Identification of causal variants for milk protein composition using sequence data in dairy cattle

Marie-Pierre Sanchez, Armelle Gion Govignon-Gion, Pascal Croiseau, Anne Barbat, M. Gelé, Sebastien S. Fritz, Guy Miranda, Patrice Martin, Mekki Boussaha, Mickael M. Brochard, et al.

Identification of causal variants for milk protein composition using sequence data in dairy cattle

Sanchez M.P.[1], Govignon-Gion A.[1], Croiseau P.[1], Barbat A.[1], Gelé M.[2], Fritz S.[3], Miranda G.[1], Martin P.[1], Boussaha M.[1], Brochard M.[2], Boichard D.[1]

[1]INRA, AgroParisTech, UMR1313 GABI, 78350 Jouy-en-Josas; [2]IDELE, 75012 Paris; [3]UNCEIA, 75012 Paris; France; marie-pierre.sanchez@jouy.inra.fr

A genome wide association study (GWAS) was performed at the sequence level to identify candidate mutations for major milk proteins in Montbéliarde (MO), Normande (NO) and Holstein (HO) dairy cattle breeds. The concentration in the six major proteins was estimated from Mid-Infrared (MIR) spectrometry on almost 600,000 test-day milk samples from 116,495 cows in the first three lactations (*PhénoFinlait* project). Out of these, 8,080 cows (2,967 MO, 2,737 NO and 2,306 HO) were genotyped with the Illumina 50k Beadchip. For each breed, genotypes were first imputed into the HD genotypes using HD genotypes of 522 MO, 546 NO and 776 HO bulls. The resulted HD genotypes were subsequently imputed at the sequence level using 27 millions of highly confident sequence variants selected from the latest run of the 1000 bull genomes project (1147 bulls). Phenotypes were average test-day measurements adjusted for environmental effects. Within breed association studies were performed for each sequence variant by using a mixed model including also a mean and a random polygenic effect. Numerous QTL regions were identified in each breed. Three QTL, located on chromosomes 6, 11 and 20 had very significant effects and were shared by the three breeds. Other significant effects, partially overlapping across breeds, were found on almost all autosomes. Potential candidate variants were identified in several QTL regions, including mutations previously described in LGB, DGAT1 or GHR genes. These results show that GWAS applied to whole sequence genotypes is a promising approach to the identification of QTL with a better resolution than lower density genotypes. The authors acknowledge the financial support from ANR and APIS-GENE, and the contribution of the 1000 bull genomes consortium.