

NOVEL POLYMORPHISMS IN GENES ASSOCIATED WITH MILK AND MEAT PRODUCTION AND DISEASE RESISTANCE IN THE GUZERÁ BREED IDENTIFIED BY WHOLE-GENOME SEQUENCING

IC Rosse¹, JG Assis^{1,2,3}, FS Oliveira^{1,2,3}, LR Leite^{1,2,3}, F Araujo^{2,3}, A Zerlotini⁴, BC Lopes⁵, WA Arbex⁶, MAMachado⁶, MGCD Peixoto⁶, RS Verneque⁶, MFM Guimarães⁶, MVGB Silva⁶, RS Coimbra⁷, G Oliveira^{2,3}, MRS Carvalho¹

¹ Instituto de Ciências Biológicas – UFMG

² Genomics and Computational Biology Group - FIOCRUZ-MG

³ Center for Excellence in Bioinformatics – FIOCRUZ-MG

⁴ EMBRAPA Informática Agropecuária - SP

⁵ EPAMIG – MG

⁶ EMBRAPA Dairy Cattle – MG

⁷ Biosystems Informatics, Research Center René Rachou – FIOCRUZ

Background: Currently, livestock accounts for 9% of internal product in Brazil and the herds that take part in milk selection programs have achieved gains of up to 1% per year. Genotyping arrays composed by hundred-thousands of SNPs were developed aiming the use in genome wide

association studies and genomic selection. These genotyping arrays contain variations observed in several commercial breeds, but do not include the Guzerá (*Bos taurus indicus*). Guzerá is a dualpurpose breed, well adapted to the tropical climate and that presents resistant to parasites and low susceptibility to mastitis. In this context, the objective of this work was to sequence and assemble the genome of one Guzerá bull in order to identify breed-specific variations that might be included in the genotyping arrays and be useful in breeding programs.

Results: We performed the genome assembly of the sequence of a Guzerá bull produced with SOLiD v4 second-generation sequencing using mate-paired libraries with 1-2 and 3-4 kb inserts. The sequences were mapped to the publicly available reference genome of *Bos taurus taurus* (UMD 3.1) using the LifeScope software. The average depth of coverage achieved from mapping was 26X for each chromosome and 87% of the reference genome was covered. A list of putative SNPs and INDELS were generated from the mapped reads, using the LifeScope and SAMtools. After filtering for quality, 4,200,936 SNPs and 664,704 INDELS were identified. Sixty-five percent of the SNPs and 89% of the INDELS were previously unknown. Additionally, 2,676,067 (64%) of the SNPs and 466,005 (70%) of the INDELS were homozygous and may represent true differences between Guzerá and *Bos taurus taurus*. Ninety percent of the over 4 million SNPs detected in this individual have not been included in the current commercial genotyping arrays and will be useful for the construction of more effective genotyping arrays for the Guzerá breed. Of 3,142,072 genetic differences in Guzerá, 2,041 variations were classified functionally according to NGS-SNP in nonsynonymous SNPs, splice-site variants, frameshift and coding INDELS. These variations were detected in 1,690 genes. Of the genes containing these variations, 54 were assigned as QTL for milk and meat quality and disease resistance based on dbQTL and literature search.

Conclusions: This study provided us with a large number of new SNPs, which may be used in order to enhance SNP genotyping arrays. Additionally, potentially functional variants were identified, that map to regions where QTLs were already detected and that constitute valuable markers to be used in association studies.

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.