

CNV: uma ferramenta promissora para aplicação em programas de melhoramento genético de bovinos



**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Gado de Corte
Ministério da Agricultura, Pecuária e Abastecimento**

**BOLETIM DE PESQUISA
E DESENVOLVIMENTO
45**

**CNV: uma ferramenta promissora
para aplicação em programas de
melhoramento genético de bovinos**

*Fabiane Siqueira
Poliana Fernanda Giachetto*

**Embrapa Gado de Corte
Campo Grande, MS
2019**

Exemplares desta publicação podem ser adquiridos na:

Embrapa Gado de Corte
Av. Rádio Maia, 830, Zona Rural, Campo Grande, MS,
79106-550, Campo Grande, MS
Fone: (67) 3368 2000
Fax: (67) 3368 2150
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações
da Embrapa Gado de Cortel

Presidente
Thais Basso Amaral

Secretário-Executivo
Rodrigo Carvalho Alva

Membros
Alexandre Romeiro de Araújo, André
Dominghetti Ferreira, Andréa Alves do Egito,
Kadjah Suleiman Jaghub, Liana Jank, Lucimara
Chiari, Marcelo Castro Pereira, Mariane de
Mendonça Vilela, Rodiney de Arruda Mauro,
Wilson Werner Koller

Supervisão editorial
Rodrigo Carvalho Alva

Revisão de texto
Rodrigo Carvalho Alva

Tratamento das ilustrações
Rodrigo Carvalho Alva

Projeto gráfico da coleção
Carlos Eduardo Felice Barbeiro

Editoração eletrônica
Rodrigo Carvalho Alva

Foto da capa
Maury Dorta de Souza Júnior
Arte: <a href="https://br.freepik.com/fotos-
vetores-gratis/tecnologia">Tecnologia foto
criado por kjpgarter - br.freepik.com

1ª edição
1ª impressão (2018): eletrônico

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Nome da unidade catalogadora

Siqueira, Fabiane.

CNV : uma ferramenta promissora para aplicação em progamas de
melhoramento genético de bovinos / Fabiane Siqueira, Poliana Fernanda
Giachetto. - Campo Grande, MS : Embrapa Gado de Corte, 2019.

PDF (36 p.). - (Boletim de Pesquisa e Desenvolvimento / Embrapa Gado
de Corte, ISSN 1983-9715 ; 45).

1. DNA. 2. Fenótipo. 3. Gado. 4. Gene. 5. Genética animal. 6. Marcador
molecular. 7. Variação genética. I. Giachetto, Poliana Fernanda. II. Título. III.
Série.

CDD 636.20821 (23. ed.)

Sumário

Introdução.....	7
Métodos e algoritmos para a identificação das CNVs.....	10
Estudos de CNVs na identificação de genes importantes para a bovinocultura	16
Estudos de identificação de CNVs realizados pela Embrapa ..	22
Conclusão.....	28
Referências	29

CNV: uma ferramenta promissora para aplicação em programas de melhoramento genético de bovinos

Fabiane Siqueira¹

Poliana Fernanda Giachetto²

Resumo – Variações no número de cópias de regiões específicas do DNA, ou simplesmente CNVs (*Copy Number Variations*), constituem um tipo de marcador molecular que é encontrado com frequência nos genomas de mamíferos e plantas. Em humanos, o impacto das CNVs sobre os fenótipos já foi relatado em características adaptativas, mortalidade embrionária e em algumas doenças complexas como autismo, esquizofrenia, doença de Chron, artrite reumatoide, diabetes tipo I e obesidade. Comparadas aos SNPs (*Single Nucleotide Polymorphisms*), as CNVs envolvem um número maior de sequências genômicas, apresentando, potencialmente, um impacto mais pronunciado na expressão gênica, em função de alterações na estrutura e na dosagem gênica. Assim, acredita-se que esse marcador também exerça um papel importante na determinação de fenótipos de interesse em animais de produção, o que torna o mapeamento das CNVs e a sua utilização em estudos de associação genômica ampla (*Genome-wide Association Studies - GWAS*) ações extremamente importantes para prospecção de genes relacionados com características fenotípicas de interesse econômico. Neste contexto, equipes multidisciplinares envolvendo diferentes Unidades da Embrapa têm se dedicado a estudos envolvendo a identificação de CNVs, visando sua aplicação nos programas de melhoramento genético de bovinos de corte e leite.

Termos para indexação: animais de produção, gene candidato, genoma, marcador molecular, variação estrutural.

¹ Bióloga, Doutora em Ciências Biológicas (Genética), Pesquisadora da Embrapa Gado de Corte, Laboratório de Genômica e Melhoramento Animal – GEMA, Campo Grande, MS.

² Zootecnista, Doutora em Produção Animal, Pesquisadora da Embrapa Informática Agropecuária, Laboratório de Bioinformática Aplicada – LBA, Campinas, SP.

Abstract – Copy number variations of specific regions of DNA (CNVs) are a type of molecular marker that is often found in mammalian and plant genomes. In humans, the impact of CNVs on phenotypes has been reported in adaptive traits, embryonic mortality, and in some complex diseases such as autism, schizophrenia, Chron's disease, rheumatoid arthritis, type I diabetes, and obesity. Compared to Single Nucleotide Polymorphisms (SNPs), CNVs involve a larger number of genomic sequences, having a more pronounced impact on gene expression due to changes in structure and gene dosage. Thus, it is believed that this marker also plays an important role in determining phenotypes of interest in production animals, which makes the mapping of CNVs and their use in genome-wide association studies (GWAS) extremely important actions for prospecting genes related to phenotypic traits of economic interest. In this context, multidisciplinary teams involving different Embrapa Units have been dedicated to studies involving the identification of CNVs, aiming at their application in beef and dairy cattle breeding programs.

Index terms: candidate gene, farm animals, genome, molecular marker, structural variation.

Introdução

Variações no número de cópias ou CNVs (*Copy Number Variations*) consistem em um tipo de variação estrutural caracterizado por alterações no número de cópias de regiões específicas do DNA, as quais resultam em duplicações ou deleções de trechos do genoma maiores do que 50 pares de bases (pb) (MacDonald *et al.*, 2014). As CNVs são encontradas com frequência em mamíferos, incluindo humanos (Redon *et al.*, 2006), primatas (Gazave *et al.*, 2011; Gokcumen *et al.*, 2011), roedores (Henrichsen *et al.*, 2009) e animais de produção, tais como suínos (Fadista *et al.*, 2008), aves (Crooijmans *et al.*, 2013), caprinos (Fontanesi *et al.*, 2010), ovinos (Jenkins *et al.*, 2016) e bovinos (Liu *et al.*, 2008).

De acordo com Conrad *et al.* (2010) e Clop *et al.* (2012), o número de nucleotídeos localizados em regiões de CNVs excede o número de SNPs (*Single Nucleotide Polymorphisms*) que são encontrados no genoma, o que leva ao entendimento de que esta fonte de variação genômica pode exercer um enorme impacto nos fenótipos. Estudos recentes demonstraram que no genoma humano o número de CNVs é superior a 12% (Zarrei *et al.*, 2015). Em bovinos da raça Nelore, o número de regiões contendo CNVs - as CNVRs (*Copy Number Variant Regions*), foi estimado em 13,05% do genoma (Lemos *et al.*, 2018).

As regiões compreendidas pelas CNVs muitas vezes contêm genes, ou seja, genes que se pensava ocorrer sempre em duas cópias por genoma podem ser encontrados às vezes em uma, três, ou mais cópias, o que pode levar à alteração dos níveis de expressão gênica ou da estrutura de transcritos (Wang *et al.*, 2019), em função da dosagem e da disrupção gênica (Bhanuprakash *et al.*, 2018). Por outro lado, a remoção e/ou alteração de sequências regulatórias também podem afetar a regulação e a expressão de genes fora da região de CNV (Kleinjan e van Heyningen, 2005).

Essas alterações justificam o seu efeito pronunciado nos fenótipos, contribuindo para a variação observada entre indivíduos, incluindo animais e plantas. Em humanos, o impacto das CNVs sobre os fenótipos varia desde alterações em características adaptativas do indivíduo até a mortalidade embrionária (Hurles *et al.*, 2008), passando por doenças complexas como o

autismo, esquizofrenia, doença de Chron, artrite reumatoide, diabetes tipo I e obesidade (Zhang *et al.*, 2009).

Em animais de produção, os estudos que envolvem a detecção de CNVs têm associado essa variante genética a características de importância econômica (Zhou *et al.*, 2018) e aos mecanismos evolutivos que levaram à domesticação e adaptação dos animais à diferentes condições ambientais (Bhanuprakash *et al.*, 2018; Wang *et al.*, 2015; Wang *et al.*, 2019). Em bovinos, a maior parte das CNVs identificadas até o momento afetam genes envolvidos na imunidade, lactação, reprodução e ruminação (Liu *et al.*, 2010b).

Em plantas, há crescentes evidências indicando que genes afetados por CNVs estão associados com características importantes, como resistência ao nematoide do cisto da soja (Cook *et al.*, 2012), tolerância ao alumínio em milho (Maron *et al.*, 2013), tolerância à toxicidade do boro na cevada (Sutton *et al.*, 2007), aumento no tamanho dos grãos de arroz (Shomura *et al.*, 2008), entre outras.

Diante do exposto e do número crescente de CNVs mapeadas em bovinos, tem-se a possibilidade de usar um marcador molecular alternativo aos SNPs para estudos de associação genômica ampla (*Genome-wide Associations Studies* - GWAS), visando à identificação de genes relacionados com características fenotípicas importantes (Zhou *et al.*, 2018). Dessa forma, este documento apresenta uma revisão da literatura sobre este tema, com o objetivo de despertar o interesse de pesquisadores e acadêmicos para a realização de estudos sobre identificação e análise de CNVs em bovinos, uma vez que a identificação dessas variações e sua associação com fenótipos de interesse econômico pode ser utilizada como ferramenta para aplicação nos programas de melhoramento genético, em especial quando integrada com a genotipagem de marcadores SNPs, que atualmente já é realidade em alguns programas brasileiros de melhoramento de bovinos de corte.

Mecanismos de formação das CNVs

As CNVs podem ser originadas a partir de células somáticas ou germinativas (Gu *et al.*, 2008). No primeiro caso, as CNVs geradas podem resultar em eventos esporádicos, enquanto a ocorrência em células germinativas, se não for deletéria, será herdada e, eventualmente, irá segregar na população em frequências significativas (Clop *et al.*, 2012).

As alterações no número de cópias que caracterizam as CNVs podem ser resultantes de duplicações ou deleções de trechos do genoma que ocorrem, geralmente, por meio de dois mecanismos: a recombinação homóloga não alélica e a recombinação não-homóloga.

Na recombinação homóloga não alélica (NAHR, *Non-Allelic Homologous Recombination*), uma inserção ou deleção podem ocorrer durante a meiose e mitose, devido à recombinação entre duas regiões com sequências similares entre cromossomos não homólogos (Bhanuprakash *et al.*, 2018).

Já na recombinação não homóloga, acredita-se que existam três mecanismos principais de geração de CNVs: junção de extremidades não homólogas (NHEJ, *Non-Homologous End-Joining*), parada na forquilha de replicação e troca da fita molde entre regiões com micro-homologia (FoSTeS, *Fork Stalling and Template Switching*) e retrotransposição mediada por LINEs (*Long Interspersed Nuclear Elements*) (Clop *et al.*, 2012).

O mecanismo de NHEJ é utilizado pelas células para reparar quebras no DNA dupla fita causadas por agentes exógenos, como radiação ionizante, espécies reativas de oxigênio (Lieber *et al.*, 2003), quimioterápicos, ou quando duas forquilhas de replicação convergentes encontram uma quebra no DNA (Hastings *et al.*, 2009). O processo envolve a detecção da quebra na dupla fita, estabelecimento de uma ponte molecular, modificação das extremidades e ligação (Weterings e Gent, 2004). No modelo FoSTeS, a forquilha de replicação para durante o processo em uma determinada posição e então a fita descontínua pode trocar de template, mudando para outra forquilha próxima em função de uma micro-homologia, onde recomeça a síntese de DNA (Sun *et al.*, 2013). No caso da retrotransposição, os transposons comumente envolvidos na formação de CNVs, pelo menos em seres humanos, são os retrotransposons LINE1, em associação com elementos SVA (Herrera *et al.*, 2016), os quais consistem em retroposons compostos pelos elementos SINE (*Short Interspersed Nucleotide Element*), VNTR (*Variable Number of Tandem Repeats*) e Alu (Ostertag *et al.*, 2003).

Em humanos, estudos demonstraram que as CNVs estão distribuídas em *hot spots*, sendo as duplicações segmentais consideradas os principais catalisadores da formação de CNVs (Marques-Bonet *et al.*, 2009). Por outro lado, a análise de CNVs em uma população de vacas holandesas, baseada no

sequenciamento completo do genoma de 32 animais, evidenciou uma distribuição não uniforme das variantes (Mielczarek *et al.*, 2017).

Um número maior de variantes estruturais ocorre em regiões ricas em genes, em comparação ao restante do genoma (Hou *et al.*, 2012a). Dentro dos genes, as regiões não traduzidas (UTRs – *Untranslated Regions*) e íntrons, por sua vez, apresentam maior taxa de variação estrutural do que as regiões codificadoras (Liu *et al.*, 2011). Com relação às categorias de genes, três foram observadas como sendo as mais enriquecidas com CNVs: i) genes envolvidos com imunidade e sinalização célula-célula; ii) genes que codificam proteínas relacionadas à interação com o ambiente, incluindo resposta imune e olfato e, iii) retrovírus e genes codificadores de proteínas relacionadas à transposição (Li e Olivier, 2013). Por outro lado, a ocorrência de CNVs é significativamente reduzida em genes que codificam proteínas que mediam seu efeito através de ligações proteína-proteína (Dopman e Hartl, 2007), genes sensíveis à dosagem (Schuster-Bockler *et al.*, 2010) e genes mais conservados (Li *et al.*, 2011).

Assim como ocorre para outras mutações genéticas, algumas CNVs podem ser herdadas, enquanto outras surgem espontaneamente, as denominadas CNVs *de novo*. van Ommen (2005) estimou que mais de 99% das CNVs são herdadas, enquanto o restante é gerado *de novo* durante a meiose. De acordo com Tizaoui (2018), a incidência de CNVs *de novo* é favorecida por infecções, inflamação crônica, estresse celular, radicais livres gerados pela resposta inflamatória e fatores fisiológicos diversos; variações somáticas também podem causar fragilidade genômica, aumentando a probabilidade de rearranjos genômicos. As variantes herdáveis podem reduzir ou aumentar a expressão de determinados genes, os quais, em combinação com variações somáticas, podem resultar em uma predisposição às doenças (Tzaoui, 2018).

Métodos e algoritmos para a identificação das CNVs

Uma série de técnicas experimentais podem ser utilizadas na detecção de CNVs, incluindo a hibridização fluorescente *in situ* (FISH), hibridização genômica comparativa (CGH) (Kallioniemi *et al.*, 1992), análise de microarranjos de oligonucleotídeos representacional (ROMA) (Lucito *et al.*, 2000; Lucito *et*

al., 2003), arranjos ou chips de genotipagem de SNPs (Dellinger *et al.*, 2010) e, mais recentemente, o sequenciamento de genomas. Dentre essas técnicas, as mais utilizadas têm sido os arranjos de CGH (aCGH), os chips de genotipagem de SNPs e o sequenciamento genômico baseado nas tecnologias de sequenciamento NGS (*Next Generation Sequencing*).

A hibridização genômica comparativa, que substituiu em grande parte a técnica de FISH em função de um maior poder de resolução (Pinkel e Albertson, 2005), consiste em microarranjos, geralmente elaborados a partir de sequências genômicas inseridas em BACs (*Bacterial Artificial Chromosomes*), PACs (*Bacteriophage P1 - derived Artificial Chromosomes*) e cosmídeos, fixadas em uma lâmina, a qual é hibridizada com uma mistura de duas amostras de DNA, teste e referência, onde cada uma delas é marcada com um corante fluorescente diferente (Hupé *et al.*, 2004; Ou *et al.*, 2009). A comparação da medida da intensidade do sinal da amostra teste e da referência (que deve possuir duas cópias de cada região genômica, com exceção das regiões não pseudoautosômicas em cromossomos sexuais, onde uma cópia única é esperada nos machos) permite a identificação de regiões que ganharam ou perderam DNA (Hupé *et al.*, 2004). Regiões com elevada intensidade de hibridização são inferidas como correspondentes a sequências contendo cópias duplicadas em relação à amostra referência, enquanto regiões com baixa intensidade de hibridização em relação à amostra referência, correspondem a sequências deletadas na amostra teste (Schridder e Hahn, 2010).

Atualmente, Affymetrix, Agilent Technologies e Roche NimbleGen são as principais empresas fabricantes de aCGH para ampla cobertura do genoma. No entanto, os arranjos também podem ser confeccionados de maneira customizada, enriquecidos em regiões genômicas sabidamente implicadas com o fenótipo que se deseja estudar (Constantini *et al.*, 2018).

Os microarranjos de genotipagem de SNPs de alta densidade (ou chips de SNPs) têm sido largamente utilizados na detecção das CNVs, em função, principalmente, da cobertura do genoma e custo, aliado ao amplo uso dessa tecnologia nos estudos de associação genômica ampla (GWAS). Os chips de SNPs de alta densidade possuem vantagens em relação aos aCGHs, como um maior poder de resolução do número de cópias, principalmente, no caso da existência de pequenas diferenças entre o genoma referência e a amostra teste (Redon *et al.*, 2006).

A inferência das CNVs a partir dos chips de SNPs requer a obtenção de intensidades de sinal de alta qualidade, assim como informação sobre a frequência alélica para cada loco analisado. A maioria dos métodos utilizados na análise de dados de genotipagem de SNPs para a identificação de CNVs faz a detecção a partir dos dados de LRR (\log_2 da razão de R, onde R é o valor normalizado da intensidade total obtida para os dois alelos do SNP) e BAF (frequência do alelo B; medida normalizada da relação de intensidade alélica para cada SNP). De maneira resumida, se os dois alelos de um SNP são designados A e B e as suas intensidades normalizadas, a e b , respectivamente, a deleção ou duplicação da região do DNA que os contém irá resultar em uma redução ou aumento da intensidade do sinal. Assim, para a estimativa da CNV, a e b são transformados em R, onde $R = a + b$, e θ é calculado, onde $\theta = \arctan(a/b) / (\pi/2)$. R mede a intensidade combinada do sinal dos alelos A e B e θ mede a razão de intensidade alélica relativa (Li e Olivier, 2013).

Devido à natureza dos dados gerados, algoritmos baseados em modelos ocultos de Markov (HMM, *Hidden Markov Models*) têm sido frequentemente utilizados na análise de dados de genotipagem de SNPs para a identificação de CNVs, como é o caso do QuantiSNP (Collela *et al.*, 2007), PennCNV (Wang *et al.*, 2007), GenoCN (Sun *et al.*, 2009), e MixHMM (Liu *et al.*, 2010). Nesses algoritmos, os estados ocultos representam o número de cópias de sondas (probes) adjacentes (Li e Olivier, 2013) e eles diferem na maneira como a probabilidade de transição e emissão são calculadas (Seiser e Innocenti, 2014).

Um dos programas mais utilizados na identificação de CNVs a partir de dados de genotipagem é o PennCNV, uma ferramenta de software livre, desenvolvida em Perl – *Practical Extract and Report Language*, que incorpora vários componentes em um modelo HMM. O PennCNV analisa os padrões de intensidade do sinal em todo o genoma e identifica, por meio dos sinais de marcadores SNPs consecutivos, as mudanças no número de cópias de trechos do genoma (Wang *et al.*, 2007).

O PennCNV utiliza as duas medidas de intensidade de sinal calculadas para cada SNP, o LRR e o BAF, gerados a partir das plataformas de genotipagem conforme descrito acima. A combinação dos valores de LRR e BAF pode ser usada para inferir o número de cópias de regiões do genoma, como ilustrado na Figura 1. 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). Na presença de uma duplicação, há um aumento nos valores de LRR e uma separação do genótipo heterozigoto em dois grupos.

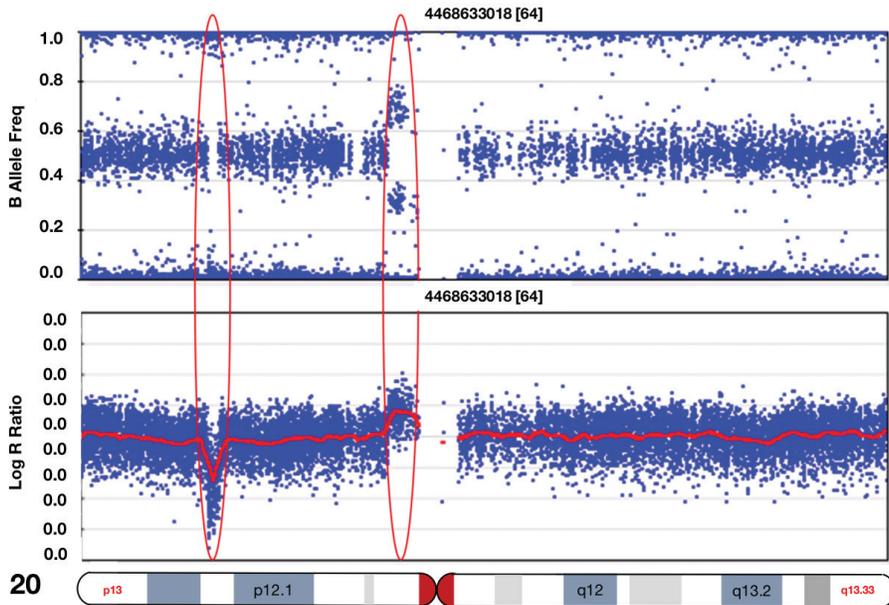


Figura 1. Valores de BAF (frequência do alelo B, painel superior) e LRR (\log_2 da razão de R, onde R é o valor normalizado da intensidade total obtida para os dois alelos do SNP, painel inferior) no cromossomo 20 de um indivíduo, mostrando uma deleção e uma duplicação. Quando se tem uma deleção, os valores de BAF se agrupam ao redor de 0 ou 1 e são ausentes ao redor de 0,5. No caso de uma duplicação, os valores de BAF se agrupam ao redor de 0, 0,33, 0,67, e 1, refletindo os genótipos AAA, AAB, ABB e BBB, respectivamente. No caso do LRR, quando ocorre uma deleção, os valores observados para os marcadores SNP da região deletada diminuem; quando ocorre uma duplicação, os valores de LRR aumentam. Adaptado de Wang *et al.* (2007).

Com o avanço das tecnologias NGS, abordagens de identificação de CNVs baseadas na análise de sequências tornaram-se mais acessíveis. Comparado aos aCGHs e aos chips de genotipagem de SNPs, onde as sondas são pré-definidas para regiões genômicas limitadas, as leituras (ou reads) geradas a partir das plataformas de sequenciamento amostram, de maneira aleatória, todo o genoma analisado (Zhao *et al.*, 2013). As vantagens do sequenciamento ainda incluem uma maior resolução e estimativa

mais acurada das CNVs, identificação mais precisa dos limites das alterações (bordas das CNVs) e também a identificação de variações mais sutis, como é o caso das inversões e pequenas inserções e deleções (Tuzun *et al.*, 2005; Zhao *et al.*, 2013).

A detecção de CNVs baseada em dados de sequenciamento NGS pode ser realizada utilizando-se diferentes abordagens, como o mapeamento de reads pareadas (MRP), profundidade de reads (RD), análise de reads divididas (SR), montagem de novo do genoma ou a combinação de todas elas (Zhao *et al.*, 2013). Cada método possui diferentes potencialidades e limitações e nenhuma ferramenta sozinha é capaz de identificar as variações no DNA em sua totalidade (Pirooznia *et al.*, 2015), sendo os algoritmos baseados em MRP e RD (Figura 2), atualmente, os métodos mais populares para a detecção de CNVs com dados de NGS (Zhang *et al.*, 2019).

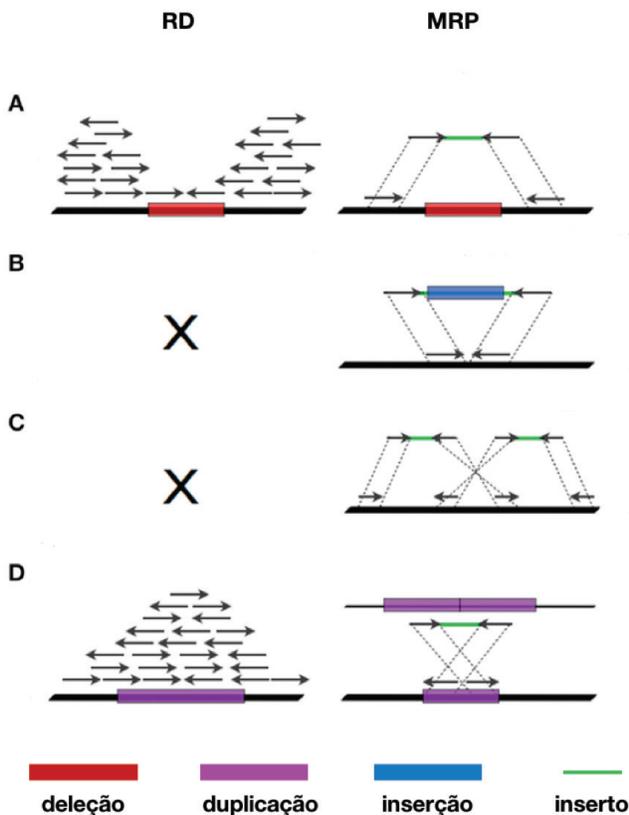


Figura 2. Padrões de CNVs identificadas a partir de dados de sequenciamento total do genoma, utilizando as estratégias de profundidade de reads (RD) e mapeamento de reads pareados (MRP), evidenciando uma deleção (A), inserção de uma nova sequência (B), inversão (C) e duplicação (D). As setas simbolizam as reads. Adaptado de Tattini *et al.* (2015).

O MRP baseia-se na análise das discrepâncias do tamanho e da orientação de múltiplas sequências de reads pareadas em um genoma referência (Tuzun *et al.*, 2005). Distâncias significativamente distintas dos tamanhos de insertos pré-determinados na construção das bibliotecas genômicas são consideradas CNVs e esse método pode identificar não somente duplicações e deleções, mas também inserções de elementos móveis, inversões e duplicações em tandem (Zhao *et al.*, 2013). Por outro lado, o MRP não identifica eventos de inserções maiores do que os tamanhos dos insertos das bibliotecas construídas para o sequenciamento (Medvedev *et al.*, 2009), assim como CNVs em regiões de baixa complexidade com duplicações segmentais (Zhao *et al.*, 2013).

Existem hoje várias ferramentas para a detecção de CNVs a partir de dados de MRP, tais como o PEMer (Korbel *et al.*, 2009), BreakDancer (Chen *et al.*, 2009) e Ulysses (Gillet-Markowska *et al.*, 2014). Essas ferramentas utilizam, geralmente, abordagens de análise baseadas em duas estratégias diferentes, o método de agrupamento e o método baseado em modelo. O método de agrupamento adota a distância pré-definida para identificar as reads discordantes, enquanto o método baseado em modelo utiliza um teste de probabilidade para identificar distâncias não usuais entre os pares de reads, em comparação com a distribuição das distâncias no genoma (Zhao *et al.*, 2013). PEMer e BreakDancer são ferramentas que utilizam o método de agrupamento, enquanto Ulysses emprega testes estatísticos baseados na cobertura relativa da variação candidata (Pirooznia *et al.*, 2015).

Métodos de identificação de CNVs baseados na profundidade das reads (RD) fundamentam-se na hipótese de que a profundidade de cobertura das reads em uma região genômica é correlacionada com o número de cópias da região (Teo *et al.*, 2012). As estratégias utilizadas na análise de RD são classificadas em três categorias, dependendo do delineamento adotado no estudo: amostra única, amostras pareadas caso-controle ou grandes populações de amostras. No caso de amostra única, a estratégia utilizada consiste em estimar a distribuição da profundidade das reads usando modelos matemáticos para identificar regiões com padrões anormais em relação à distribuição observada das reads por todo o genoma; nas amostras pareadas, o número de cópias das amostras teste em relação a controle (genoma referência) é reportado, enquanto que nos estudos baseados em população, a média global da distribuição da profundidade das reads é usada para identificar as CNVs (Zhao *et al.*, 2013).

A estimativa de CNVs por meio do método RD segue basicamente os seguintes passos: i) as reads são alinhadas ao genoma referência e contadas dentro de uma janela de tamanho pré-definida; ii) as contagens são normalizadas para remoção de eventuais vieses devidos ao conteúdo de GC e regiões repetitivas, entre outros; iii) um algoritmo de segmentação é aplicado para identificar as janelas contínuas que possuem o mesmo número de CNVs e determinar os ganhos e perdas de regiões do genoma; iv) a significância estatística das CNVs identificadas é predita e filtros são aplicados (Boeva *et al.*, 2011, Janevski *et al.*, 2012, Zhao *et al.*, 2013, Pirooznia *et al.*, 2015). Comparado aos demais métodos baseados em NGS, o RD é capaz de detectar o número exato de cópias (Salmi *et al.*, 2017) e tem uma acurácia maior na detecção de grandes CNVs (Yoon *et al.*, 2009).

Estudos de CNVs na identificação de genes importantes para a bovinocultura

Em 2008, Liu e colaboradores identificaram as primeiras CNVs em animais taurinos (Liu *et al.*, 2008) em um estudo de baixa escala (somente cinco animais) dentro do *Bovine HapMap Consortium* (<https://bovinegenome.elsiklab.missouri.edu/>). Utilizando a técnica de aCGH, os autores identificaram 25 CNVs em 16 autossomos de três bovinos da raça Holandês e dois Hereford. Em 2010, utilizando uma estratégia combinada, que envolveu a utilização de aCGH e FISH, Liu e colaboradores identificaram 117 CNVRs (*Copy Number Variation Regions*) a partir de *Bos taurus*, *Bos indicus* e raças compostas (Liu *et al.*, 2010). A análise das CNVRs evidenciou diferenças específicas entre as raças na frequência das CNVRs, as quais refletiram aspectos relacionados à ancestralidade dos animais, como a domesticação e formação das raças. A análise dos genes contidos nas CNVRs identificou funções biológicas enriquecidas, tais como imunidade, lactação, reprodução e ruminação.

Com a implementação do *Bovine SNP50 BeadChip* (Illumina), a genotipagem de SNPs em larga escala possibilitou um rápido avanço no estudo das CNVs em bovinos (Matukumalli *et al.*, 2009). Os autores identificaram 79 CNVs a partir da genotipagem de 576 animais de 21 raças bovinas, incluindo raças africanas, compostas, zebuínas e taurinas e verificaram que grande parte das CNVs preditas continham genes que desempenham papéis

importantes na determinação da divergência das subespécies, como genes de receptores olfatórios (OR), genes relacionados à resposta imune (*UL16 binding protein 2* e *ATP-binding cassette, sub-family C (CFTR/MRP), member 4*) e genes do complexo principal de histocompatibilidade (MHC - BoLA) (Matukumalli *et al.*, 2009).

A publicação da sequência do primeiro genoma bovino em 2009 (Elsik *et al.*, 2009) e os avanços na tecnologia de sequenciamento de DNA possibilitaram caracterizações mais profundas das variações genéticas presentes nesses animais (Stothard *et al.*, 2011). De acordo com Zhou *et al.* (2018c), as análises baseadas em métodos que utilizam o sequenciamento genômico são drasticamente mais sensíveis do que as metodologias baseadas em microarranjos na detecção de CNVs, além de possibilitarem a identificação de pequenas inserções e deleções e ainda, com algumas limitações, eventos de inversões e retrotransposição. Assim, observa-se, geralmente, um número maior de CNVs identificadas a partir do re-sequenciamento completo dos genomas dos animais investigados e comparação com o genoma bovino referência. Atualmente, em função do extenso nível de anotação, o genoma referência mais utilizado na identificação de CNVs em bovinos a partir de dados de sequenciamento é o UMD3.1 (<https://bovinegenome.elsiklab.missouri.edu/>), que foi utilizado nos estudos apresentados a seguir.

A partir do sequenciamento de 75 indivíduos representando seis raças bovinas nativas da China e duas raças exóticas especializadas para a produção de carne, Mei *et al.* (2019) identificaram 11.486 CNVRs. A análise de agrupamento dessa população separou os animais em três clados, incluindo *Bos taurus*, *Bos indicus* e *Bos taurus* x *Bos indicus* e o enriquecimento funcional evidenciou genes associados com características econômicas tais como adaptação, produção de carne e cor da pelagem.

Uma detalhada caracterização das CNVs identificadas por meio do sequenciamento total do genoma de 200 animais de raças francesas para corte e leite evidenciou 1.100 variantes abrangendo 1.803 genes, dos quais 86% correspondiam a genes codificadores de proteínas não conhecidas, 66 eram pseudogenes, 26 genes codificadores de micro-RNAs, 87 genes codificadores de pequenos RNAs nucleares e nucleolares, 38 RNAs ribossomais e nove genes codificadores de outros RNAs (Letaief *et al.*, 2017). Um total de

186 regiões deletadas removeram parcialmente ou totalmente um gene, refletindo possíveis alterações na expressão gênica e perturbações nas vias as quais eles pertencem. Ainda, 1.095 variantes estavam localizadas dentro de regiões de QTLs (*Quantitative Trait Loci*) associadas com importantes características em bovinos, como peso corporal, taxa de concepção e teor de gordura no leite (Letaief *et al.*, 2017).

Um total de 14.821 CNVs, incluindo 5.025 duplicações e 9.796 deleções foram encontradas com o sequenciamento de oito touros da raça Holandês com estimativas de valores genéticos extremamente altos e baixos para porcentagem de proteína e gordura no leite, sendo que 487 CNVRs diferenciavam esses dois grupos de animais (Gao *et al.*, 2017). Dentro dessas CNVRs diferenciais, 235 genes funcionais foram identificados, os quais foram significativamente enriquecidos para funções biológicas relacionadas a metabolismo lipídico e proteico, cascata de sinalização de insulina e IGF (*insulin-like growth factor*), prolactina e AMPK (*activated protein kinase*) e 95 das CNVRs estavam sobrepostas com 75 QTLs associadas com proteína e gordura do leite.

O sequenciamento total do genoma de duas raças chinesas de bovinos, Nanyang e Qinchuan, originadas a partir de *Bos indicus* e *Bos taurus*, respectivamente, foi realizado com o objetivo de identificar as características genéticas responsáveis pela variação fenotípica observada entre elas, sendo identificadas 2.907 CNVs, as quais abrangiam 495 genes funcionais, enriquecidos para sistema imune e adaptabilidade ao ambiente (Xu *et al.*, 2017). Os autores observaram que o gene codificador do receptor da leptina (LEPR) apresentou um número de cópias significativamente maior nos animais Qinchuan, o que foi relacionado com a maior deposição de gordura no músculo desses animais.

A elevada sensibilidade e resolução da técnica, associada aos custos decrescentes do sequenciamento genômico, além de aumentar a lista de CNVs identificadas, têm possibilitado a realização de estudos que avançam no sentido de se compreender os padrões de distribuição e segregação das mesmas nas populações de bovinos. Este avanço, no entanto, é criticamente dependente da qualidade dos genomas montados (Couldrey *et al.*, 2017), os quais, segundo os autores, ainda ficam atrás da qualidade observada atualmente para o genoma humano.

Tecnologias de sequenciamento que geram reads longas, que compreendem dezenas de milhares de pares de bases, como a PacBio, estão contribuindo para o fechamento das lacunas e montagem de regiões complexas das atuais sequências dos genomas referência (Rhoads e Au, 2015). Este fato, juntamente com a capacidade das tecnologias de reads longas em sequenciar extensas regiões repetitivas, fazem delas a plataforma mais indicada para o uso na identificação de CNVs, embora o seu custo ainda seja limitante (Couldrey *et al.*, 2017).

Nesse sentido, Mielczarek *et al.* (2017) sequenciaram o genoma completo de 32 vacas da raça Holandês para avaliar a variabilidade entre os indivíduos com relação à distribuição e tamanho das CNVs, além de anotar funcionalmente o efeito da posição das mesmas. Uma elevada e significativa variação foi observada no tamanho e número de CNVs entre os animais, sendo a maior parte das deleções (mais numerosas) e duplicações localizadas em regiões intergênicas e íntrons e poucas identificadas em transcritos não codificadores e regiões de *splicing*. De maneira geral, genes que apresentam baixa pressão de seleção mostraram um número significativamente maior de CNVs, sendo que o gene que codifica para uma proteína quinase dependente de cGMP do tipo I exibiu o maior número de deleções, enquanto um receptor de células T apresentou o maior número de duplicações.

Com o objetivo de caracterizar as diferenças no conteúdo gênico, regulação e estrutura entre o genoma bovino e o do búfalo de rio, Li *et al.* (2019) utilizaram o genoma referência UMD3.1 como base para comparações com o genoma de 14 bubalinos. Os autores identificaram 127 CNVRs deletadas no genoma dos animais e, por meio dessas regiões, descobriram cinco genes ausentes no genoma nos búfalos, em comparação aos bovinos. Os genes deletados foram confirmados por análises de expressão gênica dos animais e associados às variações fenotípicas observadas entre as espécies.

Utilizando uma estratégia híbrida, que envolveu o sequenciamento PacBio (*reads* longas) de um touro representativo de rebanhos leiteiros da Nova Zelândia e o sequenciamento Illumina (*reads* curtas) de 556 animais dentro da população, incluindo duos e trios, Couldrey *et al.* (2017) utilizaram uma metodologia baseada no pedigree e identificaram um grupo de CNVs segregantes nos animais, sendo observado um padrão de distribuição multimodal nas CNVs classificadas como altamente transmissíveis.

Embora a maior parte dos estudos de CNVs em animais de produção ainda estejam relacionados à sua descoberta e predição de seus efeitos sobre os fenótipos com base na função dos genes encontrados dentro das regiões de CNVs (Zhou *et al.*, 2016), uma série de estudos de associação genômica ampla (GWAS) utilizando esse tipo de marcador como alternativa aos SNPs, já têm sido publicados.

Utilizando o *Bovine SNP50 BeadChip*, Xu *et al.* (2014a) identificaram em fêmeas da raça Holandês 34 CNVs significativamente associadas com características relacionadas à produção de leite. Em animais da raça Angus, a mesma equipe identificou uma deleção de CNV significativamente associada com OPG (ovos por grama de fezes) em um estudo relacionado à resistência da população à nematoides gastrintestinais (Xu *et al.*, 2014b). Hou *et al.* (2012a) também identificaram CNVs significativamente associadas às diferenças na resistência a nematoides gastrintestinais comparando duas populações de animais Angus, sendo uma caracterizada como resistente e outra suscetível aos parasitas.

A mesma estratégia de comparação de dois grupos de animais com fenótipos extremos, no caso para eficiência alimentar, utilizando o consumo alimentar residual (CAR) de vacas Holandesas como estimativa, foi utilizada por Hou *et al.* (2012b). Os autores encontraram CNVRs específicas a cada grupo e, analisando os genes contidos nessas regiões, identificaram uma prevalência de genes relacionados à resposta inflamatória e imunidade nos animais mais eficientes para CAR, enquanto nos animais menos eficientes, genes envolvidos com desenvolvimento ósseo e de órgãos foram mais representados (Hou *et al.*, 2012b).

Ben Sassi *et al.* (2016), utilizando o *Bovine SNP50 BeadChip*, detectaram em animais da raça Holandês 15 CNVRs associadas com características de conformação e produção, tais como proteína e gordura no leite, estatura das vacas, ângulo do pé, profundidade do úbere e dias vazios. Também em animais da raça Holandesa, Durán Aguilar *et al.* (2017) utilizaram o *BovineHD Genotyping BeadChip* (Illumina) para identificar CNVs e por meio de análises de associação (GWAS) encontraram CNVRs significativamente associadas com o escore de contagem de células somáticas do leite, além de 18 genes candidatos nestas regiões.

Ainda em animais da raça Holandês, um estudo baseado em GWAS identificou 87 CNVRs significativamente associadas com características complexas em rebanhos leiteiros, relacionadas à reprodução, produção e conformação corporal, a partir da genotipagem de 47 touros utilizando a estratégia de aCGH (Liu *et al.*, 2019).

Analisando amostras de 2.230 animais Nelore com o BovineHD, Zhou *et al.* (2016) encontraram 17 CNVs significativamente associadas com a conformação, musculatura e precocidade à desmama e ao sobreano e ganho de peso à desmama. Silva *et al.* (2016) genotiparam bovinos Nelore machos e por meio de GWAS identificaram seis regiões associadas com maciez da carne, incluindo uma região do cromossomo 3 que contem genes da via do cAMP e cGMP, os quais são conhecidos por influenciar a hipertrofia, metabolismo e regeneração do músculo esquelético (Berdeaux e Stewart, 2012). Uma associação significativa de 12 CNVRs com a taxa de conversão alimentar de animais Nelore foi identificada a partir de dados de genotipagem realizada em quatro diferentes plataformas - *Bovine SNP50 BeadChip*, *BovineHD BeadChip*, *GGP Indicus Neogen HD* e *Affymetrix Axiom® Genome-Wide BOS 1 Array*, sendo os dados avaliados conjuntamente por meio de imputação. Os genes encontrados nas regiões de CNVs foram relacionados a metabolismo energético, lipídico e protéico, os quais, segundo os autores, são diretamente relacionados com eficiência alimentar em gado de corte (Santana *et al.*, 2016).

Com o objetivo de analisar CNVs que diferenciavam as linhagens, Yang *et al.* (2017) identificaram, por meio de GWAS de uma população composta por oito raças de bovinos nativos da China genotipada com o chip *BovineSNP50*, CNVs associadas à seleção divergente para adaptação às condições ambientais. Foram identificados 13.225 eventos de CNVs e 3.356 CNVRs, as quais continham 1.522 genes anotados. Estes genes foram relacionados a importantes funções moleculares e processos biológicos, tais como processos metabólicos, resposta a estímulos e sistema imune.

Estudos de identificação de CNVs realizados pela Embrapa

Dados de genotipagem de bovinos de corte e de leite utilizando chips de DNA de diferentes densidades são largamente gerados na Empresa Brasileira de Pesquisa Agropecuária (Embrapa) para estudos de associação de genótipos com características de interesse econômico. Estes estudos visam à geração de conhecimentos e de soluções de inovação que possam ser aplicadas ao melhoramento genético dos rebanhos nacionais. Com relação

à identificação e análises de CNVs, alguns trabalhos científicos já foram realizados ou ainda estão em andamento em diferentes Unidades da Embrapa.

Silva *et al.* (2016b) realizaram a primeira análise ampla de CNVs em uma população de bovinos Nelore composta por 1.717 animais genotipados em alta densidade (BovineHD com aproximadamente 770 K SNPs). Além disso, os autores sequenciaram o genoma completo de oito touros ancestrais importantes para a raça, com uma cobertura mínima de 20X. Objetivou-se com este estudo fazer uma análise de alta resolução para detectar e caracterizar CNVs na raça Nelore, além de estimar frequências de variantes específicas desta raça. Um total de 68.007 CNVs, representando 54.510 duplicações de cópia única, 1.729 duplicações de cópia dupla, 11.672 deleções de cópia única e 96 deleções de cópia dupla, foram identificadas com a análise de dados de genotipagem. O número de SNPs em cada CNV detectada variou de 20 a 1.420 e o comprimento das CNVs variou de 20,01 Kb a 7,75 Mb. Para os dados de NGS dos touros foram detectadas um total de 12.786 CNVs, representando 999 duplicações e 11.787 deleções, com tamanhos médios de $252,8 \pm 692,0$ Kb e $22,9 \pm 194,2$ Kb, respectivamente.

De acordo com os autores, diferentes métodos para condensar CNVs sobrepostas em regiões variantes de número de cópias já foram propostos (Butter *et al.*, 2013; Kim *et al.*, 2012; Glessner *et al.*, 2013). As CNVRs da raça Nelore foram identificadas usando o JM-CNV (*Java Merging Copy Number Variants*), que considera o comprimento e a frequência da CNV, remove CNVs extremamente longas ou pouco frequentes da análise inicial e resolve problemas observados no ponto de interrupção (Alkan *et al.*, 2011; Meyerson *et al.*, 2010). As 68.007 CNVs detectadas foram condensadas em 7.319 CNVRs, representando uma cobertura total de 1,56 Gb (61,91%) do genoma autossômico bovino. Um total de 2.306 duplicações, 212 deleções e 4.801 duplicações e deleções foram observadas nas CNVRs identificadas. Foi relatada uma alta correlação positiva entre o número de CNVRs detectadas e o tamanho dos cromossomos bovinos (0,98), contrariamente ao que foi observado em termos do número total de CNVs detectadas (0,34). Os autores verificaram que um total de 4.097 CNVRs (55,98%) estão localizadas em regiões do genoma contendo 10.399 genes anotados, que podem ser classificados funcionalmente como codificadores de proteínas ($n = 10.070$), microRNAs ($n = 159$), pequenos RNAs nucleolares - snoRNA ($n = 148$), pequenos RNAs nucleares - snRNA ($n = 10$), pequenos RNAs diversos que não

se enquadram nas outras definições de RNAs - miscRNA (n = 8) e RNAs ribossomais - rRNA (n = 2). A anotação automatizada desses genes com termos do *Gene Ontology* (GO) revelou categorias importantes, incluindo processos metabólicos e celulares, regulação biológica, resposta a estímulos, sinalização celular, reprodução e crescimento. Portanto, uma investigação mais aprofundada dessas regiões poderá revelar informações importantes para compreensão dos mecanismos subjacentes que afetam características econômicas importantes em bovinos.

Os resultados obtidos por Silva *et al.* (2016b) permitiram comparações diretas entre os resultados da detecção de CNVs por meio da utilização de duas plataformas distintas (*BovineHD* e dados de NGS) com resultados de estudos independentes publicados na literatura. Dessa forma, o mapa de CNVs de bovinos foi significativamente enriquecido e foi possível identificar variantes potencialmente associadas com características de interesse, principalmente nas regiões do genoma que abrigam QTLs que afetam as características de produção. Este trabalho fornece informações importantes que poderão contribuir para o desenvolvimento de estudos futuros de associação entre CNVs e características de produção importantes para o melhoramento genético em bovinos.

Silva *et al.* (2016a), também utilizando a versão HD do Bovine BeadChip (Illumina), investigaram 723 bovinos machos da raça Nelore e identificaram 49.997 CNVs e 2.649 CNVRs, usando os softwares PennCNV (Wang *et al.*, 2007) e CNVRuler (Kim *et al.*, 2012), respectivamente. A proporção cromossômica coberta pelas CNVRs variou entre os cromossomos (de 2,3% no BTA22 a 19,7% no BTA15) e o número de regiões com perda e ganho de cópias foi de 1.454 e 891, respectivamente. CNVRs com tamanho entre 5 e 50 kb representaram a maioria das regiões identificadas (71,4%), enquanto CNVRs maiores que 1 Mb raramente foram observadas (0,5%).

Os autores compararam esses resultados com estudos anteriores e observaram uma sobreposição em aproximadamente 1.400 CNVRs (> 50%). Um total de 1.155 CNVRs (43,6%) se sobrepôs a 2.750 genes, os quais se mostraram enriquecidos para processos envolvendo guanosina tri-fostato (GTP), que influenciam a morfologia e fisiologia do músculo esquelético. Os CNVRs da raça Nelore também se sobrepuseram à QTLs para maciez de carne relatados em outras raças (8,9%, 236 CNVRs) e em um estudo an-

terior realizado com essa mesma população (4,1%, 109 CNVRs – Tizioto *et al.*, 2013). Várias CNVRs polimórficas (encontradas em mais de 1% da população) se sobrepuseram aos QTLs associados com maciez de carne em bovinos no banco de dados QTLdb (Hu *et al.*, 2013). As famílias de genes das enzimas peptidase e protease enriquecidas nessas regiões têm um papel fundamental no amaciamento da carne durante o processo post-mortem, pois estão envolvidas na proteólise associada à maciez e na renovação do tecido conjuntivo (Sentandreu *et al.*, 2002). De acordo com os autores, duas CNVRs também estão localizadas próximas aos genes do metabolismo da glutatona, os quais também foram anteriormente associados à maciez de carne. Dez CNVRs que se sobrepuseram às regiões associadas com maciez de carne foram validadas com sucesso utilizando a metodologia de PCR quantitativo em Tempo Real (*Polymerase Chain Reaction Quantitative Real Time*).

Em bovinos de leite, Silva *et al.* (2014) investigaram, com base no chip *BovineHD*, a abundância e a distribuição de CNVs e CNVRs em uma população de 476 bovinos Gir brasileiros. Para identificação das CNVs foi utilizado o software PennCNV (Wang *et al.*, 2007) e para determinação das CNVRs o software CNVRuler (Kim *et al.*, 2012). Foram encontradas 26.672 CNVs, sendo em média 62 CNVs por animal. Além disso, 1.898 CNVRs foram detectadas nos cromossomos autossômicos, com 96% destes entre 1,1 Kb e 100 Kb. A ferramenta VEP (*Variant Effect Predictor*) do Ensembl, usando as informações das CNVRs como entrada, encontrou 913 regiões codificadoras, sugerindo que as regiões dos éxons estavam duplicadas.

Carmo *et al.* (2015) identificaram CNVs associadas com características reprodutivas e produtivas em 475 bovinos Gir leiteiro genotipados no chip *BovineHD* da Illumina. As PTAs (*Predicted Transmitting Ability*) para as características de produção de leite, de gordura, de proteína e produção de sólidos avaliadas em até 305 dias, além de percentual de gordura, de proteína, de sólidos e idade ao primeiro parto foram preditas pela Embrapa Gado de Leite com base na lactação das vacas Gir presentes no Arquivo Zootécnico Nacional de 2014. Um total de 735.303 SNPs foram mapeados em 29 cromossomos autossômicos na montagem do genoma bovino UMD3.1. Foram observadas 155 CNVs associadas significativamente com as oito características estudadas. Dentre os genes associados com mais de três características muitos foram previamente relatados como associados com características reprodutivas e produtivas em diversas raças bovinas. De acordo com os au-

tores, os resultados obtidos indicam que as CNVs identificadas apresentam relação funcional direta com a produção de leite e que uma mesma CNV pode estar envolvida com a manifestação de diversas características de interesse econômico em bovinos da raça Gir.

A fim de fornecer uma base para futuros estudos de associação, Chud *et al.* (2018) identificaram CNVRs na raça composta Girolando, resultante de acasalamento entre as raças Holandês (taurina) e Gir (zebuína). Um método de profundidade de leitura foi realizado usando o software CNVnator em dados NGS de dois touros Girolando, dois Gir e 10 holandeses. As CNVs individuais foram fundidas em CNVRs com base em regiões genômicas que se sobrepuseram em pelo menos 1 pb. No total, foram identificadas 1.286 CNVRs, sendo 520 deleções, 255 duplicações e 511 regiões mistas nos genomas de todas as amostras. Os autores observaram 34 CNVRs, sendo nove deleções e 25 mistas em comum (sobreposição > 50%) apenas entre Girolando e a raça Holandesa, e 181 CNVRs (20 deleções, 21 duplicações, 140 mistas) apenas entre Girolando e Gir, sugerindo herança de origem parental das raças Holandesa e Gir, respectivamente. Em uma dessas CNVRs específicas da raça Holandesa encontra-se o gene da citocina da família interleucina 6 (LIF), que está ligado à produção de gordura e características de fertilidade em animais holandeses. Genes relacionados com resistência a doenças (por exemplo, o gene CD4) também coincidiram com as CNVRs presentes apenas em bovinos Gir e Girolando, sugerindo uma origem zebuína para a CNV. Esses resultados mostraram evidências de CNVRs específicas compartilhadas entre Girolando e raças puras que poderão ser analisadas em análises futuras visando sua aplicação nos programas de melhoramento dessas raças.

Atualmente, está em fase final de conclusão na Embrapa Gado de Corte, em parceria com a Embrapa Instrumentação Agropecuária, uma atividade de pesquisa que visa identificar CNVs e genes associados com características de interesse econômico em bovinos da raça Canchim. De 2011 a 2019 foram coletadas pela Embrapa Gado de Corte, com o apoio da ABCCAN (Associação Brasileira dos Criadores de Canchim), de criadores, do Programa Embrapa de Melhoramento de Bovinos de Corte – Geneplus e da Embrapa Pecuária Sudeste, amostras biológicas e informações fenotípicas de, aproximadamente, 1.800 touros jovens participantes das Provas Canchim de Avaliação de Desempenho (PCAD). As características avaliadas foram: peso ao nascimento, peso ao sobreano, ganho de peso no confinamento, perímetro escrotal, conformação frigorífica, área de olho

de lombo, marmoreio, espessura de gordura na carcaça, pelagem, pigmentação de mucosas, aprumos e comprimento de umbigo. Até o momento, aproximadamente 1.100 animais das PCADs foram genotipados em painéis de SNPs de alta e média densidade da plataforma Illumina.

Com esse banco de dados estão sendo desenvolvidas ações de pesquisa que visam avaliar a viabilidade de implementação da seleção genômica nas avaliações genéticas desta raça e estudos de associação que objetivam identificar regiões cromossômicas, genes e marcadores moleculares que estejam relacionados com a manifestação de fenótipos de interesse econômico, com o intuito de desenvolver ferramentas genômicas que possam ser aplicadas no programa de melhoramento da raça.

O Canchim é fruto de um trabalho de pesquisa iniciado na década de 1940, na Embrapa Pecuária Sudeste, por pesquisadores brasileiros que buscavam um animal que reunisse as características produtivas das raças taurinas com a rusticidade dos zebuínos, sendo esta uma característica extremamente importante para aumentar a produtividade dos rebanhos em regiões tropicais. Para a formação da raça Canchim foram utilizados touros Charolês e fêmeas das raças Indubrasil, Nelore e Guzerá, originando um animal com a proporção genética de 5/8 Charolês e 3/8 zebuíno.

O fato de ter nascido dentro de uma Instituição de Pesquisa determinou o sucesso desta raça, já que sua formação e posterior expansão sempre foram pautadas por rigorosos critérios técnicos de seleção, sendo desde o seu princípio exaustivamente posto à prova e avaliado para que características de interesse econômico fossem aprimoradas e fixadas na raça. Passados quase 70 anos do seu desenvolvimento, o Canchim de hoje não é o mesmo 5/8 Charolês/Zebu do passado, é um animal lapidado racialmente, mais produtivo, melhor adaptado às condições ambientais e com bom potencial de ganho de peso e rendimento de carcaça, sendo uma das melhores opções para cruzamento a campo em qualquer região do Brasil.

A partir da genotipagem dos animais participantes das PCADs, utilizando os chips *BovineSNP50 BeadChip*, *BovineHD BeadChip*, *BovineLD Genotyping BeadChip* e *BovineHD Genotyping BeadChip*, foram identificadas 6.488 CNVs, representando 1.241 deleções e 5.247 duplicações. Análises de enriquecimento dos genes identificados nessas regiões estão sendo realizadas, assim como a identificação de sobreposições com QTLs previamente relatados na literatura.

Este estudo é a primeira análise ampla realizada com animais Canchim, sendo que os resultados gerados auxiliarão na compreensão dos mecanismos genéticos que controlam características fenotípicas importantes para a raça e poderão permitir o desenvolvimento de novas estratégias de seleção de animais superiores.

Conclusão

Variações genômicas estruturais como os SNPs têm sido largamente utilizadas na atualidade, na identificação de marcadores genéticos associados às características de crescimento e produção em bovinos, assim como no estudo de aspectos evolutivos e relacionados à conservação das espécies. O uso desses marcadores por sua vez, tem possibilitado a adoção de práticas em programas de melhoramento genético como a seleção genômica, ferramenta que tem levado a incrementos importantes na acurácia da seleção e aceleração do progresso genético de rebanhos.

As CNVs constituem um novo tipo de marcador molecular, que em humanos têm sido associadas com características fenotípicas importantes, principalmente, doenças complexas. Comparada aos SNPs, as CNVs envolvem um número maior de sequências genômicas, apresentando, potencialmente, um impacto mais pronunciado na expressão gênica, em função de alterações na estrutura e na dosagem gênica. Assim, acredita-se que esse marcador tenha um papel considerável na determinação de fenótipos de interesse em animais de produção, o que torna o mapeamento das CNVs, assim como a sua aplicação em estudos de associação genômica ampla, ações desejáveis em programas de melhoramento genético.

Equipes multidisciplinares envolvendo a Embrapa Gado de Corte, Embrapa Gado de Leite, Embrapa Informática Agropecuária, Embrapa Pecuária Sul e Embrapa Pecuária Sudeste têm se dedicado a estudos envolvendo a identificação de CNVs, principalmente, em bovinos das raças Canchim, Nelore e raças leiteiras. Os resultados obtidos têm mostrado a associação significativa de CNVs ou CNVRs com características como qualidade da carne e do leite, adaptabilidade e resistência a doenças e parasitas e, dada a quantidade considerável de dados já obtidos na Embrapa com a genotipagem de SNPs em larga escala e a crescente acessibilidade ao sequenciamento do genoma dos animais, vislumbra-se um incremento bastante interessante nos índices produtivos da bovinocultura brasileira de carne e leite.

Referências

- ALKAN, C.; COE, B. P.; EICHLER, E. E. Applications of next-generation sequencing genome structural variation discovery and genotyping. **Nature Reviews Genetics**, v. 12, n. 5, p. 363–375, 2011.
- BHANUPRAKASH, V.; CHHOTARAY, S.; PRUTHVIRAJ, D. R.; RAWAT, C.; KARTHIKEYAN, A.; PANIGRAHI, M. Copy number variation in livestock: A mini review. **Veterinary World**, v. 11, n. 4, p. 535, 2018.
- BEN SASSI, N.; GONZÁLEZ-RECIO, Ó.; DE PAZ-DEL RÍO, R.; RODRÍGUEZ-RAMILO, S. T.; FERNÁNDEZ, A. I. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. **Journal of Dairy Science**, v. 99, p. 6371–80, 2016.
- BERDEAUX, R.; STEWART, R. cAMP signaling in skeletal muscle adaptation: hypertrophy, metabolism, and regeneration. **American Journal of Physiology - Endocrinology and Metabolism**, v. 303, p. E1–E17, 2012.
- BOEVA, V.; ZINOVYEV, A.; BLEAKLEY, K.; VERT, J. P.; JANOUÉIX-LEROSÉY, I.; DELATTRE, O.; *et al.* Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. **Bioinformatics**, v. 27, p. 268–269, 2011.
- BUTLER, J. L.; LOCKE, M. E. O.; HILL, K. A.; DALEY, M. HD-CNV: hotspot detector for copy number variants. **Bioinformatics**, v. 29, n. 2, p. 262–263, 2013.
- CARMO, A. S.; OLIVEIRA JR., G. A.; CHUD, T. O. S.; PANETTO, J. C. C.; VERNEQUE, R. S.; MACHADO, M. A.; COLE, J. B.; SILVA, M. V. G. B. Identificação de CNVs associadas com características reprodutivas e produtivas em animais Gir leiteiro. In: XXV CONGRESSO BRASILEIRO DE ZOOTECNIA – ZOOTEC 2015, FORTALEZA, CE, 2015.
- CHEN, K.; WALLIS, J. W.; MCLELLAN, M. D.; LARSON, D. E.; KALICKI, J. M.; POHL, C. S.; *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. **Nature Methods**, v. 6, p. 677–681, 2009.
- CHUD, T. C. S.; BICKHART, D. M.; ZERLOTINI NETO, A. Z.; COLE, J. B.; SILVA, M. V. G.; MUNARI, D. P. Copy number variation in dairy cattle using next-generation sequencing. In: PLANT AND ANIMAL GENOME CONFERENCE, 26., 2018, San Diego. Abstracts... [S.l.: s.n.], 2018.
- CLOP, A.; VIDAL, O.; AMILLS, M. Copy number variation in the genomes of domestic animals. Copy number variation in the genomes of domestic animals. **Animal Genetics**, v. 43, n. 5, p. 503–517, 2012.
- COULDREY, C.; KEEHAN, M.; JOHNSON, T. *et al.* Detection and assessment of copy number variation using PacBio long-read and Illumina sequencing in New Zealand dairy cattle. **Journal of Dairy Science**, v. 100, n. 7, p. 5472–5478, 2017.
- COLELLA, S.; YAU, C.; TAYLOR, J. M.; MIRZA, G.; BUTLER, H.; CLOUSTON, P.; BASSETT, A. S.; SELLER, A.; HOLMES, C. C.; RAGOSSIS, 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.
- CONRAD, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. **Nature**, v. 464, p. 704–712, 2010.
- COOK, D. E.; LEE, T. G.; GUO, X.; *et al.* Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. **Science**, v. 338, n. 6111, p. 1206–1209, 2012.

- COSTANTINI, A.; SKARP, S.; KÄMPE, A.; MÄKITIE, R. E.; PETTERSSON, M.; MÄNNIKKÖ, M.; JIAO, H.; TAYLAN, F.; LINDSTRAND, A.; MÄKITIE, O. Rare copy number variants in array-based comparative genomic hybridization in early-onset skeletal fragility. **Frontiers in Endocrinology**, v. 9, p. 380, 2018.
- CROOIJMANS, R. P.; FIFE, M. S.; FITZGERALD, T. W.; STRICKLAND, S.; CHENG, H. H.; KAISER, P.; REDON, R.; GROENEN, M. A. Large scale variation in DNA copy number in chicken breeds. **BMC Genomics**, v. 14, n. 1, p. 398, 2013.
- DELLINGER, A. E.; SAW, S. M.; GOH, L. K.; SEIELSTAD, M.; YOUNG, T. L.; Li, Y. J. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. **Nucleic Acids Research**, v. 38, n. 9, p. e105-e105, 2010.
- DOPMAN E. B.; HARTL, D. L. A portrait of copy-number polymorphism in *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences of the United States of America USA**, v. 104, p. 19.920-19.925, 2007.
- DURÁN AGUILAR, M.; ROMÁN PONCE, S. I.; RUIZ LÓPEZ, F. J.; GONZÁLEZ PADILLA, E.; VÁSQUEZ PELÁEZ, C. G.; BAGNATO, A.; *et al.* Genome-wide association study for milk somatic cell score in Holstein cattle using copy number variation as markers. **Journal of Animal Breeding and Genetics**, v. 134, p. 49-59, 2017.
- ELSIK, C. G.; TELLAM, R. L.; WORLEY, K. C. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v. 324, n. 5926, p. 522-528, 2009.
- FADISTA, J.; NYGAARD, M.; HOLM, L. E.; THOMSEN, B.; BENDIXEN, C. A snapshot of CNVs in the pig genome. **PLoS One**, v. 3, p. e3916, 2008.
- FONTANESI, L.; MARTELLI, P. L.; BERETTI, F.; RIGGIO, V.; DALL'OLIO, S.; COLOMBO, M.; CASADIO, R.; RUSSO, V.; PORTOLANO, B. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome. **BMC Genomics**, v. 11, n. 1, p. 639, 2010.
- 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, n. 1, p. 265, 2017.
- GAZAVE, E.; DARRÉ, F.; MORCILLO-SUAREZ, C.; PETIT-MARTY, N.; CARREÑO, A.; MARIGORTA, U. M.; RYDER, O. A.; BLANCHER, A.; ROCCHI, M.; BOSCH, E.; BAKER, C. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. **Genome Research**, v. 21, n. 10, p. 1626-1639, 2011.
- GILLET-MARKOWSKA, A.; RICHARD, H.; FISCHER, G.; LAFONTAINE, I. Ulysses: accurate detection of low-frequency structural variations in large insert-size sequencing libraries. **Bioinformatics**, v. 31, p. 801-808, 2014.
- GLESSNER, J. T.; LI, J.; HAKONARSON, H. ParseCNV integrative copy number variation association software with quality tracking. **Nucleic Acids Research**, v. 41, n. 5, p. e64, 2013.
- GOKCUMEN, O.; BABB, P. L.; ISKOW, R. C.; ZHU, Q.; SHI, X.; MILLS, R. E.; IONITA-LAZA, I.; VALLENDER, E. J.; CLARK, A. G.; JOHNSON, W. E.; LEE, C. Refinement of primate copy number variation hotspots identifies candidate genomic regions evolving under positive selection. **Genome Biology**, v. 12, n. 5, p. R52, 2011.
- GU, W.; ZHANG, F.; LUPSKI, J. R. Mechanisms for human genomic rearrangements. **Pathogenetics**, v. 1, n. 1, p. 4, 2008.
- HASTINGS, P. J.; LUPSKI, J. R.; ROSENBERG, S. M.; Ira, G. Mechanisms of change in gene copy number. **Nature Reviews Genetics**, v. 10, n. 8, p. 551-564, 2009.

- HENRICHSEN, C. N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v. 18, p. R1-R88, 2009.
- HERRERA, R. J.; GARCIA-BERTRAND, R.; SALZANO, F. M. **Genes and genomics dynamics**. In: Genomes, Evolution, and Culture. Past, Present and Future of Humankind. p. 87-123, 2016. John Wiley & Sons. USA: Wiley-Blackwell. Wiley, Chichester 254 pp.
- HOU, Y. *et al.* 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.; CHUNG, H.; HUTCHISON, J. L.; NORMAN, H. D.; CONNOR, E. E.; *et al.* Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. **Functional & Integrative Genomics**, v. 12, p. 717-23, 2012b.
- HU, Z-L.; PARK, C. A.; WU, X-L.; REECY, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, p. D871–D879, 2013.
- JANEVSKI, A.; VARADAN, V.; KAMALAKARAN, S.; BANERJEE, N.; DIMITROVA, N. Effective normalization for copy number variation detection from whole genome sequencing. **BMC Genomics**, v. 13 (Suppl. 6), p. S16, 2012.
- JENKINS, G. M.; GODDARD, M. E.; BLACK, M. A.; BRAUNING, R.; AUVRAY, B.; DODDS, K. G.; KIJAS, J. W.; COCKETT, N.; MCEWAN, J. C. Copy number variants in the sheep genome detected using multiple approaches. **BMC Genomics**, v. 17, n. 1, p. 441, 2016.
- KALLIONIEMI, A.; KALLIONIEMI, O. P.; SUDAR, D.; RUTOVITZ, D.; GRAY, J. W.; WALDMAN, F.; PINKEL, D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. **Science**, v. 258, p. 818-821, 1992.
- 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, n. 13, p. 1790–1792, 2012.
- KLEINJAN, D. A.; VAN HEYNINGEN, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. **American Journal of Human Genetics**, v. 76, p. 8-32, 2005.
- KORBEL, J. O.; URBAN, A. E.; AFFOURTIT, J. P.; GODWIN, B.; GRUBERT, F.; SIMONS, J. F.; *et al.* Paired-end mapping reveals extensive structural variation in the human genome. **Science**, v. 318, p. 420-426, 2007.
- LEMOIS, M. V. A. D.; BERTON, M. P.; PERIPOLLI, E.; GREGÓRIO, M.; PERIPOLLIA, E.; RAFAEL, M.; BIANCA, F.; ALINE, S.; ANGÉLICA, S.; LUCIA, G.; *et al.* Copy number variation regions in Nellore cattle: evidences of environment adaptation. **Livestock Science**, v. 207, p. 51-48, 2018.
- LETAIEF, R.; REBOURS, E.; GROHS, C.; MEERSSEMAN, C.; FRITZ, S.; TROUILH, L.; ESQUERRÉ, D.; BARBIERI, J.; KLOPP, C.; PHILIPPE, R.; BLANQUET, V. Identification of copy number variation in French dairy and beef breeds using next-generation sequencing. **Genetics Selection Evolution**, v. 49, n. 1, p. 77, 2017.
- LI, Y.; ZHENG, H.; LUO, R.; WU, H.; ZHU, H.; LI, R.; CAO, H.; WU, B.; HUANG, S.; SHAO, H.; MA, H. Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. **Nature biotechnology**, v. 29, n. 8, p.723, 2011.

- LI, W.; OLIVIER, M. Current analysis platforms and methods for detecting copy number variation. **Physiological Genomics**, v. 45, n. 1, p. 1-16, 2013.
- LI, W.; BICKHART, D. M.; RAMUNNO, L.; IAMARTINO, D.; WILLIAMS, J. L.; LIU, G. E. Comparative sequence alignment reveals River Buffalo genomic structural differences compared with cattle. **Genomics**, v. 111, n. 3, p. 418-425, 2019.
- LIEBER, M. R.; MA, Y.; PANNICKE, U.; SCHWARZ, K. Mechanism and regulation of human non-homologous DNA end-joining. **Nature Reviews Molecular Cell Biology**, v. 4, p. 712-720, 2003.
- LIU, G. E.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; LI, R. W.; ALEXANDER, L. J.; KEELE, J. W.; MATUKUMALLI, L. K.; SMITH, T. P.; GASBARRE, L. C. Detection of germline and somatic copy number variations in cattle. In: **Animal Genomics for Animal Health Developments in Biologicals**, v. 132, p. 231-237, 2008.
- LIU, Z.; LI, A.; SCHULZ, V.; CHEN, M.; TUCK, D. MixHMM: inferring copy number variation and allelic imbalance using SNP arrays and tumor samples mixed with stromal cells. **PLoS One**, v. 5, n. 6, p. e10909, 2010.
- LIU, G. E. *et al.* Analysis of copy number variations among diverse cattle breeds. **Genome Research**, v. 20, p. 693-703, 2010b.
- LIU, M.; FANG, L.; LIU, S.; PAN, M. G.; SEROUSSI, E.; COLE, J. B.; MA, L.; CHEN, H.; LIU, G. E. Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. **BMC Genomics**, v. 20, n. 1, p. 181, 2019.
- LUCITO, R.; HEALY, J.; ALEXANDER, J.; REINER, A.; ESPOSITO, D.; CHI, M.; RODGERS, L.; BRADY, A.; SEBAT, J.; TROGE, J.; WEST, J. A.; *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. **Genome Research**, v. 13, p. 2291-2305, 2003.
- LUCITO, R.; WEST, J.; REINER, A.; ALEXANDER, D.; ESPOSITO, D.; MISHRA, B.; POWERS, S.; NORTON, L.; WIGLER, M. Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. **Genome Research**, v. 10, p. 1726-1736, 2000.
- MACDONALD, J. R.; ZIMAN, R.; YUEN, R. K.; FEUK, L.; SCHERER, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. **Nucleic Acids Research**, v. 42, p. D986-D992, 2014.
- MARON, L. G.; GUIMARAES, C. T.; KIRST, M.; *et al.* Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 13, p. 5241-5246, 2013.
- MARQUES-BONET, T.; GIRIRAJAN, S.; EICHLER, E. E. The origins and impact of primate segmental duplications. **Trends in Genetics**, v. 25, p. 443-454, 2009.
- MATUKUMALLI, L. K.; LAWLEY, C. T.; SCHNABEL, R. D.; TAYLOR, J. F.; ALLAN, M. F.; HEATON, M. P.; O'CONNELL, J.; MOORE, S. S.; SMITH, T. P.; SONSTEGARD, T. S.; VAN TASSELL, C. P. Development and characterization of a high-density SNP genotyping assay for cattle. **PLoS One**, v. 4, n. 4, p. e5350, 2009.
- MEDVEDEV, P.; STANCIU, M.; BRUDNO, M. Computational methods for discovering structural variation with next-generation sequencing. **Nature Methods**, v. 6, p. S13-20, 2009.
- MEI, C. *et al.* Copy number variation detection in Chinese indigenous cattle by whole genome sequencing. **Genomics**, p. 1-6, 2019.

MEYERSON, M.; GABRIEL, S.; GETZ, G. Advances in understanding cancer genomes through second-generation sequencing. **Nat Rev Genet.**, v. 11, n. 10, p. 685–696, 2010.

MIELCZAREK, M.; FRĄSZCZAK, M.; GIANNICO, R.; *et al.* Analysis of copy number variations in Holstein-Friesian cow genomes based on whole-genome sequence data. **Journal of Dairy Science**, v. 100, n. 7, p. 5515-5525, 2017.

OSTERTAG, E. M.; GOODIER, J. L.; ZHANG, Y.; KAZAZIAN Jr, H. H. SVA elements are nonautonomous retrotransposons that cause disease in humans. **The American Journal of Human Genetics**, v. 73, n. 6, p. 1444-1451, 2003.

OU, Z.; KANG, S. H. L.; SHAW, C. A.; CARMACK, C. E.; WHITE, L. D.; PATEL, A.; BEAUDET, A. L.; CHEUNG, S. W.; CHINAULT, A. C. Bacterial artificial chromosome-emulation oligonucleotide arrays for targeted clinical array-comparative genomic hybridization analyses. **Genetics in Medicine**, v. 10, n. 4, p.278, 2008.

PINKEL, D.; ALBERTSON, D. G. Comparative genomic hybridization. **Annual Review of Genomics and Human Genetics**, v. 6, n. 331-354, 2005.

PIROOZANIA, M.; GOES, F. S.; ZANDI, P. P. Whole-genome CNV analysis: advances in computational approaches. **Frontiers in Genetics**, v. 6, p. 138, 2015.

REDON, R.; ISHIKAWA, S.; FITCH, K. R.; *et al.* Global variation in copy number in the human genome. **Nature**, v. 444, p. 444-454, 2006.

RHOADS, A.; Au, K. F. PacBio sequencing and its applications. **Genomics Proteomics Bioinformatics**, v. 13, p. 278–289, 2015.

SALMI, A.; JADID, S. E.; JAMAIL, I.; BENSELLAK, T.; PHILIPPE, R.; BLANQUET, V.; MOUSSA, A. CNV-LDC: An optimized CNV detection method for low depth of coverage data. In: Proceedings of 10th International Joint on Biomedical Engineering Systems and Technology - BIOSTEC 2017, p. 37–42, 2017.

SANTANA, M. H. A.; JUNIOR, G. A. O.; CESAR, A. S. M.; FREUA, M. C.; DA COSTA GOMES, R.; SILVA, S. D. L.; LEME, P. R.; FUKUMASU, H.; CARVALHO, M. E.; VENTURA, R. V.; COUTINHO, L. L. Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. **Journal of Applied Genetics**, v. 57, n. 4, p. 495-504, 2016.

SCHRIDER, D. R.; HAHN, M. W. Gene copy-number polymorphism in nature. **Proceedings of the Royal Society B: Biological Sciences**, v. 277, p. 3213-3221, 2010.

SCHUSTER-BOCKLER, B.; CONRAD, D.; BATEMAN, A. Dosage sensitivity shapes the evolution of copy-number varied regions. **PLoS One**, v. 5, p. e9474, 2010.

SEISER, E. L.; INNOCENTI, F. Hidden markov model-based CNV detection algorithms for Illumina genotyping microarrays. **Cancer Informatics**, v. 13, p. CIN-S16345, 2014.

SENTANDREU, M.; COULIS, G.; OUALI, A. Role of muscle endopeptidases and their inhibitors in meat tenderness. **Trends Food Sci Technol.**, v. 13, p. 400–421, 2002.

SHOMURA, A.; IZAWA, T.; EBANA, K.; EBITANI, T.; KANEGAE, H.; KONISHI, S.; MASAHIRO, Y. Deletion in a gene associated with grain size increased yields during rice domestication. **Nature Genetics**, v. 40, n. 8, p. 1023–1028, 2008.

SILVA, M. V. G. B.; OLIVEIRA JUNIOR, G. A.; REY, F. S. B.; GIACHETTO, P. F.; MACHADO, M. A.; VERNEQUE, R. S.; FERRAZ, J. B. S. Analysis of Copy Number Variation Regions in a Population of Dairy Gyr Cattle. In: WORLD CONGRESS OF GENETICS APPLIED TO

LIVESTOCK PRODUCTION, 10., 2014, Vancouver. Proceedings... Champaign: American Society of Animal Science, 2014.

da SILVA, V. H.; REGITANO, L. C. 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. **PLoS One**, v. 1, n. 6, p. e0157711, 2016a.

da SILVA, J. M.; GIACHETTO, P. F.; SILVA, L. O. C.; 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 QTLs affecting production traits. **BMC Genomics**, v. 17, p. 454, 2016b.

STANKIEWICZ, P.; LUPSKI, J. R. Genome architecture, rearrangements and genomic disorders. **Trends in Genetics**, v. 18, p. 74-82, 2002.

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

SUN, W.; WRIGHT, F. A.; TANG, Z.; NORDGARD, S. H.; LOO, P. V.; YU, T.; KRISTENSEN, V. N.; PEROU, C. M. Integrated study of copy number states and genotype calls using high-density SNP arrays. **Nucleic Acids Research**, v. 37, n. 16, p. 5365-5377, 2009.

SUN, Z.; LIU, P.; JIA, X.; WITHERS, M. A.; JIN, L.; LUPSKI, J. R.; ZHANG, F. Replicative mechanisms of CNV formation preferentially occur as intrachromosomal events: evidence from Potocki-Lupski duplication syndrome. **Human Molecular Genetics**, v. 22, n. 4, p. 749-756, 2013.

SUTTON, T.; BAUMANN, U.; HAYES, J.; *et al.* Boron-toxicity tolerance in barley arising from efflux transporter amplification. **Science**, v. 318, n. 5855, p. 1446-1449, 2007.

TATTINI, L.; D'AURIZIO, R.; MAGI, A. Detection of genomic structural variants from next-generation sequencing data. **Frontiers in Bioengineering and Biotechnology**, v. 3, p. 1-8, 2015.

TEO, S. M.; PAWITAN, Y.; KU, C. S.; CHIA, K. S.; SALIM, A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. **Bioinformatics**, v. 28, p. 2711-2718, 2012.

TIZAOU, K. *De novo* vs. inherited copy number variations in multiple sclerosis susceptibility. **Cellular & Molecular Immunology**, v. 15, n. 9, p. 812, 2018.

TIZIOTO, P. C.; DECKER, J. E.; TAYLOR, J. F.; SCHNABEL, R. D.; MUDADU, M.; SILVA, F. L.; *et al.* Genome scan for meat quality traits in Nelore beef cattle. **Physiol Genomics**, v. 45, p. 1012-1020, 2013.

TUZUN, E.; SHARP, A. J.; BAILEY, J. A.; *et al.* Fine-scale structural variation of the human genome. **Nature Genetics**, v. 37, p. 727-732, 2005.

VAN OMMEN, G. J. B. Frequency of new copy number variation in humans. **Nature Genetics**, v. 37, n. 4, p. 333-334, 2005.

WANG, H.; CHAI, Z.; HU, D.; JI, Q.; XIN, J.; ZHANG, C.; ZHONG, J. A global analysis of CNVs in diverse yak populations using whole-genome resequencing. **BMC Genomics**, v. 20, n. 61, p. 1-12, 2019.

WANG, K.; LI, M.; HADLEY, D.; LIU, R.; GLESSNER, J.; GRANT, S. F.; HAKONARSON, H.; BUCAN, M. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v. 17, n. 11, p. 1665-1674, 2007.

WETERINGS, E.; VAN GENT, D. C. The mechanism of non-homologous end-joining: a synopsis of synapsis. **DNA Repair**, v. 3, p. 1425–1435, 2004.

XU, L.; COLE, J. B.; BICKHART, D. M.; HOU, Y.; SONG, J.; VAN RADEN, P. M.; *et al.* Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. **BMC Genomics**, v. 15, p. 683, 2014a.

XU, L.; HOU, Y.; BICKHART, D. M.; SONG, J.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; LIU, G. E. A genome-wide survey reveals a deletion polymorphism associated with resistance to gastrointestinal nematodes in Angus cattle. **Functional & Integrative Genomics**, v. 14, n. 2, p. 333-339, 2014b.

XU, Y.; JIANG, Y.; SHI, T.; CAI, H.; LAN, X.; ZHAO, X.; PLATH, M.; CHEN, H. Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds. **PLoS One**, v. 12, n. 8, p. e0183921, 2017.

YANG, L.; XU, L.; ZHU, B.; *et al.* Genome-wide analysis reveals differential selection involved with copy number variation in diverse Chinese Cattle. **Scientific Reports**, v. 7, n. 1, p. 1-12, 2017.

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.

ZARREI, M.; MACDONALD, J. R.; MERICO, D.; SCHERER, S. W. A copy number variation map of the human genome. **Nature Reviews Genetics**, v. 16, n. 3, p. 172-83, 2015.

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.

ZHANG, L.; BAI, W.; YUAN, N.; DU, Z. Comprehensively benchmarking applications for detecting copy number variation. **PLoS Computational Biology**, v. 15, n. 5, p. e1007069, 2019.

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, n. 11, p. S1, 2013.

ZHOU, Y.; UTSUNOMIYA, Y. T.; XU, L.; *et al.* Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. **BMC Genomics**, v. 17, p. 1-9, 2016.

ZHOU, Y.; CONNOR, E. E.; WIGGANS, G. R.; LU, Y.; TEMPELMAN, R. J.; SCHROEDER, S. G.; CHEN, H.; LIU, G. E. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. **BMC Genomics**, v. 19, n. 1, p. 314, 2018.

ZHOU, B.; HO, S. S.; ZHANG, X.; PATTNI, R.; HARAKSINGH, R. R.; URBAN, A. E. Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. **Journal of Medical Genetics**, v. 55, p. 735-743, 2018c.

Embrapa

Gado de Corte



MINISTÉRIO DA
AGRICULTURA, PECUÁRIA
E ABASTECIMENTO



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

CGPE 15152