Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Descriptive Analysis of Copy Number Variation Regions in a Population of Dairy Gyr Cattle

***M. V. G. B. da Silva*[*], G. A Oliveira Junior[†], F. S. B. Rey[‡], P. F. Giachetto[§], M. A. Machado[*], R. S. Verneque[*], J. B. S. Ferraz[†]**

[*]Embrapa Dairy Cattle; [†]University of São Paulo; [‡]São Paulo State University; [§]Embrapa Agricultural Informatics

ABSTRACT: The aim of this work was to investigate, based on a high density BovineHD SNP array, the abundance and distributions of CNVs and CNVR in a Gyr cattle population from Brazil. Genotype data of representative bulls were recorded, totaling 476 Gyr animals. For CNV identification was used the PennCNV software and the CNVRs were determined by the CNVRuler software. A total of 26,672 CNVs were found, being on average 62 CNV per animal. Also, 1,898 CNVRs were detected on the autosomal chromosomes. Also, 1,898 CNVRs were detected on the autosomal chromosomes with 96% of these between 1.1 Kb to 100 Kb. The Ensembl's VEP tool, using the CNVRs information as input, found 913 coding regions, suggesting that exon regions were duplicated. In summary, the results help to better understand the Gyr genome and suggest that CNVRs might have some relationship with production traits.

Keywords:
Bos indicus
Dairy cattle
Genomic
SNP

Introduction

The Gyr cattle (*Bos indicus*) is a very important dairy breed in tropical countries like Brazil, mainly because of its tolerance to heat and parasites and because it is used in crossbreeding schemes with other specialized dairy breeds, such as Holstein. However, most of the economically important traits in dairy cattle are complex, being influenced by multiple genes or genomic regions.

In recent years, the advances made in the genomic area enable the use of dense single nucleotide polymorphism (SNP) arrays, which cover all the bovine genome and explain a majority of the genetic variations in important traits in dairy cattle. Golden et al. (2011) have stated that more than half of the increase in milk production in Holstein animals is due to improvements in the genetic area.

The 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 positional changes as inversions and translocations (Redon et al. (2006); Scherer et al. (2007); Liu et al. (2010)). For Redon et al. (2006), CNV can vary from being simple in structure, such as tandem duplication, to complex gains or losses of homologous sequences at multiple sites in the genome.

Most of the cattle CNVs are related to genomic regions for specific biological functions, such as immunity, lactation, reproduction, and rumination, exerting influence directly or indirectly on the expression of genes within and close to the rearranged region (Henrichsen et al. (2009); Zhang et al. (2009)).

Almost 15,000 CNV loci covering about one-third of the genome have been identified in humans (Seroussi et al. (2010)). For Manolio et al. (2009) the use of CNV could be an effective way to clarify the unexplained variations of traits, which are incompletely assessed by SNP information. Redon et al. (2006) discussed that CNVs could be a major source of heritable variation in complex traits.

Regions of copy number variation (CNVRs) represent the independently overlapped CNVs that 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 Gyr cattle still have been little explored, more studies about these genetic rearranges being necessary. In this context, our purpose was to investigate, based on a high density BovineHD SNP array, the abundance and distributions of CNVs and CNVR in a Gyr cattle population from Brazil.

Materials and Methods

Data. Genotype data were recorded for 476 Gyr sires from commercial partner breeders in Brazil, deriving from different regions, containing samples of the most representative bulls of the Brazilian herd. These animals were genotyped by the Illumina High-Density Bovine BeadChip with more than 777,692 informative SNPs.

CNV and CNVR identification. For CNV identification, the luminosity measure of Log R

Ration (LRR) and B allele frequency (BAF), both predicted from the BeadStudio 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 for the alleles A and B of the SNPs. A PennCNV perl script (filter_cnv.pl) was used 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). Also, samples with call rate below 90% were discarded.

The CNVRs were determined by aggregating adjacent or overlapping CNVs identified across all samples by the CNVRuler software (Kim et al. (2012)). Although PennCNV gives six different classifications for CNV, CNVRuler supports 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 the region based on the genotype (gain or loss of copy number) instead of composing regions ignoring the event type, was used. The CNVR output was analyzed with the Variant Effect Predictor (VEP) tool from the Ensembl website (http://www.ensembl.org/).

Results and Discussion

After the PennCNV quality control, we found a total of 26,672 CNVs in 430 animals, being on average 62 CNV per animal. The average number of CNV per chromosome was 919.7, varying from 29 (BTA 25) to 3,681 (BTA12). The overall CNV mean size was 60.4 Kb covering a total of 263,293 SNPs. These results demonstrated that CNVs were widespread throughout the bovine genome, as discussed for Cicconardi et al. (2013).

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. In Bae et al. (2010), who used the BovineSNP50 BeadChip in 265 *Bos taurus coreanae* 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. Probably, the different numbers of CNV found in this study and in Hou et al. (2012a) and Bae et al. (2010) were explained by differences in the resolution of the platforms, algorithms and animal populations that were utilized to infer the CNVs.

A total of 1,898 CNVRs were detected on all the autosomal chromosomes, having an average size of 26.08 Mbs per chromosome and with 96% of the CNVRs between 1.1 Kb to 100 Kb (Figure 1). The major region was on chromosome 1 (6.4 Mbp of length size). These CNVRs represent approximately 2% of all autosomal chromosomes which was estimated to be around 2Gbp. Similarly, Liu et al. (2010) found 177 CNVRs covering almost 1.07% of the genome, on 168 animals from different breeds, including Gyr.

The pattern of the different types of CNVRs (loss, gain and mixed) were specific for each chromosome, with on average more ‘gain’ regions (1,138) than ‘loss’ (627) and ‘mixed’ region (133). These differences were almost the same on chromosome 27 (just one ‘gain’ region more than ‘loss’), representing 50% of the chromosome (Figure 2). The type ‘mixed’ means that the boundary of CNVR is consistent with ‘gain’ and ‘loss’ of CNVs, being rarer than the other rearrange types. Hou et al. (2012a) found 443 CNVR but with more loss (251) than gain (144) and mixed (48).

The CNVRs identified were submitted to the VEP tool of Ensembl, and a total of 913 coding regions were found, suggesting that exon regions were duplicated. Also, 260 regions were in “upstream” or “downstream” regions, 1,107 in intragenic, 310 in intron variant and 38 in no 3’ or 5’ primer variant, that were related to no coding exons. These results show that 48% of the CNVR identified are in DNA coding regions that might influence important traits in dairy cattle.

Conclusion

The results could help to better understanding of the Gyr genome structure. The CNVRs might have an important relationship with productive traits, highlighting the importance of further studies on this area.

Literature Cited

Cicconardi, F., Chillemi, G., Tramontano, A., et al. (2013). BMC Genomics 14, 124–138.

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.
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.
Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., 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.

Figure 1. Size range distribution of the CNVRs detected

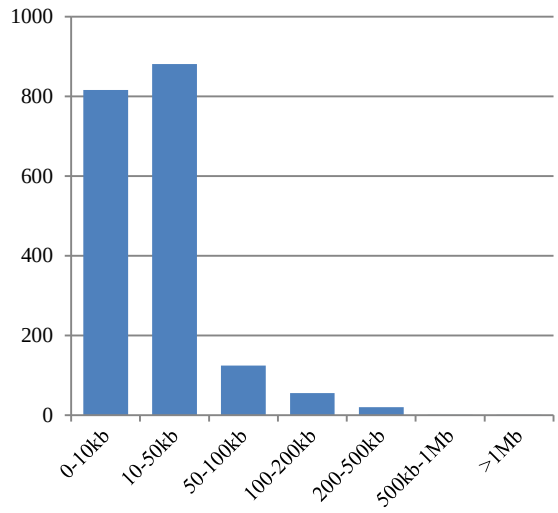

Figure 2. Relative size per type of CNVRs along the autosomal chromosomes