61° CONGRESSO BRASILEIRO DE GENÉTICA
8 a 11 de setembro de 2015 Hotel Monte Real Resort, Águas de Lindóia – SP
Sociedade Brasileira de Genética
www.sbg.org.br – ISBN 978-85-89109-06-2

IDENTIFICATION AND FUNCTIONAL ANALYSIS *IN SILICO* OF POLYMORPHISMS IN MILK PROTEIN GENES OF GUZERÁ AND GIR CATTLE

Matosinho, CGR[1]; Rosse, IC[1]; Fonseca, PAS[1]; Assis, JG[1,2]; Oliveira, FS[1,2]; Araujo, F[2]; Salim, A[2]; Lopes, BC[3]; Arbex, WA[4], Machado, MA[4], Peixoto, MGCD[4]; Verneque, RS[4]; Martins, MF[4]; Coimbra, RS[5], Silva, MVGB[4]; Oliveira, G[2,6]; Carvalho, MRS[1].

1-Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, MG.
2-Grupo de Genômica e Biologia Computacional-FIOCRUZ-Minas.
3-EPAMIG – MG.
4-Embrapa Gado de Leite – MG.
5-Neurogenômica, Centro de Pesquisas René Rachou – FIOCRUZ-Minas.
6-Vale Technology Institute, PA.

mraquel@icb.ufmg.br

Keywords: milk protein, guzerá, gir, zebu.

Brazil has one at the largest commercial cattle herds in the world and is the fourth largest milk producer worldwide, yielding approximately 35 billion kg of milk per year. Brazilian herds consist of taurine breeds (*Bos taurus*), indicine breeds (*Bos indicus*) and their crossbreeds. Guzerá and Gir are the main indicine dairy cattle breeds in the national agriculture industry, contributing to milk and meat production and supplying of genetic material for breeding purposes. It is noteworthy that the milk produced by these indicine breeds contains a higher concentration of fat and lactose in comparison with of the milk produced by taurine animals. Moreover, cows' milk, in general, contains substances that cause in humans intolerance, allergies, and, putatively, some autoimmune diseases. However, the genetic basis underlying these features and/or differences is only partially known. Thus, identifying polymorphisms in the genes involved in milk composition can be the first step in the production of healthier milk. In this context, the objective of the present study was to identify and perform functional *in silico* analysis, of variations in the genes coding for milk proteins from Guzerá and Gir cattle, such as αS1-casein (*CSN1S1*), αS2-casein (*CSN1S2*), β-casein (*CSN2*), κ-casein (*CSN3*), α-lactoalbumin (*LALBA*), β-lactoglobulin (*LGB*) and lactotransferrin (*LTF*). We performed the whole genome sequencing of three Guzerá and three Gir bulls using the SOLiD and HiSeq platforms. The sequences were mapped to the reference genome of *Bos taurus* (UMD 3.1) using the LifeScope and BWA-MEM software. A list of putative SNPs and INDELs was generated from the mapped reads using LifeScope and SAMtools. Variations shared by the six samples were classified functionally according to NSG-SNP. As a result, 44 SNPs and 5 INDELs were identified in the seven genes referred above. Forty percent of the SNPs and 80% of the INDELs were previously unknown and may be associated with the main differences of indicine and taurine milk composition. Most of the SNPs discovered in this study map to intronic (61%), downstream (12%), and upstream (23%) regions. We found one (2%) missense variant in the coding region and one (2%) 3'-UTR variant. In relation to the INDELs, analysis of the functional classes showed that four (80%) were located in intronic regions and one (20%) in the upstream region. Among these seven genes, the *LTF* gene was the one with the highest number of variations (33 SNPs and 4 INDELs). This study provided a large number of new SNPs and INDELs, which may be responsible for the main differences in the milk proteins composition between indicine and taurine breeds.

Supported by: CAPES, CNPq, FAPEMIG (CBB-1181/0 and TCT 12.093/10), NIH-USA (TW007012), CAPES/CDTS-FIOCRUZ, FIOCRUZ-MG, PDTIS-FIOCRUZ - Platform RPT04B, Bioinformatics BH, Embrapa.