

Detection of potential genetic variants affecting gene function in Guzerat cattle

Adhemar Zerlotini^{1¶}, Nedenia Bonvino Stafuzza^{2¶}, Francisco Pereira Lobo¹, Michel Eduardo Beleza Yamagishi¹, Tatiane Cristina Seleguim Chud², Alexandre Rodrigues Caetano³, Danísio Prado Munari², Dorian J. Garrick⁴, Marco Antonio Machado⁵, Marta Fonseca Martins⁵, Maria Raquel Carvalho⁶, Marcos Vinicius Gualberto Barbosa da Silva⁵

¹Embrapa Informática Agropecuária, Campinas, São Paulo, Brazil, ²Departamento de Ciências Exatas, Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal, São Paulo, Brazil, ³Embrapa Recursos Genéticos e Biotecnologia, Brasília, Distrito Federal, Brazil, ⁴Department of Animal Science, Iowa State University, Ames, Iowa, United States of America, ⁵Embrapa Gado de Leite, Juiz de Fora, Minas Gerais, Brazil, and ⁶Universidade Federal de Minas Gerais, Minas Gerais, Brazil

Guzerat is a dual-purpose breed recognized for important traits to its adaptation to adverse tropical environments such as resistance to parasites, heat tolerance and ability to intake forage with low nutritional value. Once genetic variation responsible for this traits has so far not been well characterized, the aim of this study was to identify single nucleotide variants (SNVs) and insertion/deletions (Indels) in Guzerat cattle breed from whole genome re-sequencing in order to characterize loss-of-function variants which could be associated with complex traits in this cattle breed. The genetic variants identified using HiSeq 2000 sequencing platform were classified using the Ensembl Variant Effect Predictor (VEP) tool, based on a subjective classification of the severity of the variant consequence in transcripts defined by the Sequence Ontology (high, moderate, low or modifier). The DAVID v6.7 tool was used for functional enrichment analysis using the list of genes predicted to be function or activity potentially affected by genetic variation (high and moderate effect) against the background set of bovine genes. The Gene Ontology (GO) enrichment analysis were performed considering 5% FDR threshold for significance. The KEGG pathway database was used to investigate the enriched metabolic pathways ($P < 0.01$). Approximately 16.9 million genetic variants (14,588,361 SNVs and 2,322,894 indels) were identified, of which 45,452 (43,922 SNVs and 1,530 indels) were predicted to have high and moderate effect on gene function. The data set comprised 9,922 Ensembl ID because many genes showed a high number of genetic variants. A total of nine GO terms ($FDR < 0.05$) were over represented with genetic variants, with the following molecular functions: olfactory receptor, GTPase regulator, nucleoside-triphosphatase regulator, Ras guanyl-nucleotide exchange factor and Rho guanyl-nucleotide exchange factor activity. These enriched terms may reflect the selection history of this Guzerat population and could be associated with some phenotypic traits. The KEGG pathways ($P < 0.01$) that could be affected by genetic variants were fatty acid metabolism, ECM-receptor interaction, complement and coagulation cascades, olfactory transduction, aminoacyl-tRNA biosynthesis and lysosome pathways. Because only three animals were sequenced, the over-represented GO terms and KEGG pathways should be carefully interpreted. However, these results provide important genomic information to investigate the genetic mechanisms underlying traits of interest in Guzerat cattle and to improve genomics-based breeding tools.