Use of whole sequence GWAS to improve genomic evaluation in dairy cattle

Pascal Croiseau, Thierry Tribout, Didier Boichard, Marie-Pierre Sanchez, Sebastien Fritz

Use of whole sequence GWAS to improve genomic evaluation in dairy cattle

Pascal, Thierry, Didier, Marie Pierre, Sébastien

The 1000 bull genomes reference population makes it possible to impute whole genome sequences for a large number of animals. Even if the imputation accuracy is still limited for rare variants, it becomes possible to directly pinpoint the causal mutations, or at least variants in very high linkage disequilibrium with them.

However, sequence data (ie tens of millions of variants for several thousands of animals) cannot be used in routine genomic evaluation and only the most predictive part of them should be identified and used. In this study, we propose to construct virtual chips containing 50k variants (V50K) selected from the whole genome sequence, compatible with a routine use.

About 3,000 bulls from each of the French Montbéliarde and Holstein breeds were studied for a panel of ten traits related to milk production and composition, fertility, and conformation. Each population was divided in training and validation sets. GWAS analyses were carried out at the sequence level on the training population using the GCTA software. V50K were then constructed considering different criteria: GWAS results, minor allele frequency, distribution of the variants on the genome, and functional annotations.

GBLUP and BayesC approaches were applied on these V50K and compared with the results of the Illumina Bovine SNP50 Beadchip® (IB50K). GEBV accuracy was estimated through correlations between DYD and GEBV in the validation set. Most of the tested scenarios failed to improve accuracy of GEBV obtained with the IB50K. Nevertheless, a good choice of the variants can outperform results of the standard IB50K.