UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CAMPUS DE JABOTICABAL

IDENTIFICAÇÃO DE REGIÕES COM VARIAÇÕES NO NÚMERO DE CÓPIAS DOS SEGMENTOS DE DNA EM BOVINOS DE LEITE

Tatiane Cristina Seleguim Chud

Zootecnista

2018

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CAMPUS DE JABOTICABAL

IDENTIFICAÇÃO DE REGIÕES COM VARIAÇÕES NO NÚMERO DE CÓPIAS DOS SEGMENTOS DE DNA EM BOVINOS DE LEITE

Tatiane Cristina Seleguim Chud

Orientador: Prof. Dr. Danísio Prado Munari

Coorientadores: Prof. Dr. Fernando Sebastian Baldi Rey

Dr. Marcos Vinicius Gualberto Barbosa da Silva

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

2018

Chud, Tatiane Cristina Seleguim
C559i Identificação de regiões com variações no número de cópias dos segmentos de DNA em bovinos de leite / Tatiane Cristina Seleguim Chud. – Jaboticabal, 2018
iv, 113 p.: il. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2018
Orientador: Danísio Prado Munari
Co-orientador: Fernando Sebastián Baldi Rey, Marcos Vinícius Gualberto Barbosa da Silva
Banca examinadora: Adriana Santana do Carmo, Daniel Guariz Pinheiro, Humberto Tonhati, Marcos Eli Buzanskas
Bibliografia

1.*Bos taurus indicus*. 2.*Bos taurus taurus* 3. Genômica. 4. Sequenciamento de nova-geração. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.2

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

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: IDENTIFICAÇÃO DE REGIÕES COM VARIAÇÕES NO NÚMERO DE CÓPIAS DOS SEGMENTOS DE DNA EM BOVINOS DE LEITE

AUTORA: TATIANE CRISTINA SELEGUIM CHUD
ORIENTADOR: DANISIO PRADO MUNARI
COORIENTADOR: MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA
COORIENTADOR: FERNANDO SEBASTIAN BALDI REY

Aprovada como parte das exigências para obtenção do Título de Doutora em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

Prof. Dr. DANISIO PRADO MUNARI
Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal

Prof. Dr. HUMBERTO TONHATI
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Profa. Dra. ADRIANA SANTANA DO CARMO
Departamento de Zootecnia / Universidade Federal de Goiás / Goiânia

Prof. Dr. DANIEL GUARIZ PINHEIRO
Departamento de Tecnologia / FCAV / UNESP - Jaboticabal

Prof. Adjunto MARCOS ELI BUZANSKAS
Departamento de Zootecnia / Universidade da Paraiba / Areia

Jaboticabal, 23 de fevereiro de 2018

DADOS CURRICULARES DO AUTOR

Tatiane Cristina Seleguim Chud – nascida em Porto Ferreira, SP, no dia 25 de maio de 1988, ingressou no curso de Zootecnia em março de 2007 na Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal-SP, foi bolsista de iniciação científica da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e obteve o título de bacharel em Zootecnia em março de 2012. Iniciou o curso de Mestrado pela mesma instituição de ensino, em março de 2012, sob orientação do Prof. Dr. Danísio Prado Munari e coorientação do Prof.Dr. Ricardo Vieira Ventura e do Dr. Roberto Carvalheiro. Foi bolsista de mestrado da FAPESP. Realizou estágio de pesquisa de Mestrado, sob supervisão do Prof. Dr. Flavio Schramm Schenkel, no Centre for Genetic Improvement of Livestock, University of Guelph, Canadá, no período de abril a setembro de 2013 e em fevereiro de 2014 obteve o título de Mestre em Genética e Melhoramento Animal com a dissertação intitulada “Metodologias e estratégias de imputação de marcadores genéticos em bovinos da raça Canchim”. Ingressou no curso de doutorado pela mesma instituição em março de 2014 sob orientação do Prof. Dr. Danísio Prado Munari e coorientação do Dr. Marcos Vinicius Gualberto Barbosa da Silva e do Prof. Dr. Fernando Sebastián Baldi. Foi bolsista pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior e posteriormente FAPESP. Em maio de 2016, iniciou o estágio de pesquisa de doutorado no Animal Genomics and Improvement Laboratory, United States Department of Agriculture (USDA), Beltsville, MD, Estados Unidos, sob supervisão do Dr. Derek M. Bickhart.

“Life is not easy for any of us. But what of that? We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained.”

Marie Curie

“You look at science (or at least talk of it) as some sort of demoralising invention of man, something apart from real life, and which must be cautiously guarded and kept separate from everyday existence. But science and everyday life cannot and should not be separated.”

Rosalind Franklin

“If you know you are on the right track, if you have this inner knowledge, then nobody can turn you off... no matter what they say.”

Barbara McClintock

“What spectacle can be more edifying or more seasonable, than that of liberty & learning, each leaning on the other for their mutual & surest support?”

James Madison

Dedico a todos aqueles que estiveram ao meu lado e que de alguma forma contribuíram para a realização deste trabalho, incluindo meus cachorros Bentinho e Zag.

Dedico à comunidade científica.

AGRADECIMENTOS

Pelas oportunidades que me foram concedidas ao longo de toda minha vida acadêmica.

A todos que contribuíram para execução deste trabalho de forma direta ou indireta.

Ao meu orientador e amigo Prof. Dr. Danísio pela década de convivência participando de toda minha formação profissional e pessoal. Sou extremamente grata por todo auxílio, pelas críticas e pelo incentivo ao longo desses anos. À Rose pelo carinho e receptividade.

To Dr. Derek Bickhart for his assistance, the valuable comments and all advices he has provided to me during the execution of this work. To all members of the Animal Genomics and Improvement Laboratory, in special, Dr. Paul VanRaden and Dr. John Cole, who helped me during my time in USDA, Beltsville, USA.

Ao meu coorientador Dr. Marcos Vinicius por ter cedido os dados e me auxiliado no desenvolvimento deste trabalho. Ao Prof. Dr. Fernando Baldi pela disposição em me coorientar.

Aos membros da banca do exame geral de qualificação Prof. Dr. Marcos Eli Buzanskas, Prof. Dr. Rodrigo Pelicioni Savegnago, Prof. Dr. Ricardo Vieira Ventura e Prof. Dr. Thiago Bruno Ribeiro pela colaboração e sugestões.

Aos membros da banca de defesa da tese Prof. Dr. Marcos Eli Buzanskas, Profa. Dra. Adriana Santana do Carmo, Prof. Dr. Daniel Guariz Pinheiro e Prof. Dr. Humberto Tonhati pelas considereções realizadas para a melhoria deste trabalho.

Aos meus amigos do Departamento e da vida. Le por ter compartilhado grande parte dessa fase da vida comigo, incluindo casa e “manias”. Por ter cuidado dos meus filhotes de quatro patas enquanto eu estava ausente, por ouvir meus desabafos e por estar sempre presente quando eu preciso. Ao Gui, “meu contemporâneo, mas sempre bixo” pela cumplicidade nesses anos de doutorado e claro pelas discussões inacabáveis, seja sobre ciência, política, valores e mesmo sobre a vida. À Jaque pela amizade e convivência desde a graduação seja em casa ou no departamento. Salvs, Mirele, Marcos, Ro, Dani Grossi, Pri, Elieder, Sabrina, Thiaguinho, Gerson, Ana, Rafa, Lucas, Rebeka, Eli, Alejandro, Luara, Fabrícia e Mauricio pelos cafés sem tempo de acabar e conversas sempre filosóficas. Obrigada por terem tornado esse trabalho muito mais agradável e divertido.

À Mari por ter dividido comigo as experiências do doutorado sanduíche, por aceitar sempre tomar um café no Dunkin' Donuts ou mesmo uma porção de asinhas no Buffalos e por ter sido minha família nos EUA, juntamente com a Camila e o Fernando. Obrigada por terem feito desse período um dos melhores da minha vida.

Ao meu parceiro de todas as horas Hugo, obrigada pela compreensão e paciência em todos esses anos. Obrigada por discutir meu trabalho comigo mesmo sem saber sobre o assunto, pelas palavras de incentivo quando estava desanimada ou quando dizia "está tudo errado". Por sempre estar ao meu lado, me apoiando nas decisões. Por me trazer o equilíbrio e a tranquilidade. Simplesmente obrigada por fazer parte da minha vida.

À minha mãe Cidinha por ter sido corajosa e me proporcionado condições para estudar, sempre acreditando em mim e nos meus sonhos. Ao meu pai Mauro pelo carinho.

À minha irmã-mãe Patrícia, à minha sobrinha-irmã Victória e meu cunhado-pai Rodrigo pelo amor incondicional, pelo respeito e pela dedicação. Obrigada por sempre estarem ao meu lado quando preciso ser forte! Vocês são meu exemplo!

À toda família Fantucci Matheus. Telma e Laércio meus agradecimentos por terem me acolhido como filha e terem me ajudado em vários momentos de dificuldade. Laura obrigada pela amizade e pelo carinho de irmã. Márcia, Venilton, Carol e Romero obrigada pelos momentos de descontração e pelo apoio.

À UNESP/FCAV pela formação profissional.

A todos os funcionários e professores da FCAV que eu tive o prazer de conviver nestes anos. Em especial aqueles do Departamento de Ciências Exatas pela paciência, pela disposição em ajudar e pela empatia.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão da bolsa de estudos (Processo n. 2015/08939-0) no período de julho de 2015 a fevereiro de 2018.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (PPGMA/DS) e ao projeto CAPES/EMBRAPA (15/2014) pela bolsa concedida no início do curso de doutorado.

Ao Programa de doutorado sanduíche no exterior (Processo 99999.006978/2015-00), mantido pela CAPES, pela bolsa de estudos para a realização do estágio sanduíche no exterior.

À Embrapa Gado de Leite pela concessão dos dados utilizados na tese.

Ao Laboratório Multiusuário de Bioinformática, Embrapa Informática Agropecuária, pela infraestrutura e suporte nas análises.

SUMÁRIO

Página

RESUMO iii
ABSTRACT iv
CAPÍTULO 1 - CONSIDERAÇÕES GERAIS 1
1 INTRODUÇÃO 1
2 OBJETIVOS GERAIS 2
2.1 Objetivos Específicos 2
3 REVISÃO BIBLIOGRÁFICA 3
3.1 A raça Girolando 3
3.2 Variações cromossômicas estruturais 4
3.3 Mecanismos de formação das CNVs 5
3.4 Técnicas para identificação, análise dos dados e validação das CNVs 7
3.4.1 Identificação das CNVs por meio dos painéis de SNP 8
3.4.2 Identificação por meio do sequenciamento de nova geração 10
3.4.3 Validação das CNVs 12
3.5 Estudo de associação entre CNVs e fenótipos de interesse e importância de CNVs no melhoramento genético de bovinos 13
4 REFERÊNCIAS 15
CAPÍTULO 2 IDENTIFICAÇÃO DE VARIAÇÃO NO NÚMERO DE CÓPIAS DE SEGMENTOS DE DNA EM BOVINOS GIROLANDO 24
RESUMO 24
ABSTRACT 25
1 INTRODUÇÃO 26
2 MATERIAL E MÉTODOS 28
2.1 Animais, genotipagem e ressequenciamento 28
2.2 Identificação de CNVs por meio dos painéis de genotipagem e controle de qualidade 29
2.2.1 Detecção das CNVs com base na família 31
2.3 Detecção das CNVs usando os dados de NGS 32
2.4 Validação *in silico* 32

2.5 Análise funcional por meio de prospecção dos genes e análise de enriquecimento gênico nas regiões candidatas de CNV 33
3 RESULTADOS 34
3.1 Identificação das CNVs por meio do painel de 50K e do painel de alta densidade 34
3.2 Detecção das CNVs com base na família 36
3.3 Detecção por meio dos dados de NGS 37
3.4 Regiões candidatas de CNV 39
3.4.1 Detecção utilizando o painel de 50K 39
3.4.2 Detecção pelo painel de alta densidade (HD) 43
4 DISCUSSÃO 49
5 CONCLUSÃO 56
6. REFERÊNCIAS 56
CAPÍTULO 3. VARIAÇÕES NO NÚMERO DE CÓPIAS EM TOUROS GIR, HOLANDÊS E GIROLANDO POR MEIO DE TÉCNICA DE SEQUENCIAMENTO DE NOVA GERAÇÃO 66
RESUMO 66
ABSTRACT 67
1 INTRODUÇÃO 68
2 MATERIAL E MÉTODOS 70
2.1 Animais e e ressequenciamento 70
2.2 Controle de qualidade e alinhamento das sequências no genoma referência 70
2.3 Detecção das variações no número de cópias 71
2.4 Diferenciação populacional por meio do número de cópias 72
3 RESULTADOS E DISCUSSÃO 72
3.1. Detecção de CNVs 72
3.2 Grupos de CNVRs específicos entre raças 75
3.3 Diferenciação populacional por meio dos genes anotados nas CNVs 79
4 CONCLUSÃO 82
5 REFERÊNCIAS 82
APÊNDICES 89

IDENTIFICAÇÃO DE REGIÕES COM VARIAÇÕES NO NÚMERO DE CÓPIAS DOS SEGMENTOS DE DNA EM BOVINOS DE LEITE

RESUMO - Com o avanço das tecnologias genômicas, permitiu-se detectar no genoma de humanos e animais domésticos elevado número de variações estruturais cromossômicas, como variação no número de cópias (CNV). No melhoramento genético de animais domésticos, CNVs podem auxiliar no entendimento da variabilidade genética de características de importância econômica, pois a maioria dessas regiões influenciam a expressão de genes com funções biológicas específicas. Com a finalidade de verificar possíveis relações de CNVs com características de sanidade, reprodutivas e produtivas em rebanhos leiteiros, o objetivo deste trabalho foi detectar e caracterizar CNVs em bovinos da raça composta Girolando (Gir X Holandês), identificar CNVs específicas em animais Girolando oriundas de animais da raça Gir e Holandês e investigar a diferenciação populacional entre as três raças por meio da variação no número de cópias nas regiões próximas aos genes anotados no genoma bovino. Para detecção das CNVs por meio dos dados de painel de polimorfismo de nucleotídeo único (SNP), no capítulo 2, foram utilizados registros de 1.607 vacas genotipadas com painel de SNP de média densidade (50K SNP) e de 280 touros genotipados com painel de alta densidade (HD SNP). A detecção foi realizada por meio do modelo oculto de Markov implementado pelo programa PennCNV. Foram utilizados dois touros ressequenciados para identificação das CNVs pelo método “read -depth” por meio dos dados de sequenciamento de nova-geração (NGS). Um total de 203 e 213 regiões candidatas de CNVs foram selecionadas pelos painéis de 50K e HD, respectivamente. Deleções e duplicações relacionadas a resistência a parasitas, susceptibilidade à doenças e reprodução, foram encontradas principalmente nos cromossomos BTA 5 e BTA 17. A detecção e caracterização das CNVs realizadas em bovinos de leite de raça composta demonstrou melhor entendimento das características de adaptabilidade ao clima tropical, como resistência à doenças e eficiência reprodutiva. No capítulo 3, foram utilizados cinco animais da raça Holandês, 14 animais da raça Gir e dois animais da raça Girolando. Após o alinhamento do genoma foi realizada a detecção das CNVs pela metodologia baseada em “read-depth”. A estatística V_{ST}.foi calculada pela média do número de cópias nas regiões próximas aos genes anotados no genoma bovina. Genes relacionados com fertilidade (*MEPCE, ASB3*) e susceptibilidade às doenças (*HLX, MIR-455*) foram anotados nas regiões específicas compartilhadas entre Girolando e as raças formadoras (Gir e Holandês). Os valores de V_{ST} variaram de -0,37 a 0,98. Os genes *AOX1, SLBP, TACC3* e *PRAME*, apresentaram elevado V_{ST}, indicando diferenciação populacional pelo número de cópias, possivelmente originárias durante os processos de domesticação das subespécies. Regiões especificas de CNVS foram identificadas nos animais Girolando provenientes do Gir e do Holandês. O estudo de diferenciação populacional evidenciou seleção positiva no genoma de animais da raça Gir e Girolando para características relacionadas à adaptação desses animais aos ambientes de clima tropical, possivelmente originadas do processo de domesticação.
Palavras-chave: *Bos taurus taurus*, *Bos taurus indicus*, divergência populacional, genômica, sequenciamento de nova-geração

COPY NUMBER VARIATION DISCOVERY IN DAIRY CATTLE

ABSTRACT - The advances of the genomic technologies has enabled to identify a high number of chromosomal structural variations in human and domestic animal genomes, such as copy number variation (CNV). In animal breeding, CNVs may assist to understand genetic variability of the economic important traits due most of the CNVs influence gene expression with specific biological functions. To identify possible CNVs linked to health, reproductive and productive traits in dairy cattle, the objective of this work was to detect and to describe CNVs in Girolando cattle (Gir x Holstein), to identify breed-specific CNV regions in Girolando from Gir and Holstein, and to investigate the population differentiation among the breeds using the copy number located on regions within annotated genes. In chapter 2, the CNV detection using single-nucleotide-polymorphism (SNP) panel was carried out on 1.607 females genotyped with the medium-density SNP panel (50K SNP) and 280 bulls genotyped with high-density panel (HD SNP) using Hidden Markov model implemented by PennCNV software. CNV calling also was perform using read-depth method applied on next-generation sequencing (NGS) data from two bulls resenquenced. A total of 203 and 213 CNVs candidate's regions were picked using 50K e HD panels, respectively. Deletions and duplications related to parasite resistance, to disease susceptibility, and to reproductive efficiency was observed mainly located on chromosome BTA 5 and BTA 17. The detection and characterization of the CNVs in composite dairy cattle breed demonstrated better understanding of the traits, such disease resistance and reproductive efficiency. In chapter 3, the CNV calling was carried out on three Girolando bulls, 14 Gir bulls, and five Holstein bulls resequenced using the "read-depth" method implemented by CNVnator software. The V_{ST} statistic was calculated for the average of the copy number in regions located near annotated genes. Genes linked to fertility (*MEPCE, ASB3*) and disease susceptibility (*HLX, MIR-455*) were mapped on specific regions shared between Girolando and the pure-breeds (Gir ans Holstein). V_{ST} values ranged from -0.37 to 0.98. The genes *AOX1*, *SLBP*, *TACC3* and *PRAME*, showed high V_{ST}, indicating high level of the population differentiation for the copy number located on the regions near of these genes. We found CNVR regions in Girolando specific from Gir and Holstein. The population differentiation study evidenced positive selection in genome of the Gir and Girolando animals for traits related to the adaptability of the breeds in tropical environmental may have originated from the domestication process.

Keywords: *Bos taurus taurus, Bos taurus indicus,* genomic, next-generation sequencing, population differentiation

CAPÍTULO 1 - CONSIDERAÇÕES GERAIS

1 INTRODUÇÃO

O Brasil tem produzido anualmente em torno de 24 bilhões de litros de leite, sendo uma das maiores produções do mundo. No entanto, a produtividade média por animal é de apenas 1.320 litros/ano (USDA, 2017). Com a finalidade de aumentar a produtividade do plantel, busca-se em programas de melhoramento genético de bovinos de leite aprimorar nos rebanhos brasileiros características de produção e qualidade de leite, sanidade e fertilidade (SILVA et al., 2013).

Uma das estratégias que tem sido aplicada para melhorar a eficiência produtiva e a lucratividade da cadeia do leite no país é a utilização de sistemas de acasalamentos entre raças zebuínas e taurinas. Animais cruzados e de raças compostas são responsáveis por cerca de 80% da produção de leite no Brasil, sendo a maioria bovinos da raça Girolando (COLE; SILVA, 2016). A raça Girolando foi formada por meio de cruzamentos entre bovinos da raça Gir (zebuíno) e da raça Holandês (taurino) com a finalidade de explorar os efeitos de heterose e a complementaridade entre as características favoráveis do Gir, como rusticidade e adaptabilidade e a elevada produção de leite da raça Holandesa.

Programas de melhoramento genético tem incluído informações genotípicas nas avaliações genéticas para auxiliar no aumento do progresso genético de características de interesse econômico tais como, produção de leite e fertilidade (HAYES et al., 2009; OLSON; VANRADEN; TOOKER, 2012). A integração de métodos moleculares, como marcadores genéticos, com o sistema de avaliação genética baseada no pedigree, tem permitido reduzir custos de produção, principalmente com os testes de progênie e identificar alelos letais recessivos (WIGGANS et al., 2017).

Com a disponibilidade de tecnologias genômicas, tais como painéis de genotipagem com milhares de marcadores genéticos do tipo SNP (polimorfismo de nucleotídeo único) e sequenciamento de nova-geração (NGS), permitiu-se detectar elevado número de variações estruturais cromossômicas presentes no genoma humano e de animais domésticos (FEUK et al., 2006; JIANG et al., 2013), tais como

inversões, translocações e variação no número de cópias dos segmentos de DNA (CNV), sendo esta última, o tipo de variação mais comum identificado no genoma.

CNV pode ser definido por deleções ou duplicações de segmentos cromossômicos com tamanho que variam de um quilo pares de bases (Kb) até vários megabases (Mb) (HENRICHSEN et al., 2009). Essas variações são menos frequentes no genoma em relação aos SNPs. Entretanto, devido à alta variabilidade entre os indivíduos constituiu-se como importante fonte de diversidade e variabilidade genética populacional (REDON et al., 2006). As CNVs podem influenciar na expressão gênica (MARGARETO et al., 2009; ZHANG et al., 2009), e consequentemente modificar fenótipos de interesse econômico (BICKHART et al., 2014) como susceptibilidade a doenças (WANG et al., 2012).

Estudos de identificação de regiões com variações no número de cópias dos segmentos de DNA são importantes para os programas de melhoramento genético de bovinos leiteiros, pois auxiliam na compreensão dos processos biológicos envolvidos na expressão fenotípica das características de produção (XU et al., 2014), reprodutivas (DUCOS et al., 2008) e principalmente de sanidade (ZHANG et al., 2009).

2 OBJETIVOS GERAIS

Neste trabalho o objetivo foi identificar e caracterizar regiões com variações no número de cópias nos segmentos de DNA em bovinos de leite.

2.1 Objetivos Específicos

2.1.1 Buscar genes anotados nas regiões de CNVs relacionados as características de sanidade, reprodutivas e produtivas de leite, em animais da raça Girolando por meio dos dados de painel de genotipagem.

2.1.2 Detectar deleções e duplicações por meio de dados de sequenciamento de nova-geração em touros da raça Gir, Holandês e Girolando.

2.1.3 Identificar possíveis CNVs específicas em animais Girolando provenientes de bovinos Gir e Holandês

2.1.4 Obter diferenciação populacional entre as raças por meio do número de cópias

das regiões próximas aos genes anotados no genoma bovino.

3 REVISÃO BIBLIOGRÁFICA

3.1 A raça Girolando

A formação da raça Girolando iniciou-se na década de 1940, por meio do cruzamento entre animais da raça Holandês (*Bos taurus taurus*) e Gir (*Bos taurus indicus*), com a finalidade de produzir animais altamente produtivos que se adequassem ao clima e aos diferentes sistemas de manejo e clima dos ambientes tropicais (CANAZA-CAYO et al. 2014). Somente no ano de 1996 a raça foi oficializada pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA).

Os animais considerados puros sintéticos da raça Girolando pelo MAPA, são aqueles que apresentem proporção de genes 5/8 Holandês e 3/8 Gir. Existem diversas estratégias de acasalamentos alternados entre as raças Holandês e Gir para a formação do puro sintético e são direcionados de acordo com os objetivos de seleção do produtor. A produtividade dos animais da raça Girolando dependem do processo de seleção durante a formação da raça, sendo necessário a escolha ideal dos animais de raças puras (FACÓ et al., 2005; FACÓ et al., 2008).

As características de produção de leite, idade ao primeiro parto e intervalo de partos são influenciadas pela composição genética dos animais e pelo sistema de criação (FACÓ et al., 2009). A média de produção de leite, em 305 dias, pode variar aproximadamente de 2.000 quilos até 5.000 kg (GROSSI; FREITAS, 2002; MCMANUS et al., 2008). Na literatura, foi reportado média da duração da lactação de 230 a 337 dias (GLÓRIA et al., 2002; GUIMARÃES et al., 2002). Para média de idade ao primeiro parto foi relatado valores de 32 a 42 meses (FACÓ et al., 2005; MCMANUS et al., 2008) e para intervalo de partos valores médios de 11 a 14 meses (FACÓ et al., 2005; MCMANUS et al., 2008).

Pode-se considerar que animais da raça Girolando são bem adaptados aos sistemas de produção do Brasil, com bom desempenho produtivo, principalmente em boas condições de manejo nutricional (SILVA; VELOSO, 2011).

3.2 Variações cromossômicas estruturais

O advento da genômica revelou polimorfismos de nucleotídeo único como mais abundante fonte de variação genética e fenotípica em mamíferos (SACHIDANANDAM et al., 2001; MANOLIO et al. 2009). Porém, microscópicas variações estruturais cromossômicas nos genomas de eucariotos, podem gerar alteração fenotípica por meio de mudanças na dosagem e regulação gênica (IAFRATE et al., 2004; LIU et al., 2010). Essas variações podem ser duplicações segmentares, também conhecidas como “low copy repeat”, variações no número de cópias, rearranjos de posição, como translocações e de orientação, como as inversões (FEUK et al., 2006).

As duplicações segmentares são definidas como segmentos de DNA que ocorrem em duas ou mais cópias por genoma haplóide, com as diferentes cópias compartilhando mais de 90% de similaridade de sequência (BAILEY et al., 2001). As variações no número de cópias (CNV – “Copy Number Variation”) são definidas como alterações genômicas com pelo menos um Kb de extensão, envolvendo ganhos ou perdas em comparação com um genoma de referência (IAFRATE et al., 2004; HENRICHSEN et al., 2009). Estas variações são amplamente distribuídas no genoma e influenciam na expressão gênica, variação fenotípica e adaptação por meio de desregulação e da alteração da dosagem gênica (ZHANG et al., 2009). Além disso, são importantes para identificar variabilidade genética em humanos e animais domésticos (REDON et al., 2006; STOTHARD et al., 2011; CLOP et al., 2012).

Pesquisas têm sido realizadas para avaliação das CNVs em diferentes espécies animais tais como bovinos (LIU et al., 2010; HOU et al., 2012a; BICKHART et al., 2012; JIANG et al., 2013; BOUSSAHA et al., 2015), ovinos (FONTANESI et al., 2011; LIU et al., 2013), caprinos (FONTANESI et al., 2010), suínos (CHEN et al., 2012; LI et al., 2012; WANG et al., 2013) e aves (SKINNER et al., 2009; WANG et al., 2010; LUO et al., 2013; ABERNATHY et al., 2014). Em bovinos, a facilidade de acesso a dados genômicos tem possibilitado a realização de múltiplos estudos envolvendo CNVs nessa espécie (FADISTA et al, 2010; LIU et al., 2010; HOU et al., 2012b).

De acordo com o banco de dados Genomic Variants archive (DGVa) da base EMBL-EBI (http://www.ebi.ac.uk/dgva) foram identificadas no genoma bovino 10.680

CNVs. O tamanho e número de CNVs detectados variaram entre os estudos (BAE et al., 2010; LIU et al., 2010; HOU et al., 2012a; HOU et al., 2012b; CICCONARDI et al., 2013), dependendo da técnica molecular, do algoritmo utilizado para a detecção de CNV, do tamanho da amostra e da estrutura da população (HENRICHSEN et al., 2009).

3.3 Mecanismos de formação das CNVs

Os mecanismos de recombinação homóloga não-alélica (NAHR), parada na replicação do DNA ou troca da forquilha de replicação mediadas por regiões com micro-homologia (FOSTES), junção das extremidades não homólogas (NHEJ) e retrotransposição de elementos 1, são os principais eventos de formação das variações do número de cópias nos segmentos de DNA (CNV) (CLOP et al., 2012).

A NAHR ocorre por meio do emparelhamento entre as sequências de LCRs (“low copy repeats”) ou duplicações segmentadas (GU; ZHANG; LUPSKI; 2008) que compartilham alta homologia entre si. O mecanismo FOSTES ocorre durante o processo de replicação do DNA, em que a fita tardia recém-sintetizada (“lagging strand”) pode se desprender da fita molde de DNA e se anelar em outra forquilha de replicação com micro-homologia entre as regiões, continuando a síntese da fita na forquilha invadida (LEE; CARVALHO; LUPSKI, 2007). O processo pode ocorrer em sequência, gerando múltiplas alterações no genoma (ZHANG et al., 2009). As CNVs geradas pelo processo FOSTES somente ocorrem na fase da duplicação do material genético (fase S do ciclo celular) (BICKHART; LIU, 2014).

A junção das extremidades não homólogas (NHEJ) é utilizada para reparar a quebra da dupla fita de DNA (DSB). Esse mecanismo permite a ligação das extremidades do DNA por meio da inserção ou deleção de poucos nucleotídeos (1-4 bp) nas junções das fitas quebradas, gerando uma “cicatriz informativa” no genoma (LIEBER, 2008). Este mecanismo não depende da presença de LCRs e podem ocorrer em qualquer região genômica susceptível à DSB (GU; ZHANG; LUPSKI; 2008). NHEJ está mais associado com as variações estruturais de deleções (SHAW; LUPSKI, 2005) e translocações nos cromossomos (LIEBER et al., 2010).

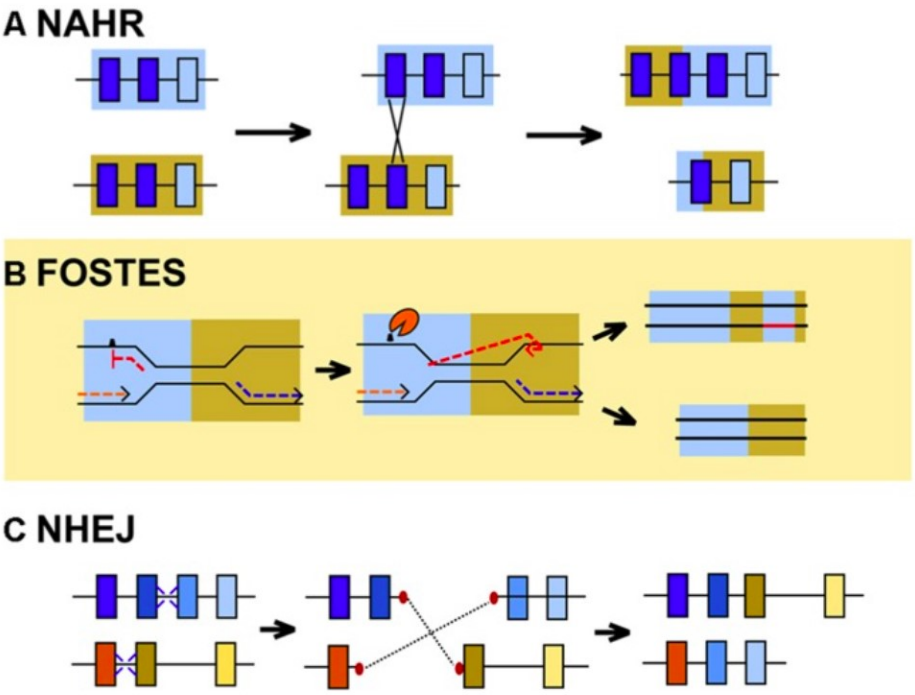


Figura 1. Mecanismos de formação de variação no número de cópias (CNV). (A) CNVs gerados pela recombinação homóloga não alélica (NAHR) em que ocorre recombinação de segmentos genômicos não-alélicos com alta similaridade (caixa azul). (B) Parada na replicação do DNA ou troca da forquilha de replicação mediadas por regiões com micro-homologia (FOSTES) ocorre quando a fita “lagging” (linha pontilhada vermelha) associa-se em uma diferente região do genoma com similaridade. (C) Quebra da dupla fita na sequência de DNA (linhas azuis). Ocorre a substituição dos nucleotídeos perdidos nos pontos de quebra e os fragmentos são ligados nos pontos de quebra das fitas. Se os fragmentos ligados são de cromossomos diferentes, podem ocorrer duplicações ou deleções. Adaptado de Bickhart e Liu (2014).

As CNVs também podem ocorrer pela inserção dos elementos intercalados longos (“Long interspersed elements” - LINE-1), ou retrotransposons. Esses elementos são abundantes no genoma de mamíferos (MORRISH et al., 2002), sua transposição no genoma ocorre por meio do intermédio de RNA, em que provavelmente é transcrito pela RNA polímerase II (BABUSHOK et al., 2007).

3.4 Técnicas para identificação, análise dos dados e validação das CNVs

As CNVs podem ser detectadas por diferentes técnicas que abrangem todo genoma, como hibridização genômica comparativa por meio de microarranjos de DNA (CGH), plataformas de genotipagem em alta densidade com marcadores genéticos do tipo SNP (BeadChip) e sequenciamento genômico de última geração (NGS) (BICKHART; LIU, 2014). A técnica CGH, como descrito por Duan et al. (2013), compara o número de cópias entre a amostra do DNA referência e outra amostra teste, identificando deleções ou duplicações de material genético em regiões específicas do DNA. Essa técnica é uma das mais acuradas para detecção de CNVs (ALKAN; COE; EICHLER, 2011). Entretanto, devido ao alto custo e grande número de amostras necessárias, recomendou-se o uso de técnicas alternativas, tais como painéis de genotipagem (WINCHESTER; YAU; RAGOUSSIS, 2009).

Devido a disponibilidade de diversos painéis de SNP comerciais para genotipagem de animais domésticos, estes têm sido amplamente utilizados na identificação, mapeamento e estudos funcionais de CNVs na pecuária (LIU et al., 2013). A detecção das CNVs por meio das plataformas de genotipagem de marcadores do tipo SNP baseia-se na medida da intensidade do sinal da fluorescência emitida na definição dos genótipos e na relação alélica de cada SNP, pois a duplicação ou deleção de determinada região irá resultar em aumento ou redução na intensidade do sinal da fluorescência. Para que a detecção das CNVs seja confiável, os painéis de SNPs devem apresentar densidade suficiente para identificação de CNVs.

Metodologias utilizando painéis de SNPs de média densidade (50K) têm detectado CNVs muito maiores em relação às plataformas de CGH e aos painéis de SNP de alta densidade, aumentando a chance de detecção viesada de CNVs (FADISTA et al., 2010, BAE et al.; 2010, JIANG et al., 2012, XU et al., 2013). Isto ocorre, pois os painéis de 50K SNP apresentam maiores distâncias entre as sondas (“probes”), dificultando a identificação de pequenas CNVs. No entanto, painéis de SNP com alta densidade promovem resultados precisos e confiáveis. No sequenciamento de nova-geração (NGS), é possível mapear todo o genoma e caracterizar detalhadamente as variações estruturais (ZHAN et al., 2011; BICKHART et al., 2015). Entretanto, a tecnologia de NGS exige ferramentas computacionais robustas e

eficientes para trabalhar com a enorme quantidade de dados gerados (DUAN et al., 2013).

Existem diversas metodologias para detecção de CNVs,a partir dos dados de painéis de SNP, implementadas em diversos programas computacionais como GADA (PIQUE-REGI et al., 2008), QuantiSNP (COLELLA et al., 2007), PennCNV (WANG et al., 2007), SVS (Golden Helix), e por meio da técnica de NGS como, CNV-seq (XIE ; TAMMI, 2009), CNVnator (ABYZOV et al., 2011), Control-FREEC (BOEVA et al., 2012), Delly ((RAUSCH et al., 2012) e Lumpy-SV (LAYER et al., 2014).

3.4.1 Identificação das CNVs por meio dos painéis de SNP

Os algoritmos para detecção de CNVs por meio de painéis de SNP utilizam a intensidade do sinal da fluorescência emitido para cada SNP, provenientes da leitura das plataformas de genotipagem. As principais medidas de avaliação da intensidade são a razão do log de R (LRR) e a frequência do alelo B (BAF). A LRR é uma medida relativa de intensidade, sendo calculada pelo logaritmo na base dois da razão da intensidade do sinal observado da amostra (R observado) pela intensidade do sinal normalizado esperado (R esperado), sendo $LRR = log_2(R_{observado}/R_{esperado})$ (PEIFFER, et al. 2006).

A duplicação ou deleção de uma região específica é determinada pela comparação do sinal das amostras em relação a uma sequência de DNA referência, em que o aumento da intensidade do sinal da fluorescência em determinada região representa amplificações ou duplicações, e redução do sinal representa deleções (CASSESE et al., 2014). A frequência do alelo B representa a medida normalizada da taxa de intensidade relativa do sinal de cada alelo (COLELLA et al., 2007).

As metodologias implementadas nos algoritmos são basicamente paramétricas, como cadeias ocultas de Markov e não-paramétricas, como segmentação binária (OLSHEN et al., 2004). Jiang et al. (2012), em estudo realizado em bovinos da raça Chinese Holstein, compararam os programas computacionais PennCNV, GADA e cnvPartition, pelos quais foram detectadas 219, 169 e 140 CNVs, respectivamente, com apenas 42 regiões em comum para os três programas. Os autores concluíram que a utilização de vários algoritmos para detecção de CNV pode

reduzir resultados falsos-positivos, entretanto pode ocorrer aumento da taxa de falsos-negativos. Segundo Winchester, Yau e Ragoussis (2009) o tamanho da amostra e o poder dos testes estatísticos são os principais fatores para obtenção de resultados confiáveis em estudos de CNV.

O programa computacional PennCNV (WANG et al., 2007) tem sido amplamente utilizado em estudos de animais domésticos (HOU et al., 2012a; SILVA et al., 2016a; MA et al., 2017). A detecção de CNVs pelo PennCNV ocorre por meio das metodologias bayesianas do modelo das cadeias ocultas de Markov utilizando o LRR e o BAF para cada marcador SNP. Quando ocorre uma deleção, há diminuição dos valores de LRR e uma ausência de heterozigotos nos valores de BAF (os agrupamentos de genótipos de SNPs localizam-se ao redor de 0 ou 1), já na presença de uma duplicação, há aumento nos valores de LRR e uma separação do genótipo heterozigoto em dois grupos (Figura 2). Alguns estudos em bovinos têm utilizado o programa Golden Helix SNP and Variation Suite (SVS) (XU et al.,2014; ZHOU et al., 2016), em que somente o LRR é utilizado para detecção das CNVs.

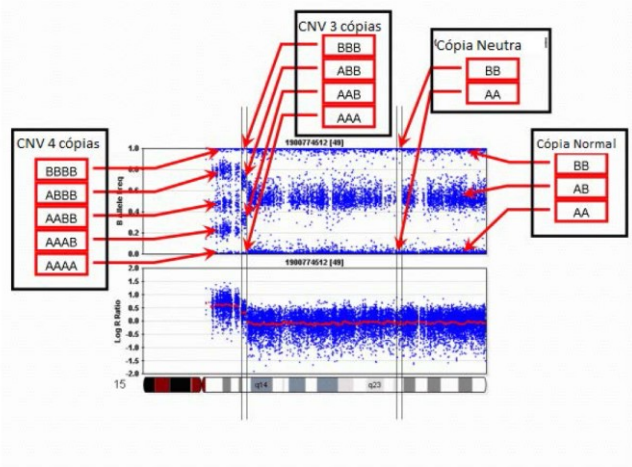


Figura 2. Ilustração dos valores da razão do log de R (LRR) e frequência do alelo B (BAF) para o cromossomo 15 de um indivíduo. Uma região normal do cromossomo tem três grupos de genótipos BAF, representados como AA, AB e BB, com valores de LRR centralizados próximo a zero. Região de cópia neutra LOH ("loss of heterozygosity"– perda de heterozigosidade) tem valores de LRR normais, mas sem o genótipo AB. O aumento no número de cópias de uma região de CNV pode ser detectado baseado no aumento do número de picos na distribuição de BAF, assim como no aumento dos valores de LRR. Os padrões de LRR e BAF para diferentes regiões de CNV, regiões com número normal de cópias (duas), regiões de LOH neutras são diferentes entre si, por isso a combinação de LRR e BAF pode ser usada para promover a identificação das CNVs (Adaptado de Wang, 2007).

3.4.2 Identificação por meio do sequenciamento de nova geração

No sequenciamento de DNA de nova geração, fragmentos do material genético são lidos como milhares de segmentos curtos, denominados de “reads” ou leituras. A montagem do genoma, consiste em alinhar todos os segmentos curtos para a reconstrução da sequência do genoma. A cobertura genômica, ou seja, o número de vezes que determinada região do genoma é coberta pelas leituras, contribui para aumentar a acurácia de identificação da sequência de DNA na região considerada. Na tecnologia do sequenciamento não existem regiões pré-definidas como nos painéis de genotipagem. Deste modo, as CNVs e os pontos de quebra (breakpoints) detectados são mais precisos, pois as leituras oriundas do NGS são aleatoriamente amostradas para todo o genoma com alta cobertura e resolução das sequências (ALKAN; COE; EICHLER, 2011; ZHAO et al., 2013)

Considerando os algoritmos de montagem de genomas, existem quatro principais métodos para identificação de CNVs por meio de dados de NGS (Figura 3). O método “paired-end mapping” (PE) (CHEN et al., 2009; HORMOZDIARI et al. 2009*),* é baseado no conceito em que as leituras possuem uma determinada distribuição para os fragmentos de DNA, em que um evento de deleção ou duplicação é detectado se as distâncias entre as extremidades dos fragmentos mapeados no genoma referência forem significativamente diferentes da média do tamanho esperado daquele fragmento de DNA. Se as extremidades dos fragmentos são mapeadas em uma orientação diferente da esperada, pode ser indicativo de um evento de inversão (FEUK, 2010). O método de “split-reads” (SR) (YE et al., 2009), consiste no mapeamento de apenas uma das “reads” no genoma referência. A leitura mapeada é utilizada como âncora para delimitar a busca pela leitura não-mapeada, sendo que a localização da leitura não-mapeada poderá ser o ponto-de-quebra da variação estrutural. A grande vantagem deste método é a identificação acurada da localização dos eventos de CNV.O método montagem *de novo*, ou seja, que não há genoma referência como base para o alinhamento das sequências, permite identificar as variações estruturais, por meio das regiões genômicas discrepantes entre a montagem *de novo* e o genoma referência.

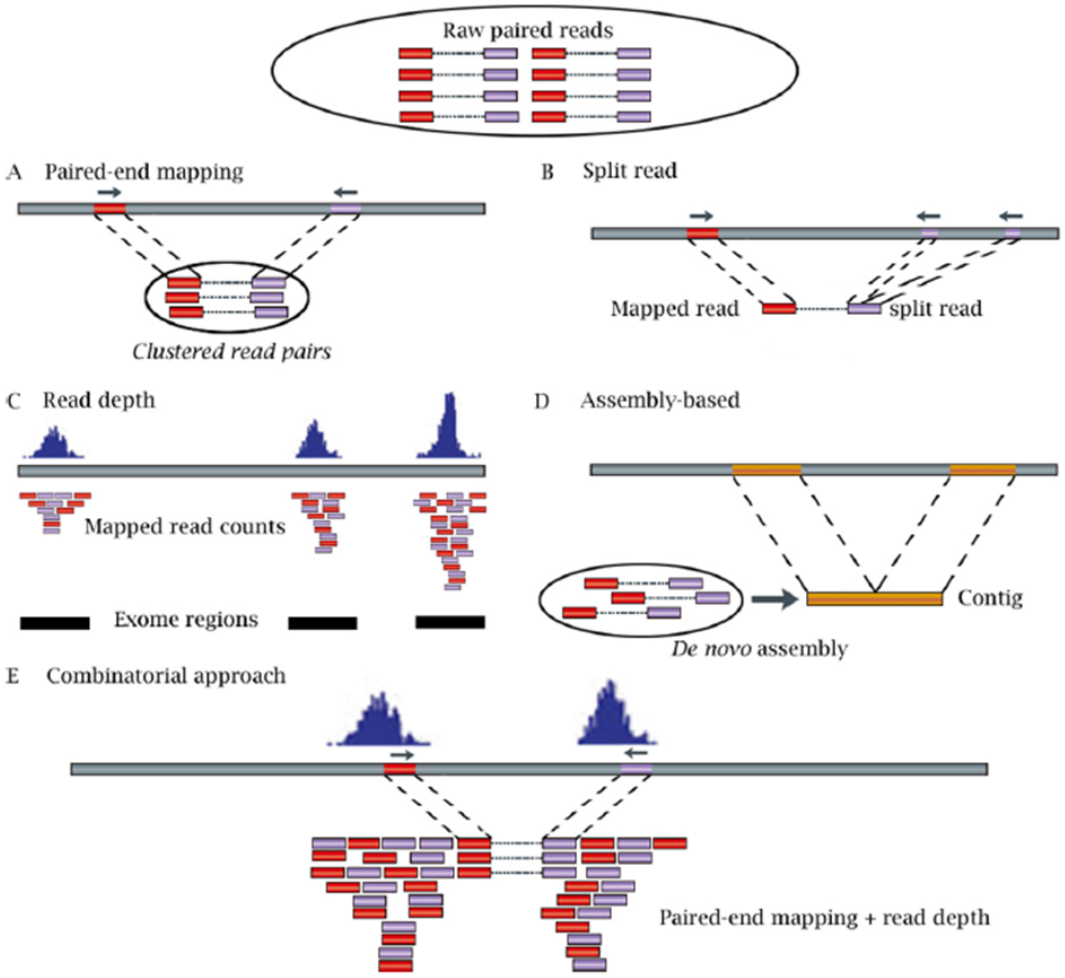


Figura 3. Diferentes métodos de detecção de CNVs por meio de dados de sequenciamento de última geração (NGS). **A.** "Paired-end mapping" (ambas extremidades de um fragmento de DNA) **B.** "Split-reads" (reads não alinhadas) **C.** "Read depth" **D.** "Assembly" **E.** Método combinando a metodologias "read-depth" e mapeamento "Paired-end". Adaptado de Zhao et al. (2013).

A detecção por meio da metodologia "read-depth" (RD) (XIE; TAMMI, 2009; MILLER et al, 2011; ABYZOV et al., 2011) depende da densidade de alinhamento das leituras ao longo dos cromossomos, em que regiões com alta cobertura são duplicações e as que apresentarem baixa cobertura são deleções. Na maioria dos algoritmos para detecção dos eventos de CNV por meio do RD é considerado o modelo de distribuição de probabilidades de Poisson (XIE;TAMMI, 2009; MILLER et al. 2011). Comparado com os demais métodos como PEM e SR, pelo método RD é possível identificar o número de eventos de CNV, uma vez que os demais permitem a identificação somente da posição das CNVs.

Além disso, o algoritmo baseado no RD é apropriado para detectar longas variações, sendo que pelos métodos baseados no PE e no SR o poder de detecção para essas CNVs é menor (YOON et al., 2009). Essa metodologia tem sido amplamente utilizada nos estudos de detecção de CNV (BICKHART et al., 2012; WANG; NETTLETON; YING, 2014; GAO et al, 2017; ZARE et al. 2017) devido a capacidade de identificar os eventos com alta-resolução. Há diversos programas que empregam o algoritmo baseado em “read-depth” tais como, SegSeq (CHIANG et al., 2009), CNVnator (ABYZOV et al. 2011) e Read-depth (MILLER et al., 2011).

3.4.3 Validação das CNVs

Após a detecção das CNVs, se faz necessário a validação das regiões encontradas por meio de técnicas moleculares, como qPCR (PCR em tempo real) (LIU et al, 2010; HOU et al., 2012; JIANG et al., 2013; XU et al., 2014). A reação quantitativa em cadeia da polimerase, é amplamente utilizada para quantificar ácidos nucleicos (DNA/RNA) pela amplificação de determinada sequência alvo, permitindo maior sensibilidade de detecção. O fragmento amplificado é monitorado em tempo real, por meio da intensidade de fluorescência emitida. Vários reagentes de amplificação estão disponíveis no mercado, incluindo sondas de hidrólise (por exemplo, sondas TaqMan®), ou moléculas de ligação que emitem fluorescência quando ligadas a cadeia dupla do DNA em regiões não-específicas, como o corante SYBR Green® (VANGUILDER et al., 2008).

Devido ao custo da análise de qPCR, quantidade suficiente de material biológico e tempo necessário para as análises, alguns autores têm considerado a validação *in silico* das CNVs em animais domésticos (SILVA et al., 2016b; LETAIEF et al., 2017). Esse tipo de validação consiste em verificar CNVs em comuns detectadas por meio de diferentes técnicas, como exemplo painéis de genotipagem e NGS. Também é possível realizar a validação por meio da informação genotípica de família, considerando que a maioria das CNVs na progênie, deve ser herdada dos pais (LOCKE et al., 2006). Entretanto, esse método de validação ainda é limitado devido as diferenças nas tecnologias, como resolução dos eventos detectados e técnicas de análises.

3.5 Estudo de associação entre CNVs e fenótipos de interesse e importância de CNVs no melhoramento genético de bovinos

A associação de regiões das CNV (CNVRs) com fenótipos de interesse, como susceptibilidade às doenças, têm sido foco de estudos para pesquisadores de diversas áreas, com o objetivo de entender melhor os processos biológicos envolvidos na expressão destas características. As CNVRs são determinadas pelo agrupamento das CNVs identificadas em diferentes amostras com sobreposição de no mínimo um par de base (REDON et al., 2006). Recomenda-se a utilização de CNVRs em estudos de associação com fenótipos, pois CNV identificadas individualmente podem detectar grande número de resultados falsos-positivos devido à alta taxa de ruído na intensidade do sinal (CASSESE et al., 2014). As análises de associação entre CNVs e características de interesse ainda não são bem estabelecidas como os estudos de associação com SNP (GWAS) (KIM et al., 2012). Vários métodos estatísticos podem ser empregados para análises de associação com CNVRs, como por exemplo regressão linear (XU et al., 2014; ZHOU et al., 2016).

O principal objetivo do estudo de CNVs na pecuária é identificar a associação destas regiões com características de interesse econômico, como produção de leite, fertilidade e doenças que afetam o desempenho produtivo. Muitas CNVs têm sido localizadas em regiões conservadas do genoma, identificadas em diferentes indivíduos e espécies, sugerindo que regiões de CNV podem estar sob seleção positiva ou negativa, permitindo a sua utilização em programas de melhoramento.

Seroussi et al. (2010), encontraram regiões de CNV localizadas no cromossomo 18 (BTA18) associadas a produção de proteína, gordura e permanência no rebanho em bovinos da raça Holstein. Nesta mesma raça, Jiang et al. (2013) relataram 367 regiões de CNV e encontraram 610 genes anotados nestas regiões com diferentes funções moleculares. Na raça Angus, Hou et al. (2012a) detectaram CNVs associadas à susceptibilidade de infestação por nematoides intestinais.

Xu et al. (2014), identificaram em bovinos da raça Holstein genotipados com o painel de 50K SNP, 34 CNVs associadas com QTL (“quantitative trait loci”) para características de produção de leite, gordura e proteína e porcentagem de gordura e proteína. Estes mesmos autores encontraram nas regiões de CNV os genes *DGAT1*

e *VPS28*, relacionados a produção de leite. Zhou et al. (2016), em bovinos da raça Nelore (*Bos taurus indicus*) encontraram 231 CNVs, sendo 17 regiões associadas com características de crescimento, com precocidade e ganho de peso ao desmame. Silva et al. (2016a) também na raça Nelore, relataram regiões de CNV associadas com maciez da carne.

Bickhart et al. (2012), detectaram 1.265 regiões de CNVs no genoma bovino utilizando dados de sequenciamento (NGS) de seis indivíduos (5 *Bos taurus taurus* e 1 *Bos taurus indicus*). Estes autores relataram que genes relacionados com resistência a patógenos e parasitas (como *CATHL4* e *ULBP17*) estavam presentes em maior número de cópias no indivíduo da raça Nelore em comparação com os bovinos taurinos, enquanto que os bovinos taurinos (Angus e Holstein) apresentaram maior número de cópias para os genes envolvidos no transporte e metabolismo de lipídeos. Boussaha et al. (2015), também utilizando dados provenientes de NGS, identificaram 547 deleções e 410 duplicações em 62 touros (27 Holstein, 17 Montbéliarde e 18 Normande). Estes autores encontraram regiões de CNV dentro ou sobrepostas a regiões de QTL para a produção de leite e escore de células somáticas. Gao et al. (2017) identificaram 6.015 CNVRs e relataram 10 genes candidatos (como *IGF2*, *FOXO3*, *SCD5*) relacionados à produção de gordura e proteína do leite, utilizando dados de NGS de oito touros da raça Holstein.

Os resultados das pesquisas têm revelado complexidade das CNVs no genoma bovino, originando questões sobre o impacto dessas variações estruturais na seleção de características produtivas em animais domésticos (BICKHART; LIU, 2014). Estudos de identificação de regiões com variações no número de cópias dos segmentos de DNA são importantes para os programas de melhoramento genético de bovinos leiteiros, pois auxiliam na compreensão dos processos biológicos envolvidos na expressão fenotípica das características de produção, reprodutivas e sanidade. Entretanto, alguns fatores ainda devem ser melhorados, como o mapa do genoma referência bovino, a anotação dos genes e as técnicas de detecção e validação.

4 REFERÊNCIAS

ABERNATHY, J.; LI, X.; JIA, X.;CHOU; W.; LAMONT, S. J.; CROOIJMANS, R.; ZHOU, H. Copy number variation in Fayoumi and Leghorn chickens analyzed using array comparative genomic hybridization. **Animal Genetics**, v. 45, p. 400-411, 2014.

ABYZOV, A.; URBAN, A. E.; SNYDER, M.; GERSTEIN, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. **Genome Research**, v. 21, p. 974-984, 2011.

ALKAN, C.; COE, B. P.; EICHLER, E. E. Genome structural variation discovery and genotyping. **Nature Review Genetics**, v. 12, p. 363-376, 2011

BAE, J.S.; CHEONG, H.S.; KIM, L.H.; NAMGUNG, S.; PARK, T.J.; CHUN, J.; KIM, J. Y.; PASAJE, C. F. A.; LEE, J. S.; SHIN, H. D. Identification of copy number variations and common deletion polymorphisms in cattle. **BMC Genomics.** v. 232, p. 232, 2010.

BABUSHOK, D. V.; KAZAZIAN, H. H. Progress in understanding the biology of the human mutagen LINE-1. **Human Mutation**, v. 28, n. 6, p. 527-539, 2007.

BAILEY, J. A.; YAVOR, A. M.; MASSA, H. F.; TRASK, B. J.; EICHLER, E. E. Segmental duplications: organization and impact within the current human genome project assembly. **Genome Research**, v.11, n.6, p. 1005-1017, 2001.

BICKHART, D. M.; HOU, Y.; SCHROEDER, S. G.; ALKAN, C.; CARDONE, M. F.; MATUKUMALLI, L. K.; SONG, J. ; SCHNABEL, R. D. ; VENTURA, M.; TAYLOR, J. F. ; GARCIA, J. F.; VAN TASSELL, C. P. ; SONSTEGARD, T. S.; EICHLER, E. E.; LIU, G. E.; HUGHES, H. Copy number variation of individual cattle genomes using next-generation sequencing. **Genome Research**, v. 22, p. 778–790, 2012.

BICKHART, D. M.; LIU, G. E. The challenges and importance of structural variation detection in livestock. **Frontiers in Genetics**, v. 5, p. 1-14, 2014.

BICKHART, D. M; HUTCHISON, J. L; XU, L.; SCHNABEL, R.D.; TAYLOR J. F.; REECY, J. M.; SCHROEDER, S.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; LIU, G. E. RAPTR-SV: a hybrid method for the detection of structural variants. **Bioinformatics**, v. 31, n. 13, p. 2084-2090, 2015.

BOEVA, V.; POPOVA, T.; BLEAKLEY, K.; CHICHE, P.; CAPPO, J.; SCHLEIERMACHER, G.; JANOUEIX-LEROSEY, I.; DELATTRE, O.; BARILLOT, E. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. **Bioinformatics**, v. 28, n. 3, p. 423-425, 2012.

BOUSSAHA, M; ESQUERRE, D.; BARBIERI, J.; DJARI,A.;PINTON,A.; LETAIEF,R.;SALIN, G.; ESCUDIÉ, F.; ROULET,A.; FRITZ,S.; SAMSON,F.; GROHS,C.; BERNARD,M.; KLOPP,C.; BOICHARD,D.; ROCHA, D. Genome-wide study of structural variants in bovine Holstein, Montbéliard e Normande dairy breeds**.** **Plos One**, v. 10, e0135931, 2015.

CASSESE, A.; GUINDANI, M.; TADESSE, M. G.; FALCIANI, F.; VANNUCCI, M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. **Annals of Applied Statistics**, v. 8, p.148-175, 2014.

CANAZA-CAYO, A.W.; LOPES, P. S.; SILVA, M. V. G. B.; COBUCI, J. A.; TORRES, R. A.; MARTINS, M. F.; ARBEX, W. A. Estrutura populacional da raça Girolando. **Ciência Rural**, Santa Maria, v.44, n.11, p. 2072-2077, 2014.

COLE, J. B.; SILVA; M.V.G. B. Invited Review Genomic selection in multi-breed dairy cattle populations. **Revista Brasileira de Zootecnia**, Viçosa, v. 45, n. 4, p.195–202, 2016.

CHEN, K.; WALLIS, J. W.; MCLELLAN, M. D.; LARSON, D. E.; KALICKI, J. M.; POHL, C. S.; MCGRATH, S. D.; WENDL, M. C.; ZHANG, Q. Y.; LOCKE, D. P.; LOCKE, D. P.; SHI, X.; FULTON, R. S., LEY, T. J.; WILSON, R. K.; DING, L., MARDIS, E. R. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. **Nature Methods,** v. 6, p. 677-681, 2009.

CHEN, C.; QIAO, R.; WEI, R.; GUO, Y.; AI, H.; MA, J.; REN, J.; HUANG, L. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. **BMC Genomics**, v. 13, p. 733, 2012.

CHIANG, D. Y.; GETZ, G.; JAFFE, D. B.; O'KELLY, M. J.; ZHAO, X.; CARTER, S. L.; RUSS, C.; NUSBAUM, C.; MEYERSON, M.; LANDER, E. S. High-resolution mapping of copy-number alterations with massively parallel sequencing. **Nature Methods**, v. 6, p. 99-103, 2009.

CICCONARDI, F.; CHILLEMI, G.; TRAMONTANO, A.; MARCHITELLI, C.; VALENTINI, A.; AJMONE-MARSAN, P.; NARDONE, A. Massive screening of copy number population-scale variation in Bos taurus genome. **BMC Genomics**, v.14, p. 124-138, 2013.

CLOP; A.; VIDAL O. ; AMILLS, M. Copy number variation in the genomes of domestic animals. **Animal Genetics**, v. 43,p. 503–17, 2012.

COLELLA, S.; YAU, C.; TAYLOR, J. M.; MIRZA, G.; BUTLER, H.; CLOUSTON, P.; BASSETT, A.S.; SELLER, A.; HOLMES, C. C.; RAGOUSSIS, J.QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. **Nucleic Acids Research**, v. 35, n. 6, p. 2013-2025, 2007.

DUAN, J.; ZHANG, J. G.; DENG, H. W.; WANG, Y. P. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. **PLos One,** v. 8, e59128, 2013

DUCOS, A.; REVAY, T.; KOVACS, A.; HIDAS, A.; PINTON, A.; BONNET-GARNIER A.; MOLTENI, L.; SLOTA, E.; SWITONSKI, M.; ARRUGA, M. V.; VAN HAERINGEN

W.A.; NICOLAE, I.; CHAVES, R.; GUEDES-PINTO, H.; ANDERSSON, M.; IANNUZZI, L. Cytogenetic screening of livestock populations in Europe: an overview**. Cytogenetics and Genome Research**, v. 120, p. 26-41.

FACÓ, O.; LÔBO, R. N. B.; MARTINS FILHO, R.; LIMA, F. A. M. Idade ao Primeiro Parto e Intervalo de Partos de Cinco Grupos Genéticos Holandês x Gir no Brasil. **Revista Brasileira de Zootecnia**, v. 34, p.1920-1926, 2005.

FACÓ, O.; LÔBO, R. N. B.; MARTINS FILHO, R.; MARTINS, G. A.; OLIVEIRA, S. M. P.; AZEVÊDO, D. M. M. R. Efeitos Genéticos Aditivos eNão-Aditivos para Características Produtivas e Reprodutivas em Vacas Holandês x Gir. **Revista Brasileira de Zootecnia**, Viçosa, v.37, n.1, p.48-53, 2008.

FACÓ, O.; MARTINS FILHO, R.; LÔBO, R. N. B.; AZEVÊDO, D. M. M. R.; OLIVEIRA, S. M. P. Efeito da Redução da Variação da Duração de Lactação na Avaliação Genética de Bovinos Leiteiros Mestiços**. Revista Ciência Agronômica**, v.40, p.287-292, 2009.

FADISTA, J.; THOMSEN, B.; HOLM, L. E.; BENDIXEN, C. Copy number variation in the bovine genome. **BMC Genomics**, v. 11, p. 284, 2010.

FEUK, L.;CARSON, A. R., SCHERER, S. W. Structural variation in the human genome. **Nature Review Genetics**, v. 7, p. 85-97, 2006.

FEUK, L. Inversion variants in the human genome: role in disease and genome architecture. **Genome Medicine**, v. 2, p. 11, 2010.

FONTANESI, L.; MARTELLI, P. L.; BERETTI, F.; RIGGIO, V.; DALL'OLIO, S.; COLOMBO, M.; CASADIO, R.; RUSSO, V.A ; PORTOLANO, B. An initial comparative map of copy number variations in the goat (Capra hircus) genome**. BMC Genomics**, v. 11, p. 639, 2010.

FONTANESI, L.; BERETTI, F.; MARTELLI, P. L.; COLOMBO, M.; DALL'OLIO, S.;OCCIDENTE, M.; PORTOLANO, B.; CASADIO, R.; MATASSINO, D.; RUSSO, V.A first comparative map of copy number variations in the sheep genome. **Genomics**, v. 97, p. 158-65, 2011.

GAO, Y.; JIANG, J.; YANG, S.; HOU, Y.; LIU, G. E.; ZHANG, S.; ZHANG, Q.; SUN, D. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. **BMC Genomics**, v. 18, p. 265, 2017.

GLÓRIA, J. R.; BERGMANN, J. A. G.; REIS, R. B.; COELHO, M. S.; SILVA, M. A. Efeito da Composição Genética e de Fatores sobre aProdução de Leite, Duração da Lactação e a Produção de Leite por Dia deIntervalo de Partos de Vacas Mestiças Holandês – Gir. **Arquivo Brasileirode Medicina Veterinária e Zootecnia**, v. 58, p.1139-1148, 2006.

GROSSI, S. F.; FREITAS, M. A. R. Eficiência Reprodutiva e Produtiva em Rebanhos Leiteiros Comerciais Monitorados por Sistema Informatizado. **Revista Brasileira de Zootecnia**, v.31, p.1362-1366, 2002.

GU, W.; ZHANG, F.; LUPSKI, J. R. Mechanisms for human genomic rearrangements. **Pathogenetics**, v. 1, p. 4, 2008.

GUIMARÃES, J. D.; ALVES, N. G.; COSTA, E. P.; SILVA, M. R.; COSTA, F. M. J.; ZAMPERLINI, B. Eficiências Reprodutivas e Produtivas em Vacas das Raças Gir, Holandês e Cruzadas Holandês x Zebu. **Revista Brasileira de Zootecnia,** v. 31, p.641-647, 2002.

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

HENRICHSEN, C.N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v.18, p. R1-8, 2009.

HORMOZDIARI, F; ALKAN, C.; EICHLER, E. E.; SAHINALP, S. C. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. **Genome Research**, v. 19, p. 1270-1278, 2009.

HOU, Y.; LIU, G. E.; BICKHART, D. M.; MATUKUMALLI, L. K.; LI, C.; SONG, J.; GASBARRE, L. C.; VAN TASSELL, C. P.; SONSTEGARD, T. S. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. **Functional & Integrative Genomics**. v.12, p. 81-92, 2012a.

HOU, Y.; BICKHART, D. M.; HVINDEN, M. L.; LI, C.; SONG, J.; BOICHARD, D. A.; FRITZ, S.; EGGEN, A.; DENISE, S.; WIGGANS, G. R.; SONSTEGARD, T.S.; VAN TASSEL, C. P.; LIU, G.E. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. **BMC Genomics**. v.13, p. 376, 2012b.

IAFRATE A.J.; FEUK L.;RIVERA M.N.; LISTEWNIK M.L.; DONAHOE P.K.; QI Y.; SCHERER S.W.; LEE C. Detection of large-scale variation in the human genome. **Nature Genetics**, v.36, p. 949–995, 2004.

JIANG, L.; JIANG, J.; WANG, J.; DING, X.; LIU, J.; ZHANG, Q. Genome wide identification of copy number variations in Chinese Holstein. **Plos One,** v**.** 7, e48732, 2012.

JIANG, L.; JIANG, J.;YANG,J.; LIU, X; WANG, H.; DING, X; LIU, J., ZHANG, Q. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. **BMC Genomics**, v.14, p.131, 2013.

KIM, J.H.; HU, H.J.; YIM, S.H.; BAE, J.S.; KIM, S.Y.; CHUNG, Y. J. CNVRuler: a copy number variation-based case-control association analysis tool**. Bioinformatics**. v. 28, p. 1790-2, 2012.

LAYER, R. M.; CHIANG, C.; QUINLAN, A. R.; HALL, I. M. LUMPY: a probabilistic framework for structural variant discovery. **Genome biology**, v.15, n. 6, 2014. Disponível em: < https://doi.org/10.1186/gb-2014-15-6-r84>

LEE, J. A.; CARVALHO, C. M.; LUPSKI, J.R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. **Cell Press**, v.131, p. 1235-1247, 2007.

LI, Y.; MEI, S.; ZHANG, X.; PENG, X.; LIU, G.; TAO, H.; WU, H.; JIANG, S.; XIONG, Y.; LI, F. Identification of genome-wide copy number variations among diverse pig breeds by array CGH. **BMC Genomics**, v. 13, p. 725, 2012.

LIEBER, M. R. The mechanism of human nonhomologous DNA end joining. **Journal of Biological Chemistry**, v. 283, p. 1-5, 2008.

LIEBER, M. R.; GU, J.; LU, H.; SHIMAZAKI,N.; TSAI, A. G. (2010). Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. **Subcellular Biochemistry**, v. 50, p. 279–296, 2010.

LIU, G. E.; HOU, Y.; ZHU, B.; CARDONE, M. F.; JIANG, L.; CELLAMARE, A.; MITRA, A.; ALEXANDER, L. J.; COUTINHO, L. L.; DELL'AQUILA, M. E; GASBARRE, L. C.; LACALANDRA, G.; LI, R. W; MATUKUMALLI, L. K.; NONNEMAN, D.; REGITANO, L. C. A.; SMITH, T. P. L.; SONG, J.; SONSTEGARD, T. S.; VAN TASSEL, C. P.; VENTURA, M.; EICHLER, E. E.; MCDANELD,T. G.; KEELE, J. W. Analysis of copy number variation among diverse cattle breeds. **Genome Research**, v. 20, p. 693-703, 2010.

LIU, J.; ZHANG, L.; LINGYANG, X.; HANGXING, R. ; LU, ZHANG, X. ZHANG, S. ZHOU, X. WEI, C.; ZHAO, F; DU, L. Analysis of copy number variation in the sheep genome using 50K SNP BeadChip array. **BMC Genomics**, v**.**14, p.229, 2013.

LOCKE, D. P., SHARP, A. J., MCCARROLL, S. A., MCGRATH, S. D., NEWMAN, T. L., CHENG, Z., SCHWARTZ, S.; ALBERTSON, D. G.; PINKEL, D.; ALTSHULER, D. M.; EICHLER, E. E. Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome. **American Journal of Human Genetics**, v. 79, n. 2, p. 275–290, 2006.

LUO, J.; YU, Y.; MITRA, A.; CHANG, S.; ZHANG, H.; LIU, G.; YANG, N.; SONG, J. Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. **G3: Genes - Genomes – Genetics,** v. 3, p. 217–223, 2013.

MA, Q.; LIU, X.; PAN, J.;MA, L.; MA, Y.; HE, X.; ZHAO, Q.; PU, Y.; LI, Y.; JIANG, L. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600 K SNP array. **Scientific Reports**, v. 7, 2017.

MANOLIO, T.A.; COLLINS, F.S.; COX, N.J.; GOLDSTEIN, D.B.; HINDORFF, L. A. HUNTER, D. J.; MCCARTHY, M. I.; RAMOS, E.M.; CARDON, L. R.; CHAKRAVARTI, A.; CHO, J. H.; GUTTMACHER, A. E.; KONG, A.; KRUGLYAK, L.; MARDIS, E.; ROTIMI, C. N.; SLATKIN, M.; VALLE, D.; WHITTEMORE, A. S.; BOEHNKE, M.; CLARK, A. G.; EICHLER, E.E.; GIBSON, G.; HAINES, J. L.; MACKAY, T. F.; MCCARROLL, S. A.; VISSCHER, P. M. Finding the missing heritability of complex diseases, **Nature,** v.461, n.7265, p. 747-753, 2009.

MARGARETO, J.; LEIS, O.; LARRARTE, E.; POMPOSO, I. C.; GARIBI, J. M.; LAFUENTE, J. V. DNA copy number variation and gene expression analyses reveal the implication of specific oncogenes and genes in GBM. **Cancer Investigation**, v. 27, p. 541–548, 2009.

MCMANUS, C.; TEIXEIRA, R. A.; DIAS, L. T.; LOUVANDINI, H.; OLIVEIRA, E. M. B. Características Produtivas e Reprodutivas de Vacas Holandesas e Mestiças Holandês x Gir no Planalto Central. **Revista Brasileira de Zootecnia**, v.37, p.819-823, 2008

MILLER, C. A.; HAMPTON, O.; COARFA, C., MILOSAVLJEVIC, A.ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads. **PLoS One**, v. 6, e16327, 2011.

MORRISH, T.A.; GILBERT, N.; MYERS, J. S.; VINCENT, B. J.; STAMATO,T.D.; TACCIOLI, G. E.; BATZER, M. A.; MORAN, J. V. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. **Nature Genetics**, v. 31, n. 2, p. 159-165, 2002.

OLSHEN, A.B.; VENKATRAMAN, E.S.; LUCITO, R.; WIGLER, M. Circular binary segmentation for the analysis of array-based DNA copy number data. **Biostatistics**, v. 5, p. 557–72, 2004.

OLSON, K. M.; VANRADEN, P. M. AND TOOKER, M. E. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. **Journal of Dairy Science,** Champaign, v. 95, n. 9 p. 5378-5383, 2012.

PEIFFER, D.A.; LE, J. M.; STEEMERS, F. J.; CHANG, W.; JENNIGES, T.; GARCIA, F.; HADEN, K.; LI, J.; SHAW, C.A.; BELMONT, J.; CHEUNG, S. W.; SHEN, R. M.; BARKER, D. L.; GUNDERSON, K. L. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. **Genome Research**, v. 16, p. 1136-1148, 2006

PIQUE-REGI, R.; MONSO-VARONA, J.; ORTEGA, A.; SEEGER, R. C.; TRICHE, T. J.; ASGHARZADEH, S. Sparse representation and Bayesian detection of genome copy number alterations from microarray data. **Bioinformatics**, v. 24, p. 309-318, 208.

RAUSCH, T.; ZICHNER, T.; SCHLATTL, A.; STÜTZ, A. M.; BENES, V.; KORBEL, J. O. DELLY: structural variant discovery by integrated paired-end and split-read analysis. **Bioinformatics**, Oxford, v. 28, n.18, p. i333 – i339, 2012

REDON, R.; ISHIKAWA, S., FITCH, K. R.; FEUK, L.; PERRY, G. H.; ANDREWS, T. D.; FIEGLER, H.; SHAPERO, M. H.; CARSON, A. R.; CHEN, W.; CHO, E. K.;

DALLAIRE, S.; FREEMAN, J. L.; GONZÁLEZ, J. R; GRATACÓS, M.; HUANG, J.; KALAITZOPOULOS, D.; KOMURA, D.; MACDONALD, J. R.; MARSHALL, C. R.; MEI, R.; MONTGOMERY, L.; NISHIMURA, K.; OKAMURA, K.; SHEN, F.; SOMERVILLE, M. J.; TCHINDA, J.; VALSESIA, A.; WOODWARK, C.; YANG, F.; ZHANG, J.; ZERJAL, T.; ZHANG, J.; ARMENGOL, L. L.; CONRAD, D. F.; ESTIVILL, X.; TYLER-SMITH, C.; CARTER, N. P.; ABURATANI, H.; LEE, C.; JONES, K. W.; SCHERER, S. W.; HURLES, M. E. Global variation in copy number in the human genome. **Nature**, v. 444, p. 444–454, 2006.

SACHIDANANDAM R, WEISSMAN D, SCHMIDT S C, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. **Nature,** v.409, n. 6822, 928-933, 2001

SEROUSSI, E.; GLICK, G.; SHIRAK, A.; YAKOBSON, E.; WELLER, J. I.; EZRA, E.; ZERON, Y. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. **BMC Genomics,** v.11, p. 673, 2010

SHAW, C. J.; LUPSKI, J. R. Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. **Human Genetics**, v.116, p. 1–7, 2005.

SILVA, J. C. P. M.; VELOSO, C. M. **Raças de Gado Leiteiro**. 1.ed. Viçosa: Aprenda Fácil, 2011. 149p

SILVA, M. V. G. B; MARTINS, M. F.; PAIVA, L. C.; CEMBRANELLI, M. A. R; ARBEX, W. A.; SANTO, K. C. L.; PANETTO, J. C. C.; COSTA, C. N.; CARVALHO, B. C. Programa de Melhoramento Genético da Raça Girolando – Sumário de Touros – Resultado do Teste de Progênie – Julho 2013. Juiz de Fora : Embrapa Gado de Leite, 2013. 52 p. (Embrapa Gado de Leite. Documentos, 163).

SILVA, V. H. da; REGITANO, L. C. de A.; GEISTLINGER, L.; PÉRTILLE, F.; GIACHETTO, P. F.; BRASSALOTI, R. A.; MOROSINI, N. S.; ZIMMER, R.; COUTINHO, L. L. Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. **PlosOne**, v. 11, n. 6, e0157711, 2016a

SILVA, J. M. da; GIACHETTO, P. F.; SILVA, L. O. da; CINTRA, L. C.; PAIVA, S. R.; YAMAGISHI, M. E. B.; CAETANO, A. R. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTL affecting production traits. **BMC Genomics,** v. 17, p.454, 2016b.

SKINNER, B. M.; ROBERTSON, L. B.; TEMPEST, H. G.; LANGLEY, E. J.; IOANNOU, D.; FOWLER, K. E.; CROOIJMANS, R. P.; HALL, A. D.; GRIFFIN, D. K.; VOLKER, M. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. **BMC Genomics**, v.10, p. 357, 2009.

STOTHARD, P.; CHOI, J. W. ; BASU U.; SUMNER-THOMSON, J. M.; MENG, Y .; LIAO, X.; MOORE S. S. Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. **BMC Genomics**, v.12, p.559, 2011.

USDA – UNITED STATES DEPARTMENT OF AGRICULTURE. Dairy: World Markets

and Trade. 2017. Disponível em : http://usda.mannlib.cornell.edu/usda/current/dairy-market/dairy-market-07-20-2017.pdf. Acessado em : 12/01/2017.

VANGUILDER, H. D.; VRANA, K. E.; FREEMAN, W. M. Twenty-five years of quantitative PCR for gene expression analysis. **Biotechniques**, v. 44, n. 5, p.619-26, 2008.

VASCONCELLOS, B. F.; PÁDUA, J. T.; MUÑOZ, M. F. C.; TONHATI, H. Efeitos Genéticos e Ambientais sobre a Produção de Leite, o Intervalo de Partos e a Duração da Lactação em um Rebanho Leiteiro com Animais Mestiços, no Brasil. **Revista Universidade Rural,** v.23, p.39-45, 2003.

WANG, K.; LI, M.; HADLEY, D.; LIU, R.; GLESSNER, J.; GRANT, S. F. A.; HAKONARSON, H.; BUCAN, M. PennCNV: an integrated hidden Markov model designed for highresolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v.17, p.1665–1674, 2007.

WANG, X.; NAHASHON, S.; FEASTER, T. K.; BOHANNON-STEWART, A.; ADEFOPE, N. An initial map of chromosomal segmental copy number variations in the chicken**. BMC Genomics**, v.11, p.351, 2010.

WANG, J.; WANG, H.; JIANG, J.; KANG, H.; FENG, X.; ZHANG, Q.; LIU, J. Identification of genome – wide copy number variations among diverse pig breeds using SNP genotyping arrays. **Plos one**, v.8, e68683, 2013.

WANG, H.; NETTLETON, D.; YING, K. Copy number variation detection using next generation sequencing read counts. **BMC Bioinformatics**, v. 15, p. 109, 2014.

WIGGANS, G. R.; COLE, J. B.; HUBBARD, S. M.; SONSTEGARD, T. S. Genomic Selection in Dairy Cattle: The USDA experience**. Annual Review of Animal Biosciences**, v. 5, p. 309 -327, 2017.

WINCHESTER, L.; YAU, C.; RAGOUSSIS, J. Comparing CNV detection methods for SNP arrays. **Briefings in Function Genomics and Proteomics**, v. 8, p. 353-366, 2009.

XIE, C.; TAMMI, M. T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. **BMC Bioinformatics**, v. 10, p. 80, 2009.

XU. L.; HOU, Y.; BICKHART, D.M.; SONG, J.; LIU, G. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. Microarrays, v. 2, p. 171-185, 2013.

XU, L.; COLE, J. B.; BICKHART, D. M.; HOU, Y.; SONG, J.; VANRADEN, P. M.; SONSTEGARD, T. S.; VAN TASSELL, C. P.; LIU, G. E. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. **BMC Genomics**, v. 15, p.683, 2014.

YE, K.; SCHULZ, M. H.; LONG, Q.; APWEILER, R.; NING, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. **Bioinformatics**. v. 25, p. 2865-2871, 2009.

YOON, S.; XUAN, Z.; MAKAROV, V.; YE, K.; SEBAT, J. Sensitive and accurate detection of copy number variants using read depth of coverage**. Genome Research**, v. 19, p. 1586-1592, 2009.

ZHAN, B.; FADISTA, J.; THOMSEN B.; HEDEGAARD, J.; PANITZ, F.; BENDIXEN, C. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. **BMC Genomics**, v. 12, p. 557, 2011.

ZHANG, F.; GU, W.; HURLES, M.E.; LUPSKI, J. R. Copy number variation in human health, disease, and evolution. **Annual Review Genomics and Human Genetics,** v. 10, p. 451-481, 2009.

ZARE F.; DOW, M.; MONTELEONE, N.; HOSNY, A.; NABAVI, S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. **BMC Bioinformatics**, v. 18, n. 1, p. 286, 2017.

ZHAO, M.; WANG, Q.;WANG, Q.; JIA, P.; ZHAO, Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. **BMC Bioinformatics**, v. 14, Suppl 11:S1, 2013.

ZHOU, Y.; UTSUNOMIYA, Y.T.; XU, L.; HAY, E.H.A.; BICKHART, D.M.; ALEXANDRE, P. A.; ROSEN, B. D.; SCHROEDER, S. G.; CARVALHEIRO, R.; NEVES, H. H. R.; SONSTEGARD, T.S.; VANTASSELL C.P.; FERRAZ, J. B. S.; FUKUMASU, H. GARCIA J.F.; LIU G.E. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. **BMC Genomics**, v. 17, 2016.

CAPÍTULO 2 IDENTIFICAÇÃO DE VARIAÇÃO NO NÚMERO DE CÓPIAS DE SEGMENTOS DE DNA EM BOVINOS GIROLANDO

RESUMO – Variação no número de cópias nos segmentos de DNA (CNV), como deleções e duplicações têm sido identificadas no genoma bovino por meio de diversas técnicas, tais como painéis de genotipagem de SNPs e sequenciamento de nova geração (NGS). CNVs podem influenciar fenótipos relacionados principalmente à sanidade e reprodução. O objetivo deste estudo foi detectar regiões de CNV no genoma de bovinos Girolando (Gir X Holandês) e identificar genes candidatos próximos a essas regiões para características produtivas, de sanidade e reprodutivas. Foram utilizados registros de 1607 vacas genotipadas com painel de polimorfismo de nucleotídeo único (SNP) de média densidade (50K SNP), 280 touros genotipados com painel de alta densidade (HD SNP) provenientes da base de dados genômicos da Embrapa Gado de Leite, Juiz de Fora, MG. e dois touros ressequenciados. A identificação das CNVs utilizando painéis de SNP foi realizada por meio do modelo das cadeias ocultas de Markov implementado no programa *PennCNV* e por meio dos dados de sequenciamento de nova geração (NGS) empregando a metodologia "read-depth" no programa CNVnator. As regiões de CNV (CNVR) foram determinadas pelo agrupamento das CNVs individuais com pelo menos um par de base em sobreposição. As CNVRs candidatas foram denominadas por meio da confirmação *in silico*, em que regiões detectadas pelos painéis de 50K e HD estivessem em sobreposição recíproca de no mínimo 50% com as regiões detectadas nos touros por meio de dados NGS. Foi realizado análise de prospecção dos genes e de enriquecimentos gênico ($p < 0,05$) por meio dos termos Mesh e Gene Ontology. A partir dos dados de NGS foram confirmadas 873 CNVs detectadas em 418 fêmeas pelo painel de 50K, agrupadas em 203 CNVRs, com tamanho médio de 53,27 kb (quilobase). Para a detecção pelo painel HD, 4.402 CNVs foram confirmadas em 262 touros e agrupadas em 213 CNVRs. Um total de 184 e 198 CNVRs candidatas selecionadas pelos painéis 50K e HD, respectivamente, estavam em sobreposição com QTL conhecidos associados a características de importância econômica de bovinos de leite como produção de leite, gordura e proteína. Os principais genes candidatos anotados dentro das CNVRs foram *GBP4*, *GIMAP7, OR6C1*, *ZNF800, WC1, IGLL1* e *IGF2.* Foi encontrada uma duplicação acurada no cromossomo BTA5 relacionada a resistência a parasitas e uma deleção especifica de zebuínos associada a eficiência reprodutiva em fêmeas. No cromossomo BTA17 foi detectada uma duplicação também associada a resistência a parasitas e doenças. Também foram observados dentro das CNVRs genes da família do complexo de histocompatibilidade bovina (MHC) relacionados a resistência e susceptibilidade à mastite. A detecção e caracterização das CNVs realizadas em bovinos de leite demonstrou melhor entendimento de características como resistência a doenças e eficiência reprodutiva em bovinos de raça composta. Evidências de variações no número de cópias em genes do sistema imune adaptativo podem ter sido originarias de raças zebuínas.

Palavras-chave: *Bos taurus taurus, Bos taurus indicus*, dados de NGS, painéis de genotipagem, raça composta

CHAPTER 2 COPY NUMBER VARIATION DISCOVERY IN GIROLANDO CATTLE

ABSTRACT – Copy number variations (CNVs), such as deletions and duplications have been identified in cattle genome using single nucleotide polymorphism (SNP) arrays and next-generation sequencing data (NGS). CNVs may influence phenotypes linked mainly to heath and reproduction. The objective of this work was to identify and to describe CNVs in Girolando cattle (Gir X Holandês). A total of 1,607 Girolando females were genotyping using the medium-density SNP array (50K data), 280 bulls were genotyped using high-density SNP array (HD) and two bulls were resequenced (NGS data). The genomic data was provided by Embrapa Dairy Cattle, Juiz de Fora, MG. PennCNV software was carried out on SNP genotyping data (50K and HD data) to call single CNVs and on the NGS data was performed a read-depth approach using CNVnator software. The copy number variation regions (CNVRs) were built merging CNVs callers considering genomic regions overlapping by at least 1 base pair. Candidates CNVRs were called according to *in silico* confirmation, which CNV calls from 50K and HD overlapped reciprocally 50% with NGS CNVs calls. The gene content and QTL were found into the candidate CNVRs and enriched analysis was performed using Gene Ontology and Mesh Terms and KEGG pathways. A total of 873 CNVs from 418 females from 50K SNPs array were confirmed in NGS data, they were overlapped in 203 regions, with average length of 53.27 kb (kilobase). For HD data, 4,402 CNVs calls were confirmed through the NGS data, aggregated in 213 CNV regions, with average length of 63.90 kb. A total of 184 and 198 candidates CNVRs from 50K and HD were overlapped with known QTL associated with milk, protein and fat traits in dairy cattle. The mainly candidates' genes annotated in the candidate CNVRs were *GBP4*, *GIMAP7, OR6C1*, ZNF800, *WC1, IGLL1 and IGF2.* We found a high-confidence duplication on BTA5 linked to parasite resistance and a Zebu deletion-specific in bulls related with a reproductive efficiency in cows. On BTA17 we detected a duplication also associated with resistance parasites and diseases. We also identified a group of genes from the major histocompatibility complex (MHC) related to susceptibility and resistance of mastitis in dairy cattle. The CNV discovery in Girolando breed may improve our understanding about the phenotypic expression of the important economic traits, including reproductive efficiency and parasite susceptibility. Our findings showed evidence of specific copy number variation in indicine breeds involved with adaptive immune system.
Keywords: *Bos taurus taurus, Bos taurus indicus*, composite breed, genotyping panel, NGS data

1 INTRODUÇÃO

Com a redução dos custos nos painéis de genotipagem, avanços na tecnologia de sequenciamento de nova geração (NGS) e o progresso na montagem do genoma bovino, variações estruturais cromossômicas, como CNVs, têm sido amplamente identificadas no genoma bovino. Variação no número de cópias dos segmentos de DNA (CNV) são definidas como regiões genômicas a partir de 50 pares de bases (bp) até vários megabases(Mb) com número variável de cópias em comparação a um genoma de referência (CONRAD et al., 2010; MILLS et al., 2011). Essas variações podem ser eventos de deleções ou duplicações dos segmentos cromossômicos, sendo uma das principais fontes de variabilidade genética em humanos e animais domésticos (REDON et al., 2006; STOTHARD et al., 2011).

Embora polimorfismos de nucleotídeo único (SNP) tem sido a mais frequente forma de variação de DNA conhecida presente no genoma dos mamíferos, alterações no número de cópias causam grandes efeitos, incluindo mudança na estrutura e dosagem dos genes, pois podem resultar em alteração da regulação gênica e manifestação dos alelos recessivos (ZHANG et al., 2009) devido às diferenças nos níveis de formação do produto gênico. As CNVs influenciam diretamente ou indiretamente na expressão gênica com funções biológicas especificas, tais como imunidade, lactação e reprodução (HENRICHSEN et al., 2009, ZHANG et al. 2009) e consequentemente tem influenciado a variação fenotípica de características (WANG et al., 2007). Além disso, esse tipo de variação estrutural tem mostrado importante função em estudos de evolução das espécies (PERRY et al., 2007) e genética de populações (XU et al., 2016).

No melhoramento genético de bovinos, as CNVs podem auxiliar na tomada de decisão em processos de seleção por meio do entendimento biológico das características de produção, sanidade e reprodução. Grande número de CNVs têm sido identificadas em raças taurinas (HOU et al., 2012a; XU et al., 2014; GAO et al., 2017) e raças zebuínas (ZHOU et al., 2016a; SILVA et al., 2016a; SILVA et al., 2016b). Vários estudos têm relatado associações entre CNVs e características de importância econômica, como produção de leite em bovinos leiteiros (XU et al.,2014) e de crescimento, em bovinos de corte (ZHOU et al., 2016a).

Diferentes técnicas moleculares tais como painéis de média e alta densidade

de marcadores do tipo SNP e sequenciamento de nova geração (NGS) tem sido aplicadas para identificar os eventos de deleções ou duplicações no genoma bovino. Devido à disponibilidade para uma grande diversidade de espécies e o custo, os painéis de SNP tornaram-se as ferramentas mais utilizadas para a identificação das CNVs no genoma bovino (LIU et al., 2013). Os algoritmos aplicados para a detecção com base nos painéis de genotipagem utilizam medidas de intensidade da luminosidade, como o log da razão de R (LRR) e frequência do alelo B (COLLELA et al., 2007). No entanto, a grande desvantagem desse método é que não é possível detectar pequenas variações com precisão (BENTLEY et al., 2008).

A tecnologia de NGS também tem sido empregada em estudos de identificação de CNVS em bovinos (SILVA et al., 2016b, GAO et al., 2017; LETAIEF et al., 2017). Essa técnica apresenta vantagens em relação aos painéis de SNPs, pois permite identificar pequenas CNVs e os pontos-de-quebra das variações mais precisamente (BICKHART et al., 2012) dado que essa ferramenta molecular permite uma detecção com alta resolução e precisão. Existem diversas metodologias disponíveis na literatura para a detecção por meio de NGS, tais como, “read-depth” (CAMPBELL et al., 2008; ABYZOV et al., 2011), mapeamento "paired-end” (CHEN et al., 2009) e “split reads" (WANG et al., 2011).

Em condições de clima tropical, como no Brasil, produtores e melhoristas têm utilizado intensamente animais cruzados e raças compostas, como a raça Girolando, com a finalidade de maximizar a produção, a eficiência e a rentabilidade da cadeia produtiva do leite (VASCONCELLOS et al., 2003). Girolando é uma das mais importantes raças compostas brasileiras, provenientes de animais taurinos (Holandês) e zebuínos (Gir), sendo a proporção de genes padrão da raça 5/8 Holandês e 3/8 Gir. As vacas Girolando apresentam características morfológicas e fisiológicas importantes para a produção de leite em países tropicais, incluindo habilidade termorreguladora, duração e persistência da lactação e capacidade e suporte do úbere (SILVA & VELOSO, 2011).

Os programas brasileiros de melhoramento genético de bovinos de leite têm buscado melhorar a produção e a composição do leite, a fertilidade e a sanidade (SILVA et al., 2013). Devido a necessidade de estudos envolvendo CNVs em populações de animais cruzados e a importância desses para a indústria do leite,

principalmente em países de clima tropical, o objetivo deste estudo foi identificar regiões CNV no genoma de bovinos Girolando e identificar genes relacionados a essas regiões para características produtivas, reprodutivas e de sanidade.

2 MATERIAL E MÉTODOS

2.1 Animais, genotipagem e ressequenciamento

O DNA genômico das fêmeas foi extraído a partir de amostras de sangue e dos machos a partir de amostras de sêmen utilizando DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA), conforme recomendado pelo fabricante. O DNA extraído foi quantificado e avaliado por espectrofotometria (NanoDrop 1000, Thermo Scientific, Wilmington, DE, USA).

Foram genotipadas 690 e 917 fêmeas com o painel BovineSNP50 v2.0 (54.609 SNPs) e Zoetis Custom SNP chip ZM2 (69.022 SNPs), respectivamente. Para as análises de detecção de CNVs foram considerados somente os marcadores em comum entre os dois painéis, totalizando 48,014 SNPs (50K). A distância média e a mediana entre os marcadores considerados nas análises foi de 54,80 kb (quilobase) e 39,51 kb, respectivamente. Um total de 280 touros foram genotipados com o painel BovineHD BeadChip (HD) (Illumina Inc., San Diego, CA, USA) consistindo em 777.962 marcadores distribuídos ao longo do genoma, com distância média entre SNPs igual a 3,43 kb e mediana igual a 2,68 kb. O parentesco médio considerando a informação de pedigree observado entre todos os animais genotipados foi de 0,001.

O genoma completo de dois touros da população de Girolando foi ressequenciado por meio da plataforma Illumina HiSeq2000 (Illumina Inc., San Diego, CA, USA). As bibliotecas utilizadas para o ressequenciamento foram do tipo ‘paired-end’, ou seja, houve sequenciamento de ambas as extremidades dos fragmentos de DNA com tamanho de 500bp, em que foram produzidas leituras com tamanho de 2 x 100 bp e 200 bp. Os procedimentos de construção das bibliotecas foram realizados de acordo com os protocolos sugeridos pelo fabricante.

2.2 Identificação de CNVs por meio dos painéis de genotipagem e controle de qualidade

A identificação das CNVs utilizando os dados de genotipagem (50K e HD) foi realizada por meio da metodologia bayesiana do modelo das cadeias ocultas de Markov implementado pelo programa computacional PennCNV (WANG et al., 2007, por meio da opção -test, sem considerar a informação de pedigree. O algoritmo utiliza o valor normalizado da intensidade total da fluorescência para os dois alelos do SNP (log da razão de R - LRR) e a medida normalizada da relação de intensidade alélica (frequência do alelo B - BAF) para cada marcador SNP, obtidos a partir da leitura da plataforma de genotipagem. Os estados ocultos de Markov identificados pelo programa PennCNV e o número de cópias correspondentes estão descritos na Tabela 1.

Tabela 1. Estados ocultos de Markov, número de cópias e respectivas descrições.

Estado	Número de cópias	Descrição para os autossomos	Genótipos
1	0	Deleção dupla	NULO
2	1	Deleção de uma cópia	A, B
3	2	Normal	AA, AB, BB
4	2	Cópia neutras com perda de heterozigose (LOH)	AA, BB
5	3	Duplicação simples	AAA, AAB, ABB, BBB
6	4	Duplicação dupla	AAAA, AAAB, AABB, ABBB, BBBB

A distribuição do LRR para cada SNP dos painéis de genotipagem ao longo de todo genoma está representada na Figura 1. Pode -se observar que a variação do LRR para o painel de 50K (Figura 1A) foi maior do que para o painel HD (Figura 1B).

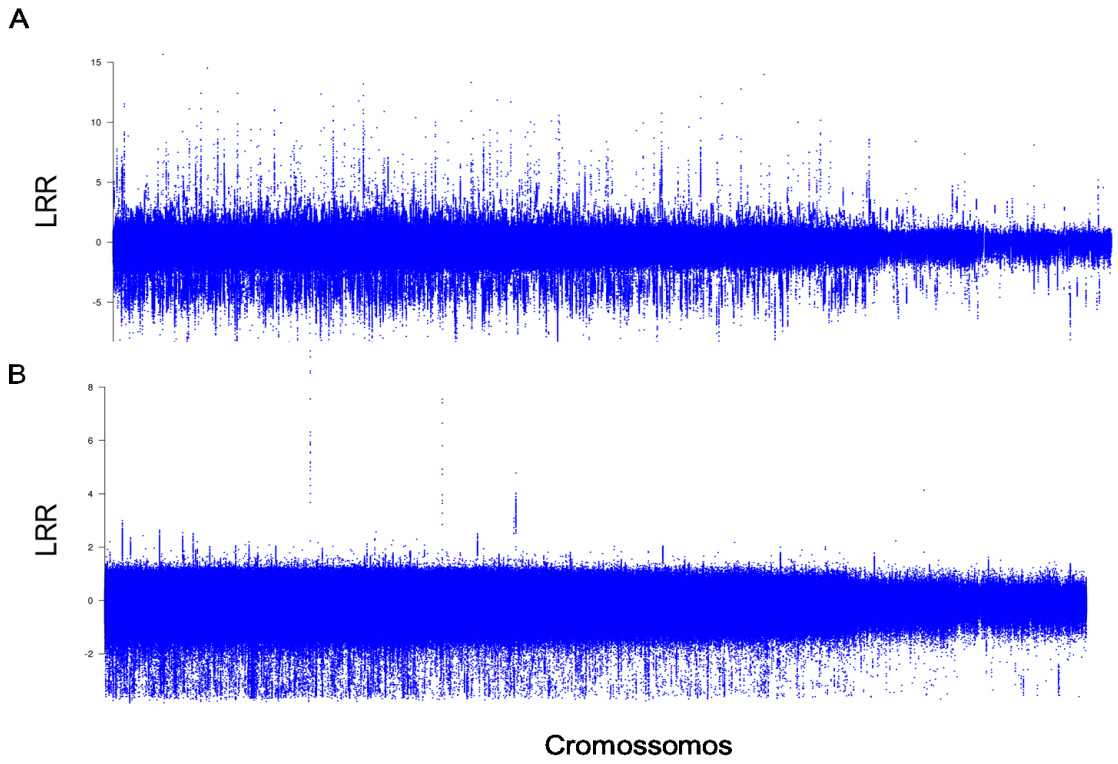


Figura 1. Representação do Log da Razão de R para os dois alelos do SNP ao longo dos cromossomos de toda a população dos indivíduos genotipados para o painel de 50K (A) e para o painel HD (B).

Foram considerados nas análises SNPs, com escore de leitura (“GenCall score”) superiores a 0,15, com posição conhecida e somente os cromossomos autossômicos. O arquivo da frequência do alelo B na população (PFB) foi criado com base no BAF de cada marcador de todas as amostras. Com a finalidade de reduzir a taxa de resultados falsos positivos, os valores de LRR de cada SNP foram ajustados para a dispersão da intensidade do sinal (“genomic waves”) ao longo das regiões genômicas (DISKIN et al. 2008) de acordo com o conteúdo de guanina e citosina (GC) esperado no genoma bovino, considerando uma região de 500 quilobase (kb) ao redor de cada marcador. Após a identificação das CNVs foi verificada a distribuição empírica das CNVs entre todos os animais e para evitar possível viés foi realizado o controle de qualidade descrito abaixo de acordo com o painel de genotipagem utilizado:

Para o painel de 50K foram utilizados os seguintes critérios de exclusão: CNVs com menos de três SNPs, amostras com mais de oito CNVs (valores discrepantes dos limites do quartil superior e inferior considerando 1,5 x Q3) e os valores de desvio-padrão do LRR maior do que 0,30, BAF “drift” maior do que 0,01 e fator de dispersão

da intensidade do LRR (“waviness factor”) maior do que 0,05 conforme critérios realizados por Hou et al., 2011 e LIU et al., 2013.

Para o painel HD foram utilizados os seguintes critérios de exclusão: CNVs com menos de 10 SNPs, amostras com mais de 82 CNVs (valores discrepantes dos limites do quartil superior e inferior) e os valores de desvio-padrão do LRR maior do que 0,30, BAF “drift” maior do que 0,01 e fator de dispersão da intensidade do LRR (“waviness factor”) maior do que 0,05.

As regiões de CNV (CNVR) foram definidas de acordo com a descrição de Redon et al. (2006). As CNVs individuais foram agrupadas considerando regiões genômicas de no mínimo um par de base (bp) em sobreposição (“overlap”) entre todas as amostras utilizando a opção *mergeBed* da ferramenta BEDtools suite (QUINLAN; HALL, 2010). As regiões de CNV foram classificadas em deleções, em que o animal apresentou uma região com perda do segmento cromossômico, duplicações em que ocorreu uma duplicação da região e mistas, ou seja, alguns indivíduos apresentaram deleções e outros duplicações na mesma região cromossômica do DNA.

2.2.1 Detecção das CNVs com base na família

Para esta análise foram considerados apenas marcadores nos cromossomos autossômicos em comum entre os três painéis de genotipagem (42.639 SNPs). Primeiramente, as CNVs foram detectadas considerando os 1.928 animais da população conforme descrito anteriormente, no entanto um novo arquivo PFB foi criado, uma vez que fêmeas e machos foram considerados juntos para a nova detecção e consequentemente alterando a frequência do alelo B na população. Após a detecção de CNVs na população foi realizada uma etapa de detecção com base na estrutura familiar em 18 trios familiares (progênie, pai e mãe).

O algoritmo utilizado para o novo procedimento de detecção das CNVs foi implementado pelo programa PennCNV como descrito por Wang et al. (2007) por meio da opção -trio. De acordo com Locke et al. (2006), a maioria das CNVs em uma progênie são herdadas dos pais. Deste modo, as CNV inferidas na progênie, mas não detectadas em nenhum dos pais, podem ser consideradas como resultados falsos positivos.

2.3 Detecção das CNVs usando os dados de NGS

As leituras dos fragmentos de DNA (“reads”) obtidas do ressequenciamento dos touros Girolando foram mapeadas ao genoma de referência UMD 3.1 por meio do algoritmo BWA-MEM (Burrows-Wheeler Aligner) (LI; DURBIN, 2009). A detecção das CNVs foi realizada por meio do programa CNVnator (ABYZOV et al., 2011) baseado na metodologia “read-depth”, em que deleções e duplicações são detectadas com base na densidade do alinhamento das leituras ao longo dos cromossomos. O tamanho da janela (“bin size”) considerado para detecção das CNVs foi de 500 bp seguindo as recomendações de Abyzov et al. (2011).

Foram consideradas apenas CNVs significativas ($p < 0,01$) para o teste estatístico t, em que a hipótese testada é se a média de sinal de profundidade das leituras é a mesma da média empírica, CNVs com fração de leituras mapeadas com baixa qualidade menores do que 0,5 ($q0 < 0,5$) e CNVs com tamanhos maiores do que um kb.

2.4 Validação *in silico*

Foi denominado neste trabalho como validação *in silico* a confirmação de regiões de CNV pela comparação entre as CNVs detectadas pelo painel de SNP e as variações detectadas nos dois touros por meio de dados de NGS. Embora o número de animais seja extremamente pequeno essa metodologia possui melhor resolução e maior sensibilidade na detecção. Foi considerada CNVs de alta-confiabilidade aquelas que apresentassem no mínimo 50% em sobreposição mútua entre a detecção pelos dados de SNP e NGS. A comparação dos arquivos foi realizada pela opção *intersect* proveniente da ferramenta BEDtools suite (QUINLAN; HALL, 2010) descrita na linha de comando abaixo:

$ intersectBed -a ngs.cnvs.sorted.bed -b 50Kcnvs.sorted.bed -f 0.50 -r > ngs.50K.cnvrs.candidates

$ intersectBed -a ngs.cnvs.sorted.bed -b HD.cnvs.sorted.bed -f 0.50 -r > ngs.HD.cnvrs.candidates

As CNVs candidatas foram agrupadas em regiões considerando um par de base em sobreposição.

2.5 Análise funcional por meio de prospecção dos genes e análise de enriquecimento gênico nas regiões candidatas de CNV

Para verificar os genes anotados dentro das CNVRs candidatas foi utilizado a plataforma Ensembl (http://www.ensembl.org/) ou RefSeq genes. Também foram verificados os QTL ("quantitative trait loci") que estavam em pelo menos 50% em sobreposição as CNVRs candidatas. O mapeamento de QTL foi realizado com base no Animal Genome database (http://128.206.12.216/drupal/sites/bovinegenome.org/files/data/umd3.1/Animalgenome_UMD3.1_QTL.gff3.gz)

O grupo de genes foi classificado de acordo com sua função pelos termos do Gene Ontology (GO) e do Medical Subject Headings (MeSH) por meio das ferramentas de anotação e mapeamento genômico (GOstats, meshr, MESH.db, MESH.Bta.eg.db) disponíveis pelo Bioconductor (http://bioconductor.org/) implementados em linguagem R. Os termos biológicos do Gene Ontology são divididos em três grupos: componente celular, processos biológicos e funções moleculares e os termos MeSH são divididos nas categorias fenômenos e processos, doenças, químicos e drogas e anatomia (MOROTA et al., 2015). Para verificar se um conjunto de genes foi enriquecido significativamente ($p < 0,05$) para determinada via biológica foi realizado um teste hipergeométrico não-paramétrico (ADAMS; SKOPEK, 1987) em que se calcula a probabilidade de que determinado conjunto de genes esteja representado mais vezes do que o esperado em comparação com a lista dos genes anotados no genoma referência em determinada via biológica (MOROTA et al., 2015).

A investigação e visualização dos mapas metabólicos paras os genes anotados identificados nas CNVRs foram realizadas pelo banco de dados Kyoto Encyclopedia of Genes and Genomes (KEGG) (KANEHISA; GOTO, 2000) por meio da plataforma DAVID Bioinformatics Resources 6.8.

3 RESULTADOS

3.1 Identificação das CNVs por meio do painel de 50K e do painel de alta densidade

Após a realização do controle de qualidade, foram identificadas 2.611 CNVs nos cromossomos autossômicos de 961 fêmeas da raça Girolando, com média de 3 eventos de CNV por animal. Eventos de deleções e duplicações em heterozigose, ou seja, deleções ou duplicações de apenas uma cópia da região do DNA, foram observados em maior frequência do que eventos em homozigose na população de fêmeas (Tabela 2; Figura 2).

Tabela 2. Número (N) dos eventos de variação no número de cópias (CNV), tamanho em quilobase (kb), mínimo (Min), médio (Médio) e máximo (Max) das CNVs, número mínimo, médio e máximo de SNPs por CNV.

Eventos	N	Min (kb)	Médio (kb)	Max (kb)	Min SNPs	Média SNPs	Max SNPs
Deleções	1.122	1,06	159,34	875,04	3	4,79	25
Deleções de duas cópias	18	36,43	140,67	299,56	3	3,55	6
Deleções com uma cópia	1.104	1,06	159,65	875,04	3	4,81	25
Duplicações	1.489	1,21	143,26	632,43	3	4,32	98
Duplicação de uma cópia	1.419	4,63	144,90	548,44	3	4,24	98
Duplicação de duas cópias	70	1,21	110,17	632,43	3	5,94	17
Total	2.611	1,06	156,18	875,04	3	4,52	98

Foram identificadas 668 CNVRs, sendo 278 deleções, 329 duplicações e 61 regiões mistas. Apenas 65 CNVRs foram identificadas em pelo menos 1% da população estudada. Entretanto foram encontradas 343 CNVRs com frequência única. As CNVRs variaram de 36,06 a 4.629,02 kb com média e mediana de 184,50 kb e 132,48 kb, respectivamente, correspondendo a 4% do genoma bovino (123,28/3.036,60 Mb). Foram encontradas quatro CNVRs em sobreposição com regiões de duplicação segmentares (SD), correspondendo a 0,42% do total da cobertura das CNVRs no genoma. A razão entre os eventos de duplicação e deleções

foi de 1,18, ou seja, foram observadas mais regiões com segmentos do genoma com cópias duplicadas do que eventos em deleções.

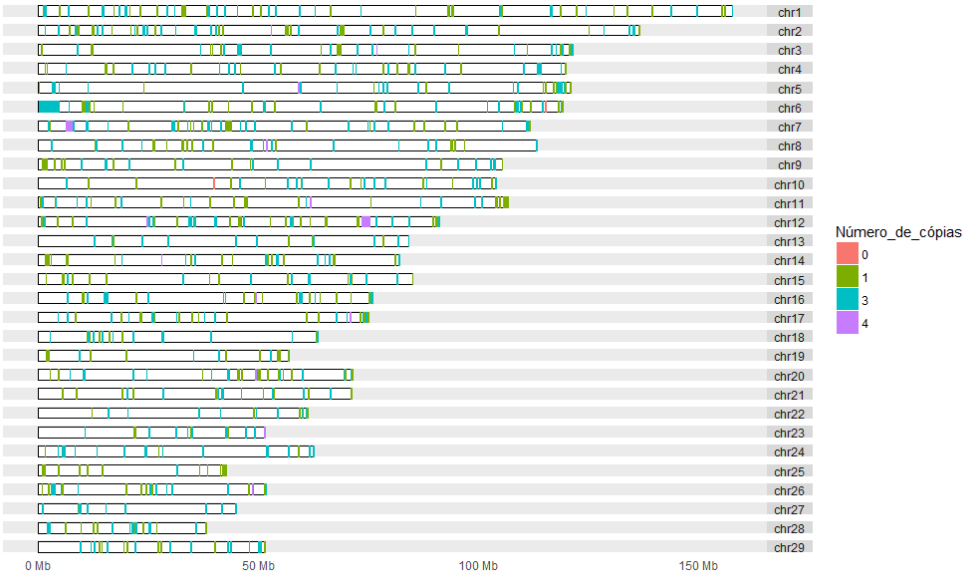


Figura 2. Distribuição do número de cópias sendo deleção dupla (0), deleção simples (1), duplicação simples (3) e duplicação de duas cópias (4), por cromossomo em vacas da raça Girolando genotipadas com painel 50K SNP.

Por meio da metodologia utilizando o painel HD, foram identificadas 2.828 CNVs ao longo dos cromossomos autossômicos de 262 touros, com média de 10 eventos por touro. Assim como na detecção pelo painel de 50K, deleções e duplicações em heterozigose também foram observadas em maior frequência do que os eventos em homozigose (Tabela 3; Figura 3). Foram identificadas 318 CNVRs, incluindo 146 deleções, 144 duplicações, e 28 regiões mistas (Figura 3). Os tamanhos das CNVRs variaram de 10,60 kb a 4.139.00 kb com média e mediana igual a 139.80 kb and 63.04 kb, respectivamente, correspondendo a 1% do genoma bovino (44.35/3,036.6 Mb). A razão entre duplicação/deleção foi de 0.98 assim o número de eventos de duplicação e deleção foi similar. Um total de 128 CNVRs foram observadas em pelo menos 1% dos touros e 123 CNVRs com frequência única.

Tabela 3. Número (N) dos eventos de variação no número de cópias (CNV), tamanho mínimo (Min) em quilobase (kb), médio (Médio) e máximo (Max) das CNVs, número mínimo, médio e máximo de SNPs por CNV.

Eventos	N	Min (kb)	Médio (kb)	Max (kb)	Min SNPs	Media SNPs	Max SNPs
Deleções	1.560	1,26	173,28	724,91	10	24,42	349
Deleções de duas cópias	698	51,90	188,20	509,30	10	21,50	107
Deleções com uma cópia	862	1,26	161,16	724,91	10	26,79	349
Duplicações	1.268	1,09	205,61	966,79	10	33,13	202
Duplicação de uma cópia	1.260	1,09	206,25	966,79	10	33,22	202
Duplicação de duas cópias	8	83,22	104,44	167,63	15	19,38	26
Total	2.828	1,09	187,77	966,00	10	28,32	349

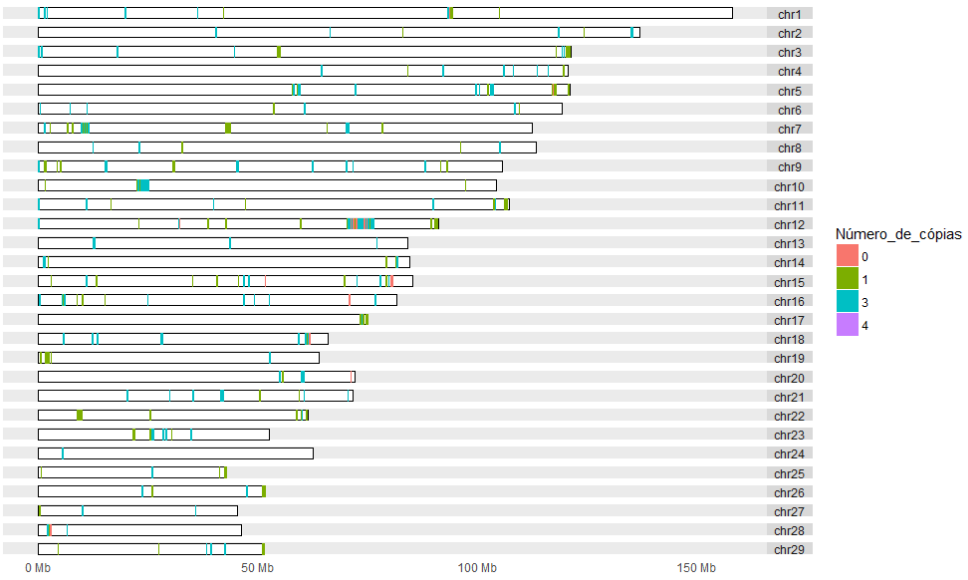


Figura 3. Distribuição do número de cópias sendo deleção dupla (0), deleção simples (1), duplicação simples (3) e duplicação de duas cópias (4), por cromossomo em touros da raça Girolando genotipadas com painel com painel de alta densidade.

3.2 Detecção das CNVs com base na família

Foram detectadas 464 CNVs nos 43 indivíduos considerados nos trios, sendo 124 CNVs consideradas de herança mendeliana na progênie, em que 68 foram provenientes da mãe e 56 oriundas do pai. O tamanho das CNVs de herança

mendeliana variou de 36,42 kb a 839,10 kb com tamanho médio de 166,70 kb e mediana de 125,60 kb. Eventos de deleções em heterozigose foram identificados em maior frequência, correspondendo a 66,93% do total das CNVs herdáveis, 32,25 % foram eventos de duplicação em heterozigose. Foi identificado somente uma duplicação em homozigose proveniente da mãe no cromossomo 11, sendo que os pontos de quebra (“breakpoints”) dessa CNV são os mesmos para mãe e progênie. As CNVs denominados *de novo*, ou seja, presentes somente na progênie, foram desconsiderados.

As 124 CNVs foram submetidas às análises de prospecção de genes e QTL, sendo que 203 possíveis genes candidatos foram anotados nessas regiões e 583 QTL relacionados a produção de leite, 301 a reprodução e 171 a características de sanidade. Os genes *FOXH1 (forkhead box H1)* e *CYHR1 (cysteine/histidine-rich 1)* estão anotados no genoma referência próximos de uma deleção mendeliana em heterozigose no cromossomo BTA14.

3.3 Detecção por meio dos dados de NGS

Embora a detecção utilizando dados de sequenciamento tenha sido realizada considerando apenas dois indivíduos da população foi identificado maior número de deleções do que duplicações conforme era esperado devido a tecnologia (Tabela 4). A razão entre eventos de duplicação e deleção foi de 0,19. A maioria das CNVs detectadas pelo programa PennCNV por meio do painel HD nos mesmos touros ressequenciados foi confirmada pelo NGS (Tabela 4). Somente uma CNV no touro A foi denominada de modo discordante entre os métodos (Tabela 5).

Também foi verificado que houve sobreposição de 16% das CNVs detectadas pelo painel de 50K usando o programa PennCNV nas filhas genotipadas do touro A presentes no banco de dados analisado, sendo essas CNVs provavelmente são de herança mendeliana e consequentemente poderá ser transmitida para as próximas gerações (Tabela 4). As CNVs detectadas nos dois touros por meio de dados de NGS, foram agrupadas em 2.395 CNVRs, sendo 2.066 deleções, 32 duplicações e 297 regiões mistas, correspondendo a 2% do genoma total bovino. Os tamanhos das

regiões variaram de 1,99 kb a 1.100,00 Kb, com média de 33,55 kb e mediana de 19,50 kb.

Tabela 4. Descrição das sequências e da variação no número de cópias (CNV) identificadas em dois touros Girolando.

Touro	Total de reads	Reads mapeadas	Cobertura	Del[a]	Dup[b]	SNP CNVs[c]	SNP CNVs filhas[d]
A	297.636.324	99,07%	10.97	1309	237	6(85%)[e]	5(16%) [e]
B	345.972.522	99,05%	12.75	1208	243	8(80%) [e]	-

[a]Número dos eventos de Deleções; [b]Números dos eventos de Duplicações; [c]Número de CNVs detectadas no individuo pelo painel de HD confirmadas (em comum) pelos dados de sequenciamento de nova-geração; [d]Número de CNVs detectadas em 14 filhas do indivíduo A pelo painel de 50K confirmadas (em comum) pelos dados de sequenciamento de nova-geração; [e]Porcentagem em relação ao total de CNVs detectadas entre parênteses.

Tabela 5. Descrição das variações no número de cópias detectadas pelo painel de genotipagem em alta densidade (HD) confirmadas por meio dos dados de sequenciamento de nova-geração (NGS).

Touro	Chr[a]	Inicio	Final	Evento HD/NGS	Genes
A					
	BTA5	59066551	59215089	Del[b] het[c]. /Del	*OR6C1*
	BTA5	59421039	59581140	Del het./Del	*LOC788242, LOC788258, ENSBTAG00000031100, LOC788285, ENSBTAG00000046446, ENSBTAG00000026078*
	BTA7	70191001	70267774	Del het/Dup[d]	*MGC137099*
	BTA12	71899472	72042304	Del hom[e]./Del	*ENSBTAG00000047383*
	BTA12	72748544	72882991	Del hom. / Del	-
	BTA12	74840021	74942860	Del hom./ Del	*LOC509854*
B					
	BTA1	93730576	93819471	Del. hom./Del	-
	BTA3	54522367	54963401	Del het. / Del	*U2, GBP4, GBP6, ENSBTAG00000037490, ENSBTAG00000037634, LOC510382, LOC100336669*, *LOC507055, ENSBTAG00000046356*
	BTA7	10366946	11091861	Del het. / Del	*ENSBTAG00000048066, LOC100298119, LOC785624, OR7A10, ENSBTAG00000048193 ENSBTAG00000046958, OR7A17 LOC788246, ENSBTAG00000034917*
	BTA12	72063255	72123747	Del het./ Del	*LOC100337053*
	BTA12	72174261	72264806	Del het. / Del	*LOC100337053*
	BTA12	72511101	72882991	Del hom./ Del	*ENSBTAG00000047181*
	BTA13	43643509	43776717	Dup het. /Dup	*AKR1C4, LOC100337056*
	BTA20	60473408	60525562	Dup het./Dup	-

[a] chr = Cromossomo; [b]Del = deleção; [c] het. = Heterozigose;[d] Dup = duplicação.; [e]hom. = homozigose

3.4 Regiões candidatas de CNV

3.4.1 Detecção utilizando o painel de 50K

Foram confirmadas 873 CNVs em 418 animais detectadas pelo painel de 50K SNPs por meio da técnica de NGS. As CNVs foram agrupadas em 203 regiões (Figura 4; Apêndice A), sendo 100 eventos de deleções e 15 duplicações concordantes em ambos os métodos. Entretanto, 63 eventos detectados como duplicações no painel de 50K foram identificados como deleção no painel NGS e sete eventos considerados duplicação pelo NGS foram detectados como deleção no painel de 50K. Sugere-se que as divergências entre os estados dos eventos possam ter sido ocasionadas devido aos animais apresentarem diferentes variações estruturais ou ainda por erros do método de detecção.

As CNVRs candidatas designadas por meio da detecção pelo painel de 50K apresentaram tamanho de 2,49 kb a 1.100,00 kb com média 53,27 kb e mediana igual a 22,50 kb. As CNVRs candidatas mais frequentes na população (> 3%) estão localizadas nos cromossomos BTA4, BTA6, BTA11, sendo que alguns indivíduos apresentaram deleções e outros duplicações para estas regiões (mistas). A CNVR_28, localizada no cromossomo BTA4, está presente em 5% das vacas Girolando. O gene *ZNF800* (“*Zinc Finger Protein 800*”) foi mapeado em região exônica na CNVR_28. Também foi mapeada uma região de QTL associada a mastite (ID = 2491) e duas regiões relatadas a porcentagem de gordura no leite (ID = 5054 e 5055).

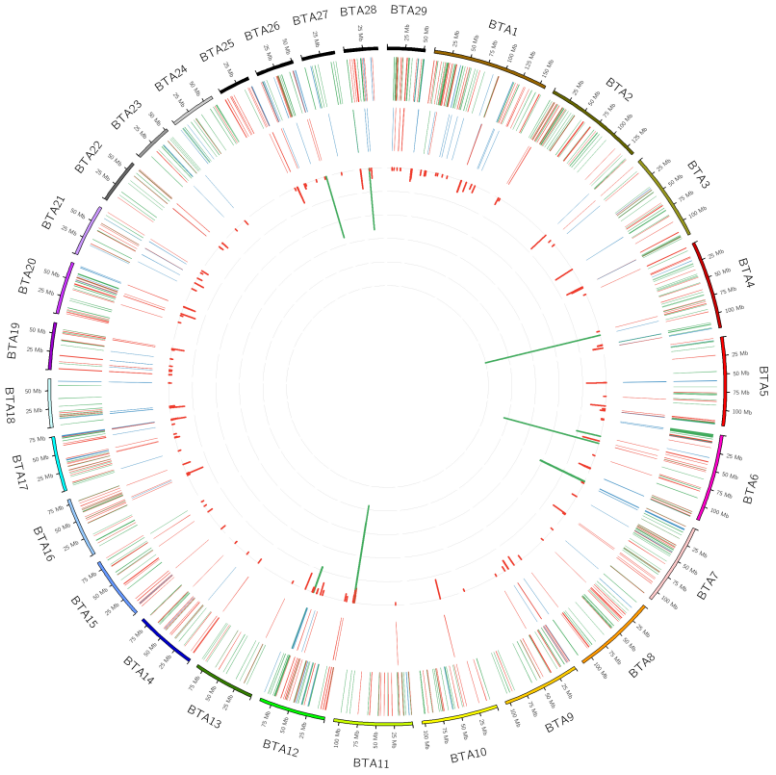


Figura 4. Representação das regiões de variações no número de cópias (CNVR) detectadas pelo painel de 50K (mais externo ao círculo) e das CNVRs candidatas confirmadas pelo sequenciamento de nova-geração (mais interno ao círculo) ao longo dos 29 cromossomos autossomos de bovinos da raça Girolando. As deleções foram representadas em vermelho, duplicações em verde e ambas (deleções e duplicações) em azul. Histograma ao centro, representa a frequências das CNVRs na população, sendo em vermelho regiões presentes em menos de 1% na população e em verde presentes em ao menos 1% da população.

Em sobreposição às CNVRs candidatas identificadas pelo painel de 50K foram mapeados157 genes anotados do genoma referência, sendo 65 CNVRs localizadas em regiões exônicas. Os genes *CD1A* e *EMSY* mapeados nas filhas do touro A também foram observados em regiões de exon nas CNVRs candidatas. O gene *IGF2 ("insulin-like growth factor 2")*, localizado no cromossomo BTA29, pode estar relacionado à uma região com duplicação no genoma de vacas da raça Girolando.

Foram encontradas 184 regiões (~90,6%) em sobreposição à 2.484 regiões de QTL ("quantitative trait loci") e às associações genéticas relacionadas com características de interesse econômico (Tabela 6). A maioria das regiões de QTL

mapeadas foram relacionadas às características de peso corporal (299 QTL), produção de leite (128 QTL), produção e porcentagem de proteína (123 e 117 QTL, respectivamente) e produção de gordura (111 QTL). Para contagem de células somáticas e mastite clinica foram observadas respectivamente 56 e 11 QTL em sobreposição às CNVRs candidatas, respectivamente.

Tabela 6. Número de regiões de QTL e associações de SNP (Associação) por classe de característica com sobreposição às regiões de variação no número de cópias (CNVRs).

Classe da Característica	Número QTL	Associação
Leite	620	43
Sanidade	178	3
Reprodução	449	22
Produção	530	100
Carne e carcaça	474	34
Conformação (Exterior)	31	0

Os genes mapeados próximos as CNVRs candidatas (Apêndice A) foram enriquecidos ($p < 0,05$) pelos termos do Gene Ontology, em 23 processos biológicos, sendo a maioria relacionados a função celular, como por exemplo regulação da divisão celular (Figura 4), para a categoria funções moleculares os termos enriquecido ($p < 0,05$) foram "monooxygenase activity" (GO:0004497), "oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen" (GO:0016705), transferase activity, transferring phosphorus-containing groups (GO:0016772), growth factor activity (GO:0008083), transcription coactivator activity (GO:0003713). Dentro da categoria componentes celulares somente o termo "SWI/SNF superfamily-type complex" (GO:0070603) foi enriquecido.

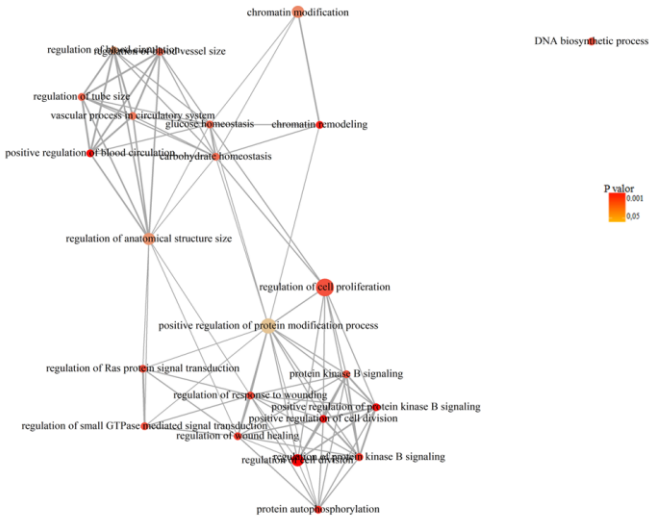


Figura 5. Mapa dos termos do Gene Ontology enriquecidos significativamente (p < 0,05) dentro da categoria Processos Biológicos considerando os genes mapeados nas regiões de variação no número de cópias candidatas detectadas pelo painel de genotipagem de 50K. Quanto maior o tamanho do círculo maior o número de genes envolvidos no processo biológico. A coloração do círculo representa o valor de p.

Importantes termos do Medical Subject Headings foram enriquecidos significativamente (p < 0,05), incluindo, prenhez (*NOS2, IGF2, OAS1Z*) e período de gestação (*NOS2, CALCRL, OAS1Z*) na categoria fenômenos e processos (Tabela 7). Na categoria químicos e drogas o termo hormônio estimulante folicular (*NOS2, IGF2*) e na categoria doenças, o termo doenças bovinas (*NOS2, IGLL1, MIR2348, MIR2389, MIR2443*). O gene *NOS2* (*"Nitric Oxide Synthase 2"*) está em sobreposição com a CNVR_148 e com a CNVR_149 identificadas no cromossomo BTA19. O *IGF2* está em sobreposição ao CNVR_202. O gene *IGLL1* ("*Immunoglobulin Lambda Like Polypeptide 1"*) foi mapeado na região de deleção CNVR_128 nesta população (Apêndice A). Todos os genes foram submetidos ao DAVID e foram significativamente enriquecidos (p < 0,05) em duas vias metabólicas por meio do banco de dados Enciclopédia de Genes e Genoma de Kioto (KEGG) (Tabela 8). A via de transdução do olfato apresentou 14 genes anotados em sobreposição das CNVRs candidatas, incluindo os genes *OR6C1* (*"Olfactory Receptor Family 6 Subfamily C Member 1"*) e *OR7A1.*

Tabela 7. Categorias dos termos MeSH (Medical Subject Headings) enriquecidos significativamente (p < 0,05) relacionados aos genes mapeados nas regiões de variações no número de cópias candidatas em bovinos Girolando.

Categoria	MeSH ID	Termo MeSH	Gene ID
Fenômenos e Processos	D008460	Carne	*NOS2, SREBF1, TFR2, IGF2*
	D002886	Cromossomos, Humano, Par 17	*NOS2, TK1*
	D005865	Período de gestação	*NOS2, CALCRL, OAS1Z*
	D011270	Prenhez, Animal	*NOS2, IGF2, OAS1Z*
	D009866	Oogenese	*NOS2, IGF2*
	D001947	Acasalamento	*SREBF1, IGLL1, IGF2*
Quimicos e Drogas	D027682	Transporte de Cálcio Proteínas	*TFR2, NOS2*
	D005640	Estimulante Folicular Hormônio do fator de crescimento	*NOS2, IGF2*
	D007334	Insulina I	*NOS2, IGF2*
Anatomia	D008198	Linfonodos	*NOS2, IGLL1*
	D054885	Células do Cumulus	*IGF2, MIRLET7B*
	D006822	Células híbridas	*NOS2, TK1, EMSY*
Doenças	D002418	Doenças Bovinas	*NOS2, IGLL1, MIR2348, MIR2389, MIR2443*

Tabela 8. Vias metabólicas (KEGG Pathway) enriquecidas (p < 0.05) por meio do Kyoto Encyclopedia of Genes and Genomes (KEGG).

KEGG Pathway	N[a]	%[b]	Valor de p [c]	Benjamini[d]
Transdução do olfato	14	9,3	0,02	0,89
Vício em cocaína	3	2,0	0,04	0,87

[a] Número de genes observados na via metabólica; [b] Porcentagem em relação ao total de genes da população referência envolvidos em cada via metabólica; [c] p valor < 0,05; [d] Benjanimi < 0,05.

3.4.2 Detecção pelo painel de alta densidade (HD)

Por meio da identificação de CNVs por dados de NGS foram confirmadas 4,402 CNVs detectadas pelo painel de genotipagem HD em 262 animais, agrupadas em 213 regiões de CNVs (Figura 6; Apêndice B). Foram confirmadas 39 deleções nos segmentos cromossômicos. As CNVRs candidatas oriundas do painel HD apresentaram em média tamanho de 63,90 kb e mediana igual a 30,50 kb e tamanho mínimo e máximo de 3,99 e 572,50 kb, respectivamente. A CNVR_5 é uma deleção localizada no cromossomo BTA1 presente em 69% da população, sendo uma deleção

com alta-confiabilidade na população (Figura 6). Nessa região foram mapeados três QTL relacionados a produção de leite e proteína (ID = 2508,4703 e 4704)), um QTL a circunferência escrotal (ID = 10644), um a taxa de concepção (ID = 3439), um relacionado a taxa de retorno da vaca (ID = 5658) três QTL associados a sanidade, incluindo susceptibilidade à ceratoconjuntivite bovina (ID = 9947), deslocamento do abomaso (ID = 5112) e à encefalopatia espongiforme bovina (ID = 1736).

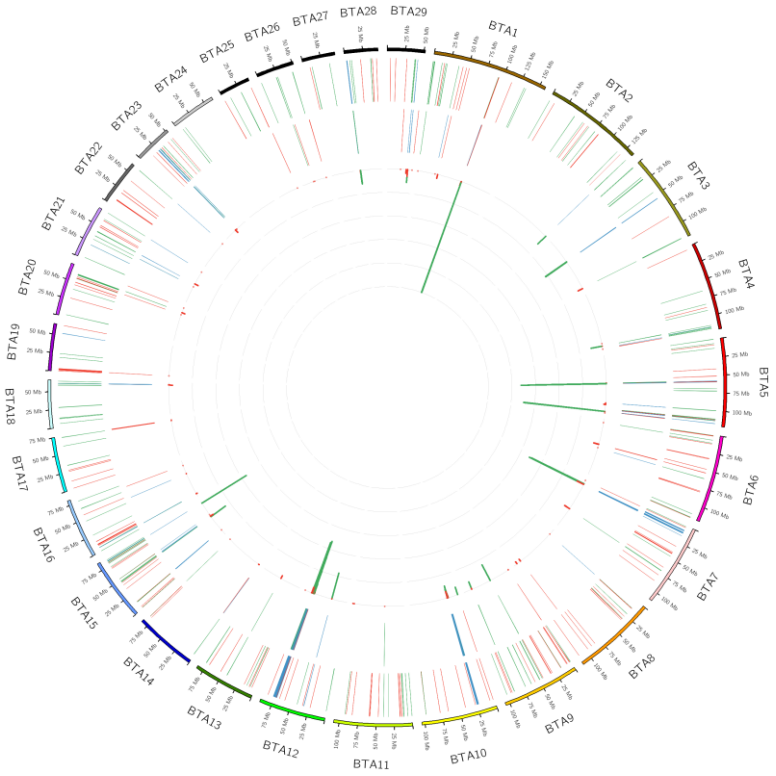


Figura 6. Representação das regiões de variações no número de cópias (CNVR) detectadas pelo painel de alta densidade (HD) e das CNVRs candidatas confirmadas pelo sequenciamento de nova geração ao longo dos 29 cromossomos autossomos de bovinos da raça Girolando. As deleções foram representadas em vermelho, duplicações em verde e mistas em azul. Histograma representa a frequências das CNVRs na população, sendo em vermelho regiões presentes em menos de 5% na população e em verde presentes em mais do que 5% da população.

No cromossomo BTA5 foram observadas três regiões (CNVR_19, CNVR_20, e CNVR_35) muito representativas na população (~45%), sendo que a CNVR_19 também foi encontrada pela detecção em painel de 50K (CNVR_39), com base na família (BTA5:59304363:59851060) herdada pela mãe e também pela detecção em

NGS. Nessa região foi anotado o gene *OR6C1* e mapeados QTL associados a características reprodutivas como facilidade de parto (ID= 10769) e período de gestação (ID=387 e ID=5388), composição de leite, como porcentagem de gordura (ID = 1545), proteína (ID = 10440, ID= 21538, ID = 10386, ID = 2511 e ID = 2428) e kappa-caseina (ID = 21537) e alpha-caseína (ID = 21535 e ID = 21536). Na CNVR_35 foi mapeado o gene *GRAMD4 ("GRAM Domain Containing 4").* Sugere-se que esse gene está associado a uma deleção em bovinos Girolando. Esse gene também está anotado na CNVR_40 detectada pelo painel de 50K. O *GRAMD4* possui importante função na ativação do gene *E2F1 ("E2F Transcription Factor 1")* envolvido na regulação da resposta inflamatória aos receptores Toll-Like. Nas CNVR_19 e CNVR_20 foram mapeados 10 QTL associados a produção de leite, gordura e proteína. Também foi encontrado um QTL associado ao deslocamento do abomaso (ID = 5117) e à contagem de células somáticas (ID = 2659).

Um total de 105 CNVRs candidatas (~49%) identificadas pelo painel HD se sobrepuseram à 123 genes anotados no genoma referência, sendo 34 CNVRs localizadas em regiões exônicas (Figura 6; Apêndice B). Na CNVR_8, localizada no cromossomo BTA3 (Apêndice B), foram mapeados os genes *GBP4* e *GBP6*, estes também foram confirmados pela detecção por meio de dados de NGS no touro B (Tabela 5). O *GIMAP7 ("GTPase, IMAP family member 7*"), na CNVR_12 (BTA4), possui importante função na sobrevivência dos linfócitos. Dentro da CNVR_55 no BTA8 foram mapeados os genes *COL27A1 ("collagen type XXVII alpha 1 chain*") e *ORM1* ("*orosomucoid 1*") envolvidos em processos de crescimentos de cartilagem e do sistema imune, respectivamente. Foram mapeados genes da família do complexo principal de histocompatibilidade (MHC) da classe II, como BolA-*DQA5* ("*major histocompatibility complex, class II, DQ alpha 5"), BolA-DQB (major histocompatibility complex, class II, DQ beta), LOC100848815 ("DQ haplotype D alpha chain-like")*, BolA-*DQA2 ("major histocompatibility complex, class II, DQ alpha"), BLA-DQB("*MHC *class II antigen*") e *BOLA ("MHC class I heavy chain*").

Foram encontradas 198 regiões (~92,9%) em sobreposição à 3.160 regiões de QTL ("quantitative trait loci") e às associações genéticas relacionadas com características de interesse econômico (Tabela 9). Assim como nas CNVRs candidatas determinadas pelo painel de 50K, a maioria das regiões de QTL mapeadas

foram relacionadas às características de peso corporal (378 QTL), produção de gordura no leite (194 QTL), produção de leite (191 QTL) e proteína (178 QTL). Para características de sanidade e reprodutivas foram mapeados 70 QTL para deslocamento do abomaso, 72 para distocia e 82 para facilidade de parto. Também foi encontrado nas CNVRs número representativo de QTL ligados a estatura corporal, sendo 119 para altura e 114 para peso de carcaça.

Tabela 9. Número de regiões de QTL e associações de SNP por classe de característica com sobreposição às regiões de variação no número de cópias (CNVRs).

Classe da Característica	Número QTL	Associação
Leite	843	21
Sanidade	175	6
Reprodução	590	15
Produção	833	48
Carne e carcaça	513	74
Conformação (Exterior)	42	0

Os genes mapeados próximos as CNVRs candidatas (Apêndice B) foram enriquecidos ($p < 0,05$) pelos termos do “Gene Ontology”, em 20 processos biológicos, assim como nas CNVRs_candidatas pelo painel de 50K (Figura 5), a maioria dos processos estão relacionados a função celular (Figura 7), para a categoria funções moleculares os termos “DNA polymerase activity” (GO:0034061), “endoribonuclease activity, producing 5'-phosphomonoesters” (GO:0016891), “collagen binding” (GO:0005518), “double-stranded RNA binding” (GO:0003725)”, “endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters” (GO:0016893) e “GDP binding” (GO:0019003) foram enriquecidos ($P < 0,05$). Dentro da categoria componentes celulares moleculares foram enriquecidos ($P < 0,05$) oito termos incluindo “nonmotile primary cilium” (GO: 0031513), "primary cilium” (GO:0072372), “midbody” (GO:0030496), “photoreceptor connecting cilium” (GO:0032391), “cytoplasmic microtubule” (GO:0005881) e “PML body” (GO:0016605).

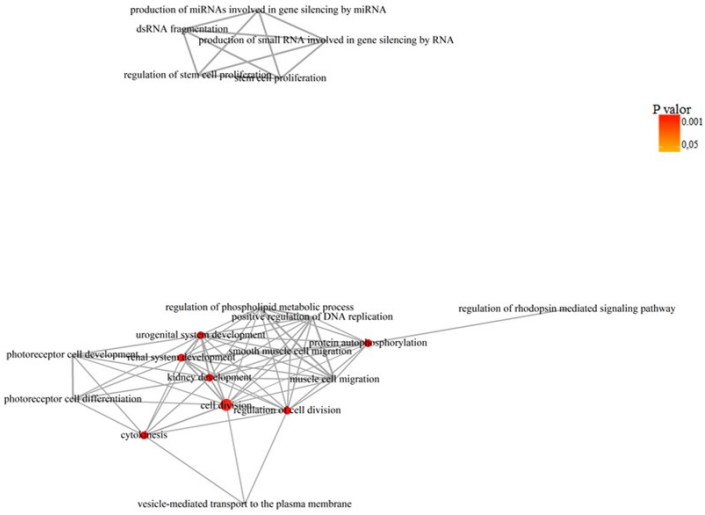


Figura 7. Mapa dos termos do “Gene Ontology” enriquecidos significativamente (P < 0,05) dentro da categoria Processos Biológicos considerando os genes mapeados nas regiões de variação no número de cópias candidatas detectadas pelo painel de genotipagem de alta densidade (HD). Quanto maior o tamanho do círculo maior o número de genes envolvidos no processo biológico. A coloração do círculo representa o valor de P.

Genes envolvidos em diversos processos biológicos do sistema imune foram enriquecidos significativamente (p < 0,05) pelos termos Mesh (Tabela 9). Na categoria doenças o termo mastite bovina foi enriquecido (p < 0,05), sendo os genes *ORM1 (BTA8) e BOLA-DQA2 (BTA23)*, enriquecidos nesse processo biológico, mapeados em sobreposição a CNVR_55 e CNVR_189 (Apêndice B). Essas regiões estão presentes em aproximadamente 2% da população de touros estudada.

Tabela 9. Categorias dos termos MeSH (Medical Subject Headings) enriquecidos significativamente (p < 0,05) relacionados aos genes mapeados nas regiões de variações no número de cópias candidatas em bovinos Girolando.

Categoria	MeSH ID	Termo MeSH	Gene ID
Fenômenos e Processos	D017951	Apresentação do Antígenos	*LOC100848815, BLA-DQB*
	D006239	Haplótipos	*LOC100848815, BLA-DQB, BOLA-DQA5, BOLA, BOLA-DQB*
	D005802	Genes, Classe II MHC	*BOLA-DQB, BLA-DQB*
	D000483	Alelos	*BLA-DQB, BOLA, BOLA-DQB, BOLA-DQA5*
	D000918	Especificidade do Anticorpo	*BLA-DQB, JSP.1,*
	D010802	Filogenia	*RAET1G, BOLA-DQA5, BOLA-DQB, GRK1, BOLA-DQA2, BOLA*
Quimicos e Drogas	D000949	Antígenos de Classe II	*BOLA-DQA2, BLA-DQB, BOLA-DQB, BOLA-DQA5, LOC100848815*
	D015395	Antígenos de Classe I	*BOLA, JSP.1, RAET1G, ULBP21*
	D055607	Receptores, células exterminadoras I	*LOC100139049, RAET1G*
	D015232	Prostaglandina	*ORM1, LOC782061*
	D011971	Receptores, Imunológicos	*LOC100139049, RAET1G*
	D020413	Regiões 3' não-traduzidas	*BOLA-DQA2, FUT7*
Anatomia	D015496	CD4-Positiva Linfócitos T	*BLA-DQB, LOC100848815*
Doenças	D008414	Mastite bovina	*ORM1, BOLA-DQA2*

Todos os genes foram submetidos ao DAVID e foram significativamente enriquecidos (p < 0,05) em 19 vias metabólicas por meio do banco de dados Enciclopédia de Genes e Genoma de Kioto (KEGG) (Tabela 10). Assim como no painel de 50K, a via metabólica de transdução do olfato foi a via mais enriquecida significativamente (p = 2,90E-18) em que 40 genes identificados nas CNVRs candidatas apresentaram participação nesse processo, representando 32,5% do total de genes localizados nas regiões.

Tabela 10. Vias metabólicas (KEGG Pathway) enriquecidas (p < 0.05) por meio do Kyoto Encyclopedia of Genes and Genomes (KEGG).

KEGG Pathway	N[a]	%[b]	Valor de p[c]	Benjamini[d]
Transdução do olfato	40	32,5	2,90E-18	1,80E-16
Doença enxerto contra hospedeiro	7	5,7	1,20E-06	3,80E-05
Processamento e apresentação do Antígeno	8	6,5	2,80E-06	5,90E-05
Rejeição de aloenxerto	7	5,7	3,00E-06	4,80E-05
Diabetes mellitus tipo I	7	5,7	4,30E-06	5,50E-05
Doença autoimmune da tireóide	7	5,7	1,10E-05	1,20E-04
Miocardite viral	7	5,7	2,30E-05	2,10E-04
Asma	5	4,1	2,50E-04	2,00E-03
Via da produção de IgA intestinal	5	4,1	1,10E-03	8,10E-03
Infecção por *Staphylococcus aureus*	5	4,1	1,50E-03	9,50E-03
Infecção HTLV-I	9	7,3	1,80E-03	1,00E-02
Molécula de adesão celular (CAMs)	7	5,7	1,80E-03	9,60E-03
Phagosome	7	5,7	2,30E-03	1,10E-02
Doença inflamatória Intestinal (IBD)	5	4,1	2,80E-03	1,30E-02
Leishmaniose	5	4,1	2,90E-03	1,20E-02
Lúpus eritematoso sistêmico	7	5,7	3,70E-03	1,50E-02
Infecção por herpes	7	5,7	5,10E-03	1,90E-02
Artrite reumatoide	5	4,1	8,30E-03	2,90E-02
Toxoplasmose	5	4,1	1,90E-02	6,20E-02

[a] Número de genes observados na via metabólica;[b] Porcentagem em relação ao total de genes da população referência envolvidos em cada via metabólica;[c] Valor de p < 0,05; [d] Benjanimi < 0,05

4 DISCUSSÃO

Neste estudo foi possível identificar CNVs em bovinos da raça Girolando utilizando painéis de média (Tabela 2; Figura 2) e alta densidade (Tabela 3; Figura 3). No entanto, os resultados devem ser analisados com cautela, uma vez que a hibridização das sondas dos painéis de genotipagem para determinar a intensidade do sinal de cada marcador são realizadas em contraste com uma população referência (WINCHESTER; YAU; RAGOUSSIS, 2009), sendo esta formada na maioria por bovinos taurinos, o que pode causar viés na identificação de CNVs em populações de bovinos provenientes de raças zebuínas (HOU et al., 2011). Consequentemente, estudos têm relatado maior número de eventos de CNVs em raças zebuínas do que em raças taurinas (LIU et al., 2010; HOU et al., 2012b; ZHOU et al., 2016b).

O número de CNVs detectado pelo painel de 50K SNP neste estudo foi inferior (2.611) ao reportado por Hou et al. (2011) (3.666 CNVs) baseado no mesmo painel

de genotipagem em uma população de 521 bovinos composta por raças taurinas, zebuínas, africanas e animais cruzados e superior ao relatado por Jiang et al. (2012) (219) em bovinos da raça Chinese Holstein. Com base na detecção pelo painel de alta densidade foi identificado neste trabalho baixo número de CNVs (2.828) comparado ao descrito por Zhou et al. (2016b) em bovinos Nelore que identificaram 110.521 eventos de CNV e Yang et al. (2017) que relataram 13.325 CNVs em bovinos de diversas localidades demográficas da China. Em relação ao método pela técnica de NGS, Bickhart et al. (2012) encontraram número de CNVs similar (1.265) em bovinos Angus, Holandês, Hereford e Nelore ao observado nos dois touros Girolando ressequenciados (~1.500) nesse trabalho.

Utilizando o painel de 50K, foi encontrado maior número de eventos de duplicações do que deleções (Tabela 2) divergindo dos resultados encontrados na literatura em bovinos, em que as deleções são mais numerosas do que duplicações, devido maior facilidade de detecção pelo algoritmo do PennCNV (HOU et al., 2011; ZHOU et al., 2016b; YANG et al., 2017). Por meio do painel HD, o número de deleções identificadas em relação ao número de duplicações foi similar (Tabela 3). Essa diferença nos resultados pode ser devida a menor resolução do painel de 50K, uma vez que as distâncias pré-definidas entre os marcadores são maiores. A detecção de CNVs por meio dos dados de NGS (Tabela 4) sugeriu que a tecnologia é mais sensível para identificação de eventos de deleções do que eventos de duplicações (razão 0,19) e essa hipótese tem sido confirmada por estudos realizados anteriormente em diferentes raças zebuínas e taurinas (SILVA et al., 2016b; LETAIEF et al., 2017).

Um total de 37 e 30 CNVRs candidatas selecionadas pelos painéis de genotipagem em 50K e HD, respectivamente, estão condizentes com as 34 CNVs reportadas por Xu et al. (2014) associadas a produção de leite e a produção e porcentagem de gordura e proteína em bovinos da raça Holandês, por exemplo CNVR_7 (BTA2:35126501-135147500), CNVR_78 (BTA11:105732001-105763500), CNVR_134 (BTA17:74627001-74642500) e CNVR_202 (BTA29:50056501-50078000) para as regiões do painel de 50K (Apêndice A) e para o painel HD as regiões CNVR_7 (BTA2:135126501-135147500), CNVR_89 (BTA11:106879501-106948500), CNVR_172 (BTA17:74998001-75058715) e CNVR_213 (BTA29:51489501-51502868) (Apêndice B). Deste modo, presume-se que estas

regiões também estejam relacionadas às características produtivas de leite na população de vacas e touros Girolando.

O gene candidato *IGF2*, anotado na CNVR_202 (50K) no BTA 29 (Apêndice A) foi reportado em outros estudos de CNVs e está associado a produção de gordura e proteína no leite (GAO et al., 2017) em bovinos taurinos e também presente em regiões de CNV em bovinos da raça Nelore (SILVA et al. 2016b). As regiões de QTL relacionadas às características de produção de leite, gordura e proteína foram principalmente identificadas em sobreposição as CNVRs candidatas nos cromossomos BTA 5 (39 QTL para o painel de 50K e 193 para o HD) e BTA 14 (74 QTL para 50K e 65 QTL para o HD). Esses resultados foram condizentes com as regiões reportadas em estudos de CNVs por meio de painéis de SNP (XU et al., 2014) e dados de NGS (BOUSSAHA et al., 2015; GAO et al., 2017).

A região CNVR_8 (painel HD) no BTA 3, foi detectada em aproximadamente 15% da população de touros Girolando. Alteração no número de cópias dos genes presentes nessa região (*GBP4* e *GBP6)* foi reportada em bovinos Nelore associada a características de carcaça (ZHOU et al., 2016a) e relacionada ao consumo alimentar residual em vacas da raça Holandesa (HOU et al., 2012c). Esses genes são da família “Guanylate Binding Protein” pertencentes a grande família das guanosina trifosfatase (GTPases) e são capazes de hidrolizar moléculas de GTP (Guanosina Trifosfato) a ambos GDP (Guanosina difosfato) e GMP (Guanosina monofosfato) e induzidos pelo Interferon (IFN) (OLSZEWSKI et al., 2006). Existe pouca informação exata sobre a função dos genes *GBP*, entretanto alguns genes desta família, como *GBP4* pode estar relacionado a proteção de patógenos intracelular (O'GORMAN, et al. 2009; VESTAL; JEYARATNAM, 2011)

Os genes *GIMAP* presentes em número de cópias duplicadas no BTA4 nas regiões candidatas CNVR_31 e CNVR_32 pelo painel 50K (Apêndice A) e CNVR_11 e CNVR_12 pelo painel HD (Apêndice B) também são pertencentes a família GTPase e preferencialmente expressos em células do sistema imune com importante função na regulação dos linfócitos T (CIUCCI; BOSSELUT, 2014). Essas regiões estão em sobreposição a regiões de QTL associadas com contagem de células somáticas e parasitas. Deste modo, sugere-se que cópias duplicadas desses genes podem estar associadas à resistência às doenças, como por exemplo a mastite. Esses genes já

foram previamente descritos em regiões de CNV em bovinos zebuínos e taurinos (LIU et al., 2010; BICKHART et al., 2012; HOU et al. 2012b). Ainda no BTA4 está a região candidata mista (CNVR_28) detectada pelo painel de 50K com maior frequência (~5%) na nossa população de vacas Girolando (Figura 4), sendo que o gene anotado nessa região (*ZNF800*) foi relacionado a contagem de células somáticas em estudo com associação de SNPs em uma população de bovinos taurinos (CHEN et al., 2015).

Genes pertencentes a família *WC1*, foram identificados em nosso estudo em quatro regiões presentes em mais de 1% da população (CNVR_24; CNVR_28, CNVR_29 e CNVR_30), como duplicações candidatas pelo NGS e como deleções ou regiões mistas pelo painel de genotipagem HD (Apêndice B). *WC1* é uma glicoproteína encontrada em grande número em mucosas, pertencente à família dos receptores "scavenger" ricos em cisteína (SRCR) e são exclusivas dos linfócitos gama/delta T, sendo esses linfócitos expressos somente em bovinos, ovinos e suínos e não presentes em humanos e ratos (HERZIG; BALDWIN, 2009). Essas variações no número de cópias e a presença em regiões de SD podem ter sido desenvolvidas como forma de sistema imune adaptado de ruminantes (LIU et al., 2009). Deste modo, supõe-se que indivíduos que apresentam regiões cromossômicas duplicadas para genes da família *WC1* provavelmente possuem um sistema imune mais diversificado e eficiente tornando-os mais resistentes às infecções.

No cromossomo BTA17, pela detecção por meio do painel de 50K, foram mapeados os genes *OAS1Z e OAS1X* na CNVR_127 (BTA17: 6365400-63673000) e o gene *IGLL1* na CNVR_128 (BTA17: 73089001:73099000) em menor número de cópias, ou seja, região de deleção (Apêndice A). O gene "*Immunoglobulin Lambda Like Polypeptide 1" (IGLL1)* foi reportado em outros estudos de CNV em bovinos taurinos e zebuínos (HOU et al., 2012b; XU et al., 2016; YANG et al., 2017). Este gene está envolvido no desenvolvimento inicial dos linfócitos B (componente principal do sistema imune humoral- anticorpos), sendo que mutações nesse gene podem resultar em células B deficientes, ocasionando agamaglobulinemia (imunodeficiência com predominância de defeitos de anticorpos) (MINEGISHI et al., 1998). Em bovinos, o *IGLL1* tem sido associado com características de produção de leite e de fertilidade (CERRI et al., 2012; MINOZZI et al., 2013) e a variação no número de cópias desse gene foi relacionada com susceptibilidade a parasitas (HOU et al., 2012b)

As regiões identificadas neste estudo nos cromossomos BTA5 (Figura 6; Apêndice B) com genes em duplicata da família WC1 em touros Girolando e a região de deleção no cromossomo BTA17 (Figura 4; Apêndice A), em que os genes *OAS1Z, OAS1X e IGLL1* estão anotados, pode estar intensamente relacionada a resistência do animal Girolando à parasitas, uma vez que Hou et al. (2012b) em bovinos da raça Angus, reportaram que animais com mais cópias dos genes *OAS e IGLL1* demonstraram-se mais susceptíveis as infecções por nematódeos gastrointestinais e aqueles indivíduos com mais cópias do gene *WC1* no genoma mais resistentes.

Devido à raça Girolando apresentar proporção de genes de zebuínos, sugere-se que as regiões de deleção e duplicação detectadas no BTA5 (pelo painel de 50K (CNVR_39 e CNVR_40) (Apêndice A) e pelo painel HD (CNVR_14 até CNVR_20) (Apêndice B) estejam associadas a características reprodutivas em Girolando, principalmente a região englobando 58Mb a 59Mb, uma vez que foi confirmada por todos os métodos de detecção utilizados nesse estudo (50K, HD, baseado na família e NGS). Essa hipótese foi embasada em estudos anteriores, em que foi encontrada a mesma região de deleção específica para raças zebuínas ou compostas (MCDANELD et al., 2014; ZHOU et al., 2016b).

De acordo com McDaneld et al. (2014), deleções em homozigose detectadas em fêmeas zebuínas apresentaram associação com a redução da eficiência reprodutiva das vacas. Entretanto, deleções presentes em machos não demonstraram ter efeito sobre a eficiência reprodutiva do touro, porém o touro poderá transmitir a cópia da região em deleção para progênie, o que poderá prejudicar a eficiência reprodutiva das suas filhas. Estudos de associação de SNPs também têm relacionado a região 25Mb a 70Mb com características reprodutivas, tais como intervalo de retorno do animal a estação de acasalamento e circunferência escrotal em zebuínos e raças compostas (HAWKEN et al., 2012; BUZANSKAS et al., 2017). O gene candidato presente nessa região é o *OR6C1*, da família dos receptores olfativos, e pode estar relacionado ao comportamento sexual e reprodutivo dos mamíferos (PARMENTIER et al.,1992; CHAMERO; LEINDERS-ZUFALL; ZUFALL, 2012).

Para análise dos termos MeSH das CNVRs candidatas oriundas do painel de HD (Tabela 9), a maioria dos genes enriquecidos ($P < 0,05$) foram da família do complexo principal de histocompatibilidade (MHC) da classe II, como BolA-*DQA5*

(*major histocompatibility complex, class II, DQ alpha 5), DQB (major histocompatibility complex, class II, DQ beta), LOC100848815 (DQ haplotype D alpha chain-like)*, BolA-*DQA2 (major histocompatibility complex, class II, DQ alpha).* Era esperado encontrar genes dessa família nas CNVRs, pois há evidências de que muitos genes do sistema imunológico adaptativo e inato estão em número de cópias variáveis dentro das espécies de vertebrados (BALAKRISHNAN et al., 2010). Zhou et al. (2016b) encontraram regiões de CNV compartilhadas entre bovinos taurinos e zebuínos envolvendo os mesmos genes do sistema MHC relatados no nosso estudo.

O sistema MHC em bovinos, denominado como antígeno leucocitário bovino (BoLA – "bovine leukocyte antigen complex"), mapeado no braço curto do cromossomo BTA23, consiste na família de genes envolvidos no processamento e apresentação de peptídeos para as células do sistema imune (STEAR, et al., 2001; PASHMI et al., 2009). Os genes do MHC encontram-se divididos em três classes, denominadas classes I, II e III, sendo que cada classe desempenha uma função. Os genes da classe II do complexo BoLA codificam proteínas expressas em células apresentadoras de antígenos, como macrófagos, para linfócitos T CD4+ (BROWN et al., 2002; MORSE et al. 2012) e são fundamentais para o sistema imune adaptativo. Em bovinos, os genes da classe II são divididos em duas classes, de acordo com o tipo de proteína codificada (DQ ou DR).

Em bovinos de leite, os genes do complexo BolA de classe II são associados à resistência ou susceptibilidade à mastite (KELM et al., 1997; SHARIF et al., 1998). Park et al. (2004) relataram que vacas que expressam maior número de diferentes alelos DQ apresentam forte resposta imune de células Th1 e, consequentemente são mais resistentes. Thompson-Crisp et al. (2014) encontraram associação dos genes do complexo BolA classe II com reposta imune humoral (AMIR) e mediada por células T (CMIR). Esses autores observaram que vacas com alta resposta imune humoral demonstraram baixa incidência de mastite clínica em comparação com vacas de baixa resposta imune humoral. Provavelmente, a razão principal para a variação no número de cópias dos genes da região MHC entre os indivíduos é a manutenção da variabilidade genética da população, protegendo-a contra patógenos. Deste modo deve-se evitar a prática de endogamia em plantéis, pois pode acarretar a perda da diversidade genética no locus MHC, e consequentemente elevar a susceptibilidade

dos rebanhos à epidemias (EIMES et al., 2011).

A via metabólica transdução do olfato foi a mais enriquecida ($p < 0,05$) por meio do KEGG para as CNVRs candidatas oriundas do painel de 50K e para o painel HD (Tabelas 7 e 10). Este resultado está condizente com resultados relados na literatura em bovinos, pois na maioria dos estudos têm sido encontrados genes dessa família enriquecidos nas CNVRs (BICKHART et al., 2012; BOUSSAHA et al., 2015; SILVA et al., 2016b; GAO et al., 2017; YANG et al., 2017). Os genes dos receptores do olfato (OR) constituem a maior família do genoma dos mamíferos dispersos na maioria dos cromossomos (LEE et al., 2013), sendo que, a maioria dos genes da família OR estão em variação no número de cópias nos genomas de diferentes espécies (LIU et al., 2010; VAN ZIFFLE; YANG; CHEHAB, 2011; WASZAK et al., 2010). Esses genes têm representado importante papel na evolução e no sistema adaptativo dos mamíferos (HASIN et al. 2008; NIIMURA, 2009). Genes da família OR ocorrem em regiões próximas a duplicações segmentares (SD) e estas regiões estão associadas com CNVs nas regiões gênicas (HOU et al., 2011; CICCONARDI et al., 2013, JIANG et al., 2013).

A detecção por meio de várias metodologias neste trabalho permitiu identificar e caracterizar acuradamente CNVRs candidatas relacionadas principalmente à resistência a doenças, como mastite e parasitas e às características reprodutivas em bovinos Girolando e auxiliar no mapa de CNVs em bovinos taurinos e zebuínos. Além disso, regiões de CNVs podem ser utilizadas em uma pré-seleção de indivíduos para o teste de progênie, por meio de identificação de regiões específicas relacionadas as características de interesse, como fertilidade e resistência a parasitas.

Entretanto, Para uma melhor avaliação dessas regiões ainda se faz necessário uma validação por PCR em tempo real e análises de associação com fenótipos. No entanto, muitos estudos verificaram os efeitos das variações no número de cópias por meio sa função dos genes anotados nas CNVRs identificadas (SEROUSSI et al., 2010; SILVA et al.2016b; YANG et al., 2017) similar ao que foi realizado neste trabalho, pois ainda há diversas especulações sobre estudos de associação ampla envolvendo CNVS com fenótipos, principalmente quantitativos (KIM et al., 2012) devido à complexidade dos genótipos gerados pela identificação de CNVs (HANDSAKER et al., 2015).

5 CONCLUSÃO

A maioria das regiões de variações no número de cópias candidatas identificadas neste estudo demonstraram que podem estar influenciando características de importância econômica para a bovinocultura de leite, principalmente relacionadas a sanidade e reprodução. Nossos resultados também evidenciaram regiões do sistema imune adaptativo com CNVs, principalmente nos cromossomos BTA5 e BTA17, provenientes exclusivamente dos animais zebuínos presentes nos animais da raça Girolando, devido essas regiões esses animais são mais resistentes e produtivos em ambientes de clima tropical.

6 REFERÊNCIAS

ABYZOV, A.; URBAN, A. E.; SNYDER, M.; GERSTEIN, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. **Genome Research**, v. 21, p. 974-984, 2011. Disponível em: < http://doi.org/10.1101/gr.114876.110 >

ADAMS, W.T.; SKOPEK, T. R. Statistical test for the comparison of samples from mutational spectra. **Journal of Molecular Biology**, v. 194, 391-396, 1987.

BALAKRISHNAN, C. N.; EKBLOM, R.; VÖLKER, M.; WESTERDAHL, H.; GODINEZ, R.; KOTKIEWICZ, H.; BURT, D. W.; GRAVES, T.; GRIFFIN, D. K.; WARREN, W. C.; EDWARDS, S. V.Gene duplication and fragmentation in the zebra finch major histocompatibility complex. **BMC Biology**, v. 8, p. 29, 2010.Disponível em: < http://doi.org/10.1186/1741-7007-8-29 > .

BENTLEY, D. R.; BALASUBRAMANIAN, S.; SWERDLOW, H. P.; SMITH, G. P.; MILTON, J.; BROWN, C. G.; … KLENERMAN, D.; DURBIN, R.; SMITH, A. J. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. V. 456, n. 7218, p. 53-59, 2008. Dísponivel em: <http://doi.org/10.1038/nature07517 >.

BICKHART, D. M.; HOU, Y.; SCHROEDER, S. G.; ALKAN, C.; CARDONE, M. F.; MATUKUMALLI, L. K.; SONG, J. ; SCHNABEL, R. D. ; VENTURA, M.; TAYLOR, J. F. ; GARCIA, J. F.; VAN TASSELL, C. P. ; SONSTEGARD, T. S.; EICHLER, E. E.; LIU, G. E.; HUGHES, H. Copy number variation of individual cattle genomes using next-generation sequencing. **Genome Research**, v. 22, p. 778–790, 2012. Disponível em: < http://doi.org/10.1101/gr.133967.111>.

BOUSSAHA, M; ESQUERRE, D.; BARBIERI, J.; DJARI,A.;PINTON,A.; LETAIEF,R.;SALIN, G.; ESCUDIÉ, F.; ROULET,A.; FRITZ,S.; SAMSON,F.; GROHS,C.; BERNARD,M.; KLOPP,C.; BOICHARD,D.; ROCHA, D. Genome-wide study of structural variants in bovine Holstein, Montbéliard e Normande dairy breeds**.** **Plos One**, v. 10, e0135931, 2015. Disponível em: < https://doi.org/10.1371/journal.pone.0135931 > .

BROWN W. C, MCGUIRE TC, MWANGI W, KEGERREIS KA, MACMILLAN H, LEWIN HA, PALMER GH. Major histocompatibility complex class II DR-restricted memory CD4(+) T lymphocytes recognize conserved immunodominant epitopes of Anaplasma marginale major surface protein 1a. **Infection and Immunity**, v. 70, p. 5521–5532, 2002. Disponível em: < http://doi.org/10.1128/IAI.70.10.5521-5532.2002 >.

BUZANSKAS, M. E. ; GROSSI, D. A. ; VENTURA, R. V. ; SCHENCKEL, F. S. ; CHUD, T. C. S.; STAFUZZA, N. B.; ROLA, L. D.; MEIRELLES, S. L. C. ; MOKRY, F. B.; MUDADU, M. A.; HIGA, R. H.; SILVA, M. V. G. B.; ALENCAR, M. M. ; REGITANO, L. C. A. ; MUNARI, D. P. Candidate genes for male and female reproductive traits in Canchim beef cattle. **Journal of Animal Science and Biotechnology ,** v. 8, n. 67, 2017. Disponível em: <http://doi.org/10.1186/s40104-017-0199-8 >.

CAMPBELL, P. J.; STEPHENS, P. J.; PLEASANCE, E. D.; O'MEARA, S.; LI, H.; SANTARIUS, T.; STEBBINGS, L. A.; LEROY, C.; EDKINS, S.; HARDY, C.; TEAGUE, J. W.; MENZIES, A.;GOODHEAD, I.; TURNER, D. J.; CLEE, C. M.; QUAIL, M. A.; COX, A.; BROWN, C.; DURBIN, R.; HURLES, M. E.; EDWARDS, P. A. W.; BIGNELL, G. R.; STRATTON, M. R.; FUTREAL, A. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nature Genetics, v. 40, p. 722-729, 2008. Disponível em: <http://doi.org/10.1038/ng.128>.

CHAMERO, P.; LEINDERS-ZUFALL, T.; ZUFALL, F. From genes to social communication: Molecular sensing by the vomeronasal organ. **Trends in Neurosciences**, v. 35, n. 10, p. 597-606, 2012. Disponível em: <http://doi.org/ 0.1016/j.tins.2012.04.011> .

CERRI, R. L. A.; THOMPSON, I. M; KIM, I. H.; EALY, A. D.; HANSEN, P. J.; STAPLES, C. R.;LI, J. L.;SANTOS, J. E. P.; THATCHER, W. W. Effects of lactation and pregnancy on gene expression of endometrium of Holstein cows at day 17 of the estrous cycle or pregnancy **Journal of Dairy Science**, v. 95, n.10, p. 5657–5675, 2012. Disponível em: <http://doi.org/10.3168/jds.2011-5114>.

CHEN, K.; WALLIS, J. W.; MCLELLAN, M. D.; LARSON, D. E.; KALICKI, J. M.; POHL, C. S.; MCGRATH, S. D.; WENDL, M. C.; ZHANG, Q. Y.; LOCKE, D. P.; LOCKE, D. P.; SHI, X.; FULTON, R. S., LEY, T. J.; WILSON, R. K.; DING, L., MARDIS, E. R. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nature Methods, v. 6, p. 677–681, 2009. Disponivel em : < https://www.doi.org/ 10.1038/nmeth.1363>.

CHEN, X.; CHENG, Z.; ZHANG, S.; WERLING, D.; WATHES, C. Combining Genome Wide Association Studies and Differential Gene Expression Data Analyses Identifies Candidate Genes Affecting Mastitis Caused by Two Different Pathogens in the Dairy Cow. **Open Journal of Animal Sciences**, v. 5, p. 358-393, 2015. Disponivel em : < http://dx.doi.org/10.4236/ojas.2015.54040>.

CICCONARDI, F.; CHILLEMI, G.; TRAMONTANO, A.; MARCHITELLI, C.; VALENTINI, A.; AJMONE-MARSAN, P.; NARDONE, A. Massive screening of copy number population-scale variation in Bos taurus genome. **BMC Genomics**, v.14, p. 124-138, 2013. Disponível em: < https://doi.org/10.1186/1471-2164-14-124 >.

CIUCCI, T., BOSSELUT, R. Gimap and T cells: a matter of life or death. **European Journal of Immunology**, v. 44, n. 2, p. 348-351, 2014. Disponível em: < https://www.doi.org/10.1002/eji.201344375>.

COLELLA, S., YAU, C., TAYLOR, J. M., MIRZA, G., BUTLER, H., CLOUSTON, P., BASSET, A. S.; SELLER, A.; HOLMES, C. C.; RAGOUSSIS, J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. **Nucleic Acids Research**, v. 35, n. 6, p. 2013-2025, 2007. Disponível em: < http://doi.org/10.1093/nar/gkm076 >

CONRAD, D. F.; PINTO, D.; REDON, R., FEUK, L.; GOKCUMEN, O.; ZHANG, Y.; AERTS, J.; ANDREWS, T. D.; BARNES, C.; CAMPBELL, P.; FITZGERALD, T.; HU, M.; IHM, C. H; KRISTIANSSON, K.; MACARTHUR, D. G.; MACDONALD, J. R.; ONYIAH, I.; PANG, A. W; ROBSON S.; STIRRUPS, K.; VALSESIA, A.; WALTER, K.; WEI, J.; WELLCOME TRUST CASE CONTROL CONSORTIUM, TYLER-SMITH, C.; CARTER, N. P.; LEE, C.; SCHERER, S. W.; HURLES, M. E. Origins and functional impact of copy number variation in the human genome. **Nature**, v. 464, n. 7289, p. 704–712, 2010. Disponivel em < http://doi.org/10.1038/nature08516>.

DISKIN, S.J.; LI, M.; HOU, C.; YANG, S.; GLESSNER, J.; HAKONARSON, H.; BUCAN, M.; MARIS, J. M.; WANG, K. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. **Nucleic Acids Research**, v. 36, p. e126, 2008. Disponível em: < http://doi.org/10.1093/nar/gkn556>.

EIMES, J. A.; BOLLMER, J. L.; WHITTINGHAM, L. A.; JOHNSON, J. A.; VAN OOSTERHOUT, C.; DUNN, P. O. Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and loss of copy number variation. **Journal of Evolutionary Biology**, v. 24, p. 1847-1856, 2011. Disponível em: < http://doi.org /10.1111/j.1420-9101.2011.02311.x >.

GAO, Yahui; Jiang, J.; YANG, S.; HOU, Y.; LIU, G. E.; ZHANG, S.; ZHANG, Q.; SUN, D. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. **BMC Genomics**, v.18, p. 265, 2017. Disponível em: <http://doi.org/10.1186/s12864-017-3636-3 >

HANDSAKER, R. E.; VAN DOREN, V.; BERMAN, J. R.; GENOVESE, G.; KASHIN, S.; BOETTGER, L. M.; MCCARROLL, S. A. Large multi-allelic copy number variations in

humans. **Nature Genetics**, v. 47, n. 3, p. 296–303, 2015. Disponível em:< http://doi.org/10.1038/ng.3200 >.

HASIN, Y. OLENDER, T.; KHEN, M.; GONZAGA-JAUREGUI, C.; KIM, P. M.; URBAN, A. E., SNYDER, M.; GERSTEIN, M. B.; LANCET, D.; KORBEL, J. O. High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. **PLoS Genetics**, v. 4, p. e1000249, 2008. Disponível em: < https://doi.org/10.1371/journal.pgen.1000249 >.

HAWKEN, R. J.; ZHANG, Y. D.; FORTES, M. R. S.; COLLIS, E.; BARRIS, W. C.; CORBET, N. J. WILLIAMS, P. J.; FORDYCE, G.; HOLROYD, R. G.; WALKLEY, J. R.; BARENDSE, W.; JOHNSTON, D. J.; PRAYAGA, K. C.; TIER, B.; REVERTER, A.; LEHNERT, S. A. Genome-wide association studies of female reproduction in tropically adapted beef cattle Journal of Animal Science, v. 90, p. 1398–1410, 2012. Disponível em: < http://doi.org/10.2527/jas.2011-4410 >.

HENRICHSEN, C.N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v.18, p.R1-8, 2009.Disponível em : < http://doi.org/ 10.1093/hmg/ddp011>

HERZIG, C. T.; BALDWIN, C. L. Genomic organization and classification of the bovine WC1 genes and expression by peripheral blood gamma delta T cells. **BMC Genomics**. v. 10, p.191, 2009. Disponível em: < http://doi.org/ 10.1186/1471-2164-10-191 >.

HOU, Y.; LIU, G. E.; BICKHART, D. M.; CARDONE, M. F.; WANG, K.; KIM, E., MATUKUMALLI, L. K.; VENTURA, M.; SONG, J.; VANRADEN, P. M.; SONSTEGARD, T. S.; VAN TASSELL, C. P. Genomic characteristics of cattle copy number variations. **BMC Genomics**, v. 12, p.127, 2011. Disponível em: <http://doi.org/10.1186/1471-2164-12-127>

HOU, Y.; LIU, G. E.; BICKHART, D. M.; MATUKUMALLI, L. K.; LI, C.; SONG, J.; GASBARRE, L. C.; VAN TASSELL, C. P.; SONSTEGARD, T. S. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. **Functional & Integrative Genomics**. v.12, p. 81-92, 2012a.Disponivel em : < http://doi.org/10.1007/s10142-011-0252-1>

HOU, Y.; BICKHART, D. M.; HVINDEN, M. L.; LI, C.; SONG, J.; BOICHARD, D. A.; FRITZ, S.; EGGEN, A.; DENISE, S.; WIGGANS, G. R.; SONSTEGARD, T.S.; VAN TASSEL, C. P.; LIU, G.E. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. BMC Genomics. v.13, p. 376, 2012b.

HOU, Y.; BICKHART, D. M.;CHUNG, H.; HUTCHISON, J. L.; NORMAN, H.D.; CONNOR, E.E.; LIU, G. E. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. **Functional & Integrative Genomics**, v.12, n. 4, p 717-723, 2012c. Disponível em: < http://doi.org/ 10.1007/s10142-012-0295-y>

.JIANG, L.; JIANG, J.; WANG, J.; DING, X.; LIU, J.; ZHANG, Q. Genome wide identification of copy number variations in Chinese Holstein**. Plos One**, v. 7, e48732, 2012. Disponível em: < http://doi.org/10.1371/journal.pone.0048732 >

JIANG, L.; JIANG, J.;YANG,J.; LIU, X; WANG, H.; DING, X; LIU, J., ZHANG, Q. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. **BMC Genomics**, v.14, p.131, 2013. Disponível em: < https://doi.org/10.1186/1471-2164-14-131 >.

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

KELM SC, DETILLEUX JC, FREEMAN AE, KEHRLI MEJ, DIETZ, AB, FOX LK, BUTLER JE, KASCKOVICS I, KELLY DH. Genetic association between parameters of innate immunity and measures of mastitis in periparturient Holstein cattle. Journal of Dairy Science, v. 80, p. 1767-1775, 1997.

KIM, J.H.; HU, H.J.; YIM, S.H.; BAE, J.S.; KIM, S.Y.; CHUNG, Y. J. CNVRuler: a copy number variation-based case-control association analysis tool**. Bioinformatics**, v. 28, p. 1790-2, 2012.

LEE, K.; NGUYEN, D. T.; CHOI, M.; CHA, S.-Y.; KIM, J.-H.; DADI, H.; SEO, H. G.; SEO, K.; CHUN, T.; PARK, C. Analysis of cattle olfactory subgenome: the first detail study on the characteristics of the complete olfactory receptor repertoire of a ruminant. **BMC Genomics**, v. 14, p. 596, 2013. Disponível em:< http://doi.org/10.1186/1471-2164-14-596 >.

LETAIEF, R.; REBOURS, E.; GROHS, C.; MEERSSEMAN, C.; FRITZ, S.; TROUILH, L.; ESQUERRÉ, D.; BARBIERI, J.; KLOPP, C.; PHILIPPE, R.; BLANQUET, V.; BOICHARD, D.; ROCHA, D.; BOUSSAHA, M. Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. Genetics Selection Evolution, v. 49, n.1, p. 77. Disponível em: < http://doi.org/10.1186/s12711-017-0352-z>.

LIU, G.E., VENTURA, M., CELLAMARE, A., CHEN, L.; CHENG, Z.; ZHU, B.; LI, C.; SONG, J.; EICHLER, E. E. Analysis of recent segmental duplications in the bovine genome, **BMC Genomics**, v. 10, p. 571, 2009. Disponível em: <https://doi.org/10.1186/1471-2164-10-571>

LIU, G. E.; HOU, Y.; ZHU, B.; CARDONE, M. F.; JIANG, L.; CELLAMARE, A.; MITRA, A.; ALEXANDER, L. J.; COUTINHO, L. L.; DELL'AQUILA, M. E; GASBARRE, L. C.; LACALANDRA, G.; LI, R. W; MATUKUMALLI, L. K.; NONNEMAN, D.; REGITANO, L. C. A.; SMITH, T. P. L.; SONG, J.; SONSTEGARD, T. S.; VAN TASSEL, C. P.; VENTURA, M.; EICHLER, E. E.; MCDANELD,T. G.; KEELE, J. W. Analysis of copy number variation among diverse cattle breeds. **Genome Research** .v. 20, p. 693–703, 2010. Disponível em: <http://doi.org/10.1101/gr.105403.110>.

LIU, J.; ZHANG, L.; LINGYANG, X.; HANGXING, R. ; LU, ZHANG, X. ZHANG, S. ZHOU, X. WEI, C.; ZHAO, F; DU, L. Analysis of copy number variation in the sheep

genome using 50K SNP BeadChip array. **BMC Genomics**, v**.** 14, p. 229, 2013. Disponível em: <https://doi.org/10.1186/1471-2164-14-229>.

LOCKE, D. P., SHARP, A. J., MCCARROLL, S. A., MCGRATH, S. D., NEWMAN, T. L., CHENG, Z., SCHWARTZ, S.; ALBERTSON, D. G.; PINKEL, D.; ALTSHULER, D. M.; EICHLER, E. E. Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome. **American Journal of Human Genetics**, v. 79, n. 2, p. 275–290, 2006.

MCDANELD, T. G.; KUEHN, L.; THOMAS, M.; POLLAK, E.; KEELE, J. Deletion on chromosome 5 associated with decreased reproductive efficiency in female cattle. **Journal of Animal Science,** v. 92, p. 1378-1384, 2014. Disponível em: <http://doi.org/10.2527/jas.2013-6821>.

MILLS, R. E.; WALTER, K.; STEWART, C.; HANDSAKER, R. E.; CHEN, K.; ALKAN, C.; 1000 GENOMES PROJECT. Mapping copy number variation by population scale genome sequencing. Nature, v. 470, n. 7332, p. 59–65, 2011. Disponível em: <http://doi.org/10.1038/nature09708>.

MINEGISHI, Y.; COUSTAN-SMITH, E.; WANG, Y.-H.; COOPER, M. D.; CAMPANA, D.; CONLEY, M. E. Mutations in the Human λ5/14.1 Gene Result in B Cell Deficiency and Agammaglobulinemia. **The Journal of Experimental Medicine**, v. 187, n. 1, p. 71-77, 1998.

MINOZZI, G.; NICOLAZZI, E. L.; STELLA, A.; BIFFANI, S.; NEGRINI, R.; LAZZARI, B.; AJMONE-MARSAN P.; WILLIAMS, J. L. Genome wide analysis of fertility and production traits in Italian Holstein cattle**. PLoS One**, v. 8, n. 11, e80219, 2013. Disponível em: < http://doi.org/10.1371/journal.pone.0080219.

MORSE K, NORIMINE J, HOPE JC, BROWN WC. Breadth of the CD4+ T cell response to Anaplasma marginale VirB9-1, VirB9-2 and VirB10 and MHC class II DR and DQ restriction elements. **Immunogenetics**, v. 64, p. 507–523, 2012. Disponível em: < http://doi.org/10.1007/s00251-012-0606-4 >.

MOROTA, G.; PEÑAGARICANO, F.; PETERSEN, J. L.; CIOBANU, D. C.; TSUYUZAKI, K.; NIKAIDO, I. An application of MeSH enrichment analysis in livestock. Animal Genetics, v. 46, p. 381–387, 2015. Disponível em: < http://doi.org/10.1111/age.12307>.

NIIMURA, Y. Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. **Human Genomics**, v. 4, n. 2, p. 107-118. Disponível em: < http://doi.org/10.1186/1479-7364-4-2-107 >.

O'GORMAN, G. M.; PARK, S.; HILL, E. W.; MEADE,K. G.; COUSSENS, P. M.; JAN NAESSENS, M. A.; KEMP, S. J.; MACHUGH, D. E. Transcriptional profiling of cattle infected with Trypanosoma congolense highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility. **BMC Genomics**, v. 10, n. 1, 2009. Disponível em: < https://doi.org/10.1186/1471-2164-10-207>.

PARK, J. Y.; MOON, J. S.; KIM, S. H.; KWON, N. H.; AHN, J.S.; DAVIS, W.C.; DAVIES, C. J. Characterization of lymphocyte subpopulations and major histocompatibility complex haplotypes of mastitis resistant and susceptible cows. **Journal of Veterinary Science**, v. 5, n. 1, p. 29-39, 2004

PARMENTIER, M.; LIBER, T F.; SCHURMANS, S.; SCHIFFMANN, S.; LEFORT, A.; EGGERICKX, D.; LEDENT, C.; MOLLEREAU, C.; GÉRARD, C.; PERRET, J.; GROOTEGOED A.; VASSART, G. Expression of members of the putative olfactory receptor gene family in mammalian germ cells. **Nature,** v. 355, p. 453-455, 1992. Disponível em: < https://doi.org/10.1038/355453a0>.

PASHMI, M.; QANBARI, S.; GHORASHI, S. A.; SHARIFI, A. R.; SIMIANER, H. Analysis of relationship between bovine lymphocyte antigen DRB3.2 alleles, somatic cell count and milk traits in Iranian Holstein population. **Journal of Animal Breeding and Genetics**, v. 126, p. 296-303, 2009. Disponível em: < http://doi.org//10.1111/j.1439-0388.2008.00783.x > .

PERRY, G. H.; DOMINY, N. J.; CLAW, K. G.; LEE, A. S.; FIEGLER, H.; REDON, R.; WERNER, J.; VILLANEA, F. A.; MOUNTAIN, J. L.; MISRA, R.; CARTER, N. P., LEE, C.; STONE, A. C. Diet and the evolution of human amylase gene copy number variation. Nature Genetics, v. 39, p. 1256-1260. 2007. Disponível em: <http://doi.org/10.1038/ng2123>

QUINLAN, A. R.; HALL, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics**, v. 26, n. 6, p. 841–842, 2010. Disponível em: <http://doi.org/10.1093/bioinformatics/btq033>

REDON, R.; ISHIKAWA, S., FITCH, K. R.; FEUK, L.; PERRY, G. H.; ANDREWS, T. D.; FIEGLER, H.; SHAPERO, M. H.; CARSON, A. R.; CHEN, W.; CHO, E. K.; DALLAIRE, S.; FREEMAN, J. L.; GONZÁLEZ, J. R; GRATACÓS, M.; HUANG, J.; KALAITZOPOULOS, D.; KOMURA, D.; MACDONALD, J. R.; MARSHALL, C. R.; MEI, R.; MONTGOMERY, L.; NISHIMURA, K.; OKAMURA, K.; SHEN, F.; SOMERVILLE, M. J.; TCHINDA, J.; VALSESIA, A.; WOODWARK, C.; YANG, F.; ZHANG, J.; ZERJAL, T.; ZHANG, J.; ARMENGOL, L. L.; CONRAD, D. F.; ESTIVILL, X.; TYLER-SMITH, C.; CARTER, N. P.; ABURATANI, H.; LEE, C.; JONES, K. W.; SCHERER, S. W.; HURLES, M. E. Global variation in copy number in the human genome. **Nature**, v. 444, p. 444-454, 2006. Disponível em: < https://www.doi.org/10.1038/nature05329>.

SEROUSSI, E.; GLICK, G.; SHIRAK, A.; YAKOBSON, E.; WELLER, J. I.; EZRA, E.; ZERON, Y. Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. **BMC Genomics,** v.11, p. 673, 2010. Diponível em: < https://doi.org/10.1186/1471-2164-11-673>.

SHARIF S, MALLARD BA, WILKIE BN, SARGEANT JM, SCOTT HM, DEKKERS JC, LESLIE KE. Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cellscore in Canadian dairy cattle. **Animal Genetics**, v. 29, 185-193,1998.

SILVA, J. C. P. M.; VELOSO, C. M. **Raças de Gado Leiteiro**. 1.ed. Viçosa: Aprenda Fácil, 2011. 149p

SILVA, M. V. G. B; MARTINS, M. F.; PAIVA, L. C.; CEMBRANELLI, M. A. R; ARBEX, W. A.; SANTO, K. C. L.; PANETTO, J. C. C.; COSTA, C. N.; CARVALHO, B. C. Programa de Melhoramento Genético da Raça Girolando – Sumário de Touros – Resultado do Teste de Progênie – Julho 2013, Juiz de Fora : Embrapa Gado de Leite, 2013, 52 p.

SILVA, V. H. da; REGITANO, L. C. de A.; GEISTLINGER, L.; PÉRTILLE, F.; GIACHETTO, P. F.; BRASSALOTI, R. A.; MOROSINI, N. S.; ZIMMER, R.; COUTINHO, L. L. Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. PlosOne, v. 11, n. 6, e0157711, 2016a. Disponível em: <https://doi.org/10.1371/journal.pone.0157711>

SILVA, J. M. da; GIACHETTO, P. F.; SILVA, L. O. da; CINTRA, L. C.; PAIVA, S. R.; YAMAGISHI, M. E. B.; CAETANO, A. R. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTL affecting production traits. **BMC Genomics,** v. 17, p.454, 2016b. Disponível em: <https://doi.org/10.1186/s12864-016-2752-9>

STEAR, M. J.; BISHOP, S. C.; MALLARD, B. A.; RAADSMA, H. The sustainability, feasibility and desirability of breeding livestock for disease resistance**. Research in Veterinary Science**, v. 71, n. 1, p. 1-7, 2001.

STOTHARD, P.; CHOI, J. W. ; BASU U.; SUMNER-THOMSON, J. M.; MENG, Y .; LIAO, X.; MOORE S. S. Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. **BMC Genomics**, v.12, p.559, 2011. Disponível em: < https://doi.org/10.1186/1471-2164-12-559>

THOMPSON-CRISPI, K.; SARGOLZAEI, M.; VENTURA, R.; ABO-ISMAIL, M.; MIGLIOR, F.; SCHENKEL, F.; MALLARD, B. A. A genome-wide association study of immune response traits in Canadian Holstein cattle. **BMC Genomics**, v. 15, p. 559, 2014. Disponível em: < http://doi.org/10.1186/1471-2164-15-559 >.

VAN ZIFFLE, J.; YANG, W.; CHEHAB, F. F. Homozygous Deletion of Six Olfactory Receptor Genes in a Subset of Individuals with Beta-Thalassemia. PLoS ONE, v. 6, n. 2, p. e17327, 2011. Disponível em: < http://doi.org/10.1371/journal.pone.0017327 >.

VASCONCELLOS, B. F.; PÁDUA, J. T.; MUÑOZ, M. F. C.; TONHATI, H. Efeitos Genéticos e Ambientais sobre a Produção de Leite, o Intervalo de Partos e a Duração da Lactação em um Rebanho Leiteiro com Animais Mestiços, no Brasil. **Revista Universidade Rural**, v.23, p.39-45, 2003

VESTAL, D. J.; JEYARATNAM, J. A. The Guanylate-Binding Proteins: Emerging Insights into the Biochemical Properties and Functions of This Family of Large Interferon-Induced Guanosine Triphosphatase. **Journal of Interferon & Cytokine Research,** v. 31, n. 1, p. 89-97, 2011. Disponível em :
< http://doi.org/10.1089/jir.2010.0102>

WANG, K.; LI, M.; HADLEY, D.; LIU, R.; GLESSNER, J.; GRANT, S. F. A.; HAKONARSON, H.; BUCAN, M. PennCNV: an integrated hidden Markov model designed for highresolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v.17, p.1665–1674, 2007.

WANG, Q.; QU, J. H.; CHENG, X. X.; KANG, Y. J.; WAN, L.; QIAN, M. P.; DENG, M. H. A study of biases of DNA copy number estimation based on PICR model. Frontiers of Mathematics in China, v. 6, p. 1203-1216, 2011. Disponível em: < https://doi.org/10.1007/s11464-011-0125-x>.

WASZAK, S. M.; HASIN, Y.; ZICHNER, T. OLENDER, T.; KEYDAR, I.; KHEN, M.; STÜTZ, A. M.; SCHLATT, A.; LANCET, D.; KORBEL, J. O.Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity. **PLoS computational biology**, v. 6, p. e1000988, 2010. Disponível em: <https://doi.org/10.1371/journal.pcbi.1000988 >.

WINCHESTER, L.; YAU, C.; RAGOUSSIS, J. Comparing CNV detection methods for SNP arrays. **Briefings in Function Genomics and Proteomics**, v. 8, p. 353-366, 2009

XU, L.; COLE, J. B.; BICKHART, D. M.; HOU, Y.; SONG, J.; VANRADEN, P. M.; SONSTEGARD, T. S.; VAN TASSELL, C. P.; LIU, G. E. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. **BMC Genomics**, v. 15, p.683, 2014. Disponível em: < http://doi.org/ 10.1186/1471-2164-15-683>

XU, L.; HOU, Y.; BICKHART, D. M.; ZHOU, Y.; ABDEL HAY, E. H.; SONG, J.; SONSTEGARD, T. S.; VAN TASSELL, C. P.; LIU, G. E. Population-genetic properties of differentiated copy number variations in cattle**. Scientific Reports,** v. 6, Article number: 2316, 2016. Disponível em: < http://doi.org/ 10.1038/srep23161>

YANG L.; XU, L.; ZHU, B.; NIU, H.; ZHANG, W.; MIAO, J., SHI, X.; ZHANG, M.; CHEN, Y.; ZHANG, L.; GAO, X.; GAO, H.; LI, L.; LIU, G. E.; LI, J. Genome-wide analysis reveals differential selection involved with copy number variation in diverse Chinese Cattle. **Scientific Reports,** v. 7, p. 14299, 2017. Disponível em: < http://doi.org/10.1038/s41598-017-14768-0>.

ZHANG, F.; GU, W.; HURLES, M.E.; LUPSKI, J. R. Copy number variation in human health, disease, and evolution. **Annual Review Genomics and Human Genetics,** v. 10, p. 451–481, 2009. Disponível em: < doi: 10.1146/annurev.genom.9.081307.164217>.

ZHOU, Y.; UTSUNOMIYA, Y.T.; XU, L.; ABDEL HAY, E. H.; BICKHART, D.M.; ALEXANDRE, P. A.; ROSEN, B. D.; SCHROEDER, S. G.; CARVALHEIRO, R.; NEVES, H. H. R.; SONSTEGARD, T.S.; VANTASSELL C.P.; FERRAZ, J. B. S.; FUKUMASU, H. GARCIA J.F.; LIU G.E. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genomics, v. 17, 2016. Disponível em: <https://doi.org/10.1186/s12864-016-2461-4>

ZHOU, Y.; UTSUNOMIYA, Y.T.; XU, L.; ABDEL HAY, E. H.; BICKHART, D.M.; SONSTEGARD, T. S.; VANTASSELL C.P.; GARCIA J.F.; LIU G.E. Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. **Scientific Reports**, v. 6, p. 29219, 2016b. Disponível em: < http://doi.org/10.1038/srep29219 >

CAPÍTULO 3. VARIAÇÕES NO NÚMERO DE CÓPIAS EM TOUROS GIR, HOLANDÊS E GIROLANDO POR MEIO DE TÉCNICA DE SEQUENCIAMENTO DE NOVA GERAÇÃO

RESUMO - A utilização de variação no número de cópias (CNV) em estudos de genética de populações poderá identificar possíveis divergências populacionais ocasionadas devido as diferenças no número de cópias (CN) de regiões ou genes e o impacto em características de interesse econômico tais como, eficiência reprodutiva e sanidade. O objetivo deste trabalho foi detectar CNVs em bovinos da raça Girolando, Gir e Holandês por meio de dados de sequenciamento de nova geração (NGS), identificar regiões de CNVs específicas em Girolando provenientes do Gir e do Holandês e verificar diferenciação populacional entre as raças por meio do número de cópias dos genes anotados nas regiões de CNVs. Foram utilizados dados de NGS oriundos de três touros da raça Girolando, quatorze touros da raça Gir e cinco touros da raça Holandês. Após o alinhamento das "reads" provenientes do ressequenciamento, as CNVs foram detectadas pela metodologia "read-depth" implementada no programa CNVnator. Para verificar a diferenciação populacional entre as raças por meio do número de cópias dos genes anotados nas regiões de CNVs, foi realizada a estatística V_{ST}. Eventos de deleções foram mais frequentes do que eventos de duplicações na maioria dos touros. Considerando todas as raças como população única, as CNVs foram agrupadas em 7.829 regiões, com frequência média de 24,01 % na população. Um total de 318 e 75 regiões de CNVs foi compartilhada exclusivamente entre Girolando e Gir e Girolando e Holandês, respectivamente. Genes relacionados com fertilidade (*MEPCE* e *ASB3*) e susceptibilidade às doenças (*HLX* e MIR-455) foram anotados nas regiões específicas compartilhadas entre Girolando e as raças formadoras (Gir e Holandês). Os valores de V_{ST} variaram de -0,37 a 0,98, quanto maior o valor para V_{ST} maior a diferenciação populacional. Nesse estudo, 70 genes apresentaram V_{ST} maior do que 0,50, indicando elevada diferenciação populacional entre as raças. O gene *AOX1* apresentou elevado V_{ST} (0,98), presente em maior número de cópias na raça Gir do que na raça Holandês. A região em que *AOX1* está localizado, pode estar segregando entre as raças zebuínas e compostas. Esse estudo permitiu concluir que CNVs são extremamente importantes para características reprodutivas e de sanidade. A variação no número de cópias teve participação na divergência genética entre bovinos zebuínos e taurinos, contribuindo para as diferenças fenotípicas entre as subespécies. O estudo de diferenciação populacional evidenciou seleção positiva no genoma de animais da raça Gir e Girolando para características relacionadas à adaptação desses animais aos ambientes de clima tropical, possivelmente originárias do processo de domesticação.
Palavras-chave: *Bos taurus taurus, Bos taurus indicus*, bovinos de leite, genética de populações

CHAPTER 3 COPY NUMBER VARIATION IN GIROLANDO, GIR, AND HOLSTEIN BULLS USING NEXT-GENERATION SEQUENCING

ABSTRACT – Applying copy number variation (CNV) in population genetics studies could identify population divergences due to alterations in copy number regions or genes and what their impact on economic interest traits, including reproductive efficiency and health traits. The objective of this study was to detect CNVs in Girolando, Gir and Holstein bulls using next-generation sequencing (NGS), to identify breed-specific CNV regions in Girolando from Gir and Holstein, and to verify the population differentiation among the breeds. We analyzed three genomes from Girolando bulls, 14 from Gir and five from Holstein.After the reads alignment, a read depth method was performed using CNVnator software to detect the CNVs. To verify the population differentiation among the breeds the Vst statistic was applied for each CNV gene. Deletions events were more frequent than duplications in most of the bulls. Considering all breeds as unique population, CNVs were merged in 7,829 regions with average frequency in the population equal to 24.01%. A total of 318 and 75 regions of CNVs were shared exclusively between Girolando and Gir and Girolando and Holstein, respectively. Genes linked to fertility (*MEPCE, ASB3*) and to disease susceptibility (*HLX, MIR-455*) were annotated in the specific regions shared between Girolando and the pure-breeds (Gir and Holstein). The Vst values ranged from -0.37 to 0.98, high Vst values indicate a high level of population differentiation. In this study, 70 genes presented Vst greater than 0.50, indicating a high population differentiation among the breeds. The *AOX1* gene showed top Vst (0.98), which it was present in a variable copy number in Gir breed. The region where *AOX1* is located may be segregating between the zebu and composite breeds. Our findings indicated that CNVs are extremely important for reproductive and health traits. The population differentiation study evidenced positive selection in genome of the Gir and Girolando animals for traits related to the adaptability of the breeds in tropical environmental may have originated from the domestication process.
Keywords: *Bos taurus taurus*, *Bos taurus indicus*, dairy cattle, population genetics

1 INTRODUÇÃO

A população de bovinos pode ser classificada em duas subespécies, sendo taurinos (sem presença de giba) e zebuínos (com presença de giba), apresentando grandes diferenças fenotípicas entre si (BRADLEY et al., 1996; THE BOVINE HAPMAP CONSORTIUM, 2009). Estudos demonstraram que essas subespécies divergiram do ancestral comum entre 0,6 e 2 milhões de anos atrás (MACHUGH et al., 1997; HIENDLEDER; LEWALSKI; JANKE, 2008). Evidências arqueológicas mostraram que taurinos e zebuínos surgiram independentemente em pelo menos dois locais de domesticação, sendo que taurinos emergiram do Oriente Próximo (Crescente Fértil) aproximadamente 8.000 a 10.000 anos atrás e os zebuínos do Vale do rio Indo (Paquistão), com difusão pela Índia ~6.000 a 8.000 atrás e mais recentemente (3.000 anos) com a introdução de machos no norte da África (LOFTUS et al., 1994; LARSON et al., 2014).

Raças bovinas localmente adaptadas com ampla diversidade genética e fenotípica foram originadas por alterações do genoma ocorridas no processo de domesticação e pela combinação dos efeitos da seleção positiva e dos eventos demográficos e de introgressão (XU et al., 2015). Para verificar as divergências genéticas entre taurinos e zebuínos e identificar possíveis origens do processo evolutivo, demográfico e de domesticação, estudos de genética de populações têm sido amplamente realizados por meio de painéis de polimorfismos de nucleotídeo único (SNP) (MAKINA et al., 2014; BUZANSKAS, et al. 2017; CAMPOS et al., 2017).Outra importante fonte de variação de diversidade genética, denominada de variação no número de cópias (CNV), tem demonstrado importante participação em processos de evolução e adaptação das espécies, como em humanos (ZHANG et al., 2009; ISKOW; GOKCUMEN; LEE, 2012), suínos (PAUDEL et al.,2013) e canídeos (BERGLUND et al., 2012). CNVs podem ser definidas como duplicações e deleções variando de 50 pares de bases até cinco megabases (Mb) na comparação entre os genomas de dois indivíduos (MILLS et al., 2011; SUDMANT et al., 2015).

Em bovinos, alguns autores têm empregado CNVs em estudos envolvendo genética de populações com a finalidade de investigar padrões de seleção, identificar a origem das diferenças genéticas entre raças de leite e de corte e principalmente entre as subespécies (taurinos e zebuínos) (BICKHART et al., 2016; XU et al., 2016;

YANG et al.,2017). Embora a evolução dos genomas bovino e a história demográfica tenham sido demasiadamente exploradas, a diversidade e propriedades genéticas das CNVs ainda não são bem estabelecidas como os SNPs (BICKHART et al, 2016). As CNVS abrangem regiões de DNA maiores e possuem efeitos na dosagem gênica, ou seja, na quantidade de produto do gene (mRNA ou proteína expressos) que poderá aumentar ou diminuir de acordo com cada cópia adicional ou reduzida do gene (INNAN; KONDRASHOV, 2010). Consequentemente, a regulação gênica será alterada e poderá ocorrer uma exposição a alelos recessivos (ZHANG et al., 2009). A aplicação de CNVs em estudos de genética de população poderá auxiliar na compreensão dessas variações e o impacto em características de interesse econômico como de eficiência reprodutiva e sanidade.

Animais provenientes de acasalamentos entre raças zebuínas e taurinas têm sido muito utilizados nos sistemas de produção de leite e de corte, principalmente em países de clima tropical, devido as diferenças genéticas e fenotípicas entre as subespécies ocasionadas principalmente pela adaptação aos diferentes ambientes e seleção, proporcionando complementariedade das raças (VASCONCELLOS et al., 2003). A raça Girolando é um exemplo de raça leiteira oriunda de cruzamentos entre taurinos (Holandês) e zebuínos (Gir), sendo criada com a finalidade de explorar as características favoráveis do Gir, como rusticidade e adaptabilidade e a elevada produção de leite da raça Holandesa.

Considerando a importância de zebuínos e animais provenientes do acasalamento entre taurinos e zebuínos para o Brasil (FAS/USDA, 2017), sendo a pecuária uma das atividades econômicas mais representativas do país (CEPEA/CNA, 2017), é imprescindível a condução de estudos envolvendo CNVs em animais de raça composta e em raças formadoras para melhor caracterização da diferenciação populacional e verificar possíveis CNVs específicos de taurinos e zebuínos. O objetivo desse trabalho foi detectar CNVs em bovinos da raça Girolando, Gir e Holandês por meio de dados de sequenciamento de nova-geração (NGS), identificar regiões de CNVs específicas em Girolando provenientes do Gir e do Holandês e verificar diferenciação populacional entre as raças por meio do número de cópias dos genes anotados nas regiões de CNVs.

2 MATERIAL E MÉTODOS

2.1 Animais e e ressequenciamento

O DNA de três touros da raça Girolando, quatorze touros da raça Gir e cinco touros da raça Holandês foi extraído a partir de amostras de sangue e sêmen utilizando DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA), conforme recomendado pelo fabricante. O DNA extraído foi quantificado e avaliado por espectrofotometria (NanoDrop 1000, Thermo Scientific, Wilmington, DE, USA). O genoma completo dos touros foi ressequenciado por meio da plataforma Illumina HiSeq2000 (Illumina Inc., San Diego, CA, USA).

As bibliotecas utilizadas para o ressequenciamento foram do tipo ‘paired-end’, ou seja, houve sequenciamento de ambas as extremidades dos fragmentos de DNA com tamanho de 500bp, em que foram produzidas leituras com tamanho de 2 x 100 bp e 200 bp. Os procedimentos de construção das bibliotecas foram realizados de acordo com os protocolos sugeridos pelo fabricante.

2.2 Controle de qualidade e alinhamento das sequências no genoma referência

A partir dos dados gerados dos sequenciamentos de nova-geração, a qualidade das sequências obtidas foi avaliada por meio da ferramenta FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Após o diagnóstico dos dados, foi realizado o controle de qualidade das leituras por meio do programa “SeqyClean” (ZHBANNIKOV et al., 2017) conforme o recomendado pelo protocolo do projeto “1000 bulls” (http://www.1000bullgenomes.com/) em que foram removidas

i. Leituras com três ou mais bases não identificadas (N) nas sequências
ii. Leituras com média de qualidade para “Phred score” inferior ou igual a 20 ou seja, a probabilidade média de que as bases estejam incorretas foi de no mínimo 0,01.
iii. Bases de baixa qualidade (com Phred score menor do que 20).
iv. Leituras com comprimento menor do que 50 bases nas sequências.

Também foram removidos das sequências adaptadores e possíveis contaminantes. A remoção dos adaptadores é importante, pois as suas sequências

podem ser incorporadas na montagem do genoma, gerando um alinhamento errôneo das bases no genoma.

Após o controle de qualidade foi realizado o alinhamento das sequências com base no genoma referência bovino UMD 3.1 (http://stothard.afns.ualberta.ca/1000_bull_genomes/reference_for_mapping/umd_3_1_reference_1000_bull_genomes.fa.gz) por meio do algoritmo BWA-MEM (LI; DURBIN, 2009) (Breve pipeline descrito no Apêndice C). As estatísticas dos alinhamentos foram realizadas por meio da ferramenta samtools flagstat (LI et al., 2009; LI, 2011).

2.3 Detecção das variações no número de cópias

A detecção das variações no número de cópias foi realizada para cada animal por meio da metodologia “read depth” (RD) implementada pelo programa computacional CNVnator (ABYZOV et al, 2011). Essa metodologia consiste em detectar as CNVs por meio da da densidade das “reads” alinhadas ao longo dos cromossomos. A estimativa de densidade de Kernel (WAND; JONES, 1995; COMANICIU; MEER, 2002), método estatístico não-paramétrico, foi utilizado para estimar a função de densidade das “reads”.

Basicamente, a metodologia consiste na comparação entre a profundidade do alinhamento das leituras entre janelas de tamanho pré-definidos (“bin size”) por meio de um valor médio central. O tamanho da janela é um limite teórico pré-definido para o ponto de quebra (“breakpoints”). Foi utilizado janelas de 250 bp e 500 bp de acordo com as recomendações dos autores (ABYZOV et al, 2011). Foram considerados nas análises somente os cromossomos autossômicos.

Para redução de resultados falsos-positivos foi realizado o controle de qualidade das CNVs detectadas. Foram consideradas apenas CNVs significativas ($p < 0,01$) para o teste estatístico t, em que a hipótese testada é se a média de sinal de profundidade das “reads” é a mesma da média empírica Também foram excluídas CNVs com fração de “reads” mapeadas com baixa qualidade menores do que 0,5 ($q0 < 0,5$) e CNVs com tamanhos maiores do que 1 quilobase (kb).

Após o controle de qualidade, o número de cópias presente em cada região foi obtido pela função *genotype* do programa CNVnator.

2.4 Diferenciação populacional por meio do número de cópias

Para verificar a estratificação das CNVs e dos genes entre os indivíduos da raça Girolando, Gir e Holandês foi utilizada a estatística denominada de V_{ST} , proposta por Redon et al. (2006). Essa estatística é equivalente ao índice de fixação (F_{ST}), estimativa de divergência genética entre subpopulações e a população total, descrito por Wright (1965), utilizada para verificar a diferenciação genética populacional, por meio das frequências alélicas.

A estatística de V_{st} foi calculada utilizando a média do número de cópias presentes nas regiões em sobreposição aos genes anotados no genoma bovino, considerando as análises aos pares das populações, conforme a equação descrita abaixo:

$$V_{ST} = \frac{V_t - V_s}{V_t}$$

Em que V_t é a variância entre os números de cópias normalizado entre todos os indivíduos e V_s é a variância media dentro de cada população ponderada pelo tamanho amostral de cada população. Quanto maior o valor de V_{ST}, maior o nível de diferenciação populacional, devido a variação no número de cópias em determinada região.

3 RESULTADOS E DISCUSSÃO

3.1. Detecção de CNVs

A cobertura média do sequenciamento, ou seja, o número de vezes que o nucleotídeo foi confirmado por leituras idênticas variou de 7,20 (Gir) à 16,46 (Girolando) (Tabela 1), permitindo poder estatístico suficiente para detecção de CNVs, uma vez que estudos anteriores relataram que 4x de cobertura já é suficiente para detectar CNVs por meio da metodologia de RD (SUDMANT et al., 2010; BICKHART et al., 2012). Na maioria dos touros, exceto para os touros E (Gir) e V (Holandês), foi

identificado maior número de deleções do que duplicações, o que já era esperado devido a sensibilidade do método para detectar duplicações (ABYZOV et al., 2011). Gao et al. (2017), em oito bovinos de leite da raça Holandês também identificaram maior número de deleções do que duplicações por meio do programa computacional CNVnator.

Tabela 1. Cobertura média das bases alinhadas (Cob.), número total de reads (Total de reads), porcentagem de reads alinhadas (map), número de deleções (DEL) e duplicações (DUP) identificadas em cada animal e tamanho médio (Tam) da variação no número de cópias detectada em quilobases(Kb).

Touro	Raça	Cob.	Total de reads	%map	Del	Dup	Tam
A	Girolando	10,97	298.378.898	99,07	1.309	237	27,61
B	Girolando	12,75	346.499.817	99,05	1.208	243	17,00
C	Girolando	16,46	448.340.078	98,90	5.278	4.410	45,00
D	Gir	15,01	408.506.315	98,98	1.192	287	26,31
E	Gir	15,49	421.943.648	98,75	8.162	11.105	87,60
F	Gir	15,10	371.824.564	99,75	3.206	556	17,65
G	Gir	15,00	362.976.826	99,76	3.667	544	20,06
H	Gir	15,80	385.658.481	99,75	3.649	588	17,26’
I	Gir	15,50	375.117.594	99,76	3.538	611	17,62
J	Gir	14,20	344.700.983	99,75	3.171	616	19,51
K	Gir	13,90	339.612.436	99,75	3.181	560	21,63
L	Gir	12,60	306.739.343	99,77	3.147	522	18,97
M	Gir	13,90	345.914.070	99,73	3.003	626	15,08
N	Gir	7,20	172.947.375	96,31	1.053	275	50,02
O	Gir	10,70	263.444.405	99,77	2.961	526	22,17
P	Gir	12,30	299.832.543	99,77	3.419	518	21,23
Q	Gir	10,20	260.469.051	99,60	774	374	38,66
R	Holandês	12,46	338.661.926	99,02	1.745	377	37,77
S	Holandês	12,05	327.629.321	99,71	1.121	238	65,25
T	Holandês	14,02	381.212.025	99,68	1.090	222	50,20
U	Holandês	10,76	292.493.299	99,71	1.205	231	58,52
V	Holandês	11,21	310.129.677	89,61	341	565	403,70

O número de CNVs detectadas nos touros Girolando (exceto touro C) foi similar ao número de CNVs identificadas na raça Holandês, porém na raça Gir foi identificado maior número de CNVs (Tabela 1). Esses resultados demonstraram que o alinhamento do genoma de animais de raça composta, como o Girolando, utilizando o genoma referência proveniente de taurinos permitiu uma detecção de CNVs confiável nesses animais. Bickhart et al. (2016) encontraram número semelhantes de CNVs entre zebuínos e taurinos em maior número de animais e diferentes raças (21

zebuínos e 54 taurinos). Estudos realizados anteriormente utilizando painéis de genotipagem para detecção encontraram maior número de CNVs em bovinos zebuínos em relação a taurinos (HOU et al., 2011; YANG et al., 2017).

Pode-se observar, duas amostras discrepantes na população em relação ao número de CNVs, sendo o Touro C da raça Girolando e o Touro E da raça Gir. Sugere-se que o elevado número de CNVs detectadas nesses indivíduos tenham ocorridos devido aos fatores laboratoriais que podem interferir na qualidade do sequenciamento, como por exemplo manipulação e qualidade do DNA e amplificação excessiva na etapa de montagem das bibliotecas, ocasionando resultados falsos-positivos. As CNVs detectadas individualmente nos animais foram agrupadas em regiões de CNVs (CNVR) considerando no mínimo um par de base em sobreposição para cada raça. Foram observadas 8.519 CNVRs únicas, ou seja, presente somente naquele individuo, para o Touro C, e 15.953 regiões para o Touro E, abrangendo 79% do total das regiões encontradas em ambas as raças (10.781 CNVRs na raça Girolando e 21.083 na raça Gir). Considerando que essas amostras eram heterogêneas em relação às demais e possivelmente a taxa de CNVs falsos-positivos era alta, as duas amostras foram removidas para as análises de diferenciação populacional.

Após a remoção das duas amostras (Touro C e Touro E), as CNVs foram novamente agrupadas em 2.505 (2.162 deleções, 318 duplicações e 25 mistas), 8.173 (7.203 deleções, 715 duplicações e 255 mistas) e 2.742 regiões (1.909 deleções, 672 duplicações e 161 mistas) para as raças Girolando, Gir e Holandês, respectivamente (Figura 1).

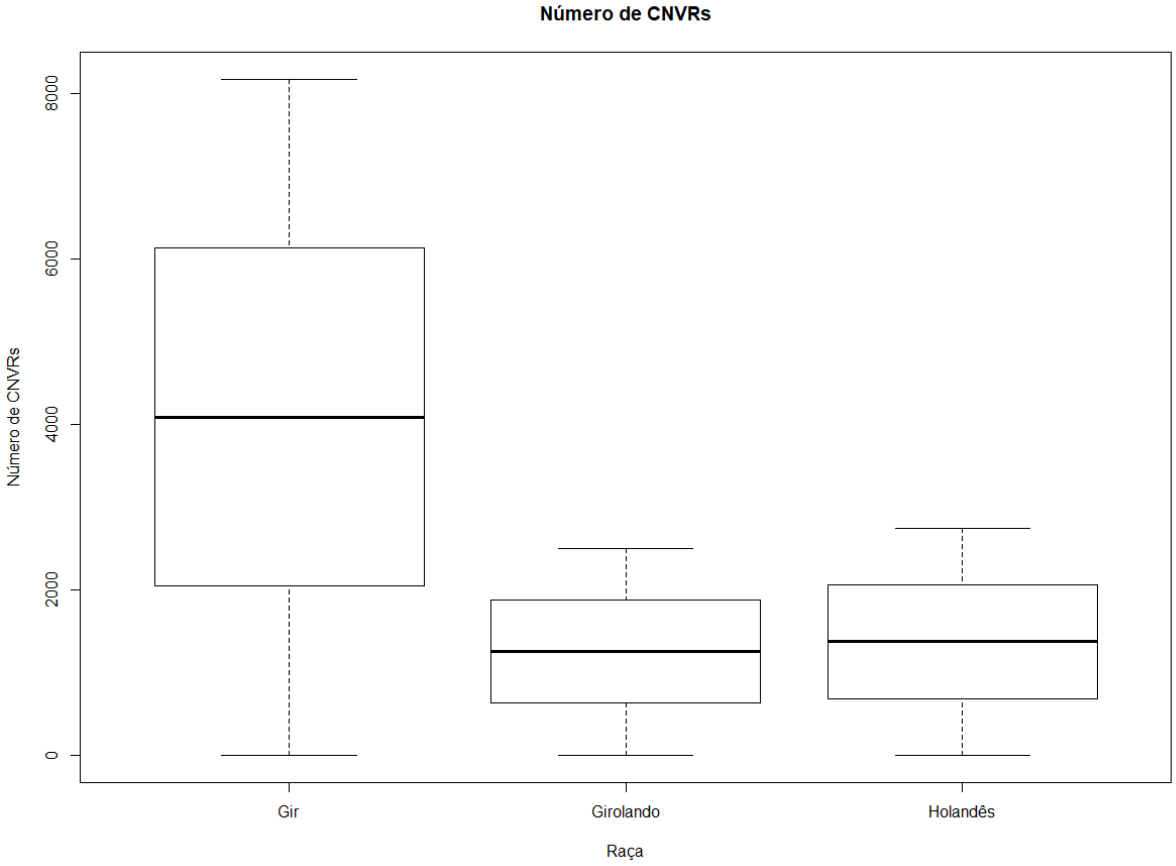


Figura 1. Número de regiões de variação no número de cópias (CNVR) dos segmentos de DNA em touros da raça Gir, Girolando e Holandês.

3.2 Grupos de CNVRs específicos entre raças

Considerando todas as raças como uma única população as CNVs foram agrupadas em 7.879 regiões (Figura 2). A frequência média de CNVRs foi de 24,01%, ou seja, em média no mínimo cinco animais apresentaram determinada CNVR.

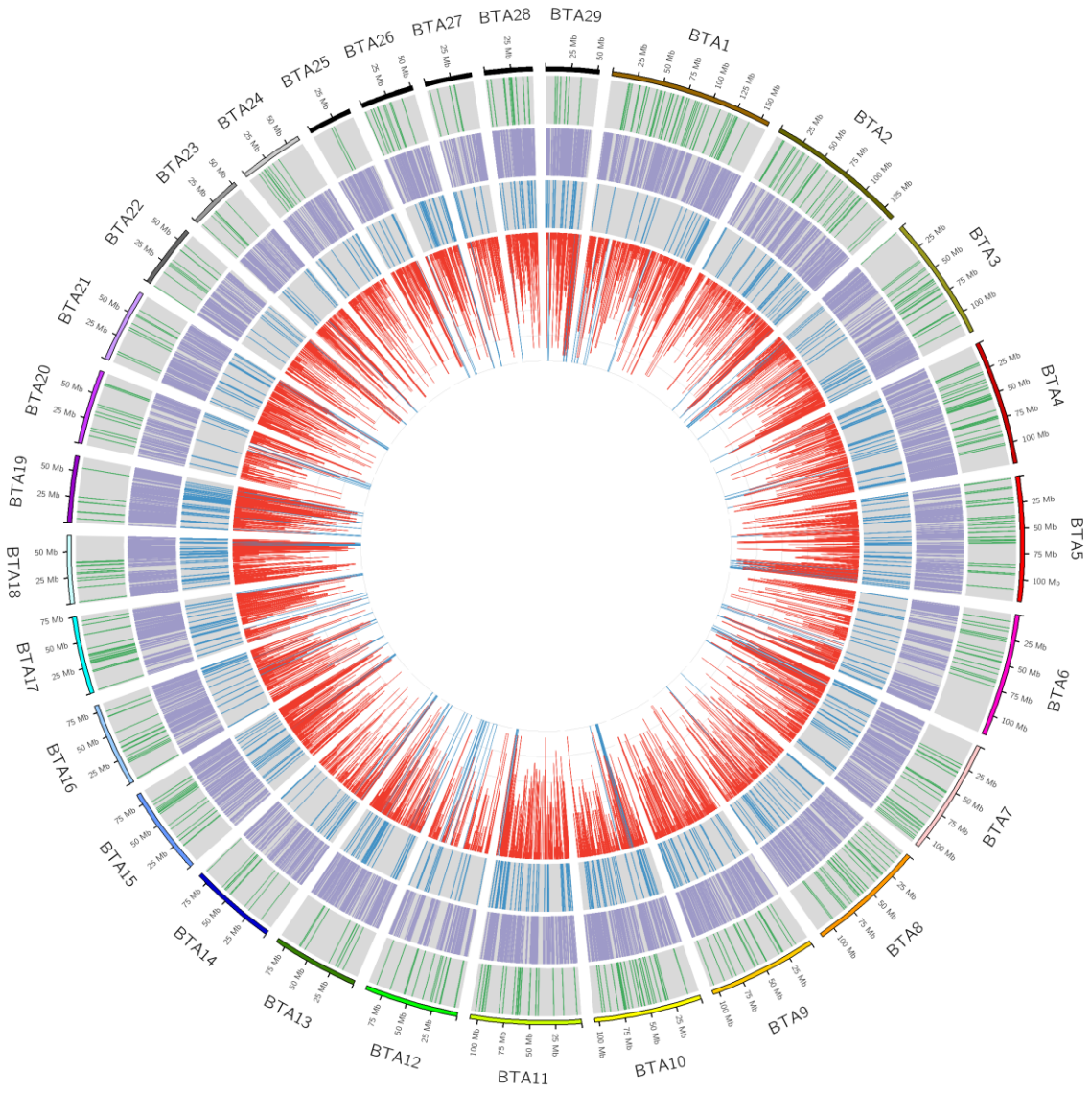


Figura 2. Mapa das regiões de variação no número de cópias (CNVRs) identificadas em bovinos da raça Girolando (em verde), Gir (em roxo), Holandês (em azul) e as CNVRs em toda a população, sendo em azul as regiões presentes em todas as amostras.

Foi observado quatro vezes mais CNVRs compartilhadas exclusivamente entre Girolando e Gir (318) do que entre Girolando e Holandês (75) (Figura 3). Sugere-se que as regiões compartilhadas entre os touros Girolando e as raças formadoras (Gir e Holandês) são provenientes de CNVs específicas de zebuínos e taurinos segregando em animais cruzados. Presumindo-se que a maioria das CNVs estão relacionadas às características de sanidade e reprodutivas (HENRICHSEN et al.,

2009, ZHANG et al. 2009), sendo zebuínos mais resistentes e adaptados ao clima tropical do que taurinos, justificou-se o maior número de CNVRs compartilhadas entre Girolando e Gir. Provavelmente, esse resultado se deve aos objetivos de seleção para as raças. A raça Girolando foi criada e tem sido selecionada para apresentar principalmente regiões do genoma provenientes do Gir associadas a resistência à parasitas, rusticidade e adaptabilidade ao clima tropical.

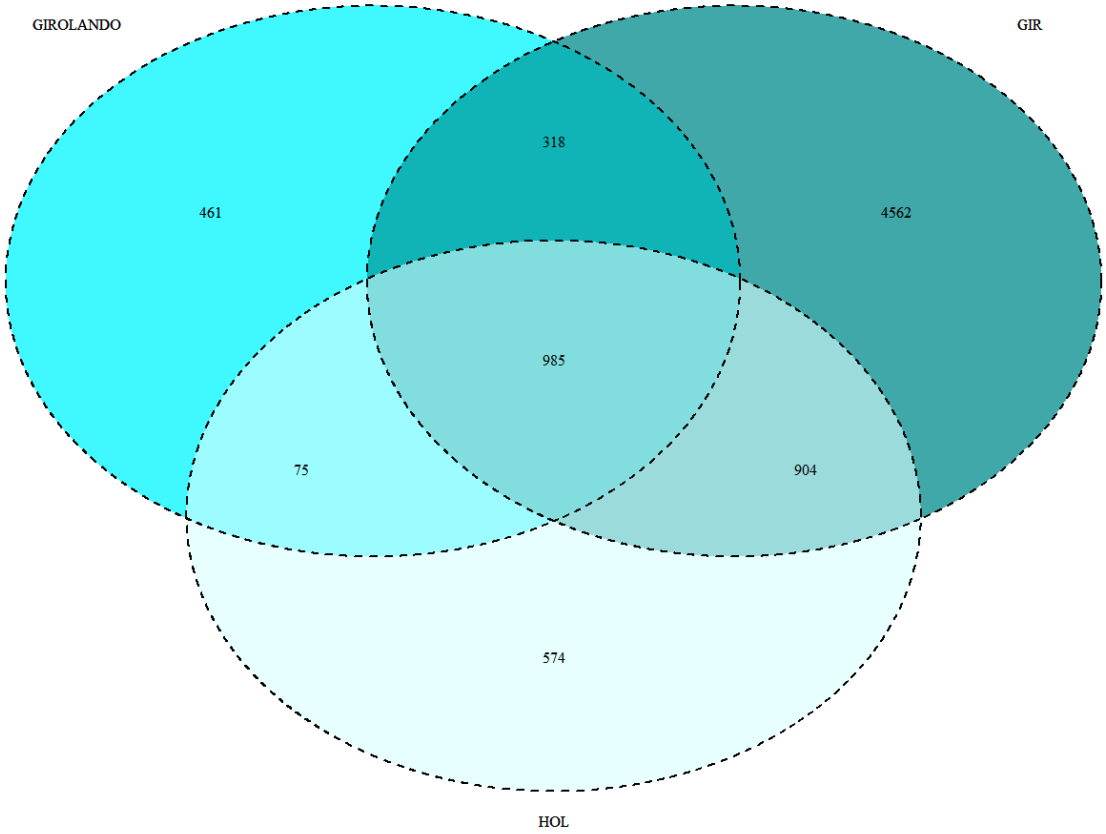


Figura 3. Número de regiões de variação no número de cópias compartilhadas entre touros da raça Girolando, Gir e Holandês (HOL).

Nas regiões específicas entre Gir e Girolando foram anotados 103 genes com base no banco de dados RefSeq Genes (Apêndice D) em sobreposição às CNVRs (20,75%). O gene *MEPCE ("Methylphosphate capping enzyme")* foi mapeado em uma CNVR, no BTA25, presente em 14 touros (70%) da raça Gir e Girolando com número de cópias médio de 1,24, ou seja, é uma região com alta confiabilidade de deleção específica de zebuínos. Esse gene pode estar associado a fertilidade de touros, uma vez que em estudo realizado por Kropp et al. (2017) embriões provenientes de touros

da raça Holandês considerados de alta fertilidade apresentaram esse gene mais expresso do que em embriões oriundos de touros de baixa fertilidade. Essa região de deleção no genoma de touros zebuínos e raças compostas pode estar afetando a fertilidade do rebanho, possivelmente influenciando na quantidade do produto do gene *MEPCE* e afetando a sua expressão.

O gene *HLX* ("*H2.0-like homeobox"*), anotado na região de CNV (deleção) no cromossomo BTA16, nos dois touros Girolando e em nove touros Gir (55% da população) pode estar relacionado a resistência a doenças, pois possui importante função na resposta do sistema imune inato em processos infecciosos. O *HLX* induz e interage com o fator de transcrição T-bet, fundamental para diferenciação das células do sistema imune do tipo Th1 (MULLEN et al., 2002). Células Th1 reconhecem antígenos de patógenos expressos na superfície das células infectadas e liberam citocinas para que os microorganismos sejam destruídos. Em ovinos esse gene foi diferencialmente expresso em ovelhas susceptíveis a helmintos gastrointestinais (PEMBERTON, et al. 2011)

Apenas 18,66% das CNVRs específicas entre Girolando e Holandes estavam em sobreposição em regiões de genes, compreendendo 20 genes anotados (Apêndice E). O gene *ASB3* ("*Ankyrin repeat and SOCS box protein 3"*), mapeado no cromossomo BTA11 em região de deleção, em três animais da raça Holandês e em um animal da raça Girolando, está envolvido nas respostas aos processos inflamatórios. A regulação desse gene pode estar relacionada à fertilidade de vacas da raça Holandesa (MOORE et al., 2016). Em humanos, disfunções no *ASB3* como mutações e inibição estão associadas com patogênese e progressão do câncer colorretal (DU et al., 2017). O micro RNA *MIR-455*, anotado no cromossomo BTA8, na CNVR presente em quatro touros Holandês e um da raça Girolando pode estar relacionado a susceptibilidade a doenças. A regulação desse microRNA está associada com a resposta ao vírus da febre aftosa e pode auxiliar no diagnóstico de animais infectados (STENFELDT et al., 2017).

3.3 Diferenciação populacional por meio dos genes anotados nas CNVs

Os valores de V_{ST} para a média do número de cópias dos genes anotados no genoma bovino variaram de -0,37 a 0,98 (Figura 4), quanto maior o V_{ST} maior a diferença genética entre as populações (Girolando, Gir e Holandês), provavelmente originárias de pressões de seleção diferentes para cada raça, indicando possíveis assinaturas de seleção. Nesse estudo foram encontrados 70 genes com V_{ST} maior do que 0,50 (Apêndice F). Esses resultados indicaram elevada diferenciação entre as raças por meio da variação do número de cópias. Xu et al. (2016) encontraram 78 CNVs de linhagem específicas (V_{ST} > 0,40) de taurinos e zebuínos.

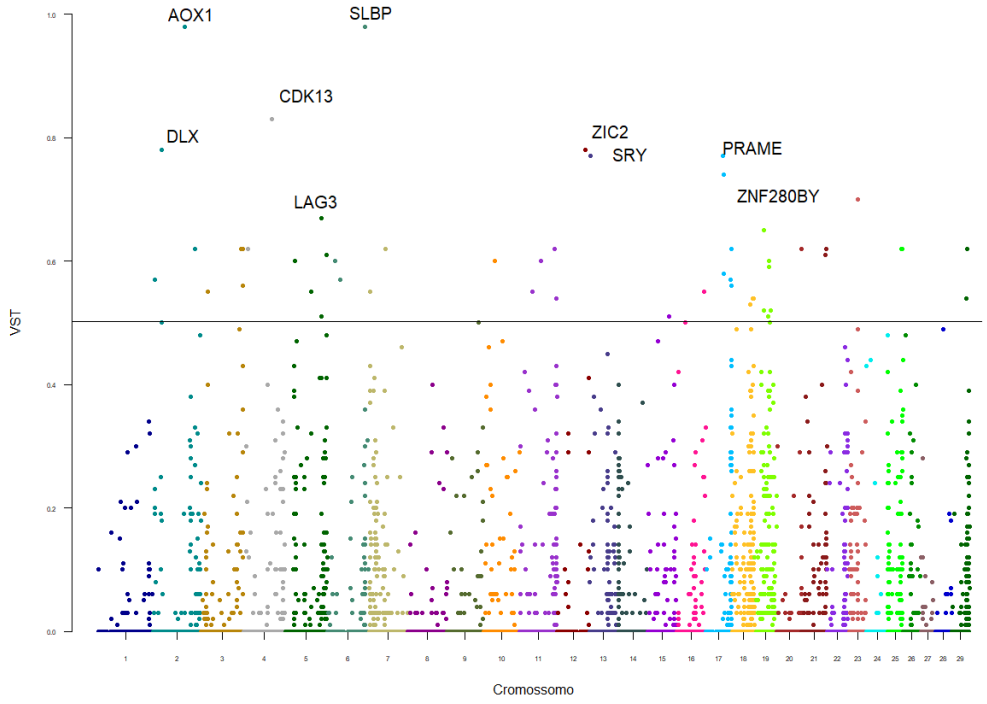


Figura 4. Diferenciação populacional entre bovinos da raça Girolando, Gir e Holandês para a média no número de cópias das regiões dos genes anotados pelo RefSeq estimadas pela estatística V_{ST} ao longo de cada cromossomo.

A CNVR (BTA2:89572251-89597250) próxima ao gene *AOX1* (“*Aldehyde Oxidase 1*”), apresentou um dos maiores V_{ST} (0,98), sendo que os touros Girolando apresentaram três cópias dessa região, touros Gir quatro cópias e os touros da raça Holandês não apresentaram variação no número de cópias. A presença de três cópias

dessa região em bovinos Girolando, sugere uma potencial região herdável de CNVs proveniente exclusivamente de zebuínos. O *AOX1* pertence à família dos aldeídos oxidase, sendo responsável pela produção do peróxido de hidrogênio e envolvido no metabolismo de fármacos e detoxificação (eliminação de toxinas das células) (BICKHART et al., 2016). Na evolução dos mamíferos, genes da família dos aldeídos estão associados primeiramente aos eventos de multiplicação desses genes e subsequente a fase de deleção ou supressão dos mesmos, de acordo com as necessidades especificas de cada espécie (GARATTINI; FRATELLI; TERAO, 2009). A duplicação na região do gene *AOX1* em zebuínos pode indicar divergências que ocorreram nos processos de domesticação dessas subespécies, como por exemplo região demográfica. Essa região específica de CNVs em zebuínos já foi confirmada em outros estudos em animais taurinos e zebuínos (HOU et al.,2012; BICKHART et al., 2012; BICKHART et al., 2016).

Os genes *SLBP* (*Stem-Loop Binding Protein*) e *TACC3* (*Transforming Acidic Coiled-Coil Containing Protein 3*), no BTA6, assim como o *AOX1,* também apresentaram V_{ST} igual a 0,98. Entretanto, apresentaram em média 1,255 cópias somente em touros da raça Holandês. Esses genes estão envolvidos nos processos de formação e crescimentos dos oócitos durante a meiose (BESSA et al., 2013; MAHDIPOUR et al., 2015), consequentemente, possuem importante função para reprodução dos bovinos. Em bovinos de corte em uma população cruzada SNPs localizados próximos a esses genes foram associados significativamente ($P < 0.0001$) com eficiência alimentar (SERÃO et al., 2013). Devido a serem em deleções relativamente grandes e possivelmente específicas de taurinos com genes de importância para reprodução e produção recomenda-se novos estudos para melhor investigação dessa região.

O gene *PRAME* (*Preferentially Expressed Antigen In Melanoma*) (V_{ST} = 0,77), foi mapeado no BTA17 dentro de uma região com 100% em sobreposição (BTA17:51252501-51660000) ao gene, com média igual a 14,51 cópias por animal somente nos animais da raça Gir. Esse gene está envolvido em processos biológicos de reprodução e imunidade e são expressos predominantemente nos testículos (CHANG et al., 2011). Yue et al. (2013) associaram negativamente o número de cópias do gene *PRAMEY* (gene *PRAME* especifico do Y) com a fertilidade de touros,

ou seja, quanto maior o CN pior é a fertilidade do touro. Esses mesmos autores também encontraram maior CN para zebuínos do que taurinos. Possivelmente, essa CNVR é especifica de zebuínos, pressupondo que em evolução, grupos de genes envolvidos nos processos reprodutivos, como *PRAME*, estão sob pressão de seleção positiva, uma vez que esses genes podem contribuir para a expansão independente e adaptação das linhagens específicas (TIAN; PASCAL; MONGET, 2009). Visando a melhoria reprodutiva dos rebanhos, sugere-se que essa CNVR seja monitorada em zebuínos e raças compostas, e no futuro poderá ser incluída em painéis de SNPs com a finalidade de auxiliar na seleção de melhores indivíduos para características reprodutivas.

Maior número de cópias (em média 41,71 cópias por animal) foi observado na região (BTA9: 91182751-91190500), com tamanho de 7,749 kb, presente em dez touros da raça Gir. Esta região está localizada próxima ao gene *RGS17* (*regulator of G protein signaling 17*), que apresentou valor de V_{st} igual a 0,5. Esse gene está envolvido no processo de regulação por sinalização da proteína G (SUURVÄLI, et al. 2013). Em humanos, mutações em genes envolvidos nas vias biológicas da proteína G estão associadas a doenças, como distúrbios no sistema endócrino (LANIA; MANTOVANI; SPADA, 2006) e diversos tipos de câncer, como de pulmão, próstata e bexiga (JAMES et al., 2009; LEE et al., 2013). Em bovinos da raça Angus, SNPs localizados próximos ao gene *RGS17*, foram associados a contagem de ovos fecal de nematoides gastrintestinais (KIM et al., 2015). Neste contexto, o elevado número de cópias da região próxima ao gene *RGS17*, identificado nos bovinos da raça Gir, pode indicar evidências de seleção positiva para animais mais resistentes e adaptados ao clima tropical.

A diferenciação populacional entre as raças Gir, Girolando e Holandês por meio da variação do número de cópias próximas aos genes relacionados à processos biológicos importantes, como reprodução e sanidade, demonstraram que CNVs desempenham importante função na divergência genética entre as raças, principalmente em processos adaptativos. A identificação de regiões de CNVs com elevado V_{ST}, próximos à genes relacionados à reprodução, podem indicar isolamento reprodutivo das raças taurinas e zebuínas durante os processos de domesticação.

4 CONCLUSÃO

As regiões compartilhadas somente entre os touros Gir e Girolando estão próximas a genes relacionados à fertilidade e à resistência às doenças, sendo CNVs extremamente importantes para características reprodutivas e de sanidade.

Genes, tais como *AOX1*, apresentaram números de cópias com elevado nível de diferenciação entre bovinos zebuínos e taurinos. Esta divergência pode ter sido originária na domesticação e no isolamento geográfico dos bovinos e, consequentemente, contribuiu para as diferenças fenotípicas entre as subespécies. As regiões de CNVs localizadas próximas aos genes com elevado nível de diferenciação populacional, indicaram possíveis evidências de seleção positiva para características de rusticidade e de adaptação aos ambientes de clima tropical, no genoma de animais da raça Gir e Girolando.

5 REFERÊNCIAS

ABYZOV, A.; URBAN, A. E.; SNYDER, M.; GERSTEIN, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. **Genome Research**, v. 21, p. 974-984, 2011. Disponível em: < http://doi.org/10.1101/gr.114876.110 >

BERGLUND, J.; NEVALAINEN, E. M.; MOLIN, A. M.; PERLOSKI, M.; THE LUPA CONSORTIUM; ANDRÉ, C.; ZODY, M.C.; SHARPE, T.; HITTE, C.; LINDBLAD-TOH, K.; LOHI, H.; WEBSTER, M. T. Novel origins of copy number variation in the dog genome, **Genome Biology**, v.13, R73, 2012. Disponível em: <http://doi.org/10.1186/gb-2012-13-8-r73>

BESSA, I.; NISHIMURA, R.; FRANCO, M.; DODE, M. Transcription Profile of Candidate Genes for the Acquisition of Competence During Oocyte Growth in Cattle. **Reproduction in Domestic Animals**, v. 48, p. 781-789, 2013. Disponível em: <http://doi.org/10.1111/rda.12162 >

BICKHART, D. M.; HOU, Y.; SCHROEDER, S. G.; ALKAN, C.; CARDONE, M. F.; MATUKUMALLI, L. K.; SONG, J. ; SCHNABEL, R. D. ; VENTURA, M.; TAYLOR, J. F. ; GARCIA, J. F.; VAN TASSELL, C. P. ; SONSTEGARD, T. S.; EICHLER, E. E.; LIU, G. E.; HUGHES, H. Copy number variation of individual cattle genomes using next-generation sequencing. **Genome Research**, v. 22, p. 778–790, 2012. Disponível em: < http://doi.org/10.1101/gr.133967.111>.

BICKHART; D. M.; XU, L.; HUTCHISON, J. L.; COLE, J. B.; NULL, D. J.; SCHROEDER, S. G.; SONG, J.; GARCIA, J. F.; SONSTEGARD, T. S.; VAN TASSELL, C. P.; SCHNABEL, R. D.; TAYLOR, J. F.; LEWIN, H. A.; LIU, G. E. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. **DNA Research**, v. 23, n. 3, p. 253–262, 2016. Disponível em: <http://doi.org/10.1093/dnares/dsw013>.

BRADLEY, D. G.; MACHUGH, D. E.; CUNNINGHAM, P.; LOFTUS, R. T. Mitochondrial diversity and the origins of African and European cattle. **Proceedings of the National Academy of Sciences**, v. 93, p. 5131-5135, 1996.

BUZANSKAS, M. E.; VENTURA, R. V.; CHUD, T. C. S.; BERNARDES, P. A.; SANTOS, D. J.; REGITANO, L. C.; ALENCAR, M. M.; MUDADU, M. A.; ZANELLA, R.; SILVA, M. V. G. B.; LI, C.; SCHENKEL, F. S.; MUNARI, D. P. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. **PLoS ONE**, v. 12, n. 2, e0171660, 2017. Disponível em: <https://doi.org/10.1371/journal.pone.0171660>

CAMPOS, B. M.; do CARMO, A. S.; do Egito, A. A.; Mariante, A. S.; Albuquerque, M. S. M., de Gouveia, J. J. S.; Malhado, C. H. M.; Verardo, L. L.; da Silva, M. V. G. B.; Carneiro, P. L. S. Genetic diversity, population structure, and correlations between locally adapted zebu and taurine breeds in Brazil using SNP markers. **Tropical Animal Health and Production**, v. 49, n. 8, p. 1677-1684, 2017. Disponível em:< https://doi.org/ 10.1007/s11250-017-1376-7>

CHANG, T.-C., YANG, Y.; YASUE, H.; BHARTI, A. K.; RETZEL, E. F.; LIU, W-S.The expansion of the PRAME gene family in Eutheria. **PLoS ONE,** v. 6, p. e16867, 2011. Disponível em:< https://doi.org/10.1371/journal.pone.0016867 >

COMANICIU, D; MEER, P. Mean shift: A robust approach toward featurespace analysis. **EEE Transactions on Pattern Analysis and Machine Intelligence,** v. 24, p.603-619, 2002.

DU, W.-Y.; LU, Z.-H.; YE, W.; FU, X.; ZHOU, Y.; KUANG, C.-M.; WU, J-X.;PAN, Z-Z.; CHEN, S.; LIU, R-Y.; Huang, W.-L. The loss-of-function mutations and down-regulated expression of ASB3 gene promote the growth and metastasis of colorectal cancer cells. **Chinese Journal of Cancer**, v. 36, p. 11, 2017. Disponível em:
< http://doi.org/10.1186/s40880-017-0180-0

GARATTINI, E.; FRATELLI, M.; TERAO, M. The mammalian aldehyde oxidase gene family. **Human Genomics**, v. 4, p. 119-130, 2009. Disponível em:
< http://doi.org/10.1186/1479-7364-4-2-119 >

GAO, Yahui; Jiang, J.; YANG, S.; HOU, Y.; LIU, G. E.; ZHANG, S.; ZHANG, Q.; SUN, D. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. **BMC Genomics**, v.18, p. 265, 2017. Disponível em:
< http://doi.org/10.1186/s12864-017-3636-3 >

HENRICHSEN, C.N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v.18, p.R1-8, 2009.Disponível em: < http://doi.org/ 10.1093/hmg/ddp011>

HIENDLEDER, S.; LEWALSKI, H.; JANKE, A. Complete mitochondrial genomes of Bos taurus and Bos indicus provide new insights into intra-species variation, taxonomy and domestication. **Cytogenetic and Genome Research**, v. 120, p. 150-156, 2008.

HOU, Y.; LIU, G. E.; BICKHART, D. M.; CARDONE, M. F.; WANG, K.; KIM, E., MATUKUMALLI, L. K.; VENTURA, M.; SONG, J.; VANRADEN, P. M.; SONSTEGARD, T. S.; VAN TASSELL, C. P. Genomic characteristics of cattle copy number variations. **BMC Genomics**, v. 12, p.127, 2011. Disponível em: <http://doi.org/10.1186/1471-2164-12-127>

HOU, Y.; BICKHART, D. M.; HVINDEN, M. L.; LI, C.; SONG, J.; BOICHARD, D. A.; FRITZ, S.; EGGEN, A.; DENISE, S.; WIGGANS, G. R.; SONSTEGARD, T.S.; VAN TASSEL, C. P.; LIU, G.E. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. **BMC Genomics**. v.13, p. 376, 2012.

INNAN, H.; KONDRASHOV, F. The evolution of gene duplications: classifying and distinguishing between models. **Nature Reviews Genetics,** v. 11, p. 97-108, 2010. Disponível em:< https://doi.org/doi: 10.1038/nrg2689.

ISKOW, R. C.; GOKCUMEN, O.; LEE, C. Exploring the role of copy number variants in human adaptation. **Trends in Genetics**., v. 28, p. 245–57, 2012. Disponível em: <https://doi.org/10.1016/j.tig.2012.03.002>

JAMES, M. A.; LU, Y.; LIU, Y.; VIKIS, H. G.; YOU, M. RGS17, an Overexpressed Gene in Human Lung and Prostate Cancer, Induces Tumor Cell Proliferation Through the Cyclic AMP-PKA-CREB Pathway. **Cancer Research**, v. 69, p. 2108-2116, 2009. Disponível em: <http://doi.org/10.1158/0008-5472.CAN-08-3495>

KIM, E. S.; SONSTEGARD, T. S.; SILVA, M. V. G. B.; GASBARRE, L. C.; TASSEL, C. P. V. Genome-Wide Scan of Gastrointestinal Nematode Resistance in Closed Angus Population Selected for Minimized Influence of MHC. **PLoS ONE**, v. 10, p. e0119380, 2015. Disponível em: < http://doi.org/10.1371/journal.pone.0119380 >

KROPP, J.; CARRILLO, J. A.; NAMOUS, H.; DANIELS, A.; SALIH, S. M.; SONG, J.; KHATIB, H. Male fertility status is associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine preimplantation embryos. **BMC Genomics,** v.18, p. 280, 2017. Disponível em: https://doi.org/10.1016/j.tig.2012.03.002

LANIA, A. G.; MANTOVANI, G.; SPADA, A. Mechanisms of disease: Mutations of G proteins and G-protein-coupled receptors in endocrine diseases. **Nature Clinical Practice. Endocrinology & Metabolism**, v. 12, p. 681–693, 2006. Disponível em: < https://doi.org/10.1038/ncpendmet0324 >

LARSON, G., PIPERNO, D. R., ALLABY, R. G., PURUGGANAN, M. D., ANDERSSON, L., ARROYO-KALIN, M., BARTON, L.; VIGUEIRA, C. C.; DENHAM, T.; DOBNEY, K.; DOUST, A. N.; GEPTS, P.; GILBERT, M. T. P.; GREMILLION, K. J.; LUCAS, L.; LUKENS, L.; MARSHALL, F. B.; OLSEN, K. M.; PIRES, J. C.; RICHERSON, P. J.; DE CASAS, R. R.;SANJUR, O. I. THOMAS, M. G.; FULLER, D. Q. Current perspectives and the future of domestication studies. **Proceedings of the National Academy of Sciences of the United States of America,** v. 111, n. 17, p. 6139-6146, 2014. Diponível em: < http://doi.org/10.1073/pnas.1323964111>

LEE, E. K.; YE, Y.; KAMAT, A. M.; WU, X. Genetic Variations in Regulator of G-Protein Signaling (RGS) Confer Risk of Bladder Cancer. **Cancer**, v. 119, p. 1643-1651, 2013. Disponível em: <http://doi.org/10.1002/cncr.27871>

Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, p. 1754-1760, 2009. Diponível em: < http://doi.org/10.1093/bioinformatics/btp324>

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R.; 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009. Disponível em: < http://doi.org/10.1093/bioinformatics/btp352>

LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, v. 27, p. 2987-2993, 2011. Disponível em:
< http://doi.org/10.1093/bioinformatics/btr509>

LOFTUS, R.T.; MACHUGH, D.E.; BRADLEY, D.G.; SHARP, P.M.; CUNNINGHAM, P. Evidence for two independent domestications of cattle. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, p. 2757-2761, 1994.

MACHUGH, D.E.; SHRIVER, M.D.; LOFTUS, R.T.; CUNNINGHAM, P.; BRADLEY, D.G. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos Taurus and Bos indicus), **Genetics**, v. 146, p. 1071-1086, 1997.

MAHDIPOUR, M.; LEITOGUINHO, A. R. C.; ZACARIAS SILVA, R. A.; VAN TOL, H. T. A.; STOUT, T. A. E.; RODRIGUES, G.; ROELEN, B. A. J. TACC3 is important for correct progression of meiosis in bovine oocytes. **PLoS ONE**, v. 10, e0132591. Disponível em:< http://doi.org/10.1371/journal.pone.0132591 >

MAKINA; S. O.; FARAI C. MUCHADEYI, F. C.; MARLE-KÖSTER, E. V.; MACNEIL, M. D.; MAIWASHE, A. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. **Frontiers in Genetics**, v. 5, p. 333, 2014. Disponível em:< http://doi.org/10.3389/fgene.2014.00333 >

MAPA. Ministério da Agricultura. Disponível em <http://www.agricultura.gov.br/animal/especies/bovinos-e-bubalinos>. Acesso em: 19/out./2017.

MILLS, R. E.; WALTER, K.; STEWART, C.; HANDSAKER, R. E.; CHEN, K.; ALKAN, C.; 1000 GENOMES PROJECT. Mapping copy number variation by population scale genome sequencing. **Nature**, v. 470, n. 7332, p. 59–65, 2011. Disponível em: <http://doi.org/10.1038/nature09708>.

MOORE, S. G.; PRYCE, J. E.; HAYES, B. J.; CHAMBERLAIN,A. J.; KEMPER, K. E.; BERRY, D. P.; MCCABE, M.; CORMICAN, P.; LONERGAN, P.; FAIR, T.; BUTLER, S. T. Differentially Expressed Genes in Endometrium and Corpus Luteum of Holstein Cows Selected for High and Low Fertility Are Enriched for Sequence Variants Associated with Fertility. **Biology of Reproduction**, v. 94, p. 1-11, 2016. Disponível em: < https://doi.org/10.1095/biolreprod.115.132951 >

MULLEN, A.C.; HUTCHINS, A.S.; HIGH, F.A.; LEE, H.W.; SYKES, K.J.; CHODOSH, L.A.; REINER, S.L. Hlx is induced by and genetically interacts with T-bet to promote heritable T(H)1 gene induction. **Nature Immunology.** v. 3, p. 652-658, 2002. Disponível em: <http://doi.org/10.1038/ni807>

PAUDEL, Y.; MADSEN, O.; MEGENS, H.J.; FRANTZ, L. A.; BOSSE, M.; BASTIAANSEN, J. W.; CROOIJMANS, R. P. M. A.; GROENEN, M. A. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. **BMC Genomics**, v. 14, p. 449, 2013. Disponível em: <http://doi.org/10.1186/1471-2164-14-449>.

PEMBERTON, J. M.; BERALDI, D.; CRAIG, B. H.; HOPKINS, J. Digital gene expression analysis of gastrointestinal helminth resistance in Scottish blackface lambs. **Molecular Ecology**, v. 20, p. 910-919, 2011. Disponível em: <http://doi.org/10.1111/j.1365-294X.2010.04992.x>

REDON, R.; ISHIKAWA, S., FITCH, K. R.; FEUK, L.; PERRY, G. H.; ANDREWS, T. D.; FIEGLER, H.; SHAPERO, M. H.; CARSON, A. R.; CHEN, W.; CHO, E. K.; DALLAIRE, S.; FREEMAN, J. L.; GONZÁLEZ, J. R; GRATACÓS, M.; HUANG, J.; KALAITZOPOULOS, D.; KOMURA, D.; MACDONALD, J. R.; MARSHALL, C. R.; MEI, R.; MONTGOMERY, L.; NISHIMURA, K.; OKAMURA, K.; SHEN, F.; SOMERVILLE, M. J.; TCHINDA, J.; VALSESIA, A.; WOODWARK, C.; YANG, F.; ZHANG, J.; ZERJAL, T.; ZHANG, J.; ARMENGOL, L. L.; CONRAD, D. F.; ESTIVILL, X.; TYLER-SMITH, C.; CARTER, N. P.; ABURATANI, H.; LEE, C.; JONES, K. W.; SCHERER, S. W.; HURLES, M. E. Global variation in copy number in the human genome. **Nature**, v. 444, p. 444-454, 2006. Disponível em: < https://www.doi.org/10.1038/nature05329>

SERÃO, N. V. L.; GONZÁLEZ-PEÑA, D.; BEEVER, J. E.; FAULKNER, D. B.; SOUTHEY, B. R.; RODRIGUEZ-ZAS S. L. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. **BMC Genetics**, v. 14, p. 94, 2013. Disponível em: < https://doi.org/10.1186/1471-2156-14-94 >

STENFELDT, C.; ARZT, J.; SMOLIGA, G.; LAROCCO, M.; GUTKOSKA, J.; LAWRENCE, P. Proof-of-concept study: profile of circulating microRNAs in Bovine serum harvested during acute and persistent FMDV infection. **Virology Journal**, v. 14, p. 71, 2017. Disponível em: < http://doi.org/10.1186/s12985-017-0743-3 >

SUDMANT, P. H.; KITZMAN, J. O.; ANTONACCI, F.; ALKAN, C.; MALIG, M.; TSALENKO, A.; SAMPAS, N.; BRUHN, L.; SHENDURE, J.; 1000 GENOMES PROJECT, EICHLER, E. E. Diversity of Human Copy Number Variation and Multicopy Genes. **Science**, v. 330, p. 641-646, 2010. Disponível em: < http://doi.org/10.1126/science.1197005 >.

SUDMANT, P. H.; RAUSCH, T.; GARDNER, E. J.; HANDSAKER, R. E.; ABYZOV, A.; HUDDLESTON, J.; ZHANG, Y.; YE, K.; JUN, G.; FRITZ, M. H.; KONKEL, M. K.; MALHOTRA, A.; STÜTZ, A. M.; SHI, X.; CASALE, F. P.; CHEN, J.; HORMOZDIARI, F.; DAYAMA, G.; CHEN, K.; MALIG, M.; CHAISSON, M. J. P.; WALTER, K.; MEIERS, S.; KASHIN, S.; GARRISON, E.; AUTON, A.; LAM, H. Y. K.; UM, X. J.; ALKAN, C.; ANTAKI, D.; BAE, T.; CERVEIRA, E.; CHINES, P.; CHONG, Z.; CLARKE, L.; DAL, E.; DING, L.; EMERY, S.; FAN, X.;MADHUSUDAN, G.; KAHVECI, F.; KIDD, J. M.; KONG, Y.; LAMEIJER, E. W.; MCCARTHY, S.; FLICEK, P.; GIBBS, R. A.; MARTH, G.; MASON, C. E.; MENELAOU, A.; MUZNY, D. M.; NELSON, B. J.; NOOR, A.; PARRISH, N. F.; PENDLETON,M.; QUITADAMO, A.; RAEDER, B.; SCHADT, E. E.; ROMANOVITCH, M.; SCHLATTL, A.; SEBRA,R.; SHABALIN, A. A.; UNTERGASSER, A.; WALKER, J. A.; WANG, M.; YU, F.; ZHANG, C.; ZHANG, J.; ZHENG-BRADLEY, X.; ZHOU, W.; ZICHNER, T.; SEBAT, J.; BATZER, M. A.; MCCARROLL, S. A.; THE 1000 GENOMES PROJECT CONSORTIUM, MILLS, R. E. GERSTEIN, M. B.; BASHIR, A.; STEGLE, O.; DEVINE, S. E.; LEE, C.; EICHLER, E. E.; KORBEL, J. O. An integrated map of structural variation in 2,504 human genomes. **Nature**, v. 526, p.75-81, 2015. Disponível em: < http://doi.org/10.1038/nature15394 >

SUURVÄLI, J.; ROBERT, J.; BOUDINOT, P.; BOUDINOT, S. R. R4 regulators of G protein signaling (RGS) identify an ancient MHC-linked synteny group. **Immunogenetics**, v. 65, p. 145-156, 2013. Disponível em:< http://doi.org/10.1007/s00251-012-0661-x >

TIAN, X.; PASCAL, G.; MONGET, P. Evolution and functional divergence of NLRP genes in mammalian reproductive systems. **BMC Evolutionary Biology**, v. 9, p. 202, 2009. Disponível em: < http://doi.org/10.1186/1471-2148-9-202 >

The Bovine HapMap Consortium. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. **Science**, v. 324, n. 5926, p. 528-532, 2009. Disponível em : <http://doi.org/10.1126/science.1167936>

WAND, M. P.; JONES, M.; C. **Kernel smothing**. New York: Chapimam & Hall. v. 60, 1995. 212p.

WRIGTH, S. The interpretation of population structure by F-statistics with special regard to systems of mating. **Evolution**, v.19, p. 395-420, 1965.

XU, L.; BICKHART, D. M; COLE, J. B.; SCHROEDER; S. G.; SONG, J.; VANTASSEL, C. P.; SONSTEGARD, T.; LIU, G. E. Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle. **Molecular Biology Evolution**, v. 32, n. 3, p. 711–725, 2015. Disponível em: < http://doi.org/10.1093/molbev/msu333>

XU, L.; HOU, Y.; BICKHART, D. M.; ZHOU, Y.; ABDEL HAY, E. H.; SONG, J.; SONSTEGARD, T. S.; VAN TASSELL, C. P.; LIU, G. E. Population-genetic properties of differentiated copy number variations in cattle. **Scientific Reports,** v. 6, Article number: 2316, 2016. Disponível em: < http://doi.org/ 10.1038/srep23161>

YANG L.; XU, L.; ZHU, B.; NIU, H.; ZHANG, W.; MIAO, J., SHI, X.; ZHANG, M.; CHEN, Y.; ZHANG, L.; GAO, X.; GAO, H.; LI, L.; LIU, G. E.; LI, J. Genome-wide analysis reveals differential selection involved with copy number variation in diverse Chinese Cattle. **Scientific Reports**, v. 7, p. 14299, 2017. Disponível em: < http://doi.org/10.1038/s41598-017-14768-0>.

YUE, X. P.; CHANG, T, C.; DEJARNETTE, J. M.; MARSHALL, C. E.; LEI, C. Z.; LIU, W-S. Copy number variation of PRAMEY across breeds and its association with male fertility in Holstein sires. **Journal of Dairy Science**, v. 96, p. 8024–8034, 2013.
Disponível em: < http://doi.org/10.3168/jds.2013-7037 >

ZHANG, F.; GU, W.; HURLES, M. E.; LUPSKI, J. R. Copy number variation in human health, disease, and evolution. **Annual Review of Genomics and Human Genetics**, v. 10, p. 451-481, 2009.
Disponível em: < https://doi.org/10.1146/annurev.genom.9.081307.164217>

ZHBANNIKOV, I.; HUNTER, S. S.; FOSTER, J. A.; SETTLES, M. L. SeqyClean: A Pipeline for High-throughput Sequence Data Preprocessing. In: ACM International Conference on Bioinformatics, Computational Biology, and Healthy Informatics, 8th.,2017, Boston, Massachusetts, USA. **Proceedings…** Boston:ACM, 2017. p. 407-416. Disponível em: <http://doi.org/10.1145/3107411.3107446>. Acesso em 10 nov. 2017.

APÊNDICES

Apêndice A. Descrição das regiões de variações no número de cópias (CNVR) candidatas em touros Girolando. Cromossomo (Chr), início e final, estado do evento por meio da detecção em sequenciamento de nova geração (NGS) e painel de genotipagem de média densidade (50K), frequência na população e os genes mapeados com base no genoma referencia UMD 3.1 na CNVR.

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_1	chr1	6721001	6754000	DUPLICAÇÃO	DUPLICAÇÃO	0.21	
CNVR_2	chr1	14653501	14659500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_3	chr1	17107001	17110500	DELEÇÃO	DELEÇÃO	0.21	
CNVR_4	chr1	38535501	38566500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_5	chr1	42379001	42382500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_6	chr1	52456501	52466500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_7	chr1	62880501	62894500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_8	chr1	93419501	93436500	DELEÇÃO	DUPLICAÇÃO	0.52	
CNVR_9	chr1	93922001	93930500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_10	chr1	94061501	94069500	DELEÇÃO	DELEÇÃO	0.52	
CNVR_11	chr1	104983001	104994500	DELEÇÃO	DUPLICAÇÃO	0.42	
CNVR_12	chr1	116663001	116672000	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_13	chr1	124383501	124421000	DELEÇÃO	DUPLICAÇÃO	0.31	
CNVR_14	chr2	8898501	8913500	DELEÇÃO	DELEÇÃO	0.31	*CALCRL*
CNVR_15	chr2	12146001	12165000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_16	chr2	12513001	12532500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_17	chr2	135126501	135147500	DELEÇÃO	DUPLICAÇÃO	0.83	
CNVR_18	chr3	637501	649000	DELEÇÃO	DELEÇÃO	0.10	*DCAF6*
CNVR_19	chr3	11991001	12022000	DUPLICAÇÃO	DELEÇÃO	0.10	*CD1A*
CNVR_20	chr3	12081001	12094500	DUPLICAÇÃO	DELEÇÃO	0.10	
CNVR_21	chr3	75976001	75986500	DELEÇÃO	DUPLICAÇÃO	0.83	
CNVR_22	chr3	118638001	118649500	DELEÇÃO	DELEÇÃO	0.73	*TWIST2*
CNVR_23	chr3	118727501	118779000	DUPLICAÇÃO	DELEÇÃO	0.10	
CNVR_24	chr3	118786501	118795500	DUPLICAÇÃO	DELEÇÃO	0.10	
CNVR_25	chr3	120666001	120739000	DELEÇÃO	MISTA	0.52	*SCMH1*
CNVR_26	chr4	1474001	1488500	DELEÇÃO	DELEÇÃO	0.10	

Apêndice A. Continuação...

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_27	chr4	84221001	84292000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_28	chr4	92420001	92461000	DELEÇÃO	DUPLICAÇÃO	5.31	*ENSBTAG00000020004*
CNVR_29	chr4	113368501	113381500	DELEÇÃO	DUPLICAÇÃO	0.10	*ENSBTAG00000010219*
CNVR_30	chr4	113431001	113474500	DELEÇÃO	DUPLICAÇÃO	0.10	*ENSBTAG00000008542, ZNF467*
CNVR_31	chr4	113648501	113685500	DUPLICAÇÃO	DUPLICAÇÃO	0.10	*GIMAP8*
CNVR_32	chr4	113719501	113799500	DUPLICAÇÃO	DUPLICAÇÃO	0.21	*GIMAP7, LOC511617*
CNVR_33	chr4	113864001	113884000	DUPLICAÇÃO	DUPLICAÇÃO	0.21	*GIMAP4*
CNVR_34	chr4	113923501	113937000	DUPLICAÇÃO	DUPLICAÇÃO	0.21	*ENSBTAG00000011240,GIMAP7*
CNVR_35	chr4	113941001	113961000	DUPLICAÇÃO	DUPLICAÇÃO	0.21	*GIMAP7,LOC512867*
CNVR_36	chr4	114024001	114042500	DUPLICAÇÃO	DUPLICAÇÃO	0.21	*ENSBTAG00000039928*
CNVR_37	chr4	119676001	119734500	DELEÇÃO	MISTA	0.42	
CNVR_38	chr4	119815501	119830000	DELEÇÃO	DELEÇÃO	0.42	
CNVR_39	chr5	58958501	59236000	DELEÇÃO	DUPLICAÇÃO	0.10	*OR6C1, ENSBTAG00000046645, LOC787945,LOC788027,LOC787991*
CNVR_40	chr5	59372501	59634500	DELEÇÃO	MISTA	0.83	*LOC788242, LOC788220, LOC788322,LOC788372,LOC788357,LOC788285, ENSBTAG00000031100, LOC788210, ENSBTAG00000046446,LOC788258*
CNVR_41	chr5	88139501	88169500	DELEÇÃO	DELEÇÃO	0.10	*C2CD5*
CNVR_42	chr5	107918001	107958500	DELEÇÃO	DUPLICAÇÃO	0.62	*B4GALNT3, NINJ2*
CNVR_43	chr5	114592501	114626500	DELEÇÃO	DELEÇÃO	0.10	*TTLL12, TSPO*
CNVR_44	chr5	117108501	117151000	DELEÇÃO	DELEÇÃO	0.10	*MIRLET7A-3, MIRLET7B, MIR3596, MIR2443*
CNVR_45	chr5	117942001	118022500	DELEÇÃO	DUPLICAÇÃO	0.21	*GRAMD4*
CNVR_46	chr6	1798501	1814000	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_47	chr6	3822501	3830000	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_48	chr6	38832001	38845500	DELEÇÃO	DELEÇÃO	1.04	*LCORL*
CNVR_49	chr6	48543501	48553500	DELEÇÃO	DELEÇÃO	0.73	
CNVR_50	chr6	53686501	53704000	DELEÇÃO	DELEÇÃO	4.16	
CNVR_51	chr6	80981001	80985000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_52	chr6	108360501	108433500	DELEÇÃO	MISTA	0.52	*ZFYVE28, ENSBTAG00000047977*
CNVR_53	chr6	108483001	108509500	DELEÇÃO	MISTA	0.42	*MXD4, HAUS3*
CNVR_54	chr6	109586501	109594500	DUPLICAÇÃO	DUPLICAÇÃO	0.31	

Apêndice A. Continuação...

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_55	chr7	6695501	7796000	DELEÇÃO	DUPLICAÇÃO	2.08	*ENSBTAG00000047589*
CNVR_56	chr7	10829001	11135000	DELEÇÃO	DUPLICAÇÃO	0.10	*LOC100847301, LOC788246,LOC786119,LOC789504,LOC512296,OR7A17*
CNVR_57	chr7	42729001	42737500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_58	chr7	85351001	85359500	DELEÇÃO	DELEÇÃO	0.62	*XRCC4*
CNVR_59	chr7	87613501	87634500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_60	chr8	2935001	2947000	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_61	chr8	32959001	32986500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_62	chr8	93794501	93830000	DELEÇÃO	DELEÇÃO	0.21	
CNVR_63	chr8	113004001	113035000	DELEÇÃO	DUPLICAÇÃO	0.42	*TSSC1*
CNVR_64	chr8	113036501	113127000	DELEÇÃO	DUPLICAÇÃO	0.21	*TSSC1*
CNVR_65	chr9	1122001	1132000	DELEÇÃO	MISTA	0.42	
CNVR_66	chr9	1189501	1195000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_67	chr9	1641501	1645500	DELEÇÃO	DELEÇÃO	0.42	
CNVR_68	chr9	9790001	9800500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_69	chr9	31092001	31106500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_70	chr9	105262001	105351500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_71	chr10	39823001	39828000	DELEÇÃO	DELEÇÃO	0.83	
CNVR_72	chr10	39834501	39838500	DELEÇÃO	DELEÇÃO	0.83	
CNVR_73	chr11	17355501	17395500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_74	chr11	103828001	103842500	DELEÇÃO	DELEÇÃO	0.31	*GPSM1*
CNVR_75	chr11	103909501	104039000	DELEÇÃO	MISTA	0.62	*SDCCAG3, PMPCA, INPP5E, ENSBTAG00000022799, SEC16A*
CNVR_76	chr11	104171501	104222000	DELEÇÃO	DUPLICAÇÃO	0.10	*FAM69B, ENSBTAG00000030281, ENSBTAG00000042482, ENSBTAG00000042930*
CNVR_77	chr11	104488001	104898000	DELEÇÃO	DELEÇÃO	0.10	*FAM163B, DBH, SARDH, ENSBTAG00000004511, ADAMTSL2, BRD3*
CNVR_78	chr11	105732001	105763500	DELEÇÃO	MISTA	4.16	*ENSBTAG00000023788*
CNVR_79	chr11	105839001	106118500	DELEÇÃO	MISTA	3.43	*RXRA, TOR4A, TUBB4B, CYSRT1, TMEM203, SSNA1, RNF224, NDOR1, GRIN1, RNF208, TMEM210*
CNVR_80	chr11	106191001	106224500	DELEÇÃO	DELEÇÃO	0.52	*FUT7, ABCA2*

Apêndice A. Continuação...

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_81	chr11	106468001	106539000	DELEÇÃO	DELEÇÃO	0.31	
CNVR_82	chr11	106572001	106584500	DELEÇÃO	DELEÇÃO	0.31	
CNVR_83	chr11	106604501	106752500	DELEÇÃO	DELEÇÃO	0.31	*OLFM1*
CNVR_84	chr11	106758501	106779500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_85	chr11	106782001	106817500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_86	chr12	745001	781000	DELEÇÃO	DELEÇÃO	0.21	
CNVR_87	chr12	4268001	4312500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_88	chr12	4477001	4485500	DELEÇÃO	DELEÇÃO	0.31	
CNVR_89	chr12	52615001	52623000	DELEÇÃO	DELEÇÃO	0.21	*MYCBP2*
CNVR_90	chr12	57569001	57577500	DELEÇÃO	DUPLICAÇÃO	0.52	
CNVR_91	chr12	65256001	65291000	DELEÇÃO	DELEÇÃO	0.21	
CNVR_92	chr12	65318001	65421000	DELEÇÃO	DELEÇÃO	0.21	
CNVR_93	chr12	72424001	72926500	MISTA	DELEÇÃO	1.14	*ENSBTAG00000047181*
CNVR_94	chr12	73484501	73717000	DELEÇÃO	DUPLICAÇÃO	0.21	*LOC100299180*
CNVR_95	chr12	73741001	73820500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_96	chr12	73852001	74253000	MISTA	DUPLICAÇÃO	0.21	*ENSBTAG00000047360, ENSBTAG00000047764*
CNVR_97	chr12	74277001	74314500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_98	chr12	74321001	74339500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_99	chr12	74374501	74448000	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_100	chr12	74467501	74590000	DUPLICAÇÃO	DUPLICAÇÃO	0.21	
CNVR_101	chr12	74633501	74651500	DUPLICAÇÃO	DUPLICAÇÃO	0.21	
CNVR_102	chr12	74652001	74661500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_103	chr12	74682001	74738500	MISTA	DUPLICAÇÃO	0.21	
CNVR_104	chr12	74739501	74747500	DUPLICAÇÃO	DUPLICAÇÃO	0.21	
CNVR_105	chr12	74760001	75139500	DELEÇÃO	DUPLICAÇÃO	0.21	*LOC509854*
CNVR_106	chr12	75150001	75177000	DUPLICAÇÃO	DUPLICAÇÃO	0.21	
CNVR_107	chr12	75177001	75202000	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_108	chr12	75223501	75238000	DELEÇÃO	DUPLICAÇÃO	0.21	

Apêndice A. Continuação...

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_109	chr12	90729501	90800500	DELEÇÃO	DELEÇÃO	0.73	*TFDP1, ATP4B, GRK1, ENSBTAG00000019965*
CNVR_110	chr12	90888501	90945500	DELEÇÃO	DELEÇÃO	0.10	*RASA3, MGC134473*
CNVR_111	chr13	16983501	16987500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_112	chr14	2870501	2921500	DELEÇÃO	DELEÇÃO	0.21	*ARC, ENSBTAG00000026340*
CNVR_113	chr14	33450001	33458500	DELEÇÃO	DUPLICAÇÃO	0.10	*ARFGEF1*
CNVR_114	chr14	66977001	67028000	DELEÇÃO	DELEÇÃO	0.10	*ENSBTAG00000012357, MIR599, MIR875*
CNVR_115	chr15	11585501	11588500	DELEÇÃO	DELEÇÃO	0.21	
CNVR_116	chr15	56681501	56689500	DELEÇÃO	DELEÇÃO	0.21	*EMSY*
CNVR_117	chr15	70789501	70800500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_118	chr15	81420501	81445500	DUPLICAÇÃO	DUPLICAÇÃO	0.10	
CNVR_119	chr16	46553501	46570500	DELEÇÃO	DELEÇÃO	0.10	*ENSBTAG00000014602*
CNVR_120	chr16	49419001	49508000	DELEÇÃO	DUPLICAÇÃO	0.73	*ENSBTAG00000021919, ENSBTAG00000046062*
CNVR_121	chr16	50702501	50777500	DELEÇÃO	DELEÇÃO	0.10	*ENSBTAG00000020839, MIR551A*
CNVR_122	chr16	70931001	70962500	DELEÇÃO	DELEÇÃO	0.31	*LGR6*
CNVR_123	chr17	6712501	6738000	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_124	chr17	35037501	35059000	DELEÇÃO	DELEÇÃO	0.21	*ENSBTAG00000000125*
CNVR_125	chr17	36332501	36337000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_126	chr17	42804501	42813000	DELEÇÃO	DELEÇÃO	0.10	*GRIA2*
CNVR_127	chr17	63654001	63673000	DUPLICAÇÃO	DELEÇÃO	0.10	*OAS1X, OAS1Z*
CNVR_128	chr17	73089001	73099000	DELEÇÃO	DELEÇÃO	0.10	*LOC100297192, IGLL1*
CNVR_129	chr17	73156501	73163000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_130	chr17	73198501	73226500	DELEÇÃO	DELEÇÃO	0.10	*ENSBTAG00000005128, MMP11,SMARCB1,ENSBTAG00000022030,CHCHD10*
CNVR_131	chr17	73693001	73749500	DELEÇÃO	DELEÇÃO	0.21	*ENSBTAG00000020566*
CNVR_132	chr17	73750501	73774500	DELEÇÃO	DUPLICAÇÃO	0.21	*ENSBTAG00000020566*
CNVR_133	chr17	74170001	74199000	DELEÇÃO	DELEÇÃO	0.62	*HIC2, ENSBTAG00000030927*
CNVR_134	chr17	74627001	74642500	DELEÇÃO	MISTA	0.31	*SLC25A1*
CNVR_135	chr18	11288501	11433000	DELEÇÃO	MISTA	0.62	
CNVR_136	chr18	11469001	11654000	DELEÇÃO	MISTA	0.62	

Apêndice A. Continuação...

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_137	chr18	14520001	14528500	DELEÇÃO	DUPLICAÇÃO	0.21	*SPG7*
CNVR_138	chr18	14530001	14574000	DELEÇÃO	DUPLICAÇÃO	0.21	*RPL13, ENSBTAG00000012215,ENSBTAG00000042786*
CNVR_139	chr18	63186001	63258500	MISTA	DUPLICAÇÃO	0.10	*ENSBTAG00000000930,ENSBTAG00000026944,ENSBTAG00000037830, ENSBTAG00000039892,ENSBTAG00000039932,ENSBTAG00000048145*
CNVR_140	chr18	63284501	63344000	DELEÇÃO	DUPLICAÇÃO	0.10	*ENSBTAG00000039691*
CNVR_141	chr19	1851001	1870500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_142	chr19	1937501	1941000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_143	chr19	2227001	2229500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_144	chr19	2425001	2444500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_145	chr19	9340501	9347000	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_146	chr19	11860501	11893500	DELEÇÃO	DELEÇÃO	0.10	*TBX4*
CNVR_147	chr19	11932001	11975000	DELEÇÃO	DELEÇÃO	0.10	*TBX2*
CNVR_148	chr19	19905001	19921000	DUPLICAÇÃO	DELEÇÃO	0.10	*NOS2,ENSBTAG00000023970*
CNVR_149	chr19	19922001	19949500	DUPLICAÇÃO	DELEÇÃO	0.10	*ULBP11, NOS2*
CNVR_150	chr19	35210501	35359500	DELEÇÃO	DUPLICAÇÃO	0.10	*TOM1L2,RAI1,SREBF1,MIR33B*
CNVR_151	chr19	54354001	54376500	DELEÇÃO	DELEÇÃO	0.10	*DNAH17*
CNVR_152	chr19	54603501	54667500	DELEÇÃO	DELEÇÃO	0.10	*TK1,TMC6,TMC8,SYNGR2,MIR2348*
CNVR_153	chr19	56614501	56703500	DELEÇÃO	DELEÇÃO	0.52	*TSEN54,TMEM94,CASKIN2,LLGL2*
CNVR_154	chr20	37273501	37300000	DELEÇÃO	DELEÇÃO	0.10	*NIPBL*
CNVR_155	chr20	51959001	51976000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_156	chr20	52047001	52062500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_157	chr20	55672501	55696000	DELEÇÃO	DELEÇÃO	0.62	
CNVR_158	chr20	71278001	71337500	DELEÇÃO	DELEÇÃO	0.52	*TERT, SLC6A18*
CNVR_159	chr20	71394501	71448000	DELEÇÃO	DELEÇÃO	0.52	*ENSBTAG00000006775,ENSBTAG00000034435*
CNVR_160	chr21	28389501	28407500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_161	chr21	41800001	41808000	DELEÇÃO	DUPLICAÇÃO	0.42	*SCFD1*
CNVR_162	chr21	53299001	53307000	DELEÇÃO	DUPLICAÇÃO	0.10	

CNVR_163	chr21	53514501	53544500	DELEÇÃO	DELEÇÃO	0.21	
CNVR_164	chr21	66169501	66298000	DELEÇÃO	DUPLICAÇÃO	0.10	*CCNK,CCDC85C,HHIPL1*

Apêndice A. Continuação...

CNVR	Chr	Inicio	Final	NGS	50K	Freq	Genes
CNVR_165	chr21	66340501	66364500	DELEÇÃO	DUPLICAÇÃO	0.10	*ENSBTAG00000013491,ENSBTAG00000035544*
CNVR_166	chr21	71029501	71051500	DELEÇÃO	DELEÇÃO	0.31	*C21H14orf79*
CNVR_167	chr22	48998001	49014500	DELEÇÃO	DELEÇÃO	0.10	*SEMA3G*
CNVR_168	chr22	49032001	49092000	DELEÇÃO	DELEÇÃO	0.10	*BAP1,PHF7,ENSBTAG00000047794*
CNVR_169	chr22	49104001	49148500	DELEÇÃO	DELEÇÃO	0.10	*GLYCTK,MIR135A-1,ENSBTAG00000047794*
CNVR_170	chr22	49468501	49554500	DELEÇÃO	DUPLICAÇÃO	0.21	*RRP9,ENSBTAG00000000671,GPR62,PCBP4,RPL29,ACY1,ABHD14B*
CNVR_171	chr22	59139001	59286000	DELEÇÃO	DELEÇÃO	0.10	*NUP210*
CNVR_172	chr22	59310501	59324500	DELEÇÃO	DELEÇÃO	0.31	
CNVR_173	chr22	59326501	59527500	DELEÇÃO	DELEÇÃO	0.21	*IQSEC1*
CNVR_174	chr23	25332501	25346500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_175	chr23	51464001	51492500	DELEÇÃO	DUPLICAÇÃO	0.21	
CNVR_176	chr25	36469001	36531500	DELEÇÃO	DELEÇÃO	0.10	*AGFG2, LRCH4, FBXO24, ACTL6B, TFR2,PCOLCE,MOSPD3,ENSBTAG00000039601, ENSBTAG00000046753*
CNVR_177	chr25	42039001	42061500	DELEÇÃO	DELEÇÃO	0.52	
CNVR_178	chr25	42065501	42177500	DELEÇÃO	DELEÇÃO	0.73	*ENSBTAG00000028976, ENSBTAG00000030721*
CNVR_179	chr25	42275001	42317000	DELEÇÃO	DELEÇÃO	0.31	*COX19, ENSBTAG00000017701, ENSBTAG00000026192,MIR339A,MIR2389,GPR146*
CNVR_180	chr25	42486501	42511000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_181	chr25	42566501	42582500	DELEÇÃO	DELEÇÃO	0.10	*PDGFA*
CNVR_182	chr26	2420001	2433000	DELEÇÃO	DELEÇÃO	0.21	
CNVR_183	chr26	2454001	2462000	DELEÇÃO	DELEÇÃO	0.21	
CNVR_184	chr26	5322001	5330500	DELEÇÃO	DUPLICAÇÃO	0.10	*PCDH15*
CNVR_185	chr26	29018501	29038500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_186	chr26	47800001	47849000	DELEÇÃO	DELEÇÃO	0.10	*PTPRE*
CNVR_187	chr26	51399501	51530500	DELEÇÃO	DELEÇÃO	2.71	*JAKMIP3, DPYSL4*
CNVR_188	chr27	19710501	19723000	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_189	chr28	2251501	2271500	DELEÇÃO	DUPLICAÇÃO	0.83	

CNVR_190	chr28	2365501	2379000	DUPLICAÇÃO	DUPLICAÇÃO	0.52	
CNVR_191	chr28	2407501	2431500	DUPLICAÇÃO	DUPLICAÇÃO	0.52	
CNVR_192	chr28	2488001	2548500	DELEÇÃO	DUPLICAÇÃO	0.52	*ENSBTAG00000036271*
CNVR_193	chr28	20780001	20786500	DELEÇÃO	DUPLICAÇÃO	2.60	
CNVR_194	chr28	21557001	21574500	DELEÇÃO	DUPLICAÇÃO	0.10	
CNVR_195	chr28	25473501	25489000	DELEÇÃO	MISTA	0.42	
CNVR_196	chr29	9635001	9668500	DELEÇÃO	DUPLICAÇÃO	0.10	*PICALM*
CNVR_197	chr29	13744501	13776500	DELEÇÃO	DELEÇÃO	0.10	
CNVR_198	chr29	20258001	20282500	DELEÇÃO	DELEÇÃO	0.10	*LUZP2*
CNVR_199	chr29	20305001	20316000	DELEÇÃO	DELEÇÃO	0.10	*LUZP2*
CNVR_200	chr29	27164501	27175500	DELEÇÃO	DELEÇÃO	0.10	*LOC100301320*
CNVR_201	chr29	27754001	27757000	DELEÇÃO	DELEÇÃO	0.10	
CNVR_202	chr29	50056501	50078000	DELEÇÃO	DUPLICAÇÃO	0.52	*IGF2, MIR483*
CNVR_203	chr29	51489501	51505500	DELEÇÃO	MISTA	0.42	

Apêndice B. Descrição das regiões de variações no número de cópias (CNVR) candidatas em touros Girolando. Cromossomo (Chr), início e final, estado do evento por meio da detecção em sequenciamento de nova geração (NGS) e painel de genotipagem de alta densidade (HD) , frequência na população e os genes mapeados com base no genoma referencia UMD 3.1 na CNVR.

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_1	chr1	41001	62500	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_2	chr1	36283001	36296500	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_3	chr1	42135501	42149500	DELEÇÃO	DELEÇÃO	3.05	
CNVR_4	chr1	93419501	93436500	DELEÇÃO	DUPLICAÇÃO	3.44	
CNVR_5	chr1	93711001	93830500	DELEÇÃO	DELEÇÃO	69.47	
CNVR_6	chr1	94061501	94069500	DELEÇÃO	DELEÇÃO	1.15	
CNVR_7	chr2	135126501	135147500	DELEÇÃO	DUPLICAÇÃO	6.87	
CNVR_8	chr3	54450001	55022500	MISTA	MISTA	14.89	*GBP4 ,LOC510382,GBP6,,LOC100336669,*
CNVR_9	chr3	119891001	119918000	DUPLICAÇÃO	DUPLICAÇÃO	3.82	*LOC615009, SCMH1*
CNVR_10	chr4	108243001	108266500	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_11	chr4	113648501	113685500	DUPLICAÇÃO	DUPLICAÇÃO	7.25	*GIMAP8*
CNVR_12	chr4	113719501	113799500	DUPLICAÇÃO	DUPLICAÇÃO	0.76	*GIMAP7, LOC511617*
CNVR_13	chr4	119676001	119734500	DELEÇÃO	DELEÇÃO	0.38	
CNVR_14	chr5	57986501	58015000	DUPLICAÇÃO	DUPLICAÇÃO	0.76	*OR2AP1*
CNVR_15	chr5	58066001	58095500	DUPLICAÇÃO	MISTA	0.76	*LOC618816, LOC529303*
CNVR_16	chr5	58127501	58153000	DELEÇÃO	MISTA	1.15	
CNVR_17	chr5	58215501	58229500	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_18	chr5	58229501	58246500	DELEÇÃO	DUPLICAÇÃO	0.76	
CNVR_19	chr5	58958501	59236000	DELEÇÃO	DELEÇÃO	43.51	*OR6C1, LOC787945, LOC788027 ,LOC787991*
CNVR_20	chr5	59372501	59634500	DELEÇÃO	MISTA	47.33	*LOC788242, LOC788322, LOC788220, LOC788372, LOC788357, LOC788285 ,ENSBTAG00000031100,LOC788210,ENSBTAG00000046446,LOC788258*
CNVR_21	chr5	99626001	99659500	DELEÇÃO	DUPLICAÇÃO	0.38	*LOC100139049*
CNVR_22	chr5	99801501	99834000	DELEÇÃO	DUPLICAÇÃO	0.38	*LOC100847738*
CNVR_23	chr5	100588001	100643500	DUPLICAÇÃO	DUPLICAÇÃO	1.15	*LOC509972,LOC784451*
CNVR_24	chr5	102456501	102583000	DUPLICAÇÃO	DELEÇÃO	1.53	*WC-7*
CNVR_25	chr5	102583001	102590500	DELEÇÃO	DELEÇÃO	1.53	
CNVR_26	chr5	102590501	102632000	DUPLICAÇÃO	DELEÇÃO	1.91	

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_27	chr5	102644001	102653000	DELEÇÃO	DELEÇÃO	1.91	
CNVR_28	chr5	103144501	103162500	DUPLICAÇÃO	MISTA	1.15	*WC.1.3*
CNVR_29	chr5	103171001	103181000	DUPLICAÇÃO	MISTA	1.15	*WC1.3*
CNVR_30	chr5	103280001	103327500	DUPLICAÇÃO	MISTA	1.15	*WC-8*
CNVR_31	chr5	103341501	103365000	DUPLICAÇÃO	MISTA	1.15	
CNVR_32	chr5	103374501	103406500	DUPLICAÇÃO	MISTA	1.15	*CD163L1*
CNVR_33	chr5	117265001	117318000	DELEÇÃO	MISTA	20.23	
CNVR_34	chr5	117392501	117647500	DELEÇÃO	MISTA	45.42	
CNVR_35	chr5	117942001	118022500	DELEÇÃO	DELEÇÃO	0.38	*GRAMD4*
CNVR_36	chr5	120705501	120726500	DELEÇÃO	DELEÇÃO	0.38	
CNVR_37	chr6	7173501	7184000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000010350*
CNVR_38	chr6	53448001	53468500	DELEÇÃO	DELEÇÃO	2.67	
CNVR_39	chr6	53493001	53526000	DELEÇÃO	DELEÇÃO	3.82	
CNVR_40	chr6	53686501	53704000	DELEÇÃO	DELEÇÃO	3.82	
CNVR_41	chr6	60749001	60782000	DELEÇÃO	DUPLICAÇÃO	0.38	*LOC539533*
CNVR_42	chr7	2728501	2756000	DELEÇÃO	DELEÇÃO	0.38	
CNVR_43	chr7	9782501	9827500	MISTA	MISTA	3.82	
CNVR_44	chr7	9828001	9916000	DUPLICAÇÃO	MISTA	3.82	*LOC100295735, LOC787882*
CNVR_45	chr7	9916001	9930000	DELEÇÃO	MISTA	3.82	
CNVR_46	chr7	9935501	9946000	DELEÇÃO	MISTA	3.82	
CNVR_47	chr7	9946001	10000500	DUPLICAÇÃO	MISTA	3.82	*LOC516132*
CNVR_48	chr7	10005001	10154500	DELEÇÃO	MISTA	32.82	*LOC512488,LOC100336881,LOC788031,LOC100299615,LOC100295806*
CNVR_49	chr7	10278501	10828000	DELEÇÃO	MISTA	27.10	*LOC541022,LOC785624,OR7A10,LOC100298119,LOC616964,LOC507378,LOC510625, LOC789554*
CNVR_50	chr7	10829001	11135000	DELEÇÃO	MISTA	33.97	*LOC100847301,LOC786119,LOC788246,LOC789504,LOC512296,OR7A17*
CNVR_51	chr7	11265501	11282500	DUPLICAÇÃO	MISTA	4.96	
CNVR_52	chr7	70191001	70275000	DUPLICAÇÃO	DUPLICAÇÃO	4.20	*HAVCR1*
CNVR_53	chr8	12452501	12498500	DELEÇÃO	DUPLICAÇÃO	0.76	
CNVR_54	chr8	96139001	96186000	DELEÇÃO	DELEÇÃO	2.29	*LOC789766*

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_55	chr8	105176501	105217000	DUPLICAÇÃO	DUPLICAÇÃO	2.29	*ORM1, COL27A1*
CNVR_56	chr9	1641501	1645500	DELEÇÃO	DELEÇÃO	1.15	
CNVR_57	chr9	45319001	45471000	DUPLICAÇÃO	DUPLICAÇÃO	10.69	*ENSBTAG00000033218, POPDC3*
CNVR_58	chr9	88207501	88217500	DUPLICAÇÃO	DUPLICAÇÃO	5.34	
CNVR_59	chr9	88281001	88325500	DUPLICAÇÃO	DUPLICAÇÃO	1.53	*RAET1G, ULBP21*
CNVR_60	chr10	1511001	1579000	DELEÇÃO	DELEÇÃO	5.34	*EPB41L4A*
CNVR_61	chr10	22504501	22557000	DUPLICAÇÃO	DELEÇÃO	0.76	*ENSBTAG00000000432, ENSBTAG00000046962*
CNVR_62	chr10	22661501	22707000	DUPLICAÇÃO	MISTA	1.53	*ENSBTAG00000000432, ENSBTAG00000045758*
CNVR_63	chr10	22707001	22792500	DELEÇÃO	MISTA	2.67	*ENSBTAG00000000432, ENSBTAG00000046464,ENSBTAG00000048302*
CNVR_64	chr10	22932001	22943500	DUPLICAÇÃO	MISTA	1.53	*ENSBTAG00000000432*
CNVR_65	chr10	23016501	23050000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000000432*
CNVR_66	chr10	23093501	23124000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000000432, ENSBTAG00000038098*
CNVR_67	chr10	23164501	23431000	DELEÇÃO	DELEÇÃO	3.44	*ENSBTAG00000000432,ENSBTAG00000036202,LOC786717,ENSBTAG00000047091 ,ENSBTAG00000047312,ENSBTAG00000048064*
CNVR_68	chr10	23436001	23497500	DELEÇÃO	MISTA	4.96	*ENSBTAG00000046477,ENSBTAG00000047683*
CNVR_69	chr10	23541001	23578000	DELEÇÃO	MISTA	6.87	*ENSBTAG00000047001,ENSBTAG00000047981*
CNVR_70	chr10	23629501	23747500	DELEÇÃO	MISTA	7.63	*ENSBTAG00000045953,ENSBTAG00000047097,ENSBTAG00000047909*
CNVR_71	chr10	23947001	23958500	DELEÇÃO	MISTA	5.73	
CNVR_72	chr10	23993001	24016500	DUPLICAÇÃO	MISTA	5.73	*ENSBTAG00000039038*
CNVR_73	chr10	24016501	24041500	DELEÇÃO	MISTA	5.73	
CNVR_74	chr10	24051001	24055000	DELEÇÃO	MISTA	5.73	
CNVR_75	chr10	24243001	24250500	DELEÇÃO	DUPLICAÇÃO	4.58	*ENSBTAG00000046546*
CNVR_76	chr10	24482001	24487500	DELEÇÃO	DUPLICAÇÃO	1.53	
CNVR_77	chr10	24493501	24513500	DELEÇÃO	DUPLICAÇÃO	1.15	
CNVR_78	chr10	24537001	24560000	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_79	chr10	24856501	24907500	DELEÇÃO	DUPLICAÇÃO	0.76	
CNVR_80	chr10	24927501	24933500	DELEÇÃO	DUPLICAÇÃO	0.76	
CNVR_81	chr10	24954501	24961000	DELEÇÃO	DUPLICAÇÃO	0.76	*ENSBTAG00000024659*
CNVR_82	chr10	25052501	25116500	DELEÇÃO	DUPLICAÇÃO	0.76	*ENSBTAG00000045863, ENSBTAG00000045883, ENSBTAG00000046993 ,ENSBTAG00000047015*

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_83	chr10	25149501	25182000	DELEÇÃO	DUPLICAÇÃO	0.76	*LOC100300850, ENSBTAG00000046995*
CNVR_84	chr11	39875501	39923500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_85	chr11	103828001	103842500	DELEÇÃO	DELEÇÃO	0.38	*GPSM1*
CNVR_86	chr11	103909501	104039000	DELEÇÃO	DELEÇÃO	0.38	*SDCCAG3, PMPCA, INPP5E, ENSBTAG00000022799, SEC16A*
CNVR_87	chr11	104171501	104222000	DELEÇÃO	DUPLICAÇÃO	0.38	*FAM69B, ENSBTAG00000030281, ENSBTAG00000042482, ENSBTAG00000042930*
CNVR_88	chr11	106191001	106224500	DELEÇÃO	DELEÇÃO	0.38	*ENSBTAG00000046140, FUT7, ABCA2*
CNVR_89	chr11	106879501	106948500	DELEÇÃO	DELEÇÃO	0.38	*ENSBTAG00000047998*
CNVR_90	chr12	32049501	32086500	DELEÇÃO	MISTA	15.65	
CNVR_91	chr12	32192501	32233000	DELEÇÃO	MISTA	1.15	
CNVR_92	chr12	70351501	70484000	DELEÇÃO	MISTA	1.91	
CNVR_93	chr12	70567001	70574500	DELEÇÃO	DELEÇÃO	3.82	
CNVR_94	chr12	70620001	70656000	DELEÇÃO	MISTA	3.82	
CNVR_95	chr12	70708501	71220000	DELEÇÃO	DUPLICAÇÃO	6.49	
CNVR_96	chr12	71252501	71449500	MISTA	MISTA	6.87	
CNVR_97	chr12	71459001	71519500	MISTA	MISTA	12.98	
CNVR_98	chr12	71526001	71611500	DELEÇÃO	MISTA	14.12	
CNVR_99	chr12	71623501	72166500	MISTA	MISTA	22.14	*LOC100337053, ENSBTAG00000047383*
CNVR_100	chr12	72174001	72270000	MISTA	DELEÇÃO	14.50	*LOC100337053*
CNVR_101	chr12	72424001	72926500	MISTA	DELEÇÃO	30.92	*ENSBTAG00000047181*
CNVR_102	chr12	72940001	72977000	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_103	chr12	72985001	73020000	DELEÇÃO	DUPLICAÇÃO	1.15	
CNVR_104	chr12	73044001	73086000	DELEÇÃO	DUPLICAÇÃO	1.15	
CNVR_105	chr12	73093001	73098500	DELEÇÃO	DUPLICAÇÃO	1.15	
CNVR_106	chr12	73110501	73192500	DELEÇÃO	DUPLICAÇÃO	1.91	
CNVR_107	chr12	73221001	73241000	DUPLICAÇÃO	DUPLICAÇÃO	1.91	
CNVR_108	chr12	73241501	73396000	MISTA	MISTA	18.70	
CNVR_109	chr12	73484501	73717000	DELEÇÃO	MISTA	19.85	*LOC100299180*

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_110	chr12	73741001	73820500	DELEÇÃO	MISTA	9.54	
CNVR_111	chr12	73852001	74253000	MISTA	MISTA	11.83	*ENSBTAG00000047360, ENSBTAG00000047764*
CNVR_112	chr12	74277001	74314500	DELEÇÃO	DUPLICAÇÃO	3.82	
CNVR_113	chr12	74321001	74339500	DELEÇÃO	MISTA	3.44	
CNVR_114	chr12	74374501	74448000	DELEÇÃO	MISTA	2.67	
CNVR_115	chr12	74467501	74590000	DUPLICAÇÃO	MISTA	1.91	
CNVR_116	chr12	74633501	74651500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_117	chr12	74652001	74661500	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_118	chr12	74682001	74738500	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_119	chr12	74739501	74747500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_120	chr12	74760001	75139500	DELEÇÃO	MISTA	30.15	*LOC509854*
CNVR_121	chr12	75150001	75177000	DUPLICAÇÃO	MISTA	27.10	
CNVR_122	chr12	75177001	75202000	DELEÇÃO	MISTA	27.10	
CNVR_123	chr12	75223501	75238000	DELEÇÃO	MISTA	0.76	
CNVR_124	chr12	75448501	75504000	DUPLICAÇÃO	DUPLICAÇÃO	5.34	*ENSBTAG00000026070*
CNVR_125	chr12	75532501	75549500	DUPLICAÇÃO	MISTA	6.11	*ENSBTAG00000026070*
CNVR_126	chr12	75589501	75609000	DELEÇÃO	MISTA	5.34	
CNVR_127	chr12	75610001	75618500	DUPLICAÇÃO	MISTA	5.34	
CNVR_128	chr12	75619001	75675000	MISTA	MISTA	5.34	*LOC523126*
CNVR_129	chr12	75802001	75813000	DELEÇÃO	MISTA	4.20	*LOC523126*
CNVR_130	chr12	75839001	75866500	DELEÇÃO	MISTA	3.82	*LOC523126*
CNVR_131	chr12	75884501	75912000	DELEÇÃO	MISTA	3.82	
CNVR_132	chr12	75957001	75987000	DELEÇÃO	DUPLICAÇÃO	1.53	
CNVR_133	chr12	75998501	76002500	DELEÇÃO	DUPLICAÇÃO	1.53	
CNVR_134	chr12	76003501	76023000	DUPLICAÇÃO	DUPLICAÇÃO	1.53	
CNVR_135	chr12	76024501	76046000	DUPLICAÇÃO	DUPLICAÇÃO	1.53	
CNVR_136	chr12	76104501	76140000	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_137	chr12	76163001	76174500	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_138	chr12	76199501	76206000	DELEÇÃO	DUPLICAÇÃO	1.15	

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_139	chr12	76217501	76233500	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_140	chr12	76233501	76249500	DELEÇÃO	DUPLICAÇÃO	1.15	
CNVR_141	chr12	76249501	76410500	DUPLICAÇÃO	DUPLICAÇÃO	1.15	
CNVR_142	chr12	76420001	76441500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_143	chr12	76442501	76459000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_144	chr12	76472501	76481500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_145	chr12	76531501	76548500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_146	chr12	76552501	76561000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_147	chr12	90729501	90800500	DELEÇÃO	DELEÇÃO	0.38	*TFDP1, ATP4B, GRK1 , ENSBTAG00000019965*
CNVR_148	chr12	90888501	90945500	DELEÇÃO	DELEÇÃO	0.38	*RASA3, MGC134473*
CNVR_149	chr13	43649501	43777500	DUPLICAÇÃO	DUPLICAÇÃO	2.67	*LOC782061, LOC100337056*
CNVR_150	chr14	1176001	1193500	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_151	chr14	2165001	2187000	DELEÇÃO	DELEÇÃO	0.38	*SCRIB, PUF60*
CNVR_152	chr14	81842501	81906500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_153	chr15	10989501	11001000	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_154	chr15	46594501	46620500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*LOC516636*
CNVR_155	chr15	46675501	46699000	DUPLICAÇÃO	DUPLICAÇÃO	1.15	*LOC511622, LOC100336980, ENSBTAG00000042988*
CNVR_156	chr15	46710001	46724500	DUPLICAÇÃO	DUPLICAÇÃO	0.76	*ENSBTAG00000021147, OR2AG2*
CNVR_157	chr15	46743001	46780500	DUPLICAÇÃO	DUPLICAÇÃO	0.76	
CNVR_158	chr15	46855001	46876000	DUPLICAÇÃO	DUPLICAÇÃO	1.91	*ENSBTAG00000037937*
CNVR_159	chr15	46945001	46952500	DELEÇÃO	DUPLICAÇÃO	10.69	
CNVR_160	chr15	47889001	47973500	DELEÇÃO	DUPLICAÇÃO	1.15	*OR52N5, ENSBTAG00000043009, LOC507428, LOC788263,*
CNVR_161	chr15	48013501	48045500	DELEÇÃO	DUPLICAÇÃO	2.67	*LOC787810, LOC787830*
CNVR_162	chr15	51709501	51720000	DUPLICAÇÃO	DELEÇÃO	0.38	
CNVR_163	chr15	80574001	80588000	DELEÇÃO	MISTA	29.39	*LOC785406*
CNVR_164	chr16	5870501	5890000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000024647, ENSBTAG00000040409*
CNVR_165	chr16	5978501	6031000	DELEÇÃO	DUPLICAÇÃO	1.53	
CNVR_166	chr16	6044501	6053000	DELEÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000039995*
CNVR_167	chr16	6069501	6119500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000039995*

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_168	chr16	52684501	52697000	DELEÇÃO	DUPLICAÇÃO	0.38	*AGRN*
CNVR_169	chr16	70931001	70962500	DELEÇÃO	DELEÇÃO	0.38	*LGR6*
CNVR_170	chr17	74170001	74199000	DELEÇÃO	DELEÇÃO	0.38	*HIC2, ENSBTAG00000030927*
CNVR_171	chr17	74627001	74642500	DELEÇÃO	DELEÇÃO	0.38	*SLC25A1*
CNVR_172	chr17	74998001	75145500	DELEÇÃO	DELEÇÃO	0.38	*DGCR8, RTN4R,*
CNVR_173	chr18	60888501	60893500	DELEÇÃO	DUPLICAÇÃO	3.05	*ENSBTAG00000045581*
CNVR_174	chr18	61474001	61553000	MISTA	MISTA	0.76	
CNVR_175	chr18	61558001	61567500	DUPLICAÇÃO	DELEÇÃO	0.38	
CNVR_176	chr18	61567501	61579000	DELEÇÃO	DELEÇÃO	0.38	
CNVR_177	chr18	61633501	61644500	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000047602*
CNVR_178	chr18	61651501	61842500	DELEÇÃO	MISTA	1.15	*ENSBTAG00000002290, LOC514552 ,ENSBTAG00000013345*
CNVR_179	chr19	1851001	1870500	DELEÇÃO	DELEÇÃO	0.38	
CNVR_180	chr20	55672501	55696000	DELEÇÃO	DELEÇÃO	0.76	
CNVR_181	chr20	60471001	60501500	DUPLICAÇÃO	DUPLICAÇÃO	2.67	
CNVR_182	chr20	60507001	60518000	DUPLICAÇÃO	DUPLICAÇÃO	2.67	
CNVR_183	chr20	71278001	71337500	DELEÇÃO	DELEÇÃO	0.38	*TERT, SLC6A18*
CNVR_184	chr21	35171001	35186000	DUPLICAÇÃO	DUPLICAÇÃO	0.76	*LOC505658, LOC509956*
CNVR_185	chr21	35218501	35233500	DELEÇÃO	DUPLICAÇÃO	0.38	*LOC508858*
CNVR_186	chr21	41800001	41808000	DELEÇÃO	DUPLICAÇÃO	3.05	*SCFD1*
CNVR_187	chr21	70579501	70594500	DELEÇÃO	DUPLICAÇÃO	0.38	
CNVR_188	chr22	60007501	60051500	DELEÇÃO	MISTA	0.76	*DNAJB8, GATA2*
CNVR_189	chr23	25332501	25441000	DELEÇÃO	MISTA	2.67	*ENSBTAG00000003352, BOLA-DQA2, BOLA-DQB, LOC100848815,BOLA-DQA5*
CNVR_190	chr23	25694501	25704500	DELEÇÃO	DUPLICAÇÃO	1.15	
CNVR_191	chr23	25704501	25879500	DELEÇÃO	MISTA	2.67	*BLA-DQB*
CNVR_192	chr23	25879501	25913000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	
CNVR_193	chr23	25913001	26239000	DELEÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000033979*
CNVR_194	chr23	26267501	26361500	DELEÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000026163*
CNVR_195	chr23	28453501	28481500	DELEÇÃO	DUPLICAÇÃO	1.15	*JSP.1*
CNVR_196	chr23	28490501	28506000	DUPLICAÇÃO	DUPLICAÇÃO	0.38	*BOLA*

Apêndice B. Continuação...

CNVR	Chr	Inicio	Fim	NGS	HD	Freq	Genes
CNVR_197	chr23	28529001	28541500	DELEÇÃO	DUPLICAÇÃO	0.38	*ENSBTAG00000037421*
CNVR_198	chr25	42486501	42511000	DELEÇÃO	DELEÇÃO	0.38	*PRKAR1B*
CNVR_199	chr25	42566501	42582500	DELEÇÃO	DELEÇÃO	0.38	*PDGFA*
CNVR_200	chr26	23618501	23659000	DUPLICAÇÃO	DUPLICAÇÃO	0.76	*ARL3*
CNVR_201	chr26	25940001	26026000	DELEÇÃO	DELEÇÃO	0.38	*TUBGCP2, ZNF511, ADAM8, ENSBTAG00000024999, MIR202*
CNVR_202	chr26	51117001	51186000	DELEÇÃO	DELEÇÃO	0.38	*ENSBTAG00000009181*
CNVR_203	chr28	2251501	2271500	DELEÇÃO	DUPLICAÇÃO	3.05	
CNVR_204	chr28	2365501	2379000	DUPLICAÇÃO	DUPLICAÇÃO	3.44	
CNVR_205	chr28	2407501	2431500	DUPLICAÇÃO	DUPLICAÇÃO	3.05	
CNVR_206	chr28	2488001	2548500	DELEÇÃO	MISTA	5.73	*ENSBTAG00000036271*
CNVR_207	chr28	2654501	2747000	DELEÇÃO	MISTA	8.02	*OR5D14, LOC787869*
CNVR_208	chr28	2786001	2867500	DELEÇÃO	MISTA	8.40	*ENSBTAG00000004217, LOC104968693*
CNVR_209	chr29	27360501	27430500	DELEÇÃO	DELEÇÃO	0.38	*ENSBTAG00000046299*
CNVR_210	chr29	39340501	39367000	DELEÇÃO	DUPLICAÇÃO	3.05	*ENSBTAG00000034244, ENSBTAG00000047419*
CNVR_211	chr29	42451001	42539000	MISTA	MISTA	8.02	*LOC618367, ENSBTAG00000037645*
CNVR_212	chr29	42539501	42577500	DUPLICAÇÃO	MISTA	4.58	
CNVR_213	chr29	51489501	51505500	DELEÇÃO	DELEÇÃO	0.38	

Apêndice C. Breve pipeline para o mapeamento das sequências no genoma referência bovino UMD 3. 1.

1. Mapeamento por meio do algoritmo BWA-MEM.

```
bwa mem -M -t 5 -R \
'@RG\tID:${SAMPLE}_${LANE}_${NUMBER}\tLN:${LANE}\tLB:${SAMPLE}\tSM:${SAMPLE}\tPL:ILLUMINA' \
umd_3_1_reference_1000_bull_genomes.fa fileName.sam
```

2. Converter arquivo em formato SAM para formato BAM.

```
samtools view -bS fileName.sam > fileName.bam
```

3. Ordenar o arquivo BAM.

```
samtools sort fileName.bam -o fileName.sorted.bam
```

4. Remover as duplicatas.

```
java jar picard/MarkDuplicates.jar \
INPUT= fileName.sorted.bam OUTPUT=fileName.sorted.mkdup.bam \
METRICS_FILE=fileName.sorted.mkdup.metrics.txt ASSUME_SORTED=true
```

5. Criar arquivo BAM indexado.

```
samtools index fileName.sorted.mkdup.bam
```

6. Juntar todas as lanes de cada animal em somente um arquivo BAM.

```
samtools merge Sample.bam *.sorted.mkdup.bam
```

7. Calcular as estatísticas do alinhamento.

```
samtools flagstat Sample.bam > Sample.flagstat
```

Apêndice D. Descrição da variação no número de cópias (CNVRs) compartilhadas entre touros da raça Gir e Girolando. Localização (cromossomo, ínicio, final), frequência (Freq) na população e genes mapeados nas regiões.

Cromossomo	Inicio	Final	Freq	Gene
chr1	52548251	52566500	2	*BBX*
chr1	88497001	88519500	2	*PIK3CA*
chr1	97833001	97848500	2	*SKIL*
chr1	146570751	146607000	2	*RRP1*
chr2	62644001	62670500	2	*CCNT2*
chr2	113370501	113388000	9	*CUL3*
chr3	22626251	22636500	9	*PRKAB2*
chr3	74446501	74470000	2	*ZRANB2*
chr3	108762001	108786500	7	*EPHA10*
chr4	32248501	32278000	3	*TRA2A*
chr4	69580001	69603500	2	*SKAP2*
chr4	114608751	114625000	8	*CHPF2; MIR671;SMARCD3*
chr5	31109501	31134250	10	*DDX23; CACNB3*
chr5	63037001	63054000	2	*TMPO*
chr5	90894501	90907500	2	*AEBP2*
chr5	103570501	103591000	3	*CLSTN3*
chr5	104362251	104371500	8	*SCNN1A*
chr5	112210001	112228750	2	*SGSM3*
chr6	23357001	23377500	2	*UBE2D3*
chr6	64961501	64971000	2	*GNPDA2*
chr6	66219501	66246500	2	*GABRG1;GABRG1*
chr7	4871501	4953500	12	*CIST1;PDE4C;RAB3A;MPV17L2*
chr7	22861001	22881000	4	*SEPT8*
chr7	27502001	27511000	2	*CTXN3*
chr9	48585501	48641000	2	*GRIK2*
chr9	75261001	75283500	3	*BCLAF1*
chr10	1511001	1581500	5	*EPB41L4A*
chr10	20728001	20746500	3	*RABGGTA;TGM1*
chr10	26673001	26686750	4	*PNP*
chr10	36021501	36034750	2	*PLCB2*

Apêndice D. Continuação...

Cromossomo	Inicio	Final	Freq	Gene
chr10	70012001	70037000	2	*NAA30*
chr11	100965501	100989750	14	*PRDM12*
chr12	19571501	19621500	2	*MIR16A;MIR15A*
chr13	17379501	17400750	2	*PFKFB3*
chr13	43649501	43777000	7	*LOC782061*
chr13	75265001	75278750	5	*WFDC3*
chr15	30470501	30487500	2	*USP2*
chr15	63352251	63394750	13	*PAX6*
chr15	64540001	64568000	4	*CSTF3*
chr15	83367751	83379500	6	*LOC518623*
chr16	25051001	25078750	11	*HLX*
chr17	6707501	6738500	5	*RPS3A*
chr17	55462501	55484500	3	*MLXIP*
chr17	64912751	64941750	7	*PLA2G1B;MSI1*
chr18	25944751	25966750	2	*CNGB1;CNGB1;CNGB1*
chr18	46020501	46069500	3	*FXYD3;LGI4;FXYD1;FXYD7;FXYD5*
chr18	55568001	55628500	9	*GRWD1*
chr18	56224001	56387750	13	*PIH1D1*
chr18	57432501	57467250	3	*KLK8;KLK9;KLK10;KLK12*
chr18	65700001	65716000	9	*ZNF329*
chr19	13594501	13622000	9	*LHX1*
chr19	18501001	18551500	2	*SUZ12*
chr19	21152001	21179000	2	*CRYBA1*
chr19	43360001	43400000	12	*CNTNAP1*
chr21	45559501	45596000	3	*BAZ1A*
chr22	53827751	53853250	11	*LOC529196;CCR1*
chr23	7393001	7439250	4	*RPS18;B3GALT4;WDR46;PFDN6;RGL2;TAPBP;MIR2376;DAXX*
chr23	13824001	13844750	3	*MOCS1;MOCS1*
chr23	16964501	16987500	3	*DLK2*
chr23	28341001	28388250	2	*BOLA-NC1*
chr25	3856001	3887000	2	*SEPT12*
chr25	27468501	27502000	4	*BCKDK;KAT8;PRSS8*

Apêndice D. Continuação...

Cromossomo	Inicio	Final	Freq	Gene
chr25	35906001	35949000	9	*COL26A1*
chr25	36597501	36660500	14	*TSC22D4;C25H7orf61;PPP1R35;MEPCE;ZCWPW1*
chr25	36768251	36798750	10	*STAG3;GPC2;GAL3ST4;LAMTOR4*
chr29	41914001	41935750	7	*CHRM1*

Apêndice E. Descrição da variação no número de cópias (CNVRs) compartilhadas entre touros da raça Holandês e Girolando. Localização (cromossomo, ínicio, final) , frequência (Freq) na população e genes mapeados nas regiões.

Cromossomo	Inicio	Fim	Freq	Gene
chr1	137910001	138353000	2	*ACAD11;DNAJC13*
chr2	98161501	98475000	2	*UNC80;ACADL*
chr3	49725001	49887500	2	*BCAR3*
chr3	70619501	70994500	2	*TNNI3K;LRRIQ3*
chr3	72169001	73533000	2	*NEGR1*
chr4	35232001	36961000	2	*SEMA3D;SEMA3A*
chr4	38940001	39449000	2	*HGF*
chr8	105138501	105228000	5	*MIR455;ORM1*
chr9	22243501	22545500	2	*IBTK*
chr9	49852501	50015500	4	*ASCC3*
chr10	26767001	26780000	2	*TEP1*
chr11	36291001	36311000	4	*ASB3*
chr17	10788501	10807500	3	*EDNRA*
chr18	35835001	35865000	3	*PRMT7;SMPD*

Apêndice F. Descrição dos genes que apresentaram V_{ST} maior do que 0.5 para cada touro da raça Girolando (A e B), Gir (D, F, G, H, I, J, K, L, M, N, O e P) e Holandês (R, S, T, U e V). Gene, V_{ST}, cromossomo (Chr), posição inicial (Inicio) e final (Fim) do gene.

Gene	VST	Chr	Inicio	Fim	A	B	D	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
AOX1	0.98	chr2	89517708	89589232	3.50	3.20	4.09	4.18	4.17	4.03	4.10	4.27	3.07	4.29	4.34	4.03	4.12	4.12	4.17	nul	nul	nul	nul	nul
SLBP	0.98	chr6	109613820	109629337	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.09	1.33	1.34	1.39	1.12
TACC3	0.98	chr6	109636202	109645421	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.09	1.33	1.34	1.39	1.12
TMEM129	0.98	chr6	109631166	109635557	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.09	1.33	1.34	1.39	1.12
CDK13	0.83	chr4	81667519	81807550	1.07	0.59	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul
DLX1	0.78	chr2	24502924	24506672	nul	nul	nul	1.18	1.35	1.37	1.28	1.29	1.39	1.27	1.26	1.30	1.30	1.37	1.09	nul	nul	nul	nul	nul
ZIC2	0.78	chr12	80719189	80723692	nul	nul	nul	1.23	1.32	1.27	1.32	1.28	1.33	1.37	1.39	1.29	1.27	1.32	1.30	nul	nul	nul	nul	nul
PRAME	0.77	chr17	51374857	51380364	nul	nul	nul	14.85	14.71	15.04	14.80	14.32	14.10	15.60	16.43	14.34	14.91	14.36	10.77	nul	nul	nul	nul	nul
SRY	0.77	chr13	5424636	5425326	nul	nul	nul	1.12	1.10	1.04	1.05	1.07	1.04	1.05	1.08	1.08	1.04	1.05	1.26	nul	nul	nul	nul	nul
ZNF280BY	0.74	chr17	51413318	51534036	nul	nul	nul	13.66	12.66	15.40	15.20	14.91	10.76	13.67	15.89	13.01	9.91	10.72	10.77	nul	nul	nul	nul	nul
BOLA	0.7	chr23	28330911	28506282	1.21	0.99	3.62	1.84	2.66	2.20	2.17	2.83	3.07	2.23	3.28	2.38	2.33	2.43	3.27	2.08	nul	nul	nul	nul
LAG3	0.67	chr5	104029080	104035920	nul	nul	0.81	1.37	0.67	0.64	0.63	0.79	0.85	0.76	nul	0.80	0.74	0.68	0.84	nul	0.61	nul	nul	nul
DLG4	0.65	chr19	27543282	27565410	nul	1.45	1.28	0.80	1.17	1.31	1.31	1.32	1.28	1.16	1.36	1.34	1.10	1.33	1.41	nul	1.30	nul	nul	nul
ADAP1	0.62	chr25	42330025	42362918	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.28	1.35	1.29	1.49
BRI3	0.62	chr25	38281389	38287455	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.29	nul	1.20	1.48	nul
CABP2	0.62	chr29	46042200	46046958	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.32	1.31	1.38	1.43	nul
CABP4	0.62	chr29	45988624	45992121	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.26	1.31	1.41	1.43	nul
CKB	0.62	chr21	69809412	69812615	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.41	1.41	1.23	1.17
DLX5	0.62	chr4	14291964	14296237	nul	nul	nul	1.31	1.28	1.40	1.23	1.25	1.25	1.17	1.27	1.26	nul	1.44	1.17	nul	nul	nul	nul	nul
GBX2	0.62	chr3	116266171	116268268	nul	nul	nul	1.09	1.25	1.36	1.13	1.28	1.20	1.11	1.19	nul	1.16	1.28	1.26	nul	nul	nul	nul	nul
GET4	0.62	chr25	42364518	42369462	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.28	1.35	1.29	1.49
MIR2323	0.62	chr17	73635820	73635896	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.28	1.36	1.30	1.31
MMD2	0.62	chr25	39609809	39672870	nul	nul	nul	nul	nul	1.30	nul	nul	nul	nul	nul	nul	nul	1.27	nul	1.39	1.34	1.40	1.42	1.43
NDUFS6	0.62	chr20	70986111	70991386	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.34	1.36	1.33	1.39
NPPC	0.62	chr2	120392625	120397261	nul	nul	1.28	1.16	1.43	1.33	1.26	1.28	1.39	1.30	1.36	nul	1.34	1.33	nul	nul	nul	nul	nul	nul
PITX1	0.62	chr7	48062844	48069063	nul	nul	1.24	1.16	1.29	nul	1.25	nul	1.39	1.24	1.30	1.26	1.32	1.16	1.27	nul	nul	nul	nul	nul
PPP1R7	0.62	chr3	120853354	120869391	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.39	1.37	1.25
PRDM12	0.62	chr11	100972720	100988364	nul	1.15	0.89	1.15	0.77	0.86	1.15	1.17	0.78	1.10	0.94	0.87	1.22	0.73	1.01	nul	nul	nul	nul	nul

Apêndice F. Continuação...

Gene	VST	Chr	Inicio	Fim	A	B	D	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
RADIL	0.62	chr25	39683975	39733712	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.39	1.34	1.40	1.42	nul
SPECC1L	0.62	chr17	73595799	73658019	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.28	1.36	1.30	1.31
TRMT61A	0.62	chr21	69817754	69824182	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.41	1.41	1.23	1.17
ARSA	0.61	chr5	120152351	120155489	nul	nul	nul	1.46	1.40	1.42	1.18	1.40	1.34	1.26	1.25	1.34	0.98	1.46	nul	nul	nul	nul	nul	nul
DIO3	0.61	chr21	68150463	68152565	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.00	0.85	nul	1.42	1.25
MIR1247	0.61	chr21	68149434	68149512	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	1.00	0.85	nul	1.42	1.25
KRT6B	0.6	chr5	27577835	27583351	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	4.16	4.05	3.71	5.19
MEIS2	0.6	chr10	32643641	32866741	nul	nul	1.33	1.20	0.90	1.14	1.33	1.24	0.93	1.02	1.23	nul	1.33	1.09	nul	nul	nul	nul	nul	nul
SLC9B2	0.6	chr6	23096103	23178027	nul	nul	nul	2.65	2.86	2.64	2.88	2.83	3.57	2.95	2.59	nul	3.25	3.35	3.70	nul	nul	nul	nul	nul
STARD3	0.6	chr19	40664259	40688903	nul	nul	1.47	1.23	1.45	1.28	1.46	1.44	1.38	1.42	nul	1.42	1.41	1.23	1.49	1.37	1.31	1.41	1.31	1.16
WDPCP	0.6	chr11	61655883	61970015	nul	nul	1.43	27.88	27.57	28.14	27.55	14.30	13.62	27.29	14.59	13.70	13.93	13.95	33.08	nul	nul	nul	nul	nul
MIR4286-1	0.59	chr19	40683919	40683995	nul	nul	1.47	1.16	1.45	1.21	1.46	1.44	1.38	1.42	nul	1.42	1.41	1.14	1.49	1.37	1.31	1.41	1.45	1.16
PNMT	0.59	chr19	40694078	40695765	nul	nul	1.47	1.16	1.45	1.21	1.46	1.44	1.38	1.42	nul	1.42	1.41	1.14	1.49	1.37	1.31	1.41	1.45	1.16
TCAP	0.59	chr19	40691157	40691886	nul	nul	1.47	1.16	1.45	1.21	1.46	1.44	1.38	1.42	nul	1.42	1.41	1.14	1.49	1.37	1.31	1.41	1.45	1.16
LOC104974629	0.58	chr17	51532235	51565832	nul	nul	nul	13.50	13.44	13.30	12.09	14.02	nul	15.84	15.46	11.69	7.08	11.39	10.77	nul	nul	nul	nul	nul
MIR2350	0.57	chr2	5398826	5398902	nul	nul	nul	1.46	1.34	1.41	1.40	1.40	1.37	1.45	1.33	1.29	1.43	1.32	1.46	1.03	1.33	1.33	1.26	0.79
SEC14L4	0.57	chr17	71606575	71630471	nul	nul	4.40	4.52	3.08	4.32	4.19	2.78	4.16	nul	4.76	1.39	2.94	4.19	5.11	nul	2.09	nul	1.40	1.43
SNCA	0.57	chr6	36285493	36432430	0.91	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	nul	3.74	3.01	nul	3.22	2.56
C17H22orf39	0.56	chr17	74711449	74715833	nul	nul	nul	nul	nul	nul	nul	nul	1.37	nul	nul	1.44	nul	nul	nul	1.41	1.31	1.25	1.30	1.41
D2HGDH	0.56	chr3	121212556	121237816	nul	nul	nul	nul	nul	nul	nul	nul	1.42	nul	nul	1.46	nul	nul	nul	1.31	1.43	1.51	1.38	1.47
ASB3	0.55	chr11	36254028	36371025	nul	0.35	nul	nul	0.05	nul	0.05	0.06	0.04	0.06	0.04	nul	nul	nul	nul	nul	nul	1.04	1.09	0.77
GUK1	0.55	chr7	2895657	2905568	nul	nul	nul	1.38	nul	nul	nul	nul	nul	nul	nul	nul	1.45	nul	nul	1.28	nul	1.31	1.48	1.29
LHX9	0.55	chr16	78788871	78801870	1.05	nul	1.21	1.47	1.11	1.25	1.16	1.15	1.19	1.25	1.07	1.35	1.09	1.17	0.77	nul	nul	nul	0.92	nul
LOC515697	0.55	chr5	74673522	74688915	2.89	nul	3.43	2.63	nul	3.52	2.79	3.24	2.79	2.96	3.46	3.40	2.76	3.64	3.69	nul	nul	nul	nul	nul
EGFL7	0.54	chr11	104128897	104133541	nul	nul	1.09	1.30	1.36	0.98	1.46	1.32	1.42	1.47	0.99	1.34	1.41	1.21	nul	0.96	1.29	1.36	1.25	1.00
MGC138914	0.54	chr18	61608934	61623944	2.58	3.70	2.53	nul	0.47	2.16	0.49	nul	nul	nul	2.20	nul	nul	nul	nul	3.43	2.65	1.94	1.88	3.04
PRDX5	0.54	chr29	43229776	43232885	nul	nul	1.42	1.35	1.38	1.23	1.34	1.27	1.33	1.46	1.21	1.37	1.38	1.36	nul	nul	1.37	nul	1.31	nul
TMEM190	0.54	chr18	62555137	62556184	nul	nul	1.03	1.45	1.44	1.04	1.04	1.43	1.44	1.11	1.47	1.37	1.37	1.45	0.82	1.18	1.35	1.33	1.17	nul
TRMT112	0.54	chr29	43228262	43229598	nul	nul	1.42	1.35	1.38	1.23	1.34	1.27	1.33	1.46	1.21	1.37	1.38	1.36	nul	nul	1.37	nul	1.31	nul
GEMIN7	0.53	chr18	53160658	53167520	nul	nul	1.44	1.39	1.35	1.40	1.46	1.37	1.34	1.39	1.49	1.34	1.32	1.43	nul	nul	1.36	nul	1.39	nul

Apêndice F. Continuação...

Gene	VST	Chr	Inicio	Fim	A	B	D	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
ZNF296	0.53	chr18	53154420	53158137	nul	nul	1.44	1.39	1.35	1.40	1.46	1.37	1.34	1.39	1.49	1.34	1.32	1.43	nul	nul	1.36	nul	1.39	nul
ACADVL	0.52	chr19	27568222	27573378	nul	1.45	1.28	1.39	1.29	1.31	1.31	1.32	1.28	1.50	nul	1.34	1.45	1.33	1.41	nul	1.30	nul	nul	nul
ARHGAP27	0.52	chr19	45679298	45695609	nul	nul	1.39	0.78	1.13	0.53	1.02	0.99	0.73	1.16	1.33	nul	1.17	0.60	nul	nul	nul	nul	nul	nul
DVL2	0.52	chr19	27573454	27581405	nul	1.45	1.28	1.39	1.29	1.31	1.31	1.32	1.28	1.50	nul	1.34	1.45	1.33	1.41	nul	1.30	nul	nul	nul
MIR324	0.52	chr19	27571340	27571429	nul	1.45	1.28	1.39	1.29	1.31	1.31	1.32	1.28	1.50	nul	1.34	1.45	1.33	1.41	nul	1.30	nul	nul	nul
HOXB7	0.51	chr19	38518636	38521061	1.26	nul	nul	1.30	1.19	1.31	1.29	1.33	0.95	1.37	1.25	1.36	1.26	1.29	1.44	nul	nul	nul	nul	nul
PAX6	0.51	chr15	63356625	63384294	1.39	nul	nul	1.26	1.34	1.33	1.27	1.44	1.34	1.40	1.29	1.19	1.36	1.39	1.34	nul	nul	nul	nul	nul
PTMS	0.51	chr5	104039249	104040501	nul	0.70	0.81	1.37	0.67	0.64	0.63	0.79	0.85	0.76	0.84	0.80	0.74	0.68	0.84	nul	0.61	nul	0.52	nul
CNTNAP1	0.5	chr19	43380113	43394560	1.49	1.50	nul	1.28	1.27	1.35	1.19	1.20	1.25	1.24	nul	1.32	1.22	1.28	nul	nul	nul	nul	nul	nul
DLX2	0.5	chr2	24489073	24491347	nul	nul	nul	1.18	1.35	1.37	1.28	1.29	1.39	1.27	nul	1.30	1.30	1.37	nul	nul	nul	nul	nul	nul
HLX	0.5	chr16	25059266	25064942	nul	nul	nul	1.19	1.36	1.42	1.36	nul	1.47	nul	1.42	1.40	1.23	1.38	1.31	nul	nul	nul	nul	nul
RGS17	0.5	chr9	91104615	91353966	nul	nul	nul	41.99	40.74	42.97	41.81	42.80	40.13	40.73	42.99	nul	41.53	41.38	nul	nul	nul	nul	nul	nul