Detection of Copy Number Variations in Nelore Beef Cattle with High-Density SNP Genotyping Data

Joaquim Manoel da Silva[1,2,3], Poliana Fernanda Giachetto[3], Luiz Otávio C. da Silva[4], Samuel R. Paiva[4], Alexandre Rodrigues Caetano[5], Michel Eduardo Beleza Yamagishi[3]

1- Unemat – State University of Mato Grosso; 2-Genetics and Molecular Biology Graduate Program, Biology Institute, State University of Campinas - Unicamp; 3-Embrapa Agricultural Informatics; 4-Embrapa Beef Cattle; 5-Embrapa Genetic Resources & Biotechnology;

E-mail: joaquimmanoel@unemat.br

INTRODUCTION

Genome-wide single nucleotide polymorphism (SNP) genotyping platforms have been widely used in studies in diverse areas, ranging from population genetics to applied genetic improvement and breeding. SNP genotyping data generated with these platforms can also be used for detecting and genotyping copy number variations (CNVs). CNVs are defined as a variable copy numbers of DNA segments ranging from 50bp to several megabases (Mbp) in comparison with a reference genome [1], and CNV regions (CNVRs) are defined as the overlapping of CNV regions across all samples [2]. Several studies have identified an abundance of CNVs in human and domestic animal genomes, where it has been shown to be involved in phenotypic variability [1-4].

MATERIAL AND METHODS

A total of 1,709 Nelore (*Bos indicus*) samples collected in Brazil were genotyped with Illumina BovineHD BeadChip. Markers unmapped to the UMD3.1 assembly were excluded from further analysis. A total of 735,242 SNPs covering the 29 bovine autosomes remained in the final dataset. PennCNV [5] was used to detect CNVs. Non-restrictive quality filters were used to exclude 233 samples with data below minimal quality standard cutoffs (LRR standard deviation <0.4 and BAF drift <0.04).

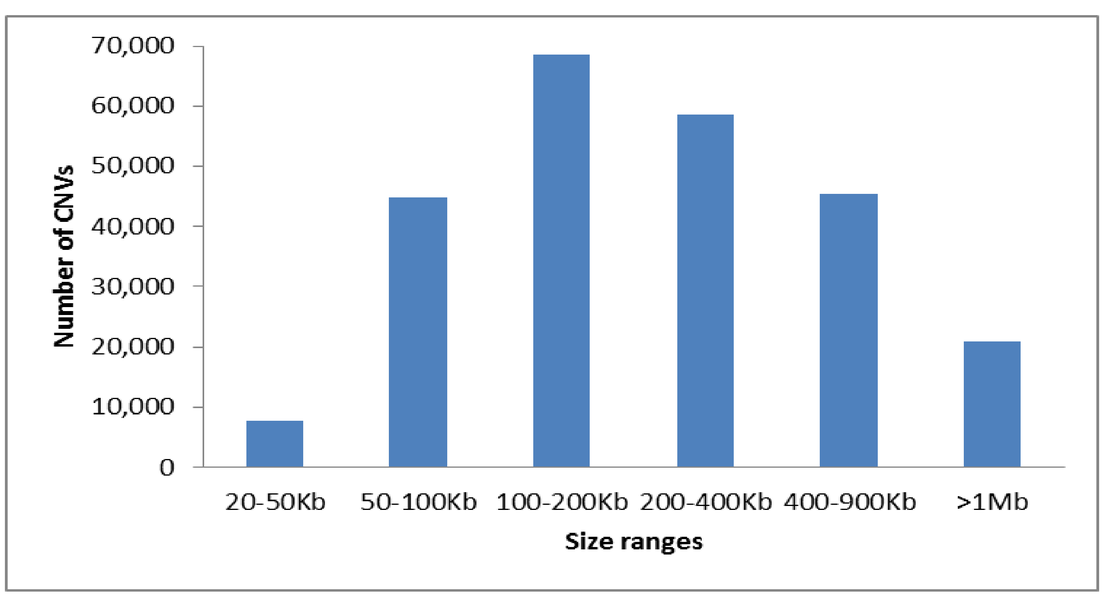

Figure 1 Size distribution of CNVs detected in this study.

RESULTS

We observed a total of 246,290 candidate CNVs distributed unevenly in all autosomal chromosomes, representing 219,997 and 26,293 gain and loss events, respectively. CNV lengths ranged from 20.02Kbp to 8.37Mbp (Figure 1), with an average of 352Kbp and a median of 204.5Kbp. The number of SNPs in each detected CNV varied from 20 to 2,116, with an average of 104 and a median of 63. A total of 138,066 CNVs were present in regions overlapping annotated genes. In particular, we observed a candidate CNV (chr16:49246631-49670918; three copies and high frequency – 347 times) that contains 8 known genes (CASP9, CELA2A, CTRC, DNAJC16, EFHD2, FBLIM1, PLEKHM2, SLC25A34). Figure 2 shows all CNVs and CNVRs within chromosome 27. We have used HD-CNV software [6] to merge the 1,044 CNVs into 128 CNVRs.

CONCLUSION

Independent published studies indicate that high-density SNP data allow for more precise CNV boundary identification, in comparison with medium-density datasets. This study represents the first published comprehensive CNV detection study with HD data in the Nelore breed. Several of the observed CNVs contain known genes with described biological roles and are thus likely to have functional impact in various metabolical functions. The same dataset is under analysis with additional algorithms to minimize Type I and Type II errors. Consistent results will be validated using qRT-PCR and next generation sequencing data.

Reference

1- Mills RE, *et al*. (2011) “Mapping copy number variation by population-scale genome sequencing”. Nature, 470(7332)59-65.

2- Redon R, *et al.* (2006) “Global variation in copy number in the human genome”. Nature, 444(7118):444-454.

3 – Iafrate AJ, *et al*. (2004) “Detection of large-scale variation in the human genome”. Nature Genetics, 36(9)949-951.

4 – Jiang L, *et al.* (2013) “Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holteins”. BMC Genomics, 14, 131-141.

5 – Wang K, *et al.* (2007) “PennCNV: Na integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data”. Genome Research, 17, 1665-1674.

6 – Butler J, *et al.* (2013) “HD-CNV: hostspot detector for copy number variants”. Bioinformatics, 29(2)262-263.

Figure 2 Candidate CNV (green – gain; red - loss) and CNVR (blue) mapped on chromosome 27.

Contact

ACKNOWLEDGEMENTS