

XII Simpósio Brasileiro de Melhoramento Animal Ribeirão Preto, SP – 12 e 13 de junho de 2017

Identificação de variação no número de cópias em bovinos leiteiros usando dados de NGS

Tatiane C Seleguim Chud^{1*}, Adhemar Zerlotini², Adriana S do Carmo³, Danísio P Munari¹, Marcos V B da Silva⁴

¹Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, SP, Brasil

²Embrapa Informática Agropecuária, Campinas, SP, Brasil

³Universidade Federal de Goiás, Goiânia, GO, Brasil

⁴Embrapa Gado de Leite, Juiz de Fora, MG, Brasil

*Autor correspondente: tatischud@gmail.com

Resumo: O objetivo deste trabalho foi identificar e caracterizar variação no número de cópias (CNV) em bovinos das raças Gir e Girolando (Holandês x Gir) utilizando dados de NGS. A detecção das CNVs em dados de NGS provenientes de cinco touros (dois da raça Gir e três da raça Girolando) foi realizada por meio do software CNVnator. CNVs individuais com no mínimo 1 pb em *overlap* entre todas as amostras foram agregadas (CNVR). Para análise funcional das CNVRs foram utilizadas a plataforma Ensembl e os termos do *Medical Subject Headings* (MeSH). Foram encontradas 667 CNVRs em comum entre as duas amostras da raça Gir, 288 regiões entre as três amostras da raça Girolando e 256 CNVRs entre as cinco amostras. Genes de interesse para a bovinocultura de leite foram mapeados nas regiões compartilhadas entre as duas raças, tais como: *NLGN1* (BTA1), *LRP11* (BTA9), *FOXH1* (BTA14) e *IGF2* (BTA29) e *PGR* (BTA 15). Termos MeSH importantes foram enriquecidos significativamente ($P < 0,05$), como diacilglicerol O-aciltransferase (D051048). Este estudo demonstrou evidências de regiões de CNVs em comum entre bovinos das raças Gir e Girolando e possíveis genes candidatos localizados nestas regiões podem estar associados às características de importância econômica como produção e composição do leite e resistência a doenças, em bovinos de leite.

Palavras-chave: *Bos primigenius indicus*, raça composta, sequenciamento

Copy number variation discovery using NGS data in dairy cattle

Abstract: The aim of this study was to detect and describe CNV regions in Gyr and Girolando cattle breeds using NGS data. CNVnator software was performed on NGS data from five bulls (two Gyr and three Girolando). The individual CNVs were merged into CNV regions (CNVRs) based on genomic regions overlapping by at least 1 bp. Gene content of cattle CNV was assessed using Ensembl genes and Medical Subject Headings terms for enrichment analysis. An account of 667 CNVRs were found in common between Gyr samples, 288 CNVRs among all three Girolando samples and 256 CNVRs among all five samples. We observed some interesting genes within the regions shared between the breeds for dairy production, including *NLGN1* (BTA1), *LRP11* (BTA9), *FOXH1* (BTA14), *PGR* (BTA 15) e *IGF2* (BTA29) and several important MeSH terms were significantly ($P < 0.05$) enriched, including Diacylglycerol O-Acyltransferase (D051048). This study showed evidences of common CNVRs between Gir and Girolando breeds and the genes found within CNVR regions may be related to important economic traits in dairy cattle, including milk yield and composition and disease resistance.

Keywords: *Bos primigenius indicus*, composite breed, sequence data

Introdução

Devido aos avanços nas ferramentas genômicas, como redução dos custos nos chips de polimorfismos de nucleotídeo único (SNP), disponibilidade do sequenciamento de nova geração e melhoria no genoma referência bovino, certas variações estruturais no cromossomo de bovinos, como os CNVs, têm sido amplamente identificadas. No melhoramento genético de bovinos leiteiros, CNVs podem auxiliar na melhoria de características produtivas, como produção de leite e de sanidade, como susceptibilidade às doenças. O objetivo deste trabalho foi identificar e caracterizar CNVs em bovinos das raças Gir e Girolando (Gir x Holandes) utilizando dados de NGS.

Material e Métodos

Foram utilizados dados de NGS provenientes de cinco touros, sendo dois da raça Gir e três da raça Girolando. Para análise da identificação das CNVs foram considerados apenas os 29 cromossomos autossômicos.

A detecção das CNVs foi realizada por meio do software CNVnator (Abyzov et al., 2011), baseado na metodologia *read-depth*, em que deleções e duplicações são detectadas com base na densidade do alinhamento das *reads* ao longo dos cromossomos. O tamanho da janela (*bin size*) considerado para detecção 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 que a região entre duas detecções não seja a mesma e CNVs com fração de *reads* mapeadas com baixa qualidade menores do que 0,5 ($q_0 < 0,5$) e CNVs com tamanhos maiores do que 1 kilobase (kb). As regiões de CNVs (CNVRs) foram definidas agregando CNVs individuais com no mínimo 1 pb em *overlap* entre todas as amostras utilizando a ferramenta BEDTools suite.

Os genes localizados dentro dos limites das CNVRs foram mapeados com base no genoma referencia UMD 3.1 por meio da plataforma Ensembl. A análise genômica funcional foi realizada utilizando os termos do *Medical Subject Headings* (MeSH), divididos nas categorias fenômenos e processos, doenças, químicos e drogas e anatomia (Morota et al., 2015).

Resultados e Discussão

Na raça Gir, foram encontradas 667 CNVRs em comum, ou seja, as mesmas regiões presentes nas duas amostras, sendo 146 deleções, 111 duplicações e 410 regiões mistas com tamanho médio de 554,80 kb. Na raça Girolando, foram identificadas 288 regiões em comum entre os três touros, com tamanho médio de 257,70, sendo 73 deleções, 71 duplicações e 144 mistas. Usando as cinco amostras, das duas raças, foram identificadas 256 regiões, com média em tamanho de 1.181,10 kb, sendo a maioria regiões mistas, ou seja, algumas amostras apresentaram duplicações enquanto outras apresentaram deleções. A distribuição das CNVRs ao longo dos cromossomos para ambas as raças, assim como o tamanho das regiões compartilhadas entre as cinco amostras estão demonstradas na Figura 1.

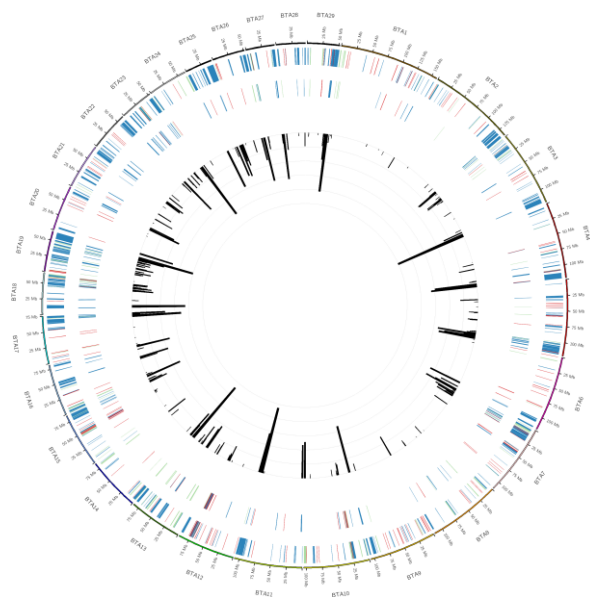
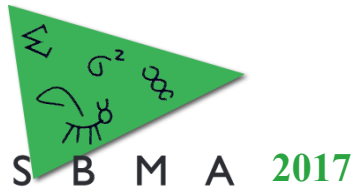


Figura 1. Distribuição ao longo dos 29 cromossomos autossomos das regiões com variação no número de cópias (CNVR) para as raças Gir e Girolando. As deleções estão em vermelho, duplicações em verde e ambas em azul. Histograma (ao centro) representa o tamanho das CNVRs, ao longo dos cromossomos, variando de 13,50 kb a 6.850 kb, em comum entre os cinco touros.



XII Simpósio Brasileiro de Melhoramento Animal Ribeirão Preto, SP – 12 e 13 de junho de 2017

Foram encontrados 5.042 genes localizados dentro das regiões em comum entre todos os touros, sendo alguns desses genes envolvidos em processos fisiológicos importantes na bovinocultura de leite, tais como, o *NLGN1* (BTA1), que está relacionado com doenças inflamatórias e fertilidade em fêmeas. Outros exemplos são o *LRP11* (BTA9), relacionado com produção de proteína do leite e susceptibilidade a patógenos que causam doenças como diarreia em bezerros e mastite, *FOXH1* (BTA14) e *IGF2*(BTA29), associados com produção e composição do leite e *PGR* (BTA 15) relacionado com termorregulação. Consequentemente, importantes termos Mesh foram enriquecidos significativamente ($p < 0,05$), incluindo, desenvolvimento ósseo e prenhez (D001846 e D011247) na categoria fenômenos e processos e diacilglicerol O-aciltransferase (D051048 e interleucina 1 (D007375) na categoria químicos e drogas. Na categoria doenças, termos relacionados a leucose (D007945, D015473) e na categoria anatomia o termo vesícula seminal (D012669), fetal (D005333) e células 3T3 (D016475).

Conclusão

Este estudo demonstrou evidências de regiões de CNVs em comum entre bovinos da raça Gir e Girolando e possíveis genes candidatos localizados nestas regiões associados com características de importância econômica para bovinos de leite em países de clima tropical, como produção e composição do leite e resistência a doenças.

Agradecimentos

Tatiane Cristina Seleguim Chud é bolsista de doutorado Fapesp processo 2015/08939-0. Danísio Prado Munari possui bolsa de produtividade em pesquisa CNPq. Marcos Vinicius G.B. da Silva teve o suporte do projeto FAPEMIG CVZ PPM 00606/15 “Identificação de Assinaturas de Seleção em Bovinos por meio de Dados Gerados por Sequenciamento de Nova Geração”

Literatura citada

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 genomic sequencing. **Genome Research**, v.21, p. 974-984, 2011.

MOROTA, G.; PEÑAGARICANO, F.; PETERSEN, J.L.; CIOBANU, D.C., TSUYUZAKI, K.; NIKAIIDO, I. An application of MeSH enrichment analysis in livestock. *Animal Genetics*, v.46, p. 381-387, 2015.