepostas de aproximadamente 1 Mb foram utilizadas nas inferências (CESAR et al., 2014; GARRICK; FERNANDO, 2013). Essa proporção foi amostrada por:

$$q_k = \frac{\tilde{\sigma}_{gw}^2}{\tilde{\sigma}_g^2} \quad [7]$$

sendo que em cada iteração MCMC $\tilde{\sigma}_{gw}^2$ é a variância genética amostrada correspondente aos SNPs contidos na janela genômica em questão (w), $\tilde{\sigma}_g^2$ a variância genética total amostrada (FERNANDO; GARRICK, 2013). As janelas cujas médias posteriores representaram mais de 1% da variância genética foram consideradas com efeito significativo, sendo possíveis regiões QTL (PETERS et al., 2012).

4.5 Busca gênica e enriquecimento funcional

As janelas genômicas que explicaram pelo menos 1% da média *a posteriori* da variância genética foram consideradas como regiões com efeito importante nas características estudadas, tendo maior probabilidade de estarem associadas com regiões QTL. Essas regiões foram expandidas em 500 Kb em cada extremidade e os genes ali presentes foram identificados pela plataforma *Ensembl* (AKEN et al., 2016). Os principais genes associados com os fenótipos com base na literatura tiveram suas funções descritas.

Para as análises de enriquecimento funcional, os termos MeSH¹³ foram detectados por meio do pacote MESHR (TSUYUZAKI et al., 2015) do software R (R CORE TEAM; R DEVELOPMENT CORE TEAM, 2013). O pacote utiliza o teste hipergeométrico para avaliar a significância do enriquecimento como descrito em Morota et al. (2015). O teste de Benjamini e Hochberg (BENJAMINI; HOCHBERG, 1995) foi utilizado para correção de múltiplos testes, considerando significativo os termos com valor de $p < 0,05$.

4.6 Imputação de touros não-genotipados

Considerando a possibilidade de aumentar o número de machos genotipados no conjunto de dados, touros sem informação genômica, que tinham pelo menos cinco progênes genotipadas no banco de dados (BERRY et al., 2014; BOISON et al., 2014), foram imputados à mesma densidade de marcadores das novilhas (~ 64K). As análises foram realizadas usando o software FImpute 2.2 (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014).

A informação dos touros previamente genotipados foram utilizados para estimar a precisão da imputação. Primeiramente, conflitos mendelianos entre pai-progênie foram

¹³ Medical Subject Headings

verificados, considerando limiar de taxa de erro mendeliana de 0,1%. Em seguida, os touros tiveram seus genótipos mascarados e imputados usando a mesma abordagem utilizada para touros não genotipados. A precisão foi quantificada a partir da correlação entre os genótipos observados e imputados (CALUS et al., 2014) usando um script R (R CORE TEAM, 2013).

Um outro estudo (Apêndice A) foi realizado explorando estratégias de imputação em animais da raça Girolando, o qual é resultante do cruzamento de animais Gir com Holandês, com composição racial variando entre 1/4 e 7/8 de Holandês. O software FImpute (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014) foi utilizado para a imputação genômica. Esse estudo colaborou de maneira significativa para a formação nesse tipo de análise, o que justifica a inserção do mesmo no corpo desta tese.

4.7 Análise de Haplótipos

Os blocos haplotípicos foram definidos considerando as regiões genômicas responsáveis por pelo menos 1% da variância genética total. Foram consideradas as informações das novilhas, bem como os touros genotipados e imputados. As análises foram realizadas considerando a versão 3 do software findhap.f90 (VANRADEN; SUN, 2014), o qual utiliza métodos determinísticos para combinar informações de pedigree e populacionais nas predições, criando uma lista de haplótipos de aproximadamente mesmo tamanho genômico.

Após a criação dos haplótipos, inconsistências mendelianas foram examinadas, verificando meios-irmãos que não compartilharam um dos dois possíveis haplótipos paternos. As regiões inconsistentes foram desconsideradas, assim como quaisquer alelos haplotípicos com menos de três cópias ou sem variação fenotípica na população.

Para estimar o efeito do haplótipo paterno, uma segunda análise com o método Bayes B foi executado, incluindo-os como efeito classificatório no mesmo modelo previamente descrito (equação 6). Em seguida, a distribuição da diferença (análise de contraste) dos efeitos *a posteriori* dos alelos haplotípicos para os touros heterozigotos foram construídas a partir das amostras MCMC, a fim de identificar os reprodutores que segregaram alelos alternativos de QTL com efeito nas características estudadas. O nível de significância adotado para esse procedimento foi 10% ($p < 0,10$).

4.8 Sequenciamento

O sequenciamento completo do genoma (WGS¹⁴) foi realizado em seis touros, totalizando 24,669,022 variantes nos cromossomos autossômicos. As sequências foram obtidas utilizando a plataforma de sequenciamento (pared-end) HiSeq 2.500, considerando os protocolos recomendados pela empresa responsável (Illumina, Inc., San Diego, CA). A média geral de cobertura foi de 14,7.

Os dados de sequenciamento foram alinhados contra o genoma de referência *Bos taurus* (UMD 3.1), utilizando o software *Burrows-Wheeler Aligner* (BWA, v0.7.15) com os parâmetros padrão (LI; DURBIN, 2010). As regiões mapeadas foram utilizadas no processo de determinação de variante genômica, com auxílio dos softwares *Picard* v2.6.0 e *Genome Analysis Toolkit* (GATK, v3.6) (MCKENNA et al., 2010). As variantes identificadas foram então filtradas pelo GATK v3.6 (MCKENNA et al., 2010), considerando o escore geral de qualidade ≥ 20 e profundidade de leitura ≥ 4 . As variantes restantes foram convertidas pelo pacote VCFTools (DANECEK et al., 2011) em uma matriz genômica codificada em 0, 1 e 2, referindo-se ao número de cópia do alelo referência.

¹⁴ Whole-genome Sequence

5. Resultados

5.1 Dados Fenotípicos

As estimativas ($\pm DP^{15}$) utilizando dados genômicos dos coeficientes de herdabilidade (h^2) foram $0,28 \pm 0,07$ e $0,49 \pm 0,09$ para a prenhez (PP) e o número de folículos antrais (NF), respectivamente. A correlação genética ($\pm DP$) foi de $-0,21 \pm 0,29$, sugerindo fraca associação genética entre as características (Tabela 2). Os valores de *Z-score* para o teste de convergência de Geweke realizado sobre a distribuição *a posteriori* dos coeficientes preditos variaram de $-1,45$ (h^2_{PP}) e $-0,22$ (h^2_{NF}), sugerindo a convergência das análises. O intervalo de alta densidade posterior (0,025% – 0,975%) predito para a correlação genética entre PP e NF ($r_{PP \times NF}$) foi notavelmente amplo, ficando entre $-0,781$ e $0,329$. Como os critérios de convergência propostos foram atendidos, esse intervalo de correlação genética pode ser atribuído ao número limitado de observações consideradas no estudo, ou sugerir ausência de associação genética entre essas características.

Tabela 2 – Estatística descritiva dos coeficientes de herdabilidade (h^2) e correlação genética (r) para prenhez precoce (PP) e o número de folículos antrais (NF) das novilhas Nelore.

Parâmetros	Média	Mediana	DP	HPD _I	HPD _S	Geweke
h^2_{PP}	0,28	0,28	0,07	0,15	0,43	-1,45
h^2_{NF}	0,49	0,49	0,09	0,32	0,68	-0,22
$r_{PP \times NF}$	-0,21	-0,20	0,29	-0,78	0,33	-0,61

DP = desvio padrão; HPD_I = limite inferior a 95% do intervalo de maior densidade posterior; HPD_S = limite superior a 95% do intervalo de maior densidade. Os valores de Geweke referem aos valores de *Z-score*.

5.2 Estudo de Associação Genômica Ampla

O LD médio ($\pm DP$) entre os marcadores adjacentes foi de $0,23 \pm 0,01$ variando de 0,20 no cromossomo 27 a 0,25 no cromossomo 5 (Figura 2), e a média do LD entre os marcadores de cada foi $0,11 \pm 0,01$, variando de 0,09 no cromossomo 27 a 0,13 no cromossomo 6 (Figura 3). O padrão de decaimento do LD em função da distância entre os marcadores está apresentado na Figura 4, com baixos valores ($r^2 < 0,01$) observados para distâncias maiores que 170 Kb.

¹⁵ Desvio Padrão

Figura 2 – Valores médios ($\pm$ desvio padrão) de desequilíbrio de ligação (r^2) entre marcadores SNP adjacentes por cromossomos na população de novilhas Nelore estudada.

Figura 3 – Média ($\pm$ desvio padrão) do desequilíbrio de ligação (r^2) entre marcadores SNP por cromossomos na população de novilhas Nelore estudada.

Figura 4 – Comportamento do desequilíbrio de ligação médio (r^2) entre pares de marcadores em função da distância.

Um total de 2.673 janelas genômicas foram construídas considerando 30 cromossomos (29 autossomos e um sexual, X), tendo em média 25,5 SNPs e 934 Kb de tamanho. Os resultados apontaram regiões significativas nos cromossomos 5, 14 e 18 para PP (Figura 5) e 2, 8, 11, 14, 15, 16 e 22 para NF (

Figura 6), as quais explicaram mais de 1% da variância genética total. Estas regiões genômicas foram utilizadas para pesquisar genes candidatos que estão descritos no Quadro 1 e Quadro 2 para PP e NF, respectivamente. Janelas genômica nos cromossomos 1, 2, 3, 5, 14, 24, 29 e X demonstraram importante influência em PP, e nos cromossomos 2, 12, 15, 19, 24, 25 e 29 para NF, explicando mais que 0,5% da variância genética (Tabela 3 e Tabela 4).

Figura 5 – Manhattan *plot* do estudo de associação genômico amplo (GWAS) para a característica prenhez precoce em novilhas da raça Nelore, considerando limite de significância (linha vermelha) em 1% da proporção da variância genética.

Figura 6 – Manhattan *plot* do estudo de associação genômico amplo (GWAS) para a característica de número total de folículos antrais em novilhas Nelore, considerando limite de significância (linha vermelha) em 1% da proporção da variância genética.

Tabela 3 – Descrição das janelas genômicas que explicaram pelos menos 0,5% da variância genética total para a característica de prenhez precoce em novilhas da raça Nelore.

Chr	SNP _i	SNP _f	Pos _i	Pos _f	Tamanho	#SNPs	%Var
5	rs109437025	rs110687761	73021518	73966890	945372	22	3,90
5	rs42561706	rs137385583	77022303	77984850	962547	27	3,31
5	rs110496647	rs136544553	81133605	81955158	821553	22	1,70
14	rs41724652	rs133297141	23007464	23984847	977383	29	1,62
18	rs136460244	rs41891085	55018279	55956772	938493	27	1,18
14	rs41624840	rs136805030	29061608	29975350	913742	23	1,02
5	rs42917128	rs136339681	60131115	60941344	810229	23	0,93
24	rs136828522	rs137238317	60015982	60980620	964638	43	0,84
5	rs137127461	rs109435449	72059606	72982750	923144	28	0,82
2	rs42509691	rs134051905	54015771	54997281	981510	23	0,80
24	rs109329309	rs135881583	52121696	52957990	836294	23	0,68
5	rs110450288	rs133794376	82043465	82961509	918044	26	0,68
2	rs133912634	rs134084039	58011152	58971614	960462	25	0,68
X	rs134685381	rs137716652	44037579	44972292	934713	21	0,67
14	rs135852767	rs42298467	25021594	25986431	964837	22	0,66
5	rs110797637	rs137576699	51113065	51990170	877105	25	0,62
1	rs136647907	rs133111309	125026356	125988072	961716	24	0,62
29	rs134769207	rs42172278	24031236	24999881	968645	24	0,52
3	rs109945234	rs42368646	3017659	3968604	950945	42	0,51
X	rs134673004	rs134676523	43104725	43892074	787349	22	0,50

Chr = Cromossomo; SNP_i = SNP presente no início da janela genômica; SNP_f = SNP presente no final da janela genômica; Pos_i = posição em par de bases do início da janela genômica; Pos_f = posição em par de bases do final da janela genômica; Tamanho = tamanho em par de bases da janela genômica; #SNPs = número de SNPs presentes na janela genômica; %Var = porcentagem da variância genética explicada pela janela genômica.

Tabela 4 – Descrição das janelas genômicas que explicaram pelos menos 0,5% da variância genética total para a característica de número de folículos antrais em novilhas da raça Nelore.

Chr	SNP _i	SNP _f	Pos _i	Pos _f	Tamanho	#SNPs	% Var
22	rs42527679	rs109315219	15007849	15974506	966657	25	5,56
11	rs109186395	rs109576486	70020906	70972289	951383	27	5,50
8	rs42970082	rs136129526	7019096	7974290	955194	22	3,58
16	rs132917040	rs133998930	71002173	71955202	953029	25	1,96
14	rs41724652	rs133297141	23007464	23984847	977383	29	1,29
2	rs136879249	rs136073319	123033476	123985696	952220	22	1,12
15	rs135765779	rs109657326	9007403	9964870	957467	24	1,11
24	rs133021126	rs109305902	62008417	62594751	586334	24	0,91
29	rs42183484	rs135436243	33000052	33914381	914329	23	0,90
12	rs109641299	rs134942057	60015627	60994461	978834	28	0,81
2	rs42321794	rs134609840	56028202	56979919	951717	21	0,79
19	rs109030622	rs135605613	61000109	61990875	990766	45	0,75
19	rs136160175	rs41919438	56008922	56945797	936875	55	0,75
19	rs137834616	rs137557574	26010131	26962473	952342	29	0,67
25	rs134031092	rs134416406	16006913	16986653	979740	21	0,60
12	rs110539435	rs109044869	30015791	30992759	976968	30	0,58
15	rs137836273	rs110334648	81018176	81864846	846670	32	0,58
12	rs137815663	rs109275028	79019126	79976465	957339	27	0,56
X	rs133820946	rs135451827	144008894	144988891	979997	35	0,50

Chr = Cromossomo; SNP_i = SNP presente no início da janela genômica; SNP_f = SNP presente no final da janela genômica; Pos_i = posição em par de bases do início da janela genômica; Pos_f = posição em par de bases do final da janela genômica; Tamanho = tamanho em par de bases da janela genômica; #SNPs = número de SNPs presentes na janela genômica; % Var = porcentagem da variância genética explicada pela janela genômica.

Quadro 1 – Genes abrigados em janelas genômicas que explicaram mais de 1% da variância genética aditivo para prenhes precoce em novilhas Nelore.

Chr	Posição (montagem UMD 3.1)	Genes	Var (%)
5	72521518 – 74466890	<i>APOL6, HMGXB4, HO-1, ISX, LARGE1, MB, MCM5, RASD2, RBFOX2, SNORA76, TOM1</i>	3,90
	76522303 – 78484850	<i>BICD1, DNMI1, FGD4, KIAA1551, PKP2, SYT10, YARS2</i>	3,31
	80633605 – 82455158	<i>CCDC91, COX7ALP1, ERGIC2, FAR2, KLHL42, PTHLH</i>	1,70

14	22507464– 24484847	<i>ATP6V1H, FAM150A, LYPLA1, MRPL15, NPBWR1, OPRK1, PCMTD1, RB1CC1, RGS20, RP1, SOX17, ST18, TCEA1, U2, U6, XKR4</i>	1,62
18	54518279 – 56456772	<i>ALDH16A1, BAX, BCAT2, BOSTAUV1R403, BOSTAUV1R404, C19orf68, C5AR1, C5AR2, CA11, CABP5, CCDC114, CCDC155, CCDC9, CD37, CRX, DBP, DHDH, DHX34, DKKL1, ELSPBP1, EMP3, FAM38E, FCGRT, FGF21, FLT3LG, FTL, FUT1, FUT2, GLTSCR1, GLTSCR2, GRIN2D, GRWD1, GYS1, HRC, HSD17B14, IZUMO1, KCNA7, KCNJ14, KDELR1, KPTN, LHB, LIG1, LIN7B, LMTK3, MAMSTR, MEIS3, NAPA, NOSIP, NPAS1, NTF4, NTN5, NUCB1, PIH1D1, PLEKHA4, PPFIA3, PPP1R15A, PRRG2, PTH2, RASIP1, RCN3, RPL18, RPS11, RUVBL2, SAE1, SEC1, SEPWI, SLC17A7, SLC6A16, SLC8A2, SNRNP70, SPACA4, SPHK2, SULT2A1, SULT2B1, SYNGR4, TEAD2, TMEM143, TMEM160, TRPM4, TULP2, ZC3H4, ZNF114, ZNF541</i>	1,18
14	28561608 – 30475350	<i>GGH, NKAIN3, TTPA, YTHDF3</i>	1,02

Chr = cromossomo; var = variância genética explicada pela janela.

Quadro 2 – Genes abrigados em janelas genômicas que explicaram mais de 1% da variância genética para a característica de número de folículos antrais em novilhas Nelore.

Chr	Posição (montagem UMD 3.1)	Genes	Var (%)
22	14507849 – 16474506	<i>ACKR2, ANO10, CCDC13, CCK, HHATL, HIGD1A, KLHL40, LYZL4, NKTR, SEC22C, SNRK, SS18L2, TCAIM, TOPAZ1, TRAK1, VIPR1, ZBTB47, ZNF197, ZNF445, ULK4</i>	5,56
11	69520906 – 71472289	<i>ALK, BRE, C11H2orf71, CLIP4, FAM179A, FOSL2, LBH, PLB1, PPP1CB, SNORD53_SNORD92, SPDYA, TRMT61B, WDR43, YPEL5</i>	5,50
8	6519096 – 8474290	<i>ADAM29, BLK, C8orf74, CTSB, DEFB134, FAM167A, FDFT1, GATA4, GLRA3, HPGD, MTMR9, NEIL2, PINX1, SOX7, TDH, XKR6</i>	3,58

16	70502173 – 72455202	<i>ARL8A, ELF3, GPR37L1, LGR6, PPP1R12B, PROX, PTPN14, PTPN7, RNPEP, RPS6KC1, SMYD2, SYT2, UBE2T</i>	1,96
14	22507464 – 24484847	<i>ATP6V1H, FAM150A, LYPLA1, MRPL15, NPBWR1, OPRK1, PCMTD1, RB1CC1, RGS20, RP1, SOX17, ST18, TCEA1, U2, U6, XKR4</i>	1,29
2	122533476 – 124485696	<i>COL16A1, FABP3, HCRTR1, LAPTM5, MATN1, NKAIN1, PEF1, PUM1, SDC3, SERINC2, SNRNP40, TINAGL1</i>	1,12
15	8507403 – 10464870	<i>ARHGAP42, CNTN5</i>	1,11

Chr = cromossomo; var = variância genética explicada pela janela.

5.3 Análises de enriquecimento funcional utilizando os termos MeSH

As análises de enriquecimento utilizando MeSH revelaram 74 termos relacionados com PP (Tabela 5) e 48 relacionados com NF (Tabela 6). Após a correção para múltiplos testes pelo procedimento de Benjamini e Hochberg, os termos que foram significativos ($p < 0,05$) associados com PP foram *Munc-18 Proteins*, *Fucose* e *Hemoglobins*, e para NF foram *Cathepsin B*, *Receptors-Neuropeptide* e *Palmitic Acid*.

Tabela 5 – Termos MeSH relacionados com prenhez precoce em novilhas Nelore com seus respectivos *p-values*.

MeSH ID	<i>p-value</i>	MeSH Terms
D051938	0.0000	Munc18 Proteins
D005643	0.0000	Fucose
D033921	0.0000	Vesicular Transport Proteins
D006454	0.0008	Hemoglobins
D043563	0.0018	Receptors, Pheromone
D052066	0.0018	N-Ethylmaleimide-Sensitive Proteins
D009419	0.0027	Nerve Tissue Proteins
D014453	0.0039	Glucuronosyltransferase
D013739	0.0059	Testosterone
D000637	0.0073	Transaminases
D000706	0.0073	Anaphylatoxins
D001663	0.0073	Bilirubin
D001664	0.0073	Biliverdine
D001880	0.0073	Boranes
D007626	0.0073	Kell Blood-Group System
D033983	0.0073	Adaptor Protein Complex mu Subunits
D037101	0.0073	Luteinizing Hormone, beta Subunit
D044087	0.0073	Receptor, Anaphylatoxin C5a
D051857	0.0073	Otx Transcription Factors
D014801	0.0074	Vitamin A

D003577	0.0091	Cytochrome P-450 Enzyme System
D020033	0.0137	Protein Isoforms
D004050	0.0145	Ditiocarb
D006427	0.0145	Hemin
D011951	0.0145	Receptors, Complement
D013762	0.0145	Tetrahydrofolate Dehydrogenase
D015926	0.0145	Complement C3a
D026121	0.0145	Indole Alkaloids
D044162	0.0145	Parathyroid Hormone-Related Protein
D050976	0.0145	Basic-Leucine Zipper Transcription Factors
D051191	0.0145	Plakophilins
D051267	0.0145	NF-E2-Related Factor 2
D004271	0.0217	DNA, Fungal
D007770	0.0217	L-Lactate Dehydrogenase
D010171	0.0217	Palmitoyl Coenzyme A
D016861	0.0217	Cyclooxygenase Inhibitors
D051140	0.0217	Selenoproteins
D013481	0.0253	Superoxides
D011994	0.0257	Recombinant Proteins
D002245	0.0288	Carbon Dioxide
D003994	0.0288	Bucladesine
D006419	0.0288	Heme Oxygenase (Decyclizing)
D009942	0.0288	Organometallic Compounds
D013880	0.0288	Thioredoxin-Disulfide Reductase
D016566	0.0288	Organoselenium Compounds
D020813	0.0288	Receptor, trkB
D008054	0.0358	Lipid Peroxides
D010281	0.0358	Parathyroid Hormone
D016211	0.0358	Transforming Growth Factor alpha
D020555	0.0358	GTP-Binding Protein Regulators
D044385	0.0358	GTP-Binding Protein alpha Subunits
D051033	0.0358	Apoptosis Inducing Factor
D051547	0.0358	Heme Oxygenase-1
D053179	0.0358	Caspase 7
D014157	0.0389	Transcription Factors
D005819	0.0417	Genetic Markers
D007986	0.0424	Luteinizing Hormone
D016166	0.0428	Free Radical Scavengers
D033962	0.0428	Adaptor Protein Complex 2
D018160	0.0458	Receptors, Cytoplasmic and Nuclear
D016601	0.0494	RNA-Binding Proteins
D018398	0.0494	Homeodomain Proteins
D001708	0.0498	Biopterin
D011496	0.0498	Protein Methyltransferases
D011961	0.0498	Receptors, Fc
D018000	0.0498	Receptors, Peptide
D020746	0.0498	Calcium Channels, L-Type
D051537	0.0498	Hepatocyte Nuclear Factor 1
D051538	0.0498	Hepatocyte Nuclear Factor 1-alpha
D051539	0.0498	Hepatocyte Nuclear Factor 1-beta

D051541	0.0498	Hepatocyte Nuclear Factor 3-alpha
D051542	0.0498	Hepatocyte Nuclear Factor 3-beta
D051543	0.0498	Hepatocyte Nuclear Factor 3-gamma
D051559	0.0498	Hepatocyte Nuclear Factor 6

Tabela 6 – Termos MeSH relacionados com número de folículos antrais em novilhas Nelore com seus respectivos *p-values*.

MeSH ID	<i>p-value</i>	MeSH Term
D002401	0.0001	Cathepsin B
D018013	0.0007	Receptors, Neuropeptide
D050991	0.0007	GATA4 Transcription Factor
D019308	0.0017	Palmitic Acid
D010206	0.0020	Papain
D015853	0.0025	Cysteine Proteinase Inhibitors
D004124	0.0045	p-Dimethylaminoazobenzene
D007626	0.0045	Kell Blood-Group System
D007980	0.0045	Levodopa
D012844	0.0045	Sincalide
D013186	0.0045	Farnesyl-Diphosphate Farnesyltransfer
D017606	0.0045	Chlorine Compounds
D033983	0.0045	Adaptor Protein Complex mu Subunits
D043886	0.0045	Receptor, Cholecystokinin A
D043887	0.0045	Receptor, Cholecystokinin B
D015259	0.0089	Dopamine Agents
D043169	0.0089	Endostatins
D050783	0.0089	Proto-Oncogene Proteins c-ets
D009479	0.0130	Neuropeptides
D010171	0.0133	Palmitoyl Coenzyme A
D019894	0.0133	Peptide YY
D039222	0.0133	Integrin alpha1beta1
D064446	0.0133	Orexin Receptors
D018996	0.0221	Myelin P2 Protein
D044385	0.0221	GTP-Binding Protein alpha Subunits
D043562	0.0225	Receptors, G-Protein-Coupled
D005755	0.0264	Gastrins
D010169	0.0264	Palmitic Acids
D033962	0.0264	Adaptor Protein Complex 2
D015513	0.0308	Oncogene Proteins
D050778	0.0308	NFATC Transcription Factors
D002766	0.0351	Cholecystokinin
D003847	0.0351	Deoxyglucose
D004299	0.0351	Dopamine beta-Hydroxylase
D006911	0.0351	Hydroxyprostaglandin Dehydrogenases
D015347	0.0351	RNA Probes
D018006	0.0351	Myogenic Regulatory Factors
D064236	0.0351	Cartilage Oligomeric Matrix Protein
D004298	0.0394	Dopamine

D006131	0.0394	Growth Inhibitors
D007004	0.0394	Hypoglycemic Agents
D050612	0.0394	Fatty Acid Transport Proteins
D064235	0.0394	Matrilin Proteins
D012333	0.0409	RNA, Messenger
D009587	0.0436	Nitrogen Isotopes
D016877	0.0436	Oxidants
D025262	0.0436	Vacuolar Proton-Translocating ATPases
D001059	0.0479	Apoproteins

5.4 Imputação dos animais não-genotipados

O teste de parentesco apontou cerca de 9% de erros no pedigree dos animais genotipados considerados no estudado, os quais foram tratados como pais desconhecidos. Com isso, 39 touros genotipados e 38 não genotipados foram mantidos no conjunto de dados.

A acurácia de imputação dos SNPs para os touros que tiveram os respectivos genótipos mascarados foi de $0,96 \pm 0,03$, considerando pelo menos cinco progênie com informação genômica. Com base nesta alta precisão de imputação, todos os touros não genotipados com cinco ou mais progênie (20 animais) foram imputados para a mesma densidade de marcadores das novilhas (64.800 SNPs).

5.5 Estudo dos Haplótipos

Três blocos haplotípicos foram considerados nas análises de haplótipos, com base em regiões que explicaram mais do que 1% da variação genética total (cromossomos 5, 14, 18). Os blocos tiveram 4,04 Mb de tamanho médio, abrangendo cerca de 98 marcadores (Tabela 7).

Tabela 7 – Blocos Haplotípicos considerando as regiões do estudo de GWAS que explicaram mais de 1% da variância genética total da característica de prenhez precoce em novilhas da raça Nelore.

Hap	Chr	SNP _i – SNP _f	Tamanho (Mb)	#SNPs
1	5	rs136575538 – rs110111162	4,10	99
2	14	rs42410923 – rs109119025	4,07	97
3	18	rs135504567 – rs109012807	3,95	97

Hap = Bloco Haplotípico; Chr = Cromossomo; SNP_i = SNP inicial; SNP_f = SNP final; #SNPs = número de SNPs contidos nos haplotipos.

As análises de contraste apontaram para 15 diferentes ($p < 0,10$) alelos nos touros heterozigotos para os loci com influência na característica de PP (Figura 7, Figura 8 e Figura 9). O alelo haplotípico denominado por “A” (hapA) do cromossomo 5 foi detectado em três touros (Touro1, Touro2 e Touro3), tendo efeito maior que o alelo alternativo em todos eles. O mesmo resultado foi observado para os alelos E, G e I nos touros 4, 5 e 6, respectivamente (Figura 7).

Seis touros (7 a 12) apresentaram diferença significativa ($p < 0,10$) entre seus alelos no haplótipo do cromossomo 14 (Figura 8). O alelo “D” (hapD) teve efeito inferior em dois touros (Touro8 e Touro9), sugerindo uma sequência genômica desfavorável para a característica.

Os touros 11 e 12 também apresentaram diferenças significativas no cromossomo 18 (Figura 9). Embora o alelo “B” (hapB) deste cromossomo apresentar um resultado favorável quando contrastado com o alelo “C” (hapC) no touro 12, obteve resultado inferior ao alelo “A” (hapA) no touro11. O hapC, por sua vez, teve valores inferiores nos animais 12 e 14 (contrastado com o alelo “F”), sugerindo ser um alelo desfavorável para a característica PP.

Figura 7 – Distribuição da diferença dos efeitos estimados *a posteriori* entre dois alelos haplotípicos do cromossomo 5 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de prenhez de novilhas.

Figura 8 – Distribuição da diferença dos efeitos estimados *a posteriori* entre dois alelos haplotípicos do cromossomo 14 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de prenhez de novilhas.

Figura 9 – Distribuição da diferença dos efeitos estimados *a posteriori* entre dois alelos haplotípicos do cromossomo 18 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de prenhez de novilhas.

Os três principais blocos haplotípicos, com base em regiões que explicaram mais do que 1% da variação genética total, também foram considerados nas análises dos haplótipos para a característica NF. Estes estavam presentes nos cromossomos 22, 11 e 8, tendo em média 3,91 Mb de tamanho e abrangendo aproximadamente 98 marcadores (Tabela 8).

Tabela 8 – Blocos Haplotípicos considerando as regiões do estudo de GWAS que explicaram mais de 1% da variância genética total da característica número de folículos antrais.

Hap	Chr	SNP _i – SNP _f	Tamanho (Mb)	#SNPs
1	22	rs42000996 – rs134939375	3,99	95
2	11	rs134339479 – rs133809754	4,24	99
3	8	rs137535033 – rs109594484	3,50	99

Hap = Bloco Haplotípico; Chr = Cromossomo; SNP_i = SNP inicial; SNP_f = SNP final; #SNPs = número de SNPs contidos nos haplótipos.

As análises de contraste entre os alelos haplotípicos dos touros heterozigotos para os loci em questão, determinaram 20 alelos que influenciaram positivamente ($p < 0,10$) a característica NF (Figura 10, Figura 11 e Figura 12). Entre estes, cinco (haps A, D, E, e I) foram do cromossomo 22, presentes em cinco diferentes animais (Figura 10).

No cromossomo 11, 14 touros apresentaram diferenças entres os alelos haplotípicos, sendo os touros 17 e 19 também presentes nas análises do cromossomo 22 (Figura 11). Também foram compartilhados alguns animais com as análises para PP, sendo os touros 1, 2, 4 e 5 presentes nos resultados para o cromossomo 5, e os touros 8 e 15 nos cromossomos 14 e 18, respectivamente. O alelo A (hapA) teve efeito positivo em cinco animais, sugerindo possível sequência favorável para a característica NF. Já o alelo F (hapF) demonstrou resultados inferiores em cinco diferentes touros, sugerindo, neste caso, ser um alelo desfavorável para a característica. Ainda, os alelos B (hapB) e C (hapC) também apresentaram efeitos inferiores aos alelos alternativos em três (touros 21, 25 e 26) e dois (touros 19 e 24) animais, respectivamente.

As análises para a região do cromossomo 8 resultaram em seis animais com diferenças entres os alelos haplotípicos. Entre estes, dois (touros 8 e 11) também apareceram nos resultados do cromossomo 14, enquanto que os touros 6 e 18 foram reportados nos cromossomos 5 e 22 para PP e NF, respectivamente (Figura 12). O alelos A (hapA) apresentou efeito favorável à NF nos touros 27 e 8, enquanto que o alelo E (hapE) foi inferior nos touros 18 e 28.

Figura 10 – Distribuição da diferença dos efeitos estimados *a posteriori* entre dois alelos haplotípicos do cromossomo 22 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de número de folículos antrais.

Figura 11 – Distribuição da diferença dos efeitos estimados *a posteriori* entre dois alelos haplotípicos do cromossomo 11 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de número de folículos antrais.

Figura 12 – Distribuição da diferença dos efeitos estimados *a posteriori* entre dois alelos haplotípicos do cromossomo 8 em touros da raça Nelore que, de forma significativa ($p < 0,10$), influenciaram a característica de número de folículos antrais.

6. Discussão

Animais da sub-espécie *indicus* (zebuínos) são importantes principalmente em países tropicais. Em comparação com *taurus* (taurinos), esses animais são mais adaptados ao clima quente e também mais resistentes a infecções parasitárias, tornando-os mais produtivos nessas condições (PORTO NETO et al., 2011; PORTO-NETO et al., 2013). Por outro lado, os zebuínos são conhecidos por serem sexualmente mais tardios que animais taurinos (ELER et al., 2004; HAWKEN et al., 2012), aumentando a importância em estudos envolvendo características relacionadas com a precocidade sexual desses animais, tais como a PP e NF.

Características reprodutivas são tidas como complexas o que, por definição, evidencia a regulação por meio da expressão de diversos genes, além de forte influência de fatores não-genéticos (FORTES et al., 2010). Devido a esta complexidade e ao baixo controle genético da variância observada nesses fenótipos, baixos avanços têm se conquistado no melhoramento genético dessas características (BERRY et al., 2012; VEERKAMP et al., 2012; MCDANELD et al., 2014).

A média *a posteriori* do h^2 (95% da maior densidade posterior) para PP obtida no presente estudo foi de 0,28 (0,15 – 0,43), sendo inferior ao resultado apresentado por Santana Jr et al. (2015) que estimaram 0,46 (0,40 – 0,52) para uma população com mais de 18 mil novilhas Nelore com informações para a característica PP. Eler et al. (2014) que também trabalharam com rebanho Nelore, reportaram 0,37 (0,33 – 0,42), sendo mais próximos aos do presente estudo. Essas diferenças são esperadas uma vez que os trabalhos se diferem na disponibilidade de informações (n amostral), na inclusão de dados genômicos e também diferenças nos parâmetros considerados na modelagem estatística.

Walsh et al. (2014) trabalhando com uma população de 445 vacas e 522 novilhas da raça Holandesa, reportaram h^2 para a característica de número de folículos antrais de $0,31 \pm 0,14$ e $0,25 \pm 0,13$ para vacas e novilhas respectivamente, sendo inferior do valor predito no presente trabalho ($0,49 \pm 0,09$). Já Snelling et al. (2012) predizeram h^2 ($\pm$ erro padrão) de $0,73 \pm 0,18$ para NF em uma análise univariada considerando apenas informações de pedigree e $0,44$ usando informações de pedigree e em uma população de 452 novilhas da raça Holandesa. Essas discrepâncias pode estar relacionada, além do ponto já discutidos para PP, a diferenças entre as raças.

Cushman et al. (2009) reportaram influência significativa do peso ao nascimento e idade dos animais sobre a característica NF. Novilhas com baixo NF (< 15 folículos) apresentaram menor tamanho de ovários e também menor taxa de prenhez. Os autores ainda observaram um

declínio em NF a partir dos 5 anos de idade, corroborando Echternkamp et al. (2007) que também relataram decaimento na taxa de prenhez entre 5 e 6 anos de idade em animais selecionados para precocidade sexual. No presente estudo, os valores médios para NF foram inferiores a 13, com correlação genética (95% da maior densidade posterior) de -0,21 (-0,78 – 0,33) com PP. Dado que o intervalo de credibilidade inclui o valor zero, não podemos concluir a correta influência genética de uma característica sobre a outra. Em Santos et al. (2015), os autores não observaram influência da característica NF sobre a taxa de concepção de vacas Nelore, apesar de apresentarem maior produção *in vitro* de embriões. Baruselli et al. (2015) também não relataram relação entre NF e prenhez tanto em vacas quanto em novilhas da raça Nelore.

6.1 Estudos de Associação Genômica Ampla

Espigolan et al. (2013), estudando uma população de 795 touros Nelore com cerca de 447 K marcadores SNP (após QC), relataram r^2 média de 0,17 e 0,16 para SNPs global e adjacentes, respectivamente. Esta diferença observada é esperada devido ao comportamento de LD sendo inversamente proporcional à distância entre marcadores, com os genótipos mais densos demonstrando alta média global de LD, como demonstrado da Figura 4. Biegelmeyer et al. (2016) encontraram $0,21 \pm 0,27$ e $0,16 \pm 0,20$ de média r^2 entre marcadores adjacentes em 391 Hereford e 2019 Braford animais, respectivamente, genotipados com cerca de 41K SNP marcadores. Conforme observado por esses autores, LD pode ser influenciado por vários fatores sendo, portanto, específicos da população. Fonseca et al. (2016), avaliando o comportamento do LD em uma população de animais Guzerá, reportaram baixos valores ($r^2 < 0,01$) para distâncias maiores que 120 Kb, sendo próximo ao observado no presente estudo (170 Kb). Valores discrepantes de LD foram observados em 800 – 900 Kb e 900 – 1000 Kb, corroborando com os resultados de Fonseca et al. (2016). Segundo os autores, isto se justifica pelo baixo número de marcadores entre essas distâncias, criando imprecisão nas estimativas.

GWAS é uma investigação de milhares de variantes genéticas em diferentes animais para verificar se qualquer marcador está associado com a característica de interesse (WANG et al., 2005). Todos os efeitos dos SNPs foram preditos simultaneamente no modelo utilizando abordagem bayesiana, o que resulta em um melhor controle das associações espúrias, com menor superestimação dos efeitos do SNPs e corrigindo possíveis efeitos de estratificação populacional (CANTOR; LANGE; SINSHEIMER, 2010; GARRICK; FERNANDO, 2013; PETERS et al., 2013).

O fato de diferentes regiões em diferentes cromossomos afetarem características relacionadas com fertilidade reflete a complexidade destas. Fortes et al. (2013a) analisaram mais de 200 regiões genômicas espalhados por todos os cromossomos relacionadas com características de fertilidade em fêmeas. Os autores também discutiram que as regiões genômicas importantes em *Bos taurus* podem ser diferentes em animais *Bos indicus*, corroborando com algumas posições diferentes relatadas neste estudo. Portanto, os loci relatados neste estudo representam informações importantes sobre a taxa de prenhez em novilhas da raça Nelore.

6.2 Genes Candidatos

6.2.1 Prenhez em Novilhas

Cromossomo 5

Três janelas genômicas no cromossomo 5 explicaram 3,9%, 3,31% e 1,70% da variância genética para a característica PP, cobrindo vinte e quatro genes. O cromossomo 5 já foi previamente associado à características reprodutivas em populações de novilhas de corte de raça composta tropical (FORTES et al., 2014; HAWKEN et al., 2012) e também em animais *Bos taurus* (HÖGLUND et al., 2015; SAHANA et al., 2010). O estudo de Kappes et al. (2000) considerou esse cromossomo como uma importante região genômica associada à taxa de ovulação em bovinos. O gene *ISX* participa do metabolismo da vitamina A pela regulação da expressão do gene *BCOI* no intestino. Lobo et al. (2013) discutiram que o β -caroteno, que é convertido em vitamina A no intestino, participa de funções tais como visão, reprodução, imunológica e diferenciação celular.

Os genes *TOM1* e *HO-1* foram relatados como importantes na janela genômica próxima dos 73 Mb e estão relacionados com gravidez em seres humanos (DUMESIC et al., 2015). Assidi et al. (2011) discutiram que o gene *TOM1* regula negativamente o fator de transcrição AP1, o qual está envolvido na via de excreção do homônimo LH em células da granulosa pré-ovulatória de bovinos. Os autores confirmaram *TOM1* como um biomarcador negativo no desenvolvimento de oócitos. A heme-oxigenase 1 (*HO-1*) é um gene relacionado ao estresse oxidativo, e seu alto nível de expressão em células cúmulos circundantes também está relacionado com a redução da fertilização de ovócitos em humanos (AIKEN et al., 2015).

Em Hawken et al. (2012), o cromossomo 5 abrigou os marcadores mais importantes associados com características reprodutivas, tais como a idade a puberdade em bovinos. O SNP

mais significativo foi localizado perto de 96 Mb, estando próximo das regiões relatadas neste estudo (72 - 83 Mb). Outros estudos (ALLAN et al., 2009; LUNA-NEVAREZ et al., 2011; SAHANA et al., 2010) também relacionaram esse cromossomo com características reprodutivas.

O fator de crescimento semelhante à insulina-1 (*IGF-I*¹⁶) está localizado próximo de 66 Mb e frequentemente é associado à características reprodutivas em bovinos (FORTES et al., 2013c; WATHES et al., 2007; ZHANG et al., 2013). O LD médio (r^2) dos marcadores adjacentes entre a localização desse gene e a janela genômica mais próxima relatada neste estudo (~ 72-74Mb) foi de $0,35 \pm 0,30$, sugerindo uma possível ligação entre a janela reportada e o gene.

Cromossomo 14

Duas janelas genômicas apresentaram efeito significativo sobre PP no cromossomo 14, sendo uma entre 22 e 25 Mb e outra entre 28 e 31 Mb, explicando 1,62% e 1,02% da variância genética e cobrindo 16 e 4 genes, respectivamente. A primeira janela (entre 22 e 25 Mb) também apresentou efeito significativo para a característica NF. Hawken et al. (2012) associaram múltiplos clusters de SNPs com características reprodutivas no cromossomo 14 em uma população de 843 Brahman e 866 novilhas de raça composta, sendo os marcadores mais significativos localizados entre 22 a 25 Mb. Fortes et al. (2011, 2012a, 2013b) também descreveram diferentes genes entre 20 e 30 Mb no cromossomo 14 influenciando características reprodutivas em bovinos. Utsunomiya et al. (2013) relacionaram SNPs entre 23 e 26 Mb que afetam o peso ao nascer em bovinos e Utsunomiya et al. (2014) associaram esta mesma região com características reprodutivas em touros da raça Nelore. Ambos os trabalhos indicaram uma possível relação entre essa janela genômica e características reprodutivas em fêmeas.

O fato de uma mesma região genômica ter sido considerada significativa para ambas as características analisadas (PP e NF), destaca a importância do cromossomo para características reprodutivas, sugerindo possível efeito pleiotrópico. Essa hipótese foi reforçada comparando a correlação dos valores genômicos (VG) preditos para PP e NF, considerando todos os cromossomos, com a correlação dos VG quando considerado apenas os marcadores presentes nessa janela genômica (COLE et al., 2009). As estimativas (intervalo de 95% de credibilidade) desse valores foram -0,14 (-0,21 – -0,08) e -0,30 (-0,36 – -0,24), respectivamente, diferindo-se estatisticamente ($p < 0,01$), reforçando assim a hipótese de efeito pleiotrópico dessa região.

¹⁶ *Insulin like growth factor 1*

SOX17 é um gene já descrito, com influência em células germinativas primordiais, que são as precursoras das células germinativas espermatozoides e óvulos (IRIE et al., 2015). Irie et al. (2015) sugeriu que o gene *SOX17* teria um papel fundamental na manutenção das células germinativas em humanos. Também é discutido que *SOX17* regula o processo de formação de glândulas endometriais (adenogênese), que começa nas fases embrionárias e se completa na puberdade, além de também regular funções glandulares em humanos (GUIMARÃES-YOUNG et al., 2016). Essas secreções glandulares do endométrio são fundamentais para fixação do embrião, secretando a maior parte das proteínas e outros componentes do fluido uterino (GRAY et al., 2001). Guimarães-Young et al. (2016) relacionaram a expressão desse gene com o gene promotor do receptor de progesterona (*PGR*), resultando na diminuição da expressão de *PGR* em ratos com alelo silenciado de *SOX17*, com consequente comprometimento da fertilidade das fêmeas.

O gene *RGS20* pertence à família dos reguladores de proteínas de sinalização G (*RGS*). É um regulador da subunidade alfa das proteínas G, induzindo-as a sua forma inativa (MARTIN-MCCAFFREY et al., 2004). O gene tem função importante na divisão celular, sendo essencial para a primeira divisão mitótica dos embriões. Zigotos de ratos com perda da expressão do gene *RGS14* não foram capazes de progredir para a fase de 2 células (MARTIN-MCCAFFREY et al., 2004). Feuerstein et al. (2012) confirmaram *RGS2* como biomarcador do desenvolvimento de oócitos, com diferenças significativas ($p < 0,05$) nos níveis de expressão deste entre mulheres grávidas e não grávidas.

O gene *PLAG1*, localizado em aproximadamente 25 Mb do cromossomo 14, é frequentemente associado a características reprodutivas em diferentes espécies, com funções relacionadas à proliferação celular por regulação de fatores de crescimento (JUMA et al., 2016). Chen et al. (2007) concluíram que a expressão desse gene no hipotálamo e glândulas pituitárias afeta a produção de ovos em frangos. Hensen et al. (2004), trabalhando com ratos com alelo silenciado do gene *PLAG1*, sugeriram associação significativa com a taxa de crescimento e redução de fertilidade de machos e fêmeas. Em bovinos, variantes genômicas desse gene também têm sido associadas a características de crescimento e fertilidade, e também mudanças no padrão de expressão do gene *IGF1* (FORTES et al., 2012a, 2012b; HAWKEN et al., 2012). Embora a localização exata do gene não tenha sido relatada em nossos resultados, a média de LD (r^2) dos marcadores adjacentes entre a localização deste gene e a região genômica descrita nesse estudo (~ 22-24Mb) foi de $0,30 \pm 0,29$, sugerindo uma possível ligação entre os marcadores.

Cromossomo 18

No cromossomo 18, a janela genômica de maior efeito está situada entre 55 – 56Mb, explicando 1,1% da variância genética total, apresentando o maior número de genes, totalizando 83. Estudos de Pryce et al. (2010) e Ashwell et al. (2004) relataram regiões genômicas significativas para características relacionadas com fertilidade em gado leiteiro entre 61 e 64 Mb e por volta de 36 Mb, respectivamente. Sahana et al. (2010) e Höglund et al. (2012) relataram regiões QTL entre 55 e 62 Mb relacionadas com características relacionadas ao parto.

O gene *FGF21* já foi relacionado com o controle do equilíbrio energético durante estágios avançados de gravidez em camundongos, embora seja quase indetectável em vacas durante esse período, estando presente no início da lactação (CUI et al., 2014). Para esses autores, o aumento da transcrição desse gene estaria relacionado ao metabolismo materno em períodos de insuficiência nutricional, ocorridos ao longo do ciclo reprodutivo, demandando energia proveniente de lipídios. O gene *FUT2* parece ser estimulado pelo hormônio estrogênio, com padrões de expressão correlacionados com o epítipo do carboidrato H-typed-1 fucozilada, presente no epitélio endometrial, envolvida na implantação embrionária (WHITE; KIMBER, 1994). O gene *FUT1* foi associado à taxa de prenhez e também ao número total de leitões nascidos (COCHRAN et al., 2013; HONGYAN et al., 2009). Ortega et al. (2016) reportaram associação significativa entre esse gene e a taxa de concepção de vacas e também a taxa de prenhez em novilhas da raça Holandesa.

O polipeptídeo beta do hormônio luteinizante (*LHB*) é conhecido por promover a ovulação, estimulando os ovários a sintetizar esteroides (BATEMAN et al., 2015). O LH é secretado pela glândula pituitária e desempenha um importante papel nas funções gonadais. Weiss et al. (1992) associaram uma mutação do gene *LHB* com atrasos em atingir a puberdade em humanos. A produção normal desse hormônio é crucial para a correta estimulação dos receptores de LH nas células da granulosa do folículo dominante em desenvolvimento (LUO et al., 2011). Bender et al. (2014) relataram correlação positiva entre a ingestão de ração e a concentração de LD, com ambos os valores extremos (altos ou baixos) prejudiciais ao desenvolvimento embrionário.

Os genes *SLC17A7*, *SLC6A16* e *SLC8A2* fazem parte da família de transportadores de solutos. O gene *SLC17A7* é dependente de estrógeno, codificando a proteína transportadora de glutamato vesicular 1 (Vglut1), a qual é expressa em células do hipocampo (BELLOCCHIO et al., 2010), podendo influenciar a produção estrogênica (SÁRVÁRI et al., 2015). O *SLC6* transporta especificamente neurotransmissores (ex.: dopamina e serotonina), aminoácidos (ex.: GABA) e osmólitos (ex.: betaína, taurina e creatina) (HÖGLUND et al., 2005). O *SLC8A2*

regula a troca eletrogênica de Ca^{2+} e íons Na^{+} através da membrana celular. Choi et al. (2014) relataram expressão de *SLC8As* no endométrio uterino em porcas grávidas. Os autores concluíram que as moléculas de extrusão de cálcio podem estar associadas com o estabelecimento e manutenção da gravidez.

Os genes da família dos genes esfingosina quinase (*SPHK*) são parte da via metabólica do esfingolípido, a qual é altamente ativada na decídua durante a gravidez (MIZUGISHI et al., 2007). Os genes isoformas *SPHK1* e *SPHK2* participaram do percurso da esfingosina lipídica 1-fosfato (*SIP*), associado à vasoconstrição das artérias uterinas humanas durante a gravidez (HUDSON et al., 2007). Mizugishi et al. (2007) relataram que camundongos mutantes para *SPHK1* e *SPHK2* produziram fêmea infértil, com produção reduzida de *SIP*.

6.2.2 Número de Folículos Antrais

Cromossomo 22

Vinte genes foram reportados na janela genômica com efeito significativo do cromossomo 22, a qual foi responsável por 5,56% da variância genética total. Baillet et al. (2011) demonstraram que o gene *TOPAZI* (gene do domínio 1 do grupo *PAZ*, específico para os testículos e ovários) é expresso nas gônadas de ovelhas e camundongos, com forte expressão em fêmeas durante o desenvolvimento das gônadas fetais. Os autores concluíram que o gene tem função importante na meiose das células germinativas. Corroborando, Luangpraseuth-Prosper et al. (2015) discutiram que o gene desempenha um importante papel na progressão da meiose do ovócito e na espermatogênese em camundongos.

Os genes *LYZL* pertencem à classe de lisozimas do tipo c, amplamente distribuídos nas espécies animais com funções bacteriolíticas, desempenhando funções importantes na defesa do hospedeiro contra patógenos (ZHANG et al., 2005). Esses autores detectaram expressão de *LYZL4* no epitélio do epidídimo humano, associando-o com funções fisiológicas da reprodução masculina. Sun et al. (2011), incubando espermatozoides com anticorpos de *LYZL4*, detectou diminuição da capacidade de fertilidade em ratos. Alguns estudos (KWON et al., 2015; MCREYNOLDS et al., 2014) sugeriram que esse gene poderia ser usado como biomarcadores para a fertilidade masculina. Esse gene pode ter função semelhante em fêmeas, participando do desenvolvimento das células germinativas.

Cromossomo 11

Uma janela genômica do chr 11, localizada entre 69 e 71 Mb, explicou 5,50% da variância genética total. Onze genes foram observados nessa região. Entre eles, o *SPDYA* é um gene de codificação de proteínas que possui vias relacionadas com a meiose dos oócitos, com o ontologia relacionado à ligação da proteína quinase (BATEMAN et al., 2015). Esse gene também foi associado à maturação de oócitos em humanos (BARNES et al., 2003) e em *Xenopus* (CHENG et al., 2005). Porter et al. (2002) concluíram que essa classe de genes é um componente essencial das vias de proliferação celular.

O gene *FOSL2* participa na regulação da esteroidogênese, a qual incluem o hormônio estrogênio (HATZIRODOS et al., 2014a). Esse gene codifica componentes da Proteína Ativadora 1 (*AP-1*¹⁷), a qual participa na diferenciação terminal de células da granulosa para células luteais (SHARMA; RICHARDS, 2000). Przygodzka et al. (2015) sugerem que os genes da família *FOS* são ativados no corpo lúteo de suínos, desempenhando um papel importante na luteólise em suínos.

O gene *YPEL5* faz parte da família *YPEL* relacionada à divisão celular e crescimento, tendo alta expressão nos ovócitos (HOSONO et al., 2010; RAJPUT et al., 2013). Hatzirodos et al. (2014b) relataram diferença significativa na expressão desse gene em folículos grandes (> 10 mm) quando comparados com folículos pequenos (< 5 mm) de animais taurinos.

Cromossomo 8

Dezesseis genes foram relatados na região genômica significativa do cromossomo 8, a qual aproximadamente 3,58% da variância genética total. No estudo de Tenghe et al. (2016), as regiões genômicas do cromossomo 8 de vacas Holandesas foram identificadas como sendo associadas com características endócrinas de fertilidade, tais como a atividade lútea. O gene *GATA4* faz parte da família *GATA* de fatores de transcrição de *zinc-finger*, com possível envolvimento no desenvolvimento gonadal em camundongos e humanos (BOUMA et al., 2007; LOURENÇO et al., 2011), sendo que uma mutação em *GATA4* pode comprometer o ciclo promotor do hormônio anti-Mülleriana (AMH). Lourenço et al. (2011) também concluíram que o *GATA4* está relacionado em casos de transtornos de desenvolvimento sexual e defeitos cardíacos congênitos em humanos.

Balboula et al. (2010) discutiram que a inibição da proteína Catepsina B (*CTSB*) interfere positivamente na produção de oócitos e aumenta o número de embriões de boa qualidade em bovinos. Concluem que o *CTBS* poderia ser utilizado como marcador para oócitos

¹⁷ Activator Protein 1

de baixa qualidade, concordando com Bettegowda et al. (2008) que também sugeriram correlação negativa entre *CTBS* e a qualidade de oócitos em bovinos. No entanto, Warzych et al. (2016), que também associou a expressão de *CTBS* em células do cúmulo bovinas com qualidade de oócitos, sugeriu que o gene pode não estar correlacionado com a má qualidade dos oócitos bovinos.

A família de genes *ADAM* (um domínio de desintegrina e metaloprotease) é conhecida por desempenhar papel fundamental no desenvolvimento de folículos e organização e funcionamento dos ovários (BROWN; PRITCHARD; RUSSELL, 2005; FENG et al., 2016; SHOZU et al., 2005) em humanos (GAO et al., 2007; PYUN; KIM; KWACK, 2014) e em mamíferos em geral (RUSSELL; BROWN; DUNNING, 2015). *ADAMST-1* tem sido profundamente estudado sendo associado com o desenvolvimento do folículo através do controle de remodelação da matriz extracelular. Brown, Pritchard e Russell (2005) relataram que os folículos antrais iniciais foram significativamente reduzidos nos ovários de camundongos com alelo silenciado do gene *ADAMST-1*. O gene também foi detectado em células germinativas e somáticas de testículos de ratos, sugerindo associação com reprodução (CHOI et al., 2004; HAN et al., 2009).

Cromossomo 16

A janela genômica associada no cromossomo 16, localizada entre 70 e 72 Mb, abrigou treze genes e explicou 1,96% da variância genética total. La Sala et al. (2015) discutiram que o gene Receptor Acoplado a Proteína G 37 (*GPR37*) está associado a diferenciações gonadais em camundongos. Os animais com alelo silenciado do gene tinham desenvolvimento retardado dos testículos, afetando a proliferação e maturação das células de Sertoli, resultando numa redução significativa da contagem de espermatozoides. Esse gene pode ter funções semelhantes fêmeas, afetando a proliferação das células da granulosa e influenciando, consequentemente, a fertilidade dos animais.

Embora o gene *LGR6* ainda não tenha função descrita, outros genes *LGR* têm sido caracterizados como parte da família do gene responsável pela insulina, a qual também incluiu importantes genes reprodutivos tais como LD, FSH e TSH (LU; LAM; MENON, 2005). Os genes *LGR7* e *LGR8* foram relacionados com o hormônio relaxina, o qual é produzido principalmente pelo corpo lúteo. Camundongos com alelo silenciado do *LGR4* foram inférteis, com desenvolvimento anormal do trato reprodutivo (MENDIVE et al., 2006).

A subunidade reguladora da proteína fosfatase 1 12B (*PPP1R12B*) foi associada à sinalização da insulina (PHAM et al., 2012). Esse gene está envolvido na contratilidade dos

músculos lisos, regulando a hipertrofia das células uterinas em estágios iniciais da gestação (LARTEY et al., 2016). Ryu, Ryu e Lee (2016) discutiram que *PPP1R12B* participa na fosforilação de proteínas que pode influenciar o sistema reprodutivo em humanos.

A proteína-tirosina fosfatase (*PTP*) regula vários processos celulares, como o metabolismo, a adesão celular, a migração celular, o crescimento celular, influenciando também a expressão do gene *TGFβ*, o qual é associado à transição epitelial-mesenquimal e também ao hormônio AMH. Van Eekelen et al. (2010) relataram 48 genes de *PTP* expressos em embriões de peixes zebra, sendo a maioria deles de origem materna.

Cromossomo 2

A janela genômica associada com NF no cromossomo 2 cobriu doze genes e explicou aproximadamente 1,12% da variância genética total. As proteínas intracelulares de ligação a ácidos graxos (*FBP*) estão associadas ao metabolismo e/ou transporte de ácidos graxos de cadeia longa, os quais têm papel importador na maturação meiótica de oócitos que ocorre antes da ovulação (SANCHEZ-LAZO et al., 2014). Os genes da família *SNRNP*¹⁸ são relacionados à maturação de oócitos, fertilização e embriogênese em ratos (PRATHER et al., 1990).

Parisi e Lin (1999) trabalharam com o gene Pumilio 1 (*PUM1*) em *Drosophila*, associando-o com embriogênese, proliferação de células germinativas primordiais, divisão de células-tronco germinativas e processo oogênico. O estudo de Nakahata et al. (2001) associaram o gene à maturação dos oócitos de *Xenopus*. Na presença de anticorpo anti-Pum ou a sobre-expressão de Pum1 tem impacto significativo na maturação do oócito (NAKAHATA et al., 2003). Mak et al. (2016) relataram função essencial desse gene no estabelecimento da foliculogênese, meiose e fertilidade em fêmeas de camundongos.

Cromossomo 15

Apenas dois genes foram cobertos pela janela genômica com efeito significativo no cromossomo 15, localiza entre 8 e 10 Mb. O gene Rho GTPase Ativador da Proteína 42 (*ARHGAP42*) já foi associado como influenciador da pressão arterial do músculo liso vascular (BAI et al., 2013). Tem ontologia genética no processo biológico relacionado à regulação negativa da proteína Rho, que faz parte de uma família de proteínas envolvidas na regulação da divisão celular (YOSHIZAKI et al., 2003). Spencer et al. (2013) associou esse gene com a idade

¹⁸ *Small nuclear ribonucleoprotein*

da menarca e idade na menopausa em mulheres afro-americanas, sugerindo possível associação com o início da puberdade.

6.3 Análises de enriquecimento funcional utilizando os termos MeSH

O termo *munc-18* (homólogo em mamífero de *unc-18*) tem associação importante com a secreção de hormônios nas células da pituitária em ratos. Weiss et al. (2007) discutiram que o estradiol afeta a produção de *munc-18*, com função central na regulação da secreção do hormônio luteinizante (LH). Korteweg et al. (2005) concluíram que *munc-18* está envolvido na secreção de hormônios peptídicos da pituitária anterior.

A Fucose, desoxiaçúcar naturalmente encontrada na superfície das células de mamíferos, foi detectada na zona pelúcida (ZP) e na superfície dos espermatozoides. Romero-Aguirregomez, Matás e Coy (2015) observaram um aumento no número da ligação e penetração de espermatozoides na ZP quando adicionado um meio de cultivo rico em fucose, sugerindo importante função desta substância na fertilização. Estudos em bovinos (LEFEBVRE; LO; SUAREZ, 1997; SUAREZ et al., 1998; TANGHE et al., 2004) confirmaram que a fucose participa de uma interação específica entre espermatozoides e o epitélio do oviduto em bovinos, desempenhando assim um papel importante antes da fertilização.

O termo *Hemoglobin* também foi significativa, sugerindo possível associação no processo de prenhez. A vascularização dentro do ovário tem sido associada à saúde dos ovócitos (VAN BLERKOM, 2000). Brown et al. (2014) encontraram expressão de hemoglobina pelas células foliculares do ovário, sugerindo influência na regulação endócrina induzindo um estágio de hipóxia que interfere o pico de LH.

A catepsina B, uma cisteína protease lisossomal, foi negativamente correlacionada com a qualidade de oócitos e embriões em bovinos (BALBOULA et al., 2013). No estudo de Balboula et al. (2010), o estresse térmico, uma condição normal para animais criados em países tropicais, pode levar ao aumento da expressão da Catepsina B e, consequentemente, prejudicar o desenvolvimento do oócito e a fertilidade feminina.

O termo *Receptors-Neuropeptide* refere-se a receptores que ligam específicos neuropeptídeos com função neural e podem afetar diferentes vias somáticas, como resposta ao estresse. O neuropeptídeo *kisspeptin* tem sido positivamente associado como a expressão do hormônio GnRH e secreções de gonadotrofinas (AMSTALDEN et al., 2014). Para esses autores, a função dos neurônios *kisspeptin* poderia ser influenciada pelo ganho de peso no

período pós-desmama de novilhas. Redmond et al. (2011) relacionaram esse neuropeptídeo com o início da puberdade em cordeiros, pela ativação do eixo hipotalâmico-adeno-hipofisário.

O ácido palmítico é um dos ácidos graxos saturados (AGS) mais comuns presentes no fluido folicular de bovinos, ovinos e suínos (MCEVOY et al., 2000). Mu et al. (2001) associaram a concentração plasmática de AGS com apoptose de células da granulosa em seres humanos, o que pode estar relacionado com distúrbios reprodutivos. Em bovinos, a alta concentração de ácido palmítico no fluido folicular teve efeito negativo sobre a maturação do ovócito, fertilização e taxa de clivagem (LEROY et al., 2005). Para Marei, Wathes e Fouladi-Nashta (2010), a suplementação dietética com AFS pode influenciar a maturação dos oócitos e também o desenvolvimento embrionário precoce. Zeron et al. (2001) relataram concentrações elevadas de fluidos foliculares de AFS em altas temperaturas (verão), o que também está negativamente associado com fertilidade (GARCÍA-ISPIERTO et al., 2007).

6.4 Imputação de touros não genotipados e teste de parentesco

Embora os erros de paternidade possam prejudicar as análises genéticas (BANOS; WIGGANS; POWELL, 2001), existem poucos relatos sobre as taxas de erros em registros de pedigree na literatura, e a maioria destes está em gado leiteiro (ADAMOV; PETKOV; ADAMOV, 2011). Visscher et al. (2002) relataram 10% em média de erros de pedigree na população de animais da raça Holandesa. Embora seja esperado que haja maiores erros de pedigree em bovinos de corte, principalmente em sistemas extensivos, em que os animais não permanecem próximos aos pecuaristas, a taxa de erro encontrada na população em estudo foi semelhante à citada por Visscher et al. (2002), aproximadamente 9%.

A imputação é uma técnica estatística que permite a determinação de genótipos de SNPs que não foram diretamente genotipados. Para isto, depende do LD entre os marcadores genômicos e proximidade genética entre os animais para inferir os genótipos não observados (CARVALHEIRO et al., 2014; VENTURA et al., 2014). Boison et al. (2014), também considerando a informação de cinco progênes de meios-irmãos, apresentaram $0,90 \pm 0,03$ de correlação entre genótipos mascarados e observados. Para Berry, McClure e Mullen (2014), a imputação considerando pelo menos a informações genômicas de cinco de meios-irmãos pode ser uma opção acurada quando os genótipos não estão disponíveis.

6.5 Estudo de Haplótipos

As informações dos SNPs são normalmente interpretadas como genótipos não faseados. Isso significa que não se sabe a origem (materno ou paterno) dos alelos dos haplótipos parentais de uma região genômica de interesse. A fase haplotípica é formada por associações não aleatórias entre dois ou mais loci, resultando em regiões genômicas que são herdadas pela população. Essa informação favorece as informações genômicas, podendo ser aplicada a diferentes interesses como imputação, detecção de erro genotípico, pontos de recombinação, identificação de regiões idênticas por descendência (IBD¹⁹), entre outros (BROWNING; BROWNING, 2011).

As análises estatísticas de contraste entre o efeito dos alelos haplotípicos paternos indicaram que alguns touros heterozigotos possuem alelos favoráveis para a PP e NF. Apesar da baixa correlação genética entre as características, alguns touros apresentaram haplótipos com efeito significativo para ambas as características. Também se observou grupos de meio irmãos carregando o mesmo alelo haplotípico, sendo este favorável ou não para a característica. Contudo, uma análise visual dos dois primeiros componentes principais da matriz G , os quais explicaram aproximadamente 14,7% da variância total, estimados a partir dos genótipos desses animais não demonstraram nenhum padrão que indicasse prováveis famílias com os alelos de interesse (Figura 13). Ferramentas genômicas poderiam ser usadas para aumentar a frequência dessas regiões nas populações, resultando em maior ganho genético dessas características. Aumentar o conhecimento sobre regiões genômicas associadas com fenótipos de interesse pode ser uma ferramenta útil para o melhoramento genético, elegendo regiões genômicas de maior influência sobre a variância fenotípica (BOICHARD et al., 2016).

¹⁹ *Identical by descent*

Figura 13 – Cluster considerando os 2 primeiros autovetores que mais explicaram a variância dos valores da matriz G (*Eigenvector*) dos touros que apresentaram blocos haplotípicos com efeito significativo ($p < 0,10$) para prenhez precoce (PP), número de folículos antrais (NF) ou ambas características.

7. Conclusão

O estudo forneceu evidências da complexidade genômica envolvida nas características reprodutivas: prenhez precoce (PP) e número de folículos antrais (NF) em novilhas da raça Nelore. Os estudos de associação genômica ampla identificaram regiões genômicas nos cromossomos 5, 14 e 18 associadas com PP, enquanto que regiões nos cromossomos 2, 8, 11, 14, 15, 16 e 22 foram associadas com NF. A janela genômica do cromossomo 14 foi associada a ambas as características, representando possível efeito pleiotrópico. Os termos MeSH *Munc-18 Proteins*, *Fucose* e *Hemoglobins* foram relacionados com PP, e *Cathepsin B*, *Receptors-Neuropeptide* e *Palmitic Acid* foram para NF, apontando para importantes processos biológicos que estão relacionados com a expressão dos fenótipos, fornecendo informações úteis para aumentar o conhecimento sobre os componentes genéticos dessas características. Alelos haplotípicos com efeito significativo sobre os fenótipos estudados foram detectados em touros heterozigotos para as regiões genômicas de interesse. Estes poderiam ser selecionados na população visando aumentar o valor genético das características estudadas.

8. Referências Bibliográfica

- ABEYGUNAWARDENA, H.; DEMATAWEWA, C. M. B. Pre-pubertal and postpartum anestrus in tropical Zebu cattle. **Animal Reproduction Science**, v. 82–83, p. 373–387, 2004.
- ADAMOV, N.; PETKOV, V.; ADAMOV, M. Microsatellite Markers for Pedigree Verification in Cattle. **Macedonian Journal of Animal Science**, v. 1, n. 1, p. 9–15, 2011.
- AGUILAR, I. et al. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of dairy science**, v. 93, n. 2, p. 743–52, 2010.
- AGUILAR, I. et al. **PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs**. 10th World Congress of Genetics Applied to Livestock Production. **Anais...IEEE**, jul. 2013Disponível em: <<http://ieeexplore.ieee.org/document/6632596/>>
- AIKEN, C. E. et al. Decreased ovarian reserve, dysregulation of mitochondrial biogenesis, and increased lipid peroxidation in female mouse offspring exposed to an obesogenic maternal diet. **The FASEB Journal**, p. 1–9, 23 dez. 2015.
- AKANNO, E. C. et al. Genome-wide association for heifer reproduction and calf performance traits in beef cattle. **Genome**, v. 58, p. 549–557, 2015.
- AKEN, B. L. et al. The Ensembl Gene Annotation System. **Database : the journal of biological databases and curation**, v. 2016, p. baw093, 2016.
- ALLAN, M. F. et al. Confirmation of quantitative trait loci using a low-density single nucleotide polymorphism map for twinning and ovulation rate on bovine chromosome 5. **Journal of Animal Science**, v. 87, n. 1, p. 46–56, 2009.
- AMSTALDEN, M. et al. Reproduction Symposium: Hypothalamic neuropeptides and the nutritional programming of puberty in heifers. **Journal of Animal Science**, v. 92, n. 8, p. 3211–3222, 2014.
- ASHWELL, M. S. et al. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. **Journal of dairy science**, v. 87, n. 2, p. 468–475, 2004.
- ASSIDI, M. et al. Biomarkers of human oocyte developmental competence expressed in cumulus cells before ICSI: A preliminary study. **Journal of Assisted Reproduction and Genetics**, v. 28, n. 2, p. 173–188, 2011.
- BAI, X. et al. The smooth muscle-selective RhoGAP GRAF3 is a critical regulator of vascular tone and hypertension. **Nature communications**, v. 4, p. 2910, 2013.
- BAILLET, A. et al. TOPAZ1, a novel germ cell-specific expressed gene conserved during evolution across vertebrates. **PLoS ONE**, v. 6, n. 11, 2011.
- BALBOULA, A. Z. et al. Cathepsin B activity is related to the quality of bovine cumulus oocyte complexes and its inhibition can improve their developmental competence. **Molecular Reproduction and Development**, v. 77, n. 5, p. 439–448, 2010.
- BALBOULA, A. Z. et al. Cathepsin B activity has a crucial role in the developmental competence of bovine cumulus-oocyte complexes exposed to heat shock during in vitro maturation. **Reproduction**, v. 146, n. 4, p. 407–417, 2013.
- BANOS, G.; WIGGANS, G. R.; POWELL, R. L. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. **Journal of dairy science**, v. 84, p. 2523–2529, 2001.
- BARBAN, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. **Nature Genetics**, v. 48, n. 12, p. 1462–1472, 31 out. 2016.

- BARNES, E. A. et al. Human Spyl promotes survival of mammalian cells following DNA damage. **Cancer Research**, v. 63, n. 13, p. 3701–3707, 2003.
- BARUSELLI, P. S. et al. Relationship between follicle population, AMH concentration and fertility in cattle. **Anim. Reprod**, v. 12, n. 3, p. 487–497, 2015.
- BATEMAN, A. et al. UniProt: A hub for protein information. **Nucleic Acids Research**, v. 43, n. D1, p. D204–D212, 2015.
- BELLOCCHIO, E. E. et al. Uptake of glutamate into synaptic vesicles by an inorganic phosphate transporter. **Science**, v. 289, p. 957–960, 2010.
- BENDER, R. W. et al. Effects of acute feed restriction combined with targeted use of increasing luteinizing hormone content of follicle-stimulating hormone preparations on ovarian superstimulation, fertilization, and embryo quality in lactating dairy cows. **Journal of Dairy Science**, v. 97, n. 2, p. 764–778, 2014.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **J. R. Statist. Soc.**, v. 57, n. 1, p. 289–300, 1995.
- BERRY, D. P. et al. Genome-wide associations for fertility traits in Holstein–Friesian dairy cows using data from experimental research herds in four European countries. **Animal**, v. 6, p. 1206–1215, 2012.
- BERRY, D. P. et al. Imputation of ungenotyped parental genotypes in dairy and beef cattle from progeny genotypes. **Animal : an international journal of animal bioscience**, v. 8, n. 6, p. 895–903, 2014.
- BERRY, D. P.; KEARNEY, J. F. Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. **Animal**, v. 5, n. 8, p. 1162–1169, 2011.
- BERRY, D. P.; MCCLURE, M. C.; MULLEN, M. P. Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. **Journal of Animal Breeding and Genetics**, v. 131, n. 3, p. 165–172, 2014.
- BETTEGOWDA, A. et al. Identification of novel bovine cumulus cell molecular markers predictive of oocyte competence: functional and diagnostic implications. **Biology of reproduction**, v. 79, n. 2, p. 301–9, 2008.
- BIEGELMEYER, P. et al. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. **BMC genetics**, v. 17, n. 1, p. 32, 2016.
- BOICHARD, D. et al. Genomic selection in domestic animals: Principles, applications and perspectives. **Comptes Rendus Biologies**, v. 339, n. 7, p. 274–277, 2016.
- BOISON, S. A. et al. Imputation of non-genotyped individuals using genotyped progeny in Nellore, a Bos indicus cattle breed. **Livestock Science**, v. 166, p. 176–189, ago. 2014.
- BOISON, S. A. et al. Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (Bos indicus) dairy cattle: Comparison of commercially available SNP chips. **Journal of dairy science**, v. 98, n. 7, p. 4969–89, 2015.
- BOLORMAA, S. et al. Genome-wide association studies for feedlot and growth traits in cattle. **Journal of animal science**, v. 89, n. 6, p. 1684–97, jun. 2011.
- BOLORMAA, S. et al. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. **Animal Genetics**, v. 46, n. 5, p. 544–556, 2015.
- BOUMA, G. J. et al. Correct dosage of Fog2 and Gata4 transcription factors is critical for fetal testis development in mice. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 38, p. 14994–9, 2007.
- BROOKES, A. J. The essence of SNPs. **Gene**, v. 234, n. 2, p. 177–186, 1999.

- BROWN, H. M. et al. Hemoglobin: A Gas-Transport Molecule That Is Hormonally Regulated in the Ovarian Follicle in Mice and Humans. **Biology of reproduction**, v. 92, n. November 2014, p. 1–19, 2014.
- BROWN, H. M.; PRITCHARD, M.; RUSSELL, D. L. 225. Early ovarian follicle dysgenesis in a disintegrin and metalloproteinase with thrombospondin motifs type 1 (ADAMTS-1) null mice. **Reproduction, Fertility and Development**, v. 17, n. 9, p. 87, 2005.
- BROWNING, B. L.; BROWNING, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. **American Journal of Human Genetics**, v. 84, n. 2, p. 210–223, 2008.
- BROWNING, S. R.; BROWNING, B. L. Haplotype phasing: existing methods and new developments. **Nature reviews. Genetics**, v. 12, n. 10, p. 703–714, 2011.
- BURNS, B. M.; FORDYCE, G.; HOLROYD, R. G. A review of factors that impact on the capacity of beef cattle females to conceive, maintain a pregnancy and wean a calf-Implications for reproductive efficiency in northern Australia. **Animal Reproduction Science**, v. 122, n. 1–2, p. 1–22, 2010.
- CALUS, M. P. L. et al. Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. **Animal : an international journal of animal bioscience**, p. 1–11, 2014.
- CAMMACK, K. M.; THOMAS, M. G.; ENNS, R. M. Review: Reproductive traits and their heritabilities. **The Professional Animal Scientist**, v. 25, p. 517–528, 2009.
- CANAVEZ, F. C. et al. Genome sequence and assembly of *bos indicus*. **Journal of Heredity**, v. 103, n. 3, p. 342–348, 2012.
- CANTOR, R. M.; LANGE, K.; SINSHEIMER, J. S. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. **American Journal of Human Genetics**, v. 86, n. 1, p. 6–22, jan. 2010.
- CARVALHEIRO, R. et al. Accuracy of genotype imputation in Nelore cattle. **Genetics Selection Evolution**, v. 46, n. 1, p. 69, 2014.
- CESAR, A. S. et al. Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. **BMC genetics**, v. 15, n. 1, p. 39, 25 mar. 2014.
- CHAN, E. K. F.; HAWKEN, R.; REVERTER, A. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. **Animal genetics**, v. 40, n. 2, p. 149–56, abr. 2009.
- CHANG, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. **GigaScience**, v. 4, n. 1, p. 7, 2015.
- CHEN, C. F. et al. Laying traits and underlying transcripts, expressed in the hypothalamus and pituitary gland, that were associated with egg production variability in chickens. **Theriogenology**, v. 68, n. 9, p. 1305–1315, 2007.
- CHENG, A. et al. Identification and comparative analysis of multiple mammalian speedy/ringo proteins. **Cell Cycle**, v. 4, n. 1, p. 155–165, 2005.
- CHOI, I. et al. Characterization and comparative genomic analysis of intronless Adams with testicular gene expression. **Genomics**, v. 83, n. 4, p. 636–646, 2004.
- CHOI, Y. et al. Calcium extrusion regulatory molecules: Differential expression during pregnancy in the porcine uterus. **Domestic Animal Endocrinology**, v. 47, n. 1, p. 1–10, 2014.
- CHUD, T. C. S. et al. Strategies for genotype imputation in composite beef cattle. **BMC genetics**, v. 16, n. 1, p. 99, 2015.
- CLEVELAND, M. A.; HICKEY, J. M. Practical implementation of cost-effective genomic selection in

- commercial pig breeding using imputation. **Journal of Animal Science**, v. 91, n. 8, p. 3583–3592, 2013.
- COCHRAN, S. D. et al. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. **BMC genetics**, v. 14, n. 1, p. 49, 2013.
- COLE, J. B. et al. Distribution and location of genetic effects for dairy traits. **Journal of dairy science**, v. 92, n. 6, p. 2931–2946, 2009.
- CUI, Y. et al. Hepatic FGF21 production is increased in late pregnancy in the mouse. **American journal of physiology. Regulatory, integrative and comparative physiology**, v. 307, n. 3, p. R290-8, 2014.
- CUSHMAN, R. A. et al. Evaluation of antral follicle count and ovarian morphology in crossbred beef cows: Investigation of influence of stage of the estrous cycle, age, and birth weight. **Journal of Animal Science**, v. 87, n. 6, p. 1971–1980, 1 jun. 2009.
- DAETWYLER, H. D. et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. **Genetics**, v. 193, n. 2, p. 347–65, fev. 2013.
- DANECEK, P. et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156–2158, 2011.
- DARWASH, A. O.; LAMMING, G. E.; WOOLLIAMS, J. A. The phenotypic association between the interval to post-partum ovulation and traditional measures of fertility in dairy cattle. **Animal Science**, v. 65, n. 1, p. 9–16, 2 set. 1997.
- DAY, M. L.; NOGUEIRA, G. P. Management of age at puberty in beef heifers to optimize efficiency of beef production. **Animal Frontiers**, v. 3, n. 4, p. 6–11, 26 set. 2013.
- DE VRIES, A. Economic Value of Pregnancy in Dairy Cattle. **Journal of Dairy Science**, v. 89, n. 10, p. 3876–3885, 2006.
- DEKKERS, J. C. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. **Journal of animal science**, v. 82 E-Suppl, 2004.
- DJARI, A. et al. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. **BMC genomics**, v. 14, n. 1, p. 307, 7 maio 2013.
- DUMESIC, D. A. et al. Oocyte environment: Follicular fluid and cumulus cells are critical for oocyte health. **Fertility and Sterility**, v. 103, n. 2, p. 303–316, 2015.
- DURLINGER, A.; VISSER, J.; THEMME, A. Regulation of ovarian function: the role of anti-Mullerian hormone. **Reproduction**, v. 124, n. 5, p. 601–609, 1 nov. 2002.
- ECHTERNKAMP, S. E. et al. Effects of ovulation rate and fetal number on fertility in twin-producing cattle. **Journal of Animal Science**, v. 85, n. 12, p. 3228–3238, 2007.
- ELER, J. P. et al. Genetic evaluation of the probability of pregnancy at 14 months for Nellore heifers. **Journal of Animal Science**, v. 80, p. 951–954, 2002.
- ELER, J. P. et al. Additive genetic relationships between heifer pregnancy and scrotal circumference in Nellore cattle. **Journal of Animal Science**, v. 82, p. 2519–2527, 2004.
- ELER, J. P. et al. Genetic correlation between heifer pregnancy and scrotal circumference measured at 15 and 18 months of age in Nellore cattle. **Genetics and Molecular Research**, v. 5, n. 4, p. 569–80, jan. 2006.
- ELER, J. P. et al. Genetic analysis of average annual productivity of Nellore breeding cows (COWPROD). **Genetics and Molecular Research**, v. 7, n. 1, p. 234–42, jan. 2008.
- ELER, J. P. et al. Genetic relationships among traits related to reproduction and growth of Nelore females. **Theriogenology**, v. 82, n. 5, p. 708–14, 15 set. 2014.

- ESPIGOLAN, R. et al. Study of whole genome linkage disequilibrium in Nellore cattle. **BMC genomics**, v. 14, p. 305, 2013.
- FANCHIN, R. et al. Serum anti-Mullerian hormone is more strongly related to ovarian follicular status than serum inhibin B, estradiol, FSH and LH on day 3. **Human Reproduction**, v. 18, n. 2, p. 323–327, 1 fev. 2003.
- FENG, L. et al. ADAM10–Notch signaling governs the recruitment of ovarian pregranulosa cells and controls folliculogenesis in mice. **Journal of Cell Science**, v. 129, n. 11, p. 2202–2212, 2016.
- FERNANDO, R. L.; GARRICK, D. J. **GenSel-User manual for a portfolio of genomic selection related analyses**, 2012.
- FERNANDO, R. L.; GARRICK, D. J. Bayesian methods applied to GWAS. In: GONDRO, C.; VAN DER WERF, J.; HAYES, B. (Eds.). **Genome-Wide Association Studies and Genomic Prediction**. Methods in Molecular Biology. Totowa, NJ: Springer, 2013. v. 1019p. 237–274.
- FEUERSTEIN, P. et al. Genomic assessment of human cumulus cell marker genes as predictors of oocyte developmental competence: Impact of various experimental factors. **PLoS ONE**, v. 7, n. 7, 2012.
- FONSECA, P. A. DE S. et al. Retelling the recent evolution of genetic diversity for Guzerá: Inferences from LD decay, runs of homozygosity and Ne over the generations. **Livestock Science**, v. 193, n. April, p. 110–117, 2016.
- FORTES, M. R. S. et al. Association weight matrix for the genetic dissection of puberty in beef cattle. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 31, p. 13642–13647, 2010.
- FORTES, M. R. S. et al. A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. **Journal of Animal Science**, v. 89, n. 6, p. 1669–1683, 2011.
- FORTES, M. R. S. et al. Finding genes for economically important traits : Brahman cattle puberty. **Animal Production Science**, v. 52, p. 143–150, 2012a.
- FORTES, M. R. S. et al. Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. **Biology of reproduction**, v. 87, n. 3, p. 58, 2012b.
- FORTES, M. R. S. et al. Genomic regions associated with fertility traits in male and female cattle: advances from microsatellites to high-density chips and beyond. **Animal reproduction science**, v. 141, n. 1–2, p. 1–19, 2013a.
- FORTES, M. R. S. et al. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. **Animal Genetics**, v. 44, n. 6, p. 636–647, 2013b.
- FORTES, M. R. S. et al. The IGF1 pathway genes and their association with age of puberty in cattle. **Animal Genetics**, v. 44, n. 1, p. 91–95, 2013c.
- FORTES, M. R. S. et al. Post-partum anoestrus in tropical beef cattle: A systems approach combining gene expression and genome-wide association results. **Livestock Science**, v. 166, n. 1, p. 158–166, 2014.
- FREIMER, N. B.; SABATTI, C. Guidelines for association studies in Human Molecular Genetics. **Human Molecular Genetics**, v. 14, n. 17, p. 2481–2483, 2005.
- FULTON, J. E. Genomic selection for poultry breeding. **Animal Frontiers**, v. 2, n. 1, p. 30–36, 1 jan. 2012.
- GABRIEL, S. B. et al. The structure of haplotype blocks in the human genome. **Science (New York, N.Y.)**, v. 296, n. 5576, p. 2225–9, 2002.
- GADDIS, K. L. P.; NULL, D. J.; COLE, J. B. Explorations in genome-wide association studies and

network analyses with dairy cattle fertility traits. **Journal of Dairy Science**, v. 99, p. 1–16, 2016.

GAO, S. et al. FSH stimulates the expression of the ADAMTS-16 protease in mature human ovarian follicles. **Molecular Human Reproduction**, v. 13, n. 7, p. 465–471, 2007.

GARCÍA-ISPIERTO, I. et al. Climate factors affecting conception rate of high producing dairy cows in northeastern Spain. **Theriogenology**, v. 67, n. 8, p. 1379–1385, 2007.

GARCÍA-RUIZ, A. et al. **Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection**. Proceedings of the National Academy of Sciences. **Anais...**2016

GARRICK, D. J.; FERNANDO, R. L. Implementing a QTL Detection Study (GWAS) Using Genomic Prediction Methodology. In: **Methods in molecular biology (Clifton, N.J.)**. [s.l: s.n.]. v. 1019p. 275–298, 2013.

GEER, L. Y. et al. The NCBI BioSystems database. **Nucleic Acids Research**, v. 38, n. SUPPL.1, p. 492–496, 2009.

GIMENES, L. U. et al. The interval between the emergence of pharmacologically synchronized ovarian follicular waves and ovum pickup does not significantly affect invitro embryo production in *Bos indicus*, *Bos taurus*, and *Bubalus bubalis*. **Theriogenology**, v. 83, n. 3, p. 385–393, 2015.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 323–30, dez. 2007.

GOLDSTEIN, D. B. Common Genetic Variation and Human Traits. **The New England Journal of Medicine**, v. v.360, p. 1696–1698, 2009.

GRAY, C. A et al. Developmental biology of uterine glands. **Biology of reproduction**, v. 65, n. 5, p. 1311–1323, 2001.

GUARINI, A. R. et al. Genetic relationship among reproductive traits in Nellore cattle. **Animal**, v. 9, n. 5, p. 760–765, 2014.

GUIMARÃES-YOUNG, A. et al. Conditional deletion of Sox17 reveals complex effects on uterine adenogenesis and function. **Developmental Biology**, v. 414, n. 2, p. 219–227, 2016.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389–97, dez. 2007.

HAN, C. et al. Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6 with an ADAM complex required for fertilization in mice. **Biology of reproduction**, v. 80, n. 5, p. 1001–8, 2009.

HATZIRODOS, N. et al. Transcriptome profiling of the theca interna from bovine ovarian follicles during atresia. **PLoS ONE**, v. 9, n. 6, 2014a.

HATZIRODOS, N. et al. Transcriptome profiling of granulosa cells of bovine ovarian follicles during growth from small to large antral sizes. **BMC Genomics**, v. 15, n. 1, p. 24, 2014b.

HAWKEN, R. J. et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. **Journal of Animal Science**, v. 90, p. 1398–1410, 2012.

HAYES, B. J. et al. Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. **Genetical research**, v. 89, n. 4, p. 215–20, ago. 2007.

HEIDARITABAR, M. et al. Accuracy of imputation using the most common sires as reference population in layer chickens. **BMC Genetics**, v. 16, n. 1, p. 101, 18 dez. 2015.

HENSEN, K. et al. Targeted disruption of the murine *Plag1* proto-oncogene causes growth retardation and

reduced fertility. **Development Growth and Differentiation**, v. 46, n. 5, p. 459–470, 2004.

HICKEY, J. M. et al. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. **Crop Science**, v. 52, n. 2, p. 654–663, 2012.

HICKEY, J. M. et al. Genotype Imputation to Increase Sample Size in Pedigreed Populations. In: [s.l: s.n.]. p. 395–410, 2013.

HILL, W. G. Estimation of effective population size from data on linkage disequilibrium. **Genetical Research**, v. 38, n. 3, p. 209, 14 dez. 1981.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, v. 38, n. 6, p. 226–231, jun. 1968.

HÖGLUND, J. K. et al. Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. **BMC genetics**, v. 13, n. 1, p. 71, 2012.

HÖGLUND, J. K. et al. Genome-wide association study for female fertility in Nordic Red cattle. **BMC genetics**, v. 16, n. 1, p. 110, 2015.

HÖGLUND, P. J. et al. The repertoire of solute carriers of family 6: Identification of new human and rodent genes. **Biochemical and Biophysical Research Communications**, v. 336, n. 1, p. 175–189, 2005.

HONGYAN, Z. et al. Polymorphism of FUT1 gene and its relationship with litter size in northeast Hebao pigs. **Animal Husbandry and Feed Science**, v. 1, p. 1–3, 2009.

HOSONO, K. et al. YPEL5 protein of the YPEL gene family is involved in the cell cycle progression by interacting with two distinct proteins RanBPM and RanBP10. **Genomics**, v. 96, n. 2, p. 102–111, 2010.

HU, Z. L.; PARK, C. A.; REECY, J. M. Developmental progress and current status of the Animal QTLdb. **Nucleic Acids Research**, v. 44, n. D1, p. D827–D833, 2016.

HUDSON, N. K. et al. Modulation of human arterial tone during pregnancy: the effect of the bioactive metabolite sphingosine-1-phosphate. **Biol Reprod**, v. 77, n. 1, p. 45–52, 2007.

IRANO, N. et al. Genome-wide association study for indicator traits of sexual precocity in Nellore cattle. **PLoS ONE**, v. 11, n. 8, p. 1–14, 2016.

IRELAND, J. J. et al. Variation in the ovarian reserve is linked to alterations in intrafollicular estradiol production and ovarian biomarkers of follicular differentiation and oocyte quality in cattle. **Biology of reproduction**, v. 80, n. 5, p. 954–64, 2009.

IRIE, N. et al. SOX17 is a critical specifier of human primordial germ cell fate. **Cell**, v. 160, n. 1–2, p. 253–268, 2015.

JIMENEZ-KRASSEL, F. et al. Concentration of anti-Müllerian hormone in dairy heifers is positively associated with productive herd life. **Journal of dairy science**, v. 98, n. 5, p. 3036–45, 2015.

JUMA, A. R. et al. Emerging role of PLAG1 as a regulator of growth and reproduction. **Journal of Endocrinology**, v. 228, n. 2, p. R45–R56, 2016.

KADARMIDEEN, H. N. The use of genetic and genomic technologies to improve reproductive performance in cattle. **AUSTRALIAN COLLEGE OF VETERINARY - SCIENTIST ANNUAL CONFERENCE**, p. 36–44, 2010.

KAPPES, S. M. et al. Initial results of genomic scans for ovulation rate in a cattle population selected for increased twinning rate. **Journal of Animal Science**, v. 78, n. 12, p. 3053–3059, 2000.

KHATKAR, M. S. S.; RANDHAWA, I. A. S. A. S.; RAADSMA, H. W. W. Meta-assembly of genomic regions and variants associated with female reproductive efficiency in cattle. **Livestock Science**, v. 166, n. 1, p. 144–157, ago. 2014.

- KORTEWEG, N. et al. The role of Munc18-1 in docking and exocytosis of peptide hormone vesicles in the anterior pituitary. **Biology of the cell / under the auspices of the European Cell Biology Organization**, v. 97, n. 6, p. 445–455, 2005.
- KWON, W.-S. et al. Discovery of Predictive Biomarkers for Litter Size in Boar Spermatozoa. **Molecular & Cellular Proteomics**, v. 14, n. 5, p. 1230–1240, maio 2015.
- LA SALA, G. et al. Modulation of Dhh signaling and altered Sertoli cell function in mice lacking the GPR37-prosaposin receptor. **FASEB journal : official publication of the Federation of American Societies for Experimental Biology**, v. 29, n. 5, p. 2059–69, 2015.
- LARTEY, J. et al. Altered Expression of Human Smooth Muscle Myosin Phosphatase Targeting (MYPT) Isovariants with Pregnancy and Labor. **Plos One**, v. 11, n. 10, p. e0164352, 2016.
- LEFEBVRE, R.; LO, M. C.; SUAREZ, S. S. Bovine sperm binding to oviductal epithelium involves fucose recognition. **Biology of reproduction**, v. 56, n. 5, p. 1198–204, 1997.
- LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of dairy science**, v. 92, n. 9, p. 4656–4663, 2009.
- LEROY, J. L. M. R. et al. Non-esterified fatty acids in follicular fluid of dairy cows and their effect on developmental capacity of bovine oocytes in vitro. **Reproduction**, v. 130, n. 4, p. 485–495, 2005.
- LEWONTIN, R. . C. The interactions of selection and linkage II. Optimus models. **Genetics**, v. 50(4), n. 1956, p. 49–67, 1964.
- LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589–595, 2010.
- LOBO, G. P. et al. Genetics and diet regulate vitamin A production via the homeobox transcription factor ISX. **Journal of Biological Chemistry**, v. 288, n. 13, p. 9017–9027, 2013.
- LOURENÇO, D. et al. Loss-of-function mutation in GATA4 causes anomalies of human testicular development. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 4, p. 1597–602, 2011.
- LU, C.; LAM, H. N.; MENON, R. K. New members of the insulin family: Regulators of metabolism, growth and now...reproduction. **Pediatric Research**, v. 57, n. 5 II, p. 70–73, 2005.
- LUANGPRASEUTH-PROSPER, A. et al. TOPAZ1, a germ cell specific factor, is essential for male meiotic progression. **Developmental Biology**, v. 406, n. 2, p. 158–171, 2015.
- LUNA-NEVAREZ, P. et al. Single nucleotide polymorphisms in the growth hormone-insulin-like growth factor axis in straightbred and crossbred Angus, Brahman, and Romosinuano heifers: Population genetic analyses and association of genotypes with reproductive phenotypes. **Journal of Animal Science**, v. 89, n. 4, p. 926–934, 2011.
- LUO, W. et al. The Role of Luteinizing Hormone in Regulating Gene Expression During Selection of a Dominant Follicle in Cattle. **Biology of Reproduction**, v. 84, n. 2, p. 369–378, 1 fev. 2011.
- MAK, W. et al. An Important Role of Pumilio 1 in Regulating the Development of the Mammalian Female Germline. **Biology of reproduction**, v. 94, n. 6, p. 134, 2016.
- MAREI, W. F.; WATHES, D. C.; FOULADI-NASHTA, A. A. Impact of linoleic acid on bovine oocyte maturation and embryo development. **Reproduction**, v. 139, n. 6, p. 979–988, 2010.
- MARTIN-MCCAFFREY, L. et al. RGS14 is a mitotic spindle protein essential from the first division of the mammalian zygote. **Developmental Cell**, v. 7, n. 5, p. 763–769, 2004.
- MCCARTHY, M. I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. **Nature Reviews Genetics**, v. 9, n. 5, p. 356–369, 2008.

- MCDANELD, T. G. et al. Genomewide association study of reproductive efficiency in female cattle. **Journal of Animal Science**, v. 92, n. 5, p. 1945–1957, 2014.
- MCEVOY, T. G. et al. Fatty acid composition of lipids in immature cattle, pig and sheep oocytes with intact zona pellucida. **J Reprod Fertil**, v. 118, n. 1, p. 163–170, 2000.
- MCKENNA, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, v. 20, n. 9, p. 1297–1303, 1 set. 2010.
- MCRAE, A. F. et al. Linkage disequilibrium in domestic sheep. **Genetics**, v. 160, n. 3, p. 1113–1122, 2002.
- MCREYNOLDS, S. et al. Toward the identification of a subset of unexplained infertility: a sperm proteomic approach. **Fertility and Sterility**, v. 102, n. 3, p. 692–699, set. 2014.
- MENDIVE, F. et al. Defective postnatal development of the male reproductive tract in LGR4 knockout mice. **Developmental Biology**, v. 290, n. 2, p. 421–434, 2006.
- MEUWISSEN, T. H.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819–29, abr. 2001.
- MEUWISSEN, T.; HAYES, B.; GODDARD, M. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, v. 6, n. 1, p. 6, 2016.
- MISZTAL, I. et al. BLUPF90 and related programs (BGF90). **Proc. 7th World Congr. Genet. Appl. Livest. Prod.**, p. 1–2, 2002.
- MIZUGISHI, K. et al. Maternal disturbance in activated sphingolipid metabolism causes pregnancy loss in mice. **Journal of Clinical Investigation**, v. 117, n. 10, p. 2993–3006, 1 out. 2007.
- MOKHTARI, M. S. et al. Genetic relationship between heifers and cows fertility and milk yield traits in first-parity Iranian Holstein dairy cows. **Livestock Science**, v. 182, n. November, p. 76–82, 2015.
- MOROTA, G. et al. An application of MeSH enrichment analysis in livestock. **Animal Genetics**, v. 46, n. 4, p. 381–387, 2015.
- MOSSA, F. et al. Low numbers of ovarian follicles ≥ 3 mm in diameter are associated with low fertility in dairy cows. **Journal of dairy science**, v. 95, n. 5, p. 2355–61, 2012.
- MOUNTAIN, J. L. et al. Evolution of Modern Humans: Evidence from Nuclear DNA Polymorphisms. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 337, n. 1280, p. 159–165, 29 ago. 1992.
- MU, Y. M. et al. Saturated FFAs, palmitic acid and stearic acid, induce apoptosis in human granulosa cells. **Endocrinology**, v. 142, n. 8, p. 3590–3597, 2001.
- NAKAHATA, S. et al. Biochemical Identification of Xenopus Pumilio as a Sequence-specific Cyclin B1 mRNA-binding Protein That Physically Interacts with a Nanos Homolog, Xcat-2, and a Cytoplasmic Polyadenylation Element-binding Protein. **Journal of Biological Chemistry**, v. 276, n. 24, p. 20945–20953, 15 jun. 2001.
- NAKAHATA, S. et al. Involvement of Xenopus Pumilio in the translational regulation that is specific to cyclin B1 mRNA during oocyte maturation. **Mechanisms of Development**, v. 120, n. 8, p. 865–880, ago. 2003.
- ORTEGA, M. S. et al. Use of single nucleotide polymorphisms in candidate genes associated with daughter pregnancy rate for prediction of genetic merit for reproduction in Holstein cows. **Animal Genetics**, v. 47, n. 3, p. 288–297, 2016.
- PARISI, M.; LIN, H. The Drosophila pumilio gene encodes two functional protein isoforms that play multiple roles in germline development, gonadogenesis, oogenesis and embryogenesis. **Genetics**, v. 153, n.

1, p. 235–250, 1999.

PEARSON, T. A; MANOLIO, T. A. How to interpret a genome-wide association study. **Journal of the American Medical Association**, v. 299, n. 11, p. 1335–44, 19 mar. 2008.

PETERS, S. O. et al. Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. **Journal of animal science**, v. 90, n. 10, p. 3398–409, out. 2012.

PETERS, S. O. et al. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. **Jornal of Animal Science**, v. 91, p. 605–612, 2013.

PHAM, K. et al. Insulin-stimulated phosphorylation of protein phosphatase 1 regulatory subunit 12B revealed by HPLC-ESI-MS/MS. **Proteome science**, v. 10, n. 1, p. 52, 2012.

PORTER, L. A. et al. Human speedy: A novel cell cycle regulator that enhances proliferation through activation of Cdk2. **Journal of Cell Biology**, v. 157, n. 3, p. 357–366, 2002.

PORTO-NETO, L. R. et al. Molecular genetic approaches for identifying the basis of variation in resistance to tick infestation in cattle. **Veterinary Parasitology**, v. 180, n. 3–4, p. 165–172, 2011.

PORTO-NETO, L. R. et al. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. **BMC genomics**, v. 14, p. 876, 2013.

PORTO-NETO, L. R. et al. Genome-wide association for the outcome of fixed-time artificial insemination of Brahman heifers in northern Australia. **J. Anim. Sci**, v. 93, p. 5119–5127, 2015.

PRATHER, R. et al. U3 snRNPs and nucleolar development during oocyte maturation, fertilization and early embryogenesis in the mouse: U3 snRNA and snRNPs are not regulated coordinate with other snRNAs and snRNPs. **Developmental Biology**, v. 138, n. 2, p. 247–255, 1990.

PRYCE, J. E. 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, n. 7, p. 3331–3345, 2010.

PRYCE, J. E.; HAYES, B. J.; GODDARD, M. E. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. **Journal of dairy science**, v. 95, n. 1, p. 377–88, 2012.

PRZYGRÓDZKA, E. et al. Expression of factors associated with apoptosis in the porcine corpus luteum throughout the luteal phase of the estrous cycle and early pregnancy: Their possible involvement in acquisition of luteolytic sensitivity. **Theriogenology**, v. 83, n. 4, p. 535–545, 2015.

PYUN, J. A.; KIM, S.; KWACK, K. B. Interaction between thyroglobulin and ADAMTS16 in premature ovarian failure. **Clinical and Experimental Reproductive Medicine**, v. 41, n. 3, p. 120–124, 2014.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.r-project.org/>>

R CORE TEAM; R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, , 2013. Disponível em: <<http://www.r-project.org>>

RAJPUT, S. K. et al. Identification of some unknown transcripts from SSH cDNA library of buffalo follicular oocytes. **Animal**, v. 7, n. 3, p. 446–454, 2013.

REDMOND, J. S. et al. Kisspeptin activates the hypothalamic-adenohypophyseal-gonadal axis in prepubertal ewe lambs. **Reproduction**, v. 141, n. 4, p. 541–548, 2011.

RESENDE, M. D. V. et al. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, v. 56, p. 63–77, 2008.

- RESENDE, M. D. V. et al. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística Espacial**. Viçosa, MG: [s.n.].
- ROMERO-AGUIRREGOMEZCORTA, J.; MATÁS, C.; COY, P. α -L-fucosidase enhances capacitation-associated events in porcine spermatozoa. **Veterinary Journal**, v. 203, n. 1, p. 109–114, 2015.
- ROYAL, M.; MANN, G. E.; FLINT, A. P. Strategies for reversing the trend towards subfertility in dairy cattle. **Veterinary journal**, v. 160, n. 1, p. 53–60, jul. 2000.
- RUSSELL, D. L.; BROWN, H. M.; DUNNING, K. R. ADAMTS proteases in fertility. **Matrix Biology**, v. 44–46, p. 54–63, 2015.
- RYU, D.; RYU, J.; LEE, C. Genome-wide association study reveals sex-specific selection signals against autosomal nucleotide variants. **Journal of human genetics**, n. October 2015, p. 1–4, 2016.
- SAATCHI, M. et al. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics, Selection, Evolution : GSE**, v. 43, n. 1, p. 40, 2011.
- SAHANA, G. et al. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. **Animal Genetics**, v. 41, n. 6, p. 579–588, 2010.
- SALES, J. N. S. et al. Effects of a high-energy diet on oocyte quality and in vitro embryo production in Bos indicus and Bos taurus cows. **Journal of dairy science**, v. 98, n. 5, p. 3086–99, 2015.
- SANCHEZ-LAZO, L. et al. Fatty acid synthesis and oxidation in cumulus cells support oocyte maturation in bovine. **Molecular endocrinology (Baltimore, Md.)**, v. 28, n. July, p. me20141049, 2014.
- SANTANA, M. H. A. H. A et al. Genome-wide association study for feedlot average daily gain in Nelore cattle (Bos indicus). **Journal of animal breeding and genetics = Zeitschrift für Tierzüchtung und Züchtungsbiologie**, v. 131, n. 3, p. 210–6, jun. 2014.
- SANTANA JR., M. L. et al. Genetic relationship between growth and reproductive traits in Nelore cattle. **animal**, v. 1, p. 1–6, 2011.
- SANTANA JR, M. L. et al. Genetic associations among average annual productivity, growth traits and stayability: a parallel between Nelore and composite beef cattle. **Journal of Animal Science**, 12 mar. 2013.
- SANTANA JR, M. L. et al. Two-trait random regression model to estimate the genetic association of scrotal circumference with female reproductive performance in Nelore cattle. **Theriogenology**, v. 83, n. 9, p. 1534–40, 2015.
- SANTOS, G. M. G. et al. 233 HIGH NUMBERS OF ANTRAL FOLLICLES INFLUENCE THE IN VITRO EMBRYO PRODUCTION, BUT NOT THE CONCEPTION RATE OF FIXED-TIME ARTIFICIAL INSEMINATION IN NELORE CATTLE. **Reproduction, Fertility and Development**, v. 27, n. 1, p. 206, 2015.
- SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, v. 15, n. 1, p. 478, jan. 2014.
- SÁRVÁRI, M. et al. Hippocampal gene expression is highly responsive to estradiol replacement in middle-aged female rats. **Endocrinology**, v. 156, n. 7, p. 2632–2645, 2015.
- SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of animal breeding and genetics**, v. 123, n. 4, p. 218–23, ago. 2006.
- SCHERER, A. **GWAS**. 2nd. ed. [s.l.] Golden Helix, 2016.
- SHARMA, S. C.; RICHARDS, J. S. Regulation of AP1 (Jun/Fos) factor expression and activation in ovarian granulosa cells. Relation of JunD and Fra2 to terminal differentiation. **Journal of Biological Chemistry**, v. 275, n. 43, p. 33718–33728, 2000.

- SHIOTSUKI, L. et al. Genetic associations of sexual precocity with growth traits and visual scores of conformation, finishing, and muscling in Nelore cattle. **Journal of animal science**, v. 87, p. 1591–1597, 2009.
- SHOZU, M. et al. ADAMTS-1 is involved in normal follicular development, ovulatory process and organization of the medullary vascular network in the ovary. **Journal of Molecular Endocrinology**, v. 35, n. 2, p. 343–355, 2005.
- SIMIANER, H. Genomic and other revolutions—why some technologies are quickly adopted and others are not. **Animal Frontiers**, v. 6, n. 1, p. 53, 2016.
- SINGH, J. et al. A simple ultrasound test to predict the superstimulatory response in cattle. **Theriogenology**, v. 62, n. 1–2, p. 227–243, 2004.
- SMITH, B. J. *boa*: An R Package for MCMC Output Convergence. **Journal of Statistical Software**, v. 21, n. 11, p. 1–37, 2007.
- SNELLING, W. M. et al. Physiology and endocrinology symposium: How single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. **Journal of Animal Science**, v. 90, n. 4, p. 1152–1165, 2012.
- SORENSEN, D.; GIANOLA, D. **Likelihood, Bayesian and MCMC methods in quantitative genetics**. New York, NY: Springer, 2002.
- SPENCER, K. L. et al. Genetic Variation and Reproductive Timing: African American Women from the Population Architecture Using Genomics and Epidemiology (PAGE) Study. **PLoS ONE**, v. 8, n. 2, 2013.
- SUAREZ, S. S. et al. Bull sperm binding to oviductal epithelium is mediated by a Ca²⁺-dependent lectin on sperm that recognizes Lewis-a trisaccharide. **Biology of reproduction**, v. 59, n. 1, p. 39–44, 1998.
- SUN, R. et al. Lyzl4, a novel mouse sperm-related protein, is involved in fertilization. **Acta Biochimica et Biophysica Sinica**, v. 43, n. 5, p. 346–353, 2011.
- TANGHE, S. et al. Carbohydrates and glycoproteins involved in bovine fertilization in vitro. **Molecular Reproduction and Development**, v. 68, n. 4, p. 492–499, 2004.
- TAYLOR, J. F. et al. Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle. **Animal Genetics**, v. 29, n. 3, p. 194–201, jun. 1998.
- TENGHE, A. M. M. et al. Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. **Journal of Dairy Science**, v. 99, n. 7, p. 5470–5485, 2016.
- TSUYUZAKI, K. et al. MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis. **BMC Bioinformatics**, v. 16, n. 1, p. 45, 2015.
- UTSUNOMIYA, Y. T. et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC genetics**, v. 14, p. 52, jan. 2013.
- UTSUNOMIYA, Y. T. et al. Genome-wide mapping of loci explaining variance in scrotal circumference in nellore cattle. **PLoS ONE**, v. 9, n. 2, 2014.
- VAN BLERKOM, J. Intrafollicular influences on human oocyte developmental competence: perfollicular vascularity, oocyte metabolism and mitochondrial function. **Human reproduction (Oxford, England)**, v. 15 Suppl 2, p. 173–88, 2000.
- VAN EEKELEN, M. et al. Identification and expression of the family of classical protein-tyrosine phosphatases in zebrafish. **PLoS ONE**, v. 5, n. 9, p. 1–16, 2010.
- VAN MELIS, M. H. et al. Additive genetic relationships between scrotal circumference, heifer pregnancy,

and stayability in Nellore cattle. **Journal of animal science**, v. 88, n. 12, p. 3809–3813, dez. 2010.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of dairy science**, v. 91, n. 11, p. 4414–23, nov. 2008.

VANRADEN, P. M.; SUN, C. **Fast Imputation Using Medium- or Low-Coverage Sequence Data**. Proc. 10th World Congr. Genet. Appl. Livest. Prod. **Anais...** Vancouver, Canada: 2014Disponível em: <http://aipl.arsusda.gov/publish/other/2014/10WCGALP_VanRaden.pdf>

VEERKAMP, R. F. et al. Genome-wide associations for feed utilisation complex in primiparous Holstein-Friesian dairy cows from experimental research herds in four European countries. **Animal : an international journal of animal bioscience**, v. 6, n. 11, p. 1738–49, nov. 2012.

VENTURA, R. V. et al. Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. **Genetics Selection Evolution**, v. 48, n. 1, p. 71, 2016.

VENTURA, R. V. et al. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. **Journal of animal science**, v. 92, n. 4, p. 1433–44, abr. 2014.

VISSCHER, P. M. et al. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. **Journal of Dairy Science**, v. 85, n. 9, p. 2368–75, 2002.

WALSH, S. W. et al. Heritability and impact of environmental effects during pregnancy on antral follicle count in cattle. **Journal of Dairy Science**, v. 97, n. 7, p. 4503–4511, 2014.

WANG, W. Y. S. et al. Genome-wide association studies: theoretical and practical concerns. **Nature Reviews Genetics**, v. 6, p. 109–118, 2005.

WARZYCH, E. et al. Cathepsin mRNA level in bovine cumulus cells fails to be a good marker of oocyte quality. **Animal Science Papers and Reports**, v. 34, n. 4, p. 339–349, 2016.

WATHES, D. C. et al. Multiple correlation analyses of metabolic and endocrine profiles with fertility in primiparous and multiparous cows. **Journal of dairy science**, v. 90, n. 3, p. 1310–1325, 2007.

WEISS, J. et al. Hypogonadism caused by a single amino acid substitution in the β subunit of luteinizing hormone. **N Engl J Med**, v. 326, p. 179–183, 1992.

WEISS, J. M. et al. Estradiol differentially modulates the exocytotic proteins SNAP-25 and munc-18 pituitary gonadotrophs. **Journal of Molecular Endocrinology**, v. 38, n. 1–2, p. 305–314, 2007.

WHITE, S.; KIMBER, S. J. Changes in alpha (1-2)-fucosyltransferase activity in the murine endometrial epithelium during the estrous cycle, early pregnancy, and after ovariectomy and hormone replacement. **Biology of Reproduction**, v. 50, n. 1, p. 73–81, 1994.

WIGGANS, G. R. et al. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. **Journal of Dairy Science**, v. 95, n. 3, p. 1552–1558, 2012.

WOLFE, M. D. et al. Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement. **The Plant Genome**, v. 9, n. 2, 2016.

XIANG, T. et al. Imputation of genotypes in Danish purebred and two-way crossbred pigs using low-density panels. **Genetics Selection Evolution**, v. 47, n. 1, p. 54, 30 dez. 2015.

YANG, J. et al. Genome-Wide Complex Trait Analysis (GCTA): Methods, Data Analyses, and Interpretations. In: [s.l: s.n.]. p. 215–236.

YOSHIZAKI, H. et al. Activity of Rho-family GTPases during cell division as visualized with FRET-based probes. **The Journal of Cell Biology**, v. 162, n. 2, p. 223–232, 21 jul. 2003.

ZERON, Y. et al. Seasonal changes in bovine fertility: Relation to developmental competence of oocytes, membrane properties and fatty acid composition of follicles. **Reproduction**, v. 121, n. 3, p. 447–454, 2001.

ZHANG, K. et al. Molecular Cloning and Characterization of Three Novel Lysozyme-Like Genes, Predominantly Expressed in the Male Reproductive System of Humans, Belonging to the C-Type Lysozyme/Alpha-Lactalbumin Family. **Biology of Reproduction**, v. 73, n. 5, p. 1064–1071, 20 jul. 2005.

ZHANG, X. et al. Effects of selection for blood serum IGF-I concentration on reproductive performance of female Angus beef cattle. **Journal of Animal Science**, v. 91, n. 9, p. 4104–4115, 2013.

APÊNDICE A

Genotype imputation in a tropical crossbred dairy cattle population

Gerson A. Oliveira Júnior,* Tatiane C. S. Chud,§ Ricardo V. Ventura,¶† Dorian J. Garrick,‡ John B. Cole,Ω Danísio P. Munari,§ José B. S. Ferraz,* Erik Mullart, ¥, Sue DeNise,Φ, Shannon Smith, Φ, Marcos Vinícius G. B. da Silva, #

* Universidade de São Paulo (USP), Pirassununga, SP, Brazil.

§ Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, SP, Brazil.

¶ Beef Improvement Opportunities, Guelph, ON N1K1E5, Canada.

† Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON N1G2W1, Canada.

‡ Iowa State University, Ames, IA, United States.

Ω Animal Genomics and Improvement Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland, USA.

¥ CRV Holding B.V., Arnhem, The Netherlands

Φ Zoetis, Kalamazoo, Michigan, USA.

Embrapa Dairy Cattle - Brazilian Corporation of Agricultural Research, Juiz de Fora, MG, Brazil.

ABSTRACT

The application of new tools, such as genomic selection and genotype imputation, still presents challenges in crossbred populations because relationships of causal variants with markers may vary across breeds. In order to make genomic selection more cost effective, cheap low density chips are often used in combination with imputation. Genotype imputation is a statistical technique that permits the inference of high-density, ungenotyped markers for an individual genotyped at low-density. It relies on a high-density genotyped reference population and offers substantial reductions in the costs of genotyping which allows for the genotyping of more animals and, thereby increasing the accuracy of genomic selection for a given investment in genotyping. The objective of this study was to investigate different strategies for genotype imputation in a population of crossbred Girolando (Gyr × Holstein) dairy cattle. The dataset consisted of 478 Girolando, 583 Gyr and 1,198 Holstein sires genotyped at high-density with the Illumina BovineHD panel, which includes ~777K markers. The accuracy of imputation from low- (20K) and medium-densities (50 and 70K) to the HD panel density and from low to 50K density were investigated. Seven scenarios with 29 variations were tested for imputing

genotypes of 166 randomly chosen Girolando animals. Imputation accuracy was measured as the correlation between observed and imputed genotypes (CORR) and also as the proportion of genotypes that were imputed correctly (CR). This is the first paper to check imputation accuracy in a Girolando population. The sample-specific imputation accuracies ranged from 0.38 to 0.97 (CORR) and from 0.49 to 0.96 (CR) imputing from low- and medium-densities to HD, and 0.41 to 0.95 (CORR) and from 0.50 to 0.94 (CR) for imputation from 20K to 50K. The highest imputation accuracies were observed for scenarios including Girolando animals in the reference population, whereas using only Gyr animals resulted in low imputation accuracies, suggesting that the haplotypes segregating in the Girolando population had higher impact in accuracy than the pure breed haplotypes. Crossbred animals (Girolando) must be included in the reference population to provide the best imputation accuracies, mainly when imputing from low density (20K) to high density panels. The results obtained in this work provide information to more cost-effectively implement genomic selection and will assist future studies involving genomic analysis in crossbred animals.

Key words: Impute, single nucleotide polymorphism (SNP), Genotype, Genomic