Analysis of copy number variation regions in a Nellore population evaluated for feed efficiency

***G. A. Oliveira Júnior*[*], F. S. B. Rey[†], M. V. G. B. da Silva[‡], M. H. A. Santana[*], P. A. Alexandre[*], J. P. Eler[*], J. B. S. Ferraz[*]**

[*]University of São Paulo, Pirassununga, Brazil, [†]São Paulo State University "Julio de Mesquita Filho", Jaboticabal, Brazil, [‡]Embrapa Gado de Leite, Juiz de Fora, Brazil

ABSTRACT: The systematic study of Nellore's genome can contribute in a unique way to the development of traits of interest, like feed efficiency. The aims of this study were to identify copy number variation (CNV) and CNV regions (CNVR) in a Nellore population and try to correlate these with feed efficiency genes. Genotype data of the Nellore bulls were recorded, in total 2,457 animals, being 640 phenotyped for residual feed intake. For CNV identification was used the PennCNV software and the CNVRs were determined by the CNVRuler software. A total of 117,175 CNVs were found with approximately 90% between 3 Kb and 1 Mbp. Also, 6,933 CNVRs were detected with average more 'loss' than 'gain' regions. In summary, the residual feed intake genes analyzed were close or within the genes, suggesting that these CNVRs might have some influence on this trait.

Keywords: beef cattle, genomic, residual feed intake, SNP

Introduction

Most of the economically relevant traits in livestock are complex, being influenced by multiple genes or genomic regions, besides several environmental factors. The feed efficiency is a relevant variable for the entire meat production system, representing important part of costs in beef cattle. According to Rolf et al. (2010), considerable effort is now being focused on reducing the costs of raising animals by improving the feed efficiency. There are many ways to measure animal efficiency, highlighting the residual feed intake (RFI, Santana et al. (2014)).

The advances made in the genomic area in the recent years, enable researchers to identify and characterize genes and markers that influence polygenic traits in livestock. The genetic gain for these traits can be improved by enhancing knowledge of the molecular architecture of interest traits, increasing the genetic gain (Ibeagha-Awemu et al. (2008)).

In addition to the single nucleotide polymorphism (SNP), DNA copy number variants (CNV) have been revealed to be a substantial source of genetic and phenotypic variation in cattle (Hou et al. (2012b); Feuk et al. (2006)). The CNV can be defined as stretches of DNA ranging from kilobase (Kb) to megabases (Mb) in size that display copy number differences in the normal populations in comparison with a reference genome, involving genomic sequences in the form of large-scale insertions and deletions, as well as positional changes in inversions and translocations (Redon et al. (2006); Scherer et al. (2007); Liu et al. (2010)). Redon et al. (2006) stated that CNV can vary from a simple structure, such as tandem duplication to complex gains or losses of homologous sequences at multiple sites in the genome.

It's believed that most cattle CNVs are within or close to genes for specific biological functions, such as immunity, lactation, reproduction, and rumination, influencing directly or indirectly on the expression of these genes (Henrichsen et al. (2009); Zhang et al. (2009)). Santana et al. (2014) have found, based on a Nellore population, genes that could affect feed efficiency traits, and would be important to try to associate these with CNVs.

Almost 15,000 CNV loci covering about one-third of the genome have been identified in humans (Seroussi et al. (2010)). As Manolio et al. (2009) affirmed, the use of CNV could be an effective way to clarify the unexplained variations of traits, which have been partly assessed by SNP information. Redon et al. (2006) discussed how CNVs could be a major source of heritable variation in complex traits. Bickhart et al. (2012) highlighted the importance of CNV identification in Nellore cattle for economically relevant traits.

Regions of copy number variation (CNVRs) represent the independently overlapped CNVs, which can occur as a segment at a fixed chromosomal position or a multiple arrangement of variant units in close proximity.

However, CNVs and CNVR in Nellore cattle have still to be more fully explored, more studies being necessary focused on these genetic arrangements. In this context, our purposes were to investigate, bases on Illumina high density BovineHD SNP array, the abundance and distribution of CNVs and CNVR and try to connect these with genes related to a feed efficiency trait, in a Nellore cattle population from Brazil.

Materials and Methods

Data. Genotypes data were recorded for 2,457 Nellore cattle from commercial farms in Brazil, deriving from different regions, containing samples of Nellore bulls of Brazilian herd. These animals were genotyped for the Illumina High-Density Bovine BeadChip with more than 777,692 informative SNPs. From the 720 Nellore animals worked per Santana et al. (2014) 640 with adjusted phenotypes, for non-genetic effects, for RFI were included on the present study.

CNV and CNVR identification. For CNV identification, the luminosity measure of Log R Ration (LRR) and B allele frequency (BAF), both predicted from the GenomeStudio software from Illumina, were used. The

intensity generated of each SNP on the chip is represented as the normalized R value. The LRR is predicted from the ration of the expected normalized intensity of a sample and observed normalized intensity, while the BAF is calculated from the difference between the expected position of the cluster group and the actual value (Winchester et al. (2009)). The algorithms based on the first-order of Hidden Markov Model (HMM) of the PennCNV software, developed by Wang et al. (2007), were used for CNV identification. Furthermore, the software incorporates into HMM the distance between neighboring SNPs and the population frequency of the B allele, that refer to the alleles A and B of the SNPs. It was used a PennCNV perl script (filter_cnv.pl) in order to eliminate calls from low quality samples, based on the standard deviation of LRR (less than 0,30), the default for BAF drift (less than 0.01) and waviness factor (less than 0.05).

The CNVRs were determined by the CNVRuler software (Kim et al. (2012)). Although PennCNV gives six different classifications for CNV, CNVRuler support only three definitions of CNV regions (gain, loss, mixed). The parameter of recurrence used was 0.1. This parameter means that areas with low density (<10% of CNVs) are excluded to compose an estimated end region, leaving more robust definition of the beginning and end of regions. Additionally, the "Gain / Loss separated regions" option, which compiles with the region based on the genotype (gain or loss of copy number) instead of composing regions ignoring the event type, was used.

Genes. It were used the 11 genes (*AHNAK2*, *ANXA10*, *BRF1*, *CCDC171*, *CDCA4*, *DDX60*, *GPR132*, *PSIP1*, *SNAPC3*, *STMN2* and *ZNF804B*) that were related to RFI described by Santana et al. (2014).

Results and Discussion

We found a total of 117,175 CNVs in 2,457 animals. The mean of CNV size was 76 Kb with approximately 90% between 3 Kb and 1 Mbp, showing a large variability. Similarly, Hou et al. (2012a) working with BovineHD SNP chip in 147 Holstein animals, detected a total of 3,706 CNVs with an average of 25 events for each sample. Bae et al. (2010), who used the BovineSNP50 BeadChip in 265 *Bos taurus* animals, found a total of 264 CNV regions with average of 3.2 CNV per sample and 149.8 Kb average length. Henrichsen et al. (2009) discussed that the boundaries of the ranges of CNV size may reflect the resolution of the platforms used as well as the power of the prediction algorithms.

A total of 6,933 CNVRs were detected (Figure 1) covering approximately 10% of all autosomal chromosomes, which was estimated to be around 2 Gbp. The average size per chromosomes was 39.7 Mbp, varying from 28.7 to 71 Mbp (Figure 2). Likewise, Hou et al. (2012b), using a high density BoniveHD SNP array on 674 animal of different breeds, have identified 3,346 CNVR, representing approximately 4.7% of the genome.

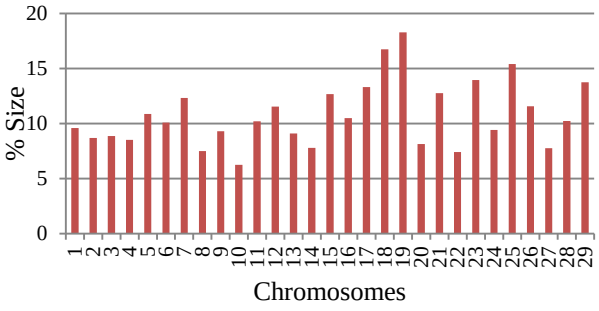

Figure 1. Relative size of copy number variations region (CNVR) in the autosomal Nellore chromosomes

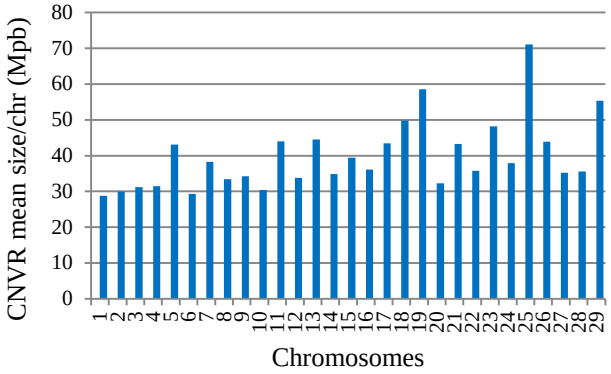

Figure 2. Average size of copy number variations region (CNVR) in the autosomal Nellore chromosomes

The pattern of the different types of CNVRs (loss, gain and mixed) were specific for each chromosome, with average more 'loss' regions than 'gain' and 'mixed' region, having 1.1 as ratio between 'loss' and 'gain'. The type 'mixed' means that the boundary of CNVR is consistent with 'gain' and 'loss' CNVs. Hou et al. (2012a) found 443 CNVR that also included more loss than gain, having 1.7 of loss and gain ration.

Among the 11 genes reported by Santana et al. (2014), three genes related with RFI were not within the CNVRs. There were CVNRs relatively close to these genes, being 26, 47 and 72 Kb the distance between the genes *PSIP1*, *CCDC171* and *STMN2* and the CNVR, respectively. In the *ANXA10* gene, that has 195,459 bp of size, there were found 51 CNVR along the gene. Likewise, the *DDX60* gene (105,555 bp) has encompassed 24 CNVRs, which were also in *ANXA10*. Contrary, there was one CNVR on chromosome 21 with 192.6 kb that covers *GPR132*, *CDCA4* and *AHNAK2* genes. The *BRF1* (51,110 bp) gene, which was also associated with one CNVR for RFI in Holstein animals by Hou et al. (2012a), had 22.7% of its size in the same CNVR that cover *GPR132*, *CDCA4* and *AHNAK2*.

Conclusion

The results could help to better understand the Nellore genome structure, once the CNVRs cover approximately 10% of the genome, considering the autosomal chromosomes. The regions that were close or contained the genes could suggest that these might exercise some influence on the trait residual feed intake.

Literature Cited

Bickhart, D. M., Hou, Y., Schroeder, S. G. et al. (2012). Genome Res. 22, 778–790.

Henrichsen, C. N., Chaignat, E., and Reymond, A. (2009). Hum. Mol. Genet. 18, 1–8.

Hou, Y., Bickhart, D. M., Chung, H. et al. (2012a). Funct. Integr. Genomics 12, 717–723.

Hou, Y., Bickhart, D. M., Hvinden, M. L. et al. (2012b). BMC Genomics 13, 376–386.

Ibeagha-Awemu, E. M., Kgwatalala, P., and Zhao, X. (2008). Mamm. Genome 19, 591–617.

Kim, J. H., Hu, H. J., Yim, S. H., et al. (2012). Bioinformatics 28, 1790–1792.

Liu, G. E., Hou, Y., Zhu, B., et al. (2010). Genome Res. 20, 693–703.

Manolio, T. A., Collins, F. S., Cox, N. J., et al. (2009). Nature 461, 747–753.

Redon, R., Ishikawa, S., Fitch, K. R., et al. (2006). Nature 444, 444–454.

Rolf, M.M., Taylor, J.F., Schnabel, R.D, et al. (2010). BMC Genet. 11, 24–34.

Santana, M. H. A., Utsunomiya, Y. T., Neves, H. H. R., et al. (2014). BMC Genet. 15, 21–29.

Scherer, S. W., Lee, C., Birney, E., et al. (2007). Nat. Genet. 39, 7–15.

Seroussi, E., Glick, G., Shirak, A., et al. (2010). BMC Genomics 11, 673–683.

Wang, K., Li, M., Hadley, D., et al. (2007). Genome Res. 17, 1665–1674.

Winchester, L., Yau, C., and Ragoussis, J. (2009). Genomics Proteomics 8, 353–366.

Zhang, F., Gu, W., Hurles, M. E. et al. (2009). Genomics Hum. Genet. 10, 451–481.