

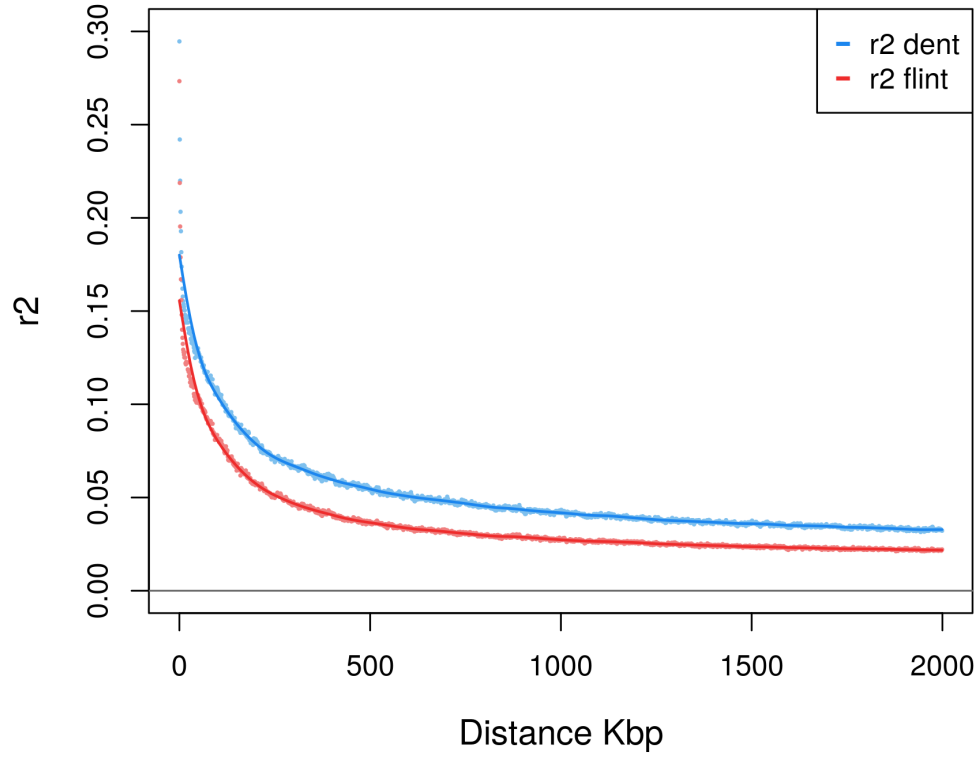
Supplementary Figures

Accounting for group-specific allele effects and admixture in genomic predictions: theory and experimental evaluation in maize

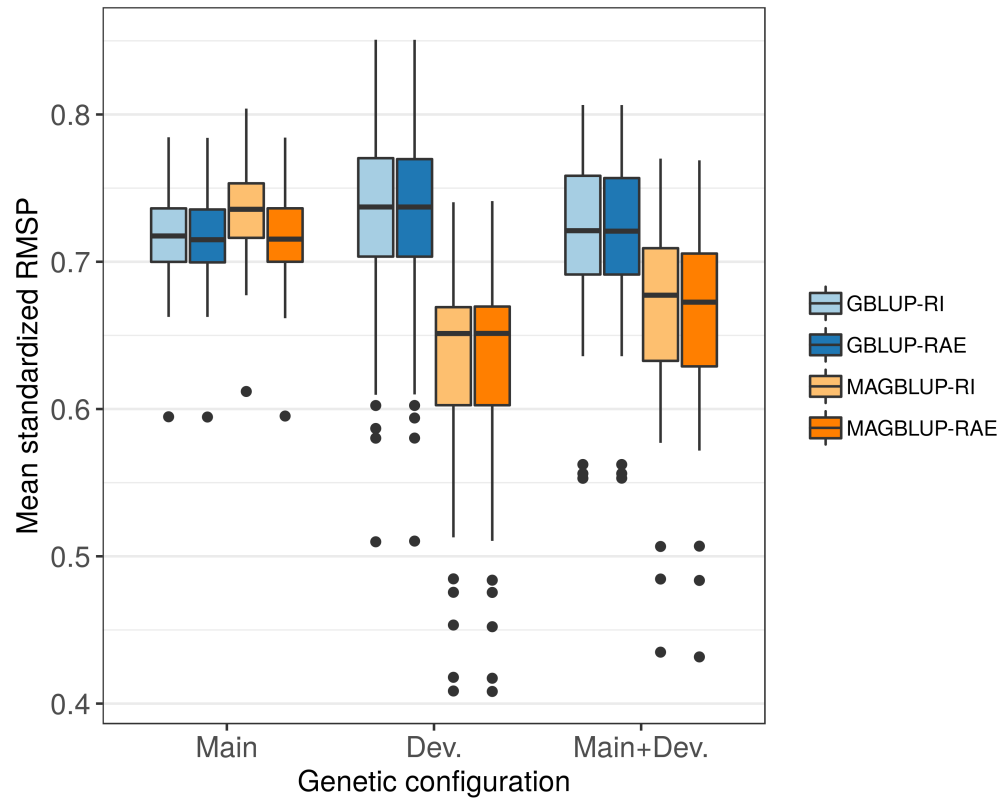
Simon Rio*, Laurence Moreau*, Alain Charcosset*, and Tristan Mary-Huard*,[†]

*Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190,
Gif-sur-Yvette, France

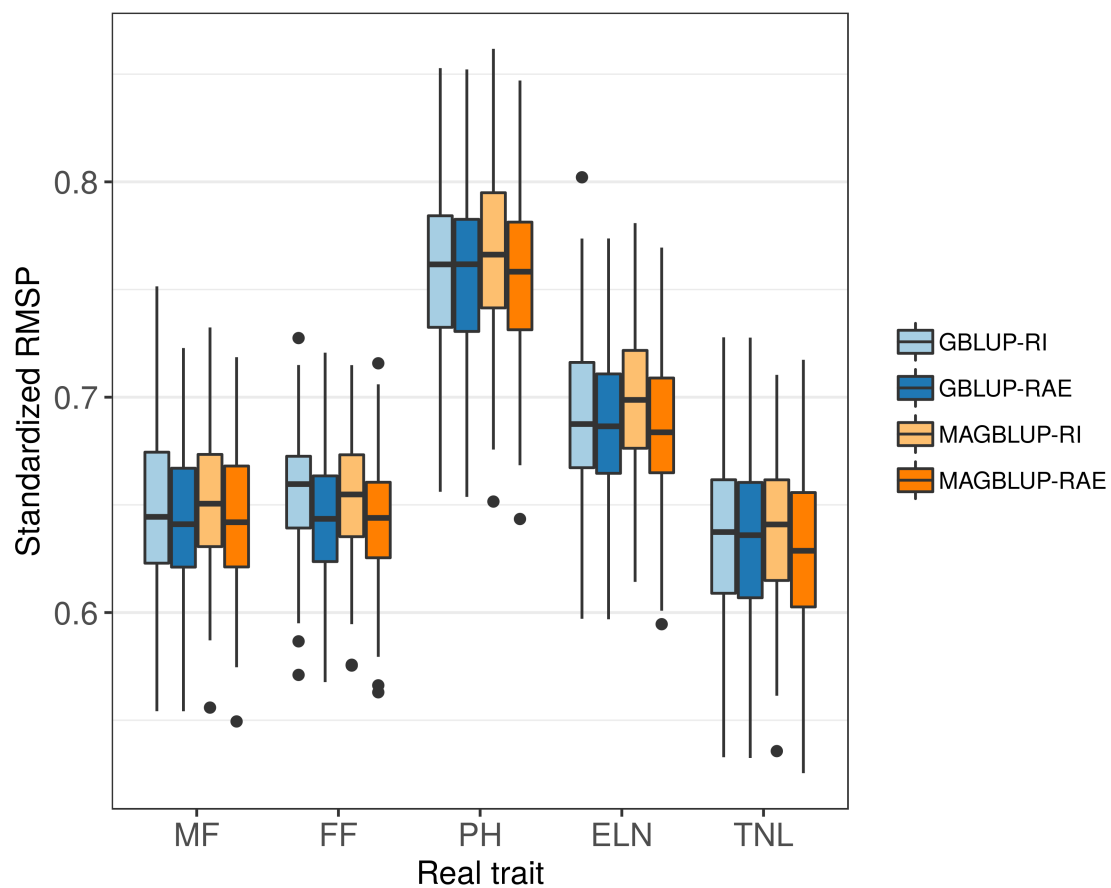
[†]MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France



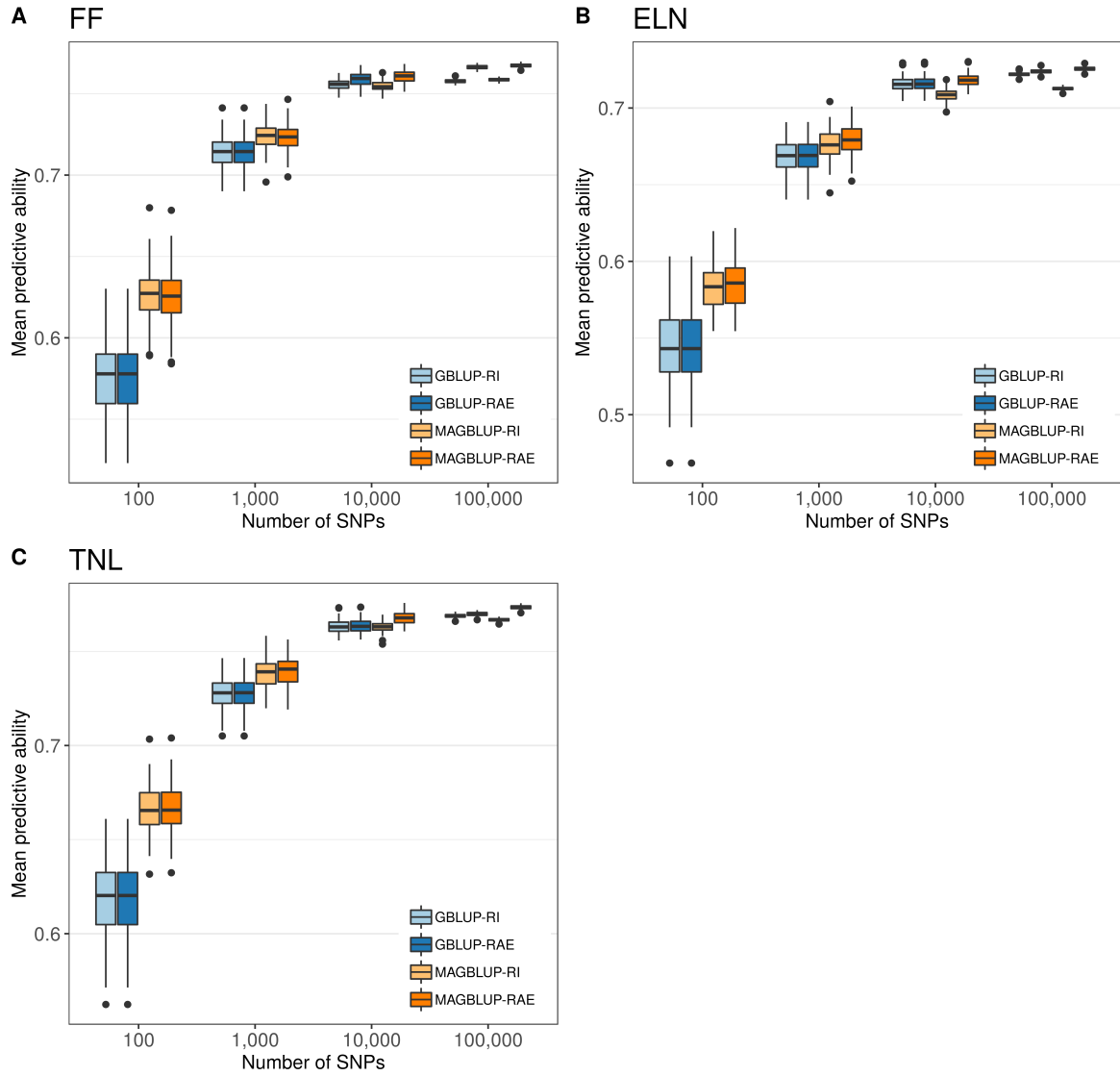
Supplementary Figure S1. LD extent estimated separately in the dent and flint genetic groups using the standard r^2 . LD was calculated and averaged for loci pairs characterized by a similar physical distance ranging from 0 to 2 Mbp, considering a sliding window of 1Kbp. A cubic smooth spline was adjusted for each group, using the R function “smooth.spline”. On these data, cumulative SNP genome coverage over chromosomes is 2.058 Gbp. On average, LD is below $r^2 = 0.1$ for SNPs that are distant by more than 113 Kbp and 56 Kbp in the dent and flint genetic groups, respectively. Based on these results, it can be estimated that around 18K SNPs and 37K SNPs are independent in the dent and flint datasets, respectively.



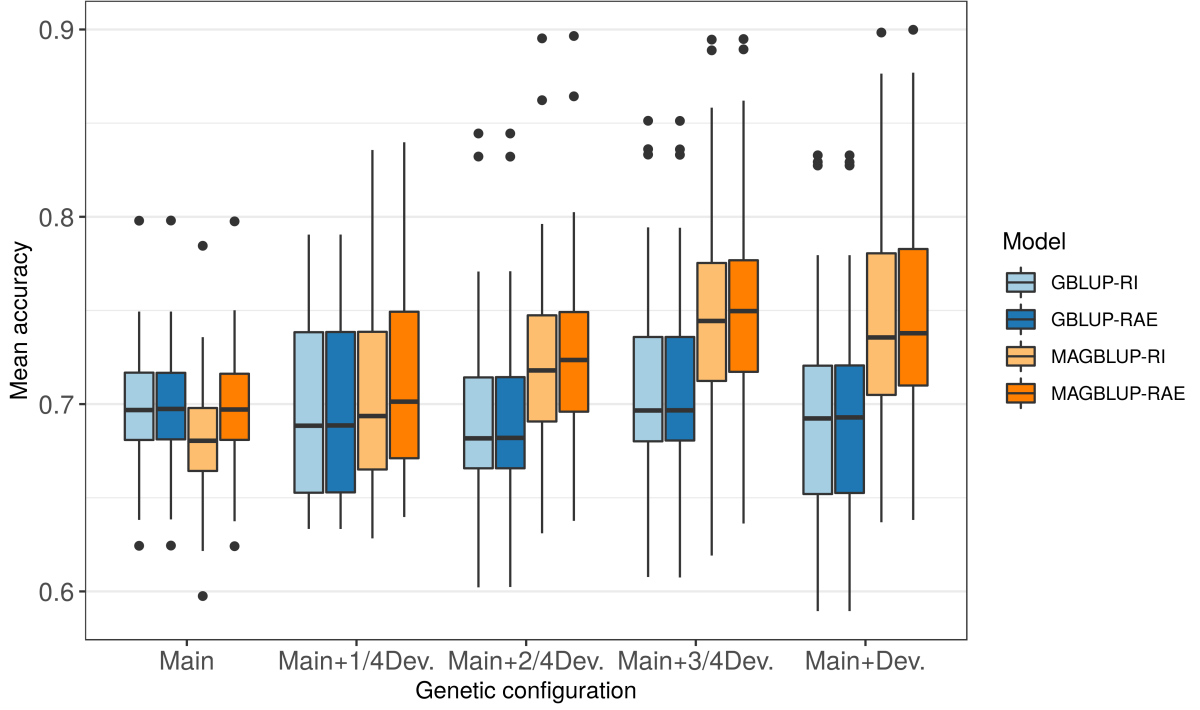
Supplementary Figure S2. Boxplots of mean standardized RMSP (evaluated by CV over 20 HO replicates) obtained using GBLUP-RI, GBLUP-RAE, MAGBLUP-RI and MAGBLUP-RAE for 50 traits simulated according to each type of genetic configuration ("Main", "Dev." and "Main+Dev.", see Table 1).



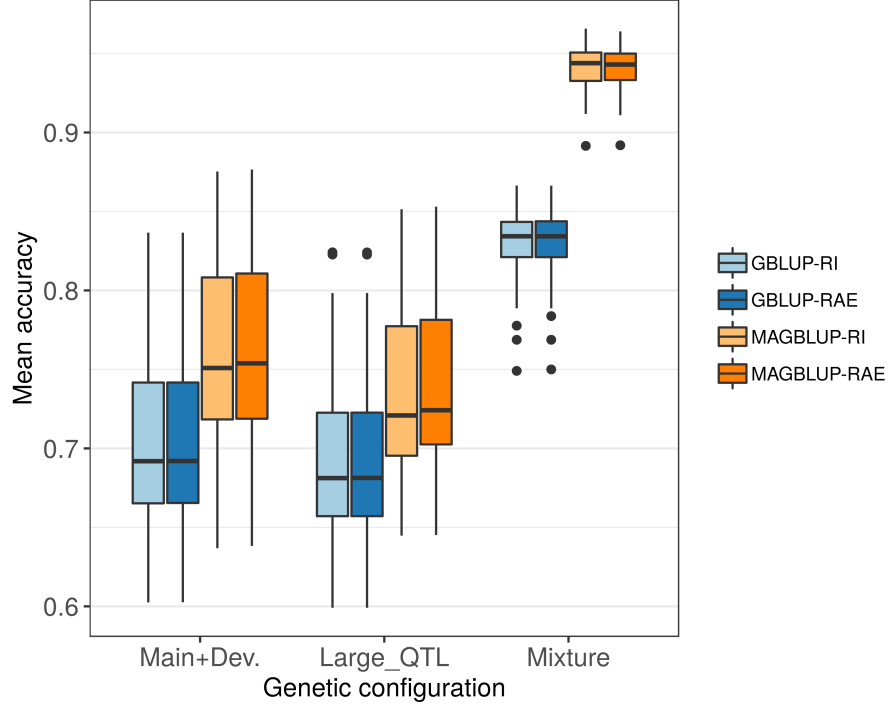
Supplementary Figure S3. Boxplots of standardized RMSP evaluated by CV (HO method with 100 replicates) for GBLUP-RI, GBLUP-RAE, MAGBLUP-RI and MAGBLUP-RAE for real traits: male flowering (MF), female flowering (FF), plant height (PH), ear leaf number (ELN) and total number of leaves (TNL).



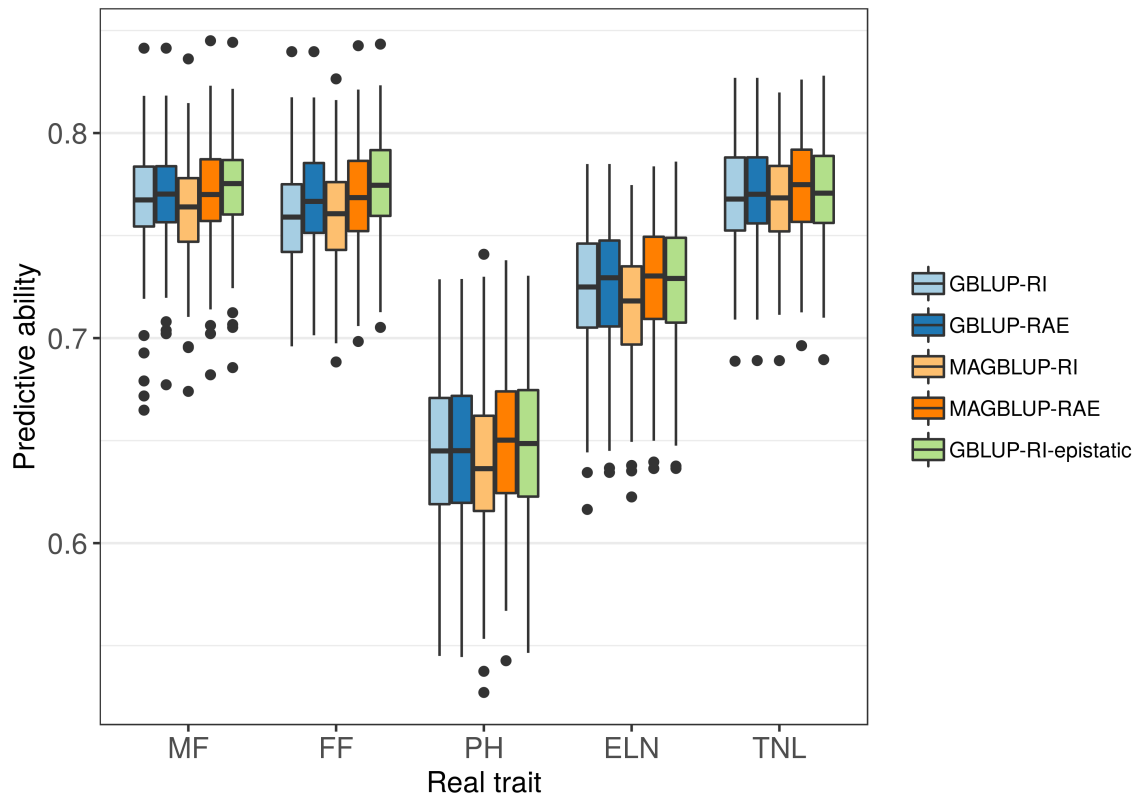
Supplementary Figure S4. Boxplots of mean predictive abilities (evaluated by CV over 100 HO replicates) obtained using GBLUP-RI, GBLUP-RAE, MAGBLUP-RI and MAGBLUP-RAE on real traits : (A) female flowering (FF), (B) ear leaf number (ELN) and (C) total number of leaves (TNL), with covariance matrices computed using SNP samples (100 replicates) of various densities: 100, 1,000, 10,000 and 100,000 SNPs.



Supplementary Figure S5. Boxplots of mean accuracies (evaluated by CV over 20 HO replicates) obtained using GBLUP-RI, GBLUP-RAE, MAGBLUP-RI and MAGBLUP-RAE for 50 traits simulated according to genetic configuration "Main", "Main+Dev." (see Table 1) and a gradient of intermediate genetic configurations called "Main+1/4Dev.", "Main+2/4Dev." and "Main+3/4Dev." where the magnitude of the variances of group-specific deviation effects is 1/4, 2/4 and 3/4 of the size of the simulated effects in the "Main+Dev." genetic configuration, respectively. For instance, in "Main+1/4Dev.", the variances of simulated allele effects are $\sigma_\gamma^2 = 2$, $\sigma_{\delta_D}^2 = \frac{1}{4} \times 1$ and $\sigma_{\delta_F}^2 = \frac{1}{4} \times 3$, instead of $\sigma_\gamma^2 = 2$, $\sigma_{\delta_D}^2 = 1$ and $\sigma_{\delta_F}^2 = 3$ in "Main+Dev."



Supplementary Figure S6. Boxplots of mean accuracies (evaluated by CV over 20 HO replicates) obtained using GBLUP-RI, GBLUP-RAE, MAGBLUP-RI and MAGBLUP-RAE for 50 traits simulated according to the genetic configuration "Main+Dev." (see Table 1) and two alternative genetic configurations: "Large_QTL" and "Mixture". "Large_QTL" is identical to "Main+Dev." except for one QTL out of 1,000 which has a large substitution effect: $\beta_{mD}^1 - \beta_{mD}^0 = 30$ and $\beta_{mF}^1 - \beta_{mF}^0 = 50$, while the substitution effect of other QTLs are 0-centered and have standard deviations equal to $\sqrt{2\sigma_\beta^2 + 2\sigma_{\delta_D}^2} = \sqrt{5}$ and $\sqrt{2\sigma_\beta^2 + 2\sigma_{\delta_F}^2} = \sqrt{10}$ for the dent and flint substitution effects, respectively. "Mixture" is a mixture distribution where 900 QTLs out of 1,000 have the same distribution as "Main+Dev." and the other 100 QTLs have an systematic effect that depends only on the allele ancestry: $\beta_{mD}^1 = \beta_{mD}^0 = 1$ and $\beta_{mF}^1 = \beta_{mF}^0 = -1$. Such effects may result from QTLs differentially fixed between groups but whose information is captured by linked SNPs.



Supplementary Figure S7. Boxplots of predictive abilities obtained by CV (HO method) for GBLUP-RI, GBLUP-RAE, MAGBLUP-RI, MAGBLUP-RAE and GBLUP-RI-epistatic for real traits: male flowering (MF), female flowering (FF), plant height (PH), ear leaf number (ELN) and total number of leaves (TNL). The GBLUP-RI-epistatic is the model described in Supplementary Table S7.