

Genome wide association study to detect chromosomal regions related to milk, fat and protein yield in Gir breed

Adam Taiti Harth Utsunomiya; UNESP Jaboticabal
Ricardo da Fonseca; UNESP Jaboticabal
Daniel Jordan de Abreu dos Santos; UNESP Jaboticabal
Marco Antônio Machado; EMBRAPA Gado de Leite
Rui da Silva Verneque; EMBRAPA Gado de Leite
Marcos Vinícius Gualberto Barbosa da Silva; EMBRAPA Gado de Leite
Corresponding author e-mail: marcos.vb.silva@embrapa.br

Abstract:

Introduction: The Genome-Wide Association Studies (GWAS) allow to identify statistically significant associations between hundreds of thousands SNP markers and phenotypes of complex traits, increasing the understanding of important biological pathways. Interesting results have been found in GWAS for taurine dairy cattle, providing confirmatory evidence of an association between the regions tested with gene databases. However, further studies involving zebu cattle are needed to check for overlapping regions already identified and described in taurine with those that can be found in zebu dairy cattle. **Objectives:** This work aims prospect information about the genes involved in milk (MY), fat (FY) and protein yield (PY) in Gyr breed, through genome wide association study. **Methods:** 1859 animals were genotyped using the Illumina BeadChip® BovineSNP50K and UMD v3.1 assembly. The Quality Control of samples and SNPs were performed in the GenABEL package of statistical software R. Samples with call rate < 0.9 were removed from the analysis as well as SNPs with MAF < 0.02, call rate < 0.98 and p-value of Fisher's exact test for HWE < 1×10^{-6} . The phenotypes used to GWAS were deregressed EBVs with accuracies greater than 0.4. Thus, three datasets were generated: MY with 1834 animals and 16693 SNPs, FY with 1793 and 16701; PY 1577 and 16459, respectively. We also performed a principal component analysis for indicate population substructures. The GRAMMAR-GAMMA method was applied to the data sets by taking into account the population substructure. The significance level of 0.05 for association's test was corrected by Bonferroni adjustment. Only SNPs with significant p-values for each trait were analyzed and, through their respective genomic coordinates were prospected candi date genes nearest from the NCBI database. **Results:** The markers Hapmap39532-BTA-102468(BTA1), BTA-70020-no-rs(BTA4), Hapmap24120-BTC-033936(BTA6) and BTA-85350-no-rs(BTA9) were significantly associated with MY. During the exploration of candidate genes were found the genes PTX3 and VEPH1 on BTA1, SEMA3E and PCLO on BTA4, FAM13A, HERC3 and LOC100847719 on BTA6 and UBE3D on BTA9. The PCLO gene has a function on insulin secretion pathway. During lactation, high-producing cows have low serum concentrations of insulin in the blood, because the mammary gland uses glucose as an energy source and for the synthesis of lactose. The HERC3 gene is related to ubiquitin-mediated proteolysis. The metabolic status of cows in early lactation (negative energy balance) can affect your body reserves (fat and protein) and activate the ubiquitin-proteasome proteolytic pathway for protein degradation in skeletal muscle. The remaining genes are involved in milk yield, for affecting the immune (resistance to mastitis) and the vascular system (irrigation mammary glands). **Conclusion:** Only SNPs associated with MY were identified and all candidate genes could be associated with some prospected metabolic pathway affecting this trait.