Evaluation of heritability partitioning approaches in livestock populations

Yuan C.[1], Gualdrón Duarte J. L.[2], Takeda H[1] , Georges M.[1] & Druet T.[1]

[1]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Avenue de l'Hôpital, 1, 4000 Liège, Belgium. [2] Walloon Breeders Association, Rue des Champs Elysées, 4, 5590-Ciney, Belgium.

Background

Heritability partitioning approaches estimate the contribution of different functional classes to the genetic variance, providing a better understanding of the genetic architecture of complex traits. However, despite the widespread adoption of these methods in livestock genetics, their performance needs to be evaluated because: methods have been developed and evaluated primarily in human genetics; livestock populations have different properties such as: high levels of relatedness, small effective population size, and long-range levels of linkage disequilibrium. In this study, we evaluated the heritability partitioning methods in livestock populations.

Method

Data. We used data from 14,762 Belgian Blue cows with imputed genotypes for > 11 million variants classified into coding sequence (CDS), untranslated regions (UTR), regions upstream or downstream of genes (UDR), intronic regions (IOR), intergenic regions (IGR), and open chromatin regions (OCR). LD and MAF categories have also been defined for LDS, MS and LDMS models.

Heritability partitioning approaches: We used GREML, with one genomic relationship matrix (GRM) $\boldsymbol{G}_S$ for each of the S categories:

$$\boldsymbol{y} = \mathbf{1}\mu + \sum_{s=1}^{S} \boldsymbol{g}_s + \boldsymbol{e} \text{ where } \boldsymbol{g}_s \sim N(0, \boldsymbol{G}_s\sigma_s^2)$$

We also used a Bayesian mixture model (BayesRR-RC) with:

$$\boldsymbol{y} = \mathbf{1}\mu + \sum_{s=1}^{S} \boldsymbol{X}_s\,\boldsymbol{\beta}_s + \boldsymbol{e}$$

and

$$\beta_{s_j} \sim \pi_{0_s}\delta_0 + \pi_{1_s}N(0,\sigma^2_{1_s}) + \pi_{2_s}N(0,\sigma^2_{2_s}) + \ldots + \pi_{L_s}N(0,\sigma^2_{L_s})$$

Where the mixture proportions (π) and variances σ_s^2 are class specific.

Simulations. Phenotypes were simulated by assigning effects to 5,000 to 10,000 causal variants (CVs). CVs were randomly selected or enriched in certain functional, LD or MAF categories according to the scenario.

Single functional annotation models and simulations

We started with simple simulations where a single functional class (OCR) was enriched in CVs. Variants in OCR accounted for 25%, 50% and 75% of the heritability. Two component (TC) models fitting one functional class versus other classes (e.g., OCR vs ~~OCR~~), were used. Estimates were unbiased despite their high levels of variations (Figure 1).

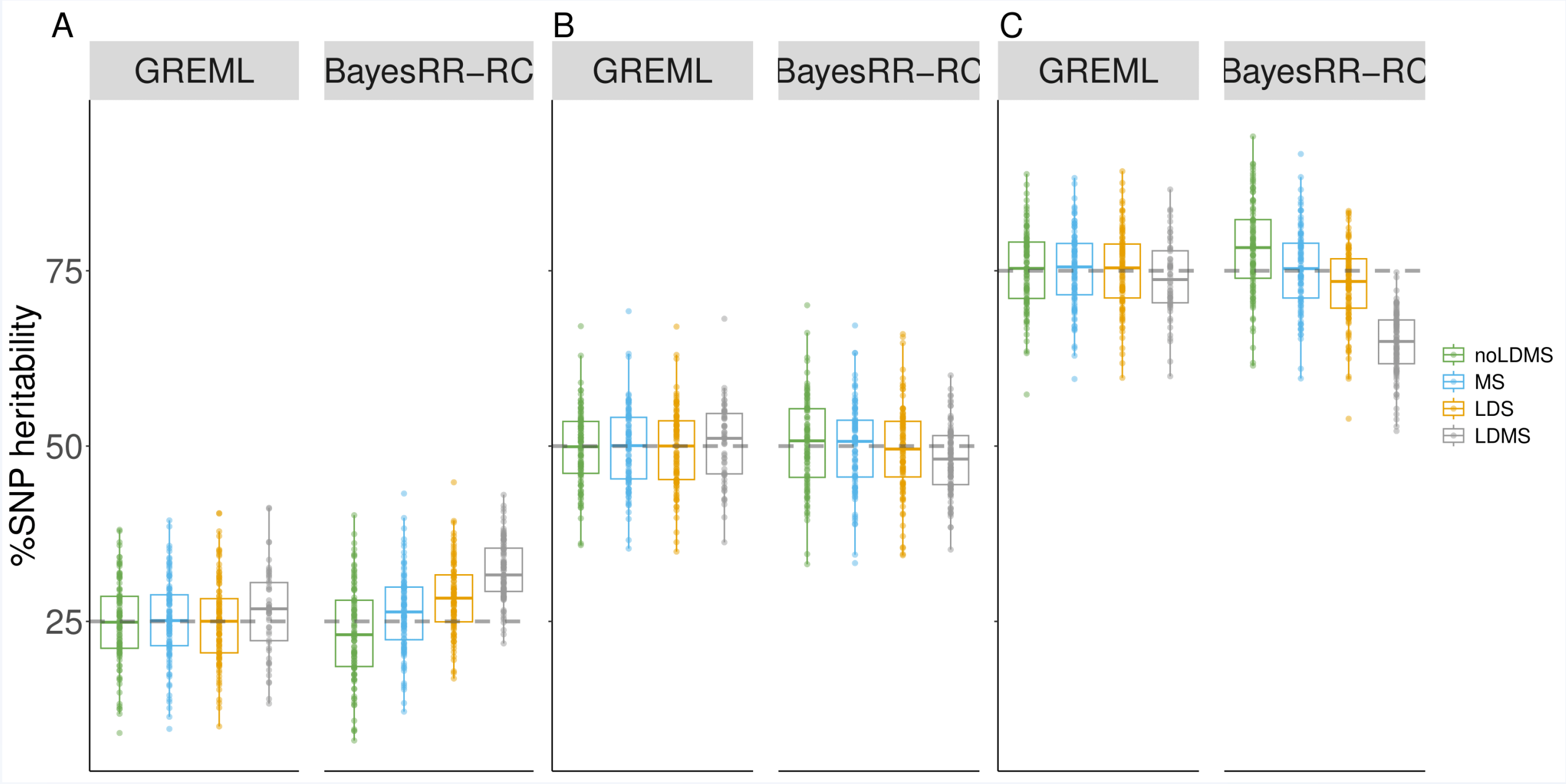


Figure 1. Estimation of %heritability when OCR accounted for 25, 50 and 75% of heritability.

We then simulated scenarios in which CVs were also enriched in some MAF or LD score categories. LDMS models were the most robust in these simulations.

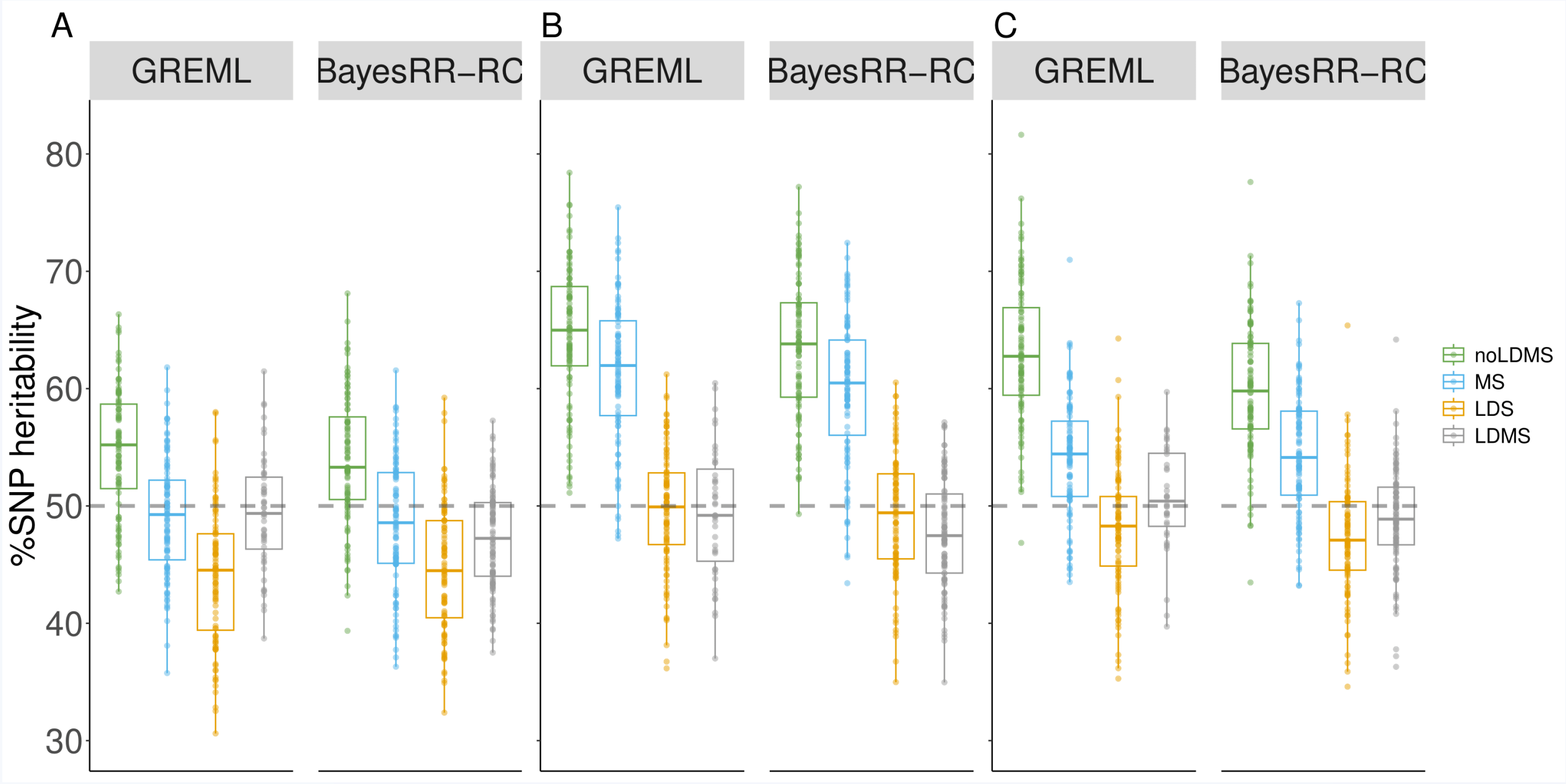
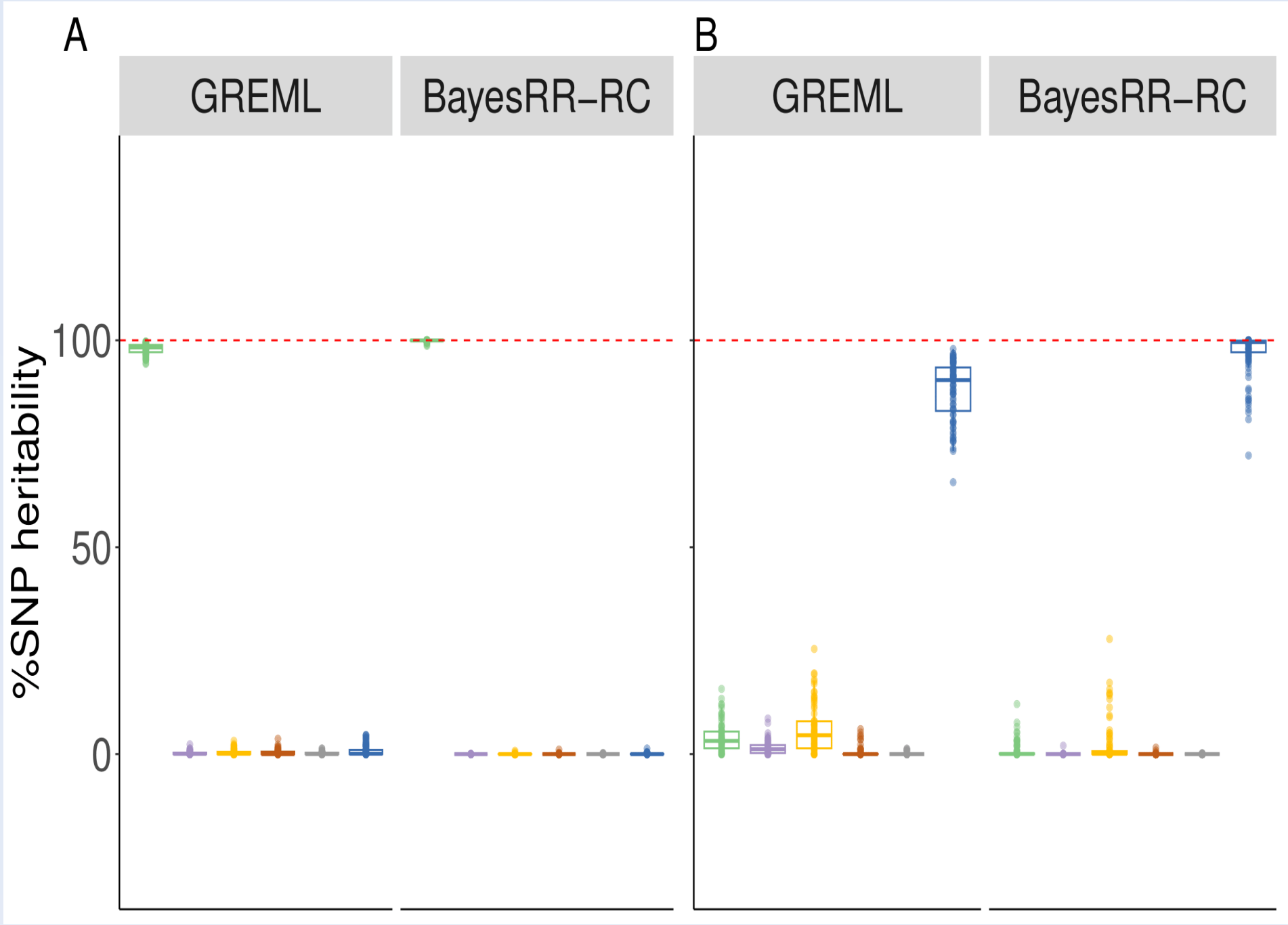


Figure 2. Estimation for simulations in which CVs were enriched in A) low MAF variants (MAF < 0.05), B) Low LD variants (LD score in the 1st quartile), and C) low MAF and low LD variants.

Evaluation of multiple functional classes models

We tested models with multiple functional classes in simulations in which CVs came from a single functional class (Figure 3) or without enrichment (Figure 4). Estimates of %SNP heritability were unbiased, but estimates from small classes had greater variation.



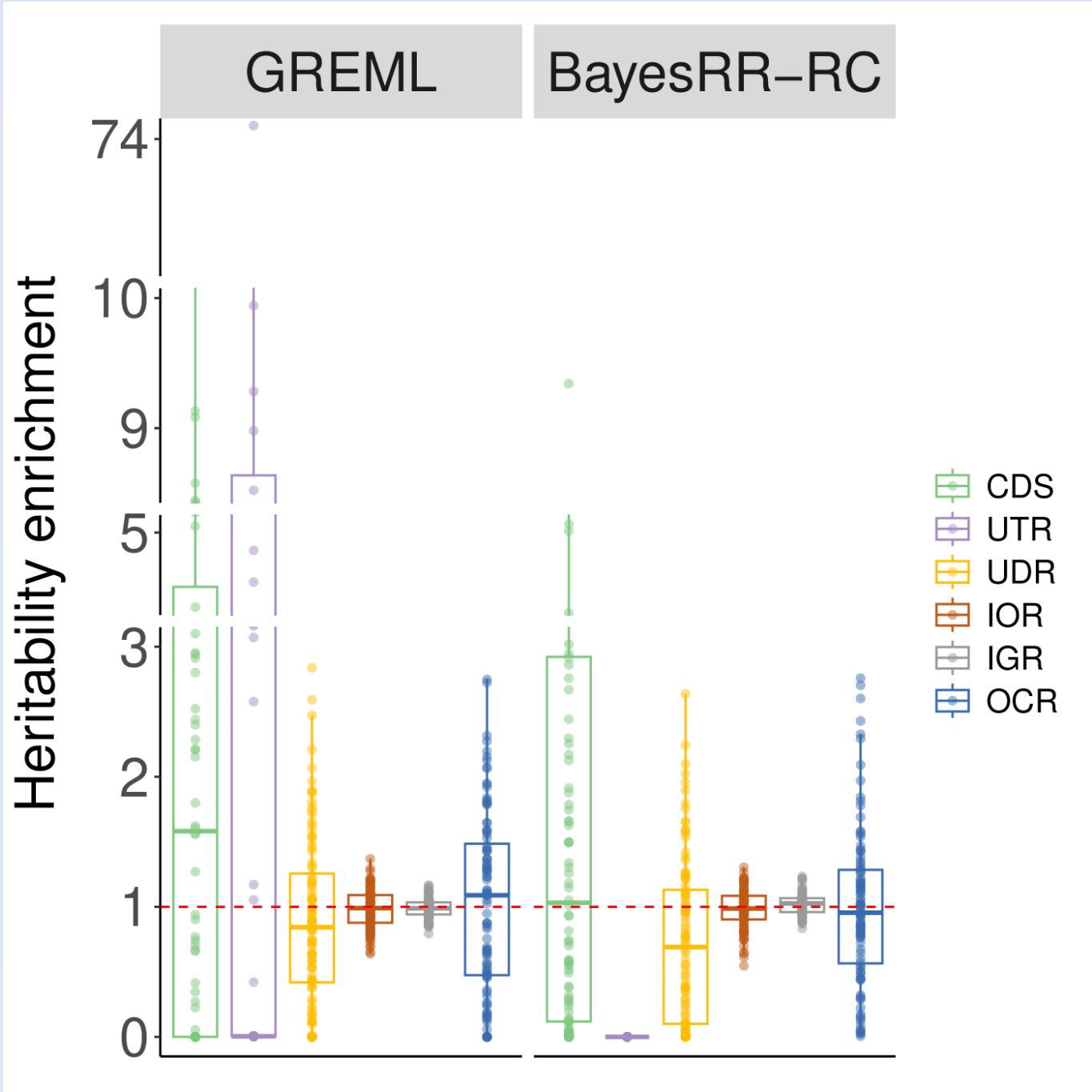


Figure 3 (left) and Figure 4 (right). Estimation of %heritability in simulations with 100% SNP heritability associated with one class, or random CVs distribution.

We then evaluated the same models in more complex simulation scenarios in which functional classes had different contributions to heritability. BayesRR-RC was more accurate than GREML in estimating the genetic contribution of each class (Figure 5). However, the estimates had low precision, were biased for some classes, and we observed confounding between classes.

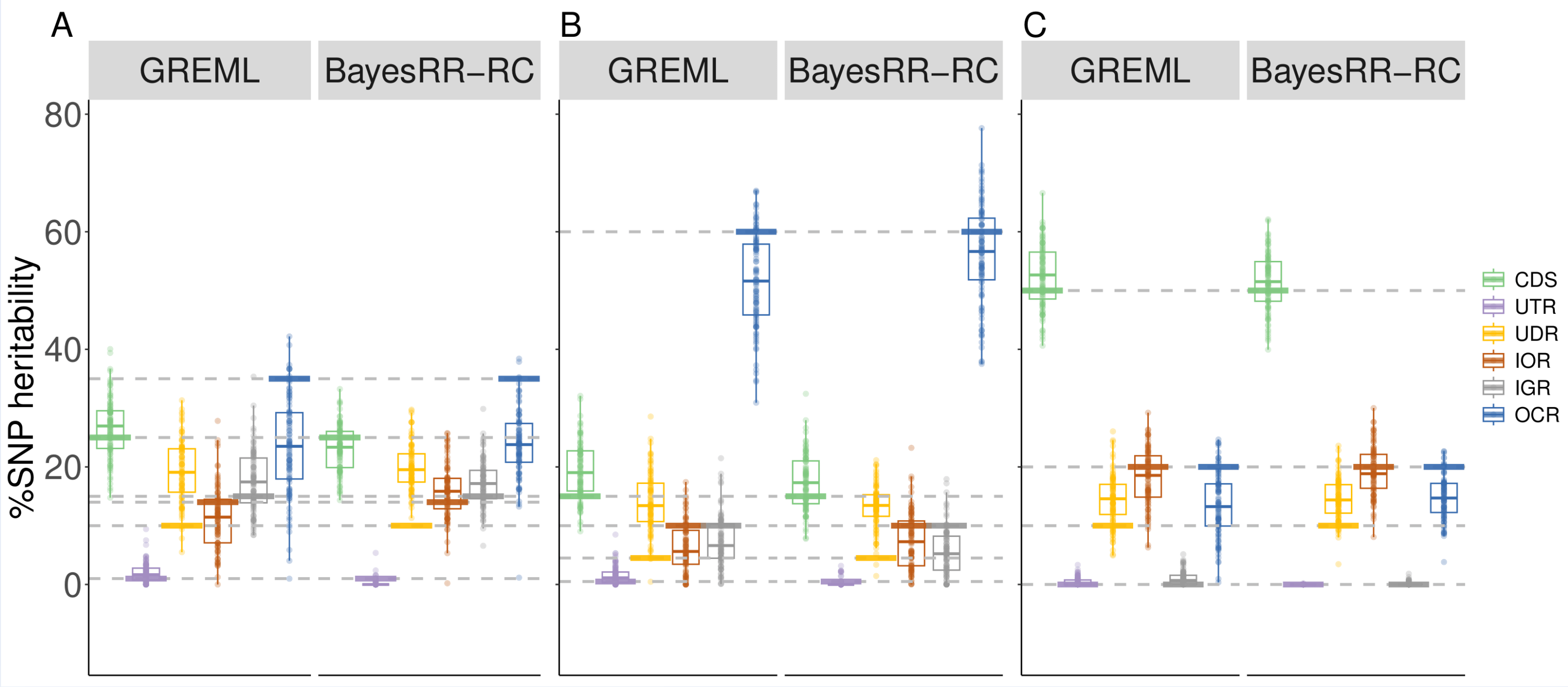


Figure 5. Estimation of %heritability using multiple component models. The simulated contribution for each category is shown by the horizontal bar.

Finally, when TC models were used in the same scenarios: the %heritability of OCR was highly overestimated, except for the scenario without enrichment (Figure 6). This overestimation was also observed for other functional categories when using the same approach.

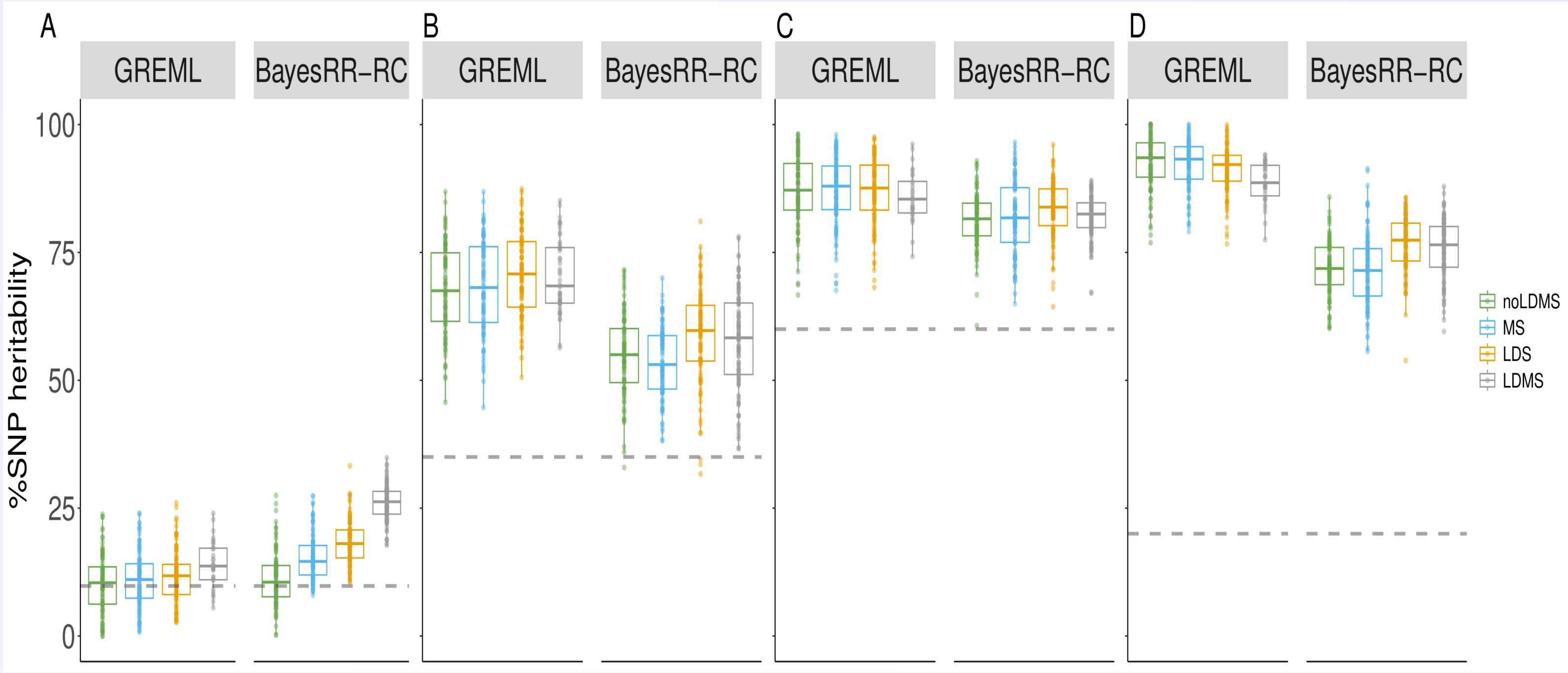


Figure 6. Estimation of %heritability using two component models in scenarios without enrichment (A) and the three complex scenarios (B-D).

Conclusions

Heritability partitioning approaches should be used with caution in livestock populations. Estimators were particularly imprecise for small categories, so models with too many and small categories should not be used. Two-component approaches that fit only one functional category at a time produced biased estimates and should not be used. Although estimates were imprecise in complex scenarios, they were still informative about the contribution of the functional classes we fitted.

More details can be found in Yuan et al. (2024) BMC Genomics 25:690

LIÈGE université
GIGA institute

Can.Yuan@uliege.be