The Use of A Priori Biological Information to Evaluate the Ability of BayesRC Model in Genomic Prediction

Baber Ali, Fanny Mollandin, Pascal Croiseau

▶ **To cite this version:**

Baber Ali, Fanny Mollandin, Pascal Croiseau. The Use of A Priori Biological Information to Evaluate the Ability of BayesRC Model in Genomic Prediction. 72st Annual Meeting of the European Association for Animal Production, Aug 2021, Davos, Switzerland. hal-03742059

HAL Id: hal-03742059

https://hal.inrae.fr/hal-03742059v1

Submitted on 2 Aug 2022

The Use of *A Priori* Biological Information to Evaluate the Ability of BayesRC Model in Genomic Prediction

Baber Ali[1,2], Fanny Mollandin[1], and Pascal Croiseau[1]

[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI , [2]Department of Animal Sciences and Center of Integrated Breeding Research, the University of Göttingen, Göttingen, Germany.

Abstract

The increase in technological advancement and decrease in SNP genotyping chip costs have led to identifying the candidate variants of complex traits, analyzing genetic architecture, and estimating genomic breeding values. The newest Bayesian genomic evaluation methods have the property to deal with all three aspects. This is the case of the BayesR which modelizes different class of genetic variance for SNPs, where SNP effect is assumed to be distributed as $\beta_i \sim \pi_1(\beta_i = 0) + \pi_2 N(0, 0.0001\sigma_g^2) + \pi_3 N(0, 0.001\sigma_g^2) + \pi_3 N(0, 0.01\sigma_g^2)$, but it cannot account for the prior biological knowledge. However, BayesRC, an extension of BayesR, accounts for this property and offers to divide SNPs into functional annotation categories as $\beta_i | annotation(i) = a \sim \pi_{1,a} N(0, \sigma_g^2) + \pi_{2,a} N(0.0001, \sigma_g^2) + \pi_{3,a} N(0.001, \sigma_g^2) +$ $\pi_{4,a} N(0.01, \sigma_g^2)$. Also, a SNP can only belong to one category of annotation. Therefore, we propose in this study to test, on a real dataset of 7483 Holstein bulls, the ability of a panel of annotation categories to improve the quality of prediction. This study used daughter yield deviation (DYD) phenotypic records for milk, fat and protein yield, clinical mastitis, and somatic cell count. We categorized SNPs into different annotation categories based on their association with a complex trait and verified them by various sources of literature and genomic databases (i.e. AnimalQTLdb: https://www.animalgenome.org/cgi-bin/QTLdb/index, and cGTEX: https://cgtex.roslin.ed.ac.uk/) . Results indicate that the same level of prediction accuracy is obtained for BayesR and BayesRC. However, BayesRC has a better ability to arrange SNPs in the effect classes because of the prior knowledge. The results also indicate that the top twenty SNPs for milk yield based on their inclusion probability in the fourth effect class ($0.01\sigma^2$g) correspond to chromosomes 1, 5, 6, 11, 14, 15, 20, and 26. These SNPs were then mapped and found to be associated with known milk production QTLs spanning genes like *DGAT1, GHR, ABCG2, PAEP, ABCG2, ...* BayesRC is quite a promising approach given its prediction accuracy and repeatability.