

Supplementary Tables

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 Table S1. Parameters estimated in the phenotypic analysis of real traits: male flowering (MF), female flowering (FF), plant height (PH), ear leaf number (ELN) and total number of leaves (TNL)

	MF	FF	PH	ELN	TNL
Row-Column 2015	AR1	AR1	AR1	AR1	AR1
Row-Column 2016	IID	IID	IID	IID	IID
μ_{2015}	64.22	65.86	190.72	11.23	15.95
μ_{2016}	71.20	72.40	182.62	11.19	16.09
μ_D	70.80	72.28	198.77	11.96	17.29
μ_F	64.95	66.39	172.77	10.48	14.90
μ_A	67.66	69.11	183.98	11.12	15.83
$\sigma_{G_D}^2$	24.15	27.65	555.17	1.48	2.28
$\sigma_{G_F}^2$	23.06	27.03	558.80	1.98	3.05
$\sigma_{G_A}^2$	16.89	20.07	695.10	1.26	1.87
$\sigma_{(G \times \beta)_{2015,D}}^2$	1.31	0.00	0.00	0.33	0.29
$\sigma_{(G \times \beta)_{2015,F}}^2$	1.35	1.99	35.00	0.10	0.04
$\sigma_{(G \times \beta)_{2015,A}}^2$	4.64	4.84	118.23	0.42	0.44
$\sigma_{(G \times \beta)_{2016,D}}^2$	1.09	2.67	88.71	0.00	0.00
$\sigma_{(G \times \beta)_{2016,F}}^2$	0.00	0.00	19.41	0.00	0.23
$\sigma_{(G \times \beta)_{2016,A}}^2$	2.11	3.51	83.21	0.00	0.13
$\sigma_{E_{2015}}^2$	2.46	2.56	101.04	0.59	0.63
$\sigma_{E_{2016}}^2$	1.54	1.94	63.76	0.24	0.34
h_D^2	0.96	0.96	0.93	0.89	0.92
h_F^2	0.96	0.96	0.94	0.94	0.94
h_A^2	0.88	0.88	0.91	0.86	0.88
r_{2015}	1.94	1.94	1.99	1.95	1.95
r_{2016}	1.99	1.99	2.00	2.00	2.00

The parameters correspond to the model presented in Rio et al. (2020). The lines "Row-Column" refer to the modeling of rows and columns as defined by the experimental design. "AR1" refers to the autoregressive model AR1, while "IID" refers to the modeling of row effects as being independent and identically distributed (IID) as well as column effects being IID. For more information, see the ASReml-R reference manual (Butler *et al.* 2009). The mean of each trial j (with $j \in \{2015, 2016\}$) was computed following: $\mu_j = \mu + \beta_j + \sum_{k=1}^3 \frac{N_k}{N} \alpha_k$ where N_k is the number of individuals (genotypes) in genetic background k (with $k \in \{D, A, F\}$) and N is the total number of individuals. The mean of each genetic background was computed following: $\mu_k = \mu + \alpha_k + \frac{1}{2} \sum_{j=1}^2 \beta_j$. The genetic variance $\sigma_{G_k}^2$ of each genetic background k and the GxE variance $\sigma_{(G \times \beta)_{jk}}^2$ of each genetic background k in each trial j were also reported. The heritabilities of each genetic background k were computed as: $h_k^2 = \sigma_{G_k}^2 \times \left(\sigma_{G_k}^2 + \frac{1}{4} \sum_{j=1}^2 \sigma_{(G \times \beta)_{jk}}^2 + \frac{1}{4} \sum_{j=1}^2 \frac{1}{\bar{r}_j} \sigma_{E_j}^2 \right)^{-1}$ where \bar{r}_j is the mean number of genotype replicates in trial j

Butler D.G., B.R. Cullis, A.R. Gilmour, B.J. Gogel, R. Thompson, ASReml-R Reference Manual Version 4. 2009. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Rio, S., T. Mary-Huard, L. Moreau, C. Bauland, C. Palaffre, et al., 2020 Disentangling group specific QTL allele effects from genetic background epistasis using admixed individuals in GWAS: 16 An application to maize flowering. PLOS Genetics 16: 1–27.

Supplementary Table S2. Mean accuracies over 50 traits simulated according to genetic configuration "Main" (see Table 1) and 20 CV replicates (SHO method), obtained using GBLUP-RI, MAGBLUP-RI and MAGBLUP-RAE

	GBLUP-RI	MAGBLUP-RI	MAGBLUP-RAE
DFA_DFA	0.49 (0.07)	0.47 (0.07)	0.49 (0.07)
A_D	0.51 (0.06)	0.49 (0.06)	0.51 (0.06)
DFA_D	0.48 (0.06)	0.46 (0.07)	0.48 (0.06)
DF_D	0.46 (0.07)	0.44 (0.08)	0.45 (0.07)
D_D	0.51 (0.07)	-	-
F_D	0.23 (0.09)	-	-
A_F	0.49 (0.09)	0.46 (0.10)	0.48 (0.09)
DFA_F	0.48 (0.09)	0.45 (0.10)	0.48 (0.09)
DF_F	0.46 (0.09)	0.44 (0.10)	0.46 (0.09)
D_F	0.20 (0.09)	-	-
F_F	0.51 (0.09)	-	-
A_A	0.52 (0.05)	0.49 (0.06)	0.52 (0.06)
DFA_A	0.51 (0.05)	0.48 (0.05)	0.50 (0.05)
DF_A	0.50 (0.05)	0.48 (0.05)	0.50 (0.05)
D_A	0.46 (0.05)	-	-
F_A	0.50 (0.05)	-	-

Standard deviations over the 50 mean accuracies (computed over 20 CV replicates) are shown between brackets

"-" indicates that a model could not be applied for the given configuration

Supplementary Table S3. Mean accuracies over 50 traits simulated according to genetic configuration "Dev." (see Table 1) and 20 CV replicates (SHO method), obtained using GBLUP-RI, MAGBLUP-RI and MAGBLUP-RAE

	GBLUP-RI	MAGBLUP-RI	MAGBLUP-RAE
DFA_DFA	0.57 (0.13)	0.64 (0.14)	0.64 (0.14)
A_D	0.29 (0.10)	0.33 (0.10)	0.33 (0.10)
DFA_D	0.28 (0.10)	0.36 (0.09)	0.36 (0.09)
DF_D	0.31 (0.11)	0.39 (0.09)	0.39 (0.09)
D_D	0.46 (0.08)	-	-
F_D	-0.01 (0.12)	-	-
A_F	0.41 (0.07)	0.47 (0.07)	0.47 (0.07)
DFA_F	0.36 (0.10)	0.44 (0.08)	0.44 (0.08)
DF_F	0.36 (0.11)	0.43 (0.09)	0.43 (0.09)
D_F	0.00 (0.10)	-	-
F_F	0.51 (0.08)	-	-
A_A	0.50 (0.06)	0.66 (0.05)	0.67 (0.05)
DFA_A	0.42 (0.05)	0.59 (0.06)	0.60 (0.06)
DF_A	0.31 (0.06)	0.34 (0.07)	0.34 (0.06)
D_A	0.17 (0.06)	-	-
F_A	0.32 (0.05)	-	-

Standard deviations over the 50 mean accuracies (computed over 20 CV replicates) are shown between brackets

"-" indicates that a model could not be applied for the given configuration

Supplementary Table S4. Mean predictive abilities over 100 CV replicates (SHO method) for female flowering (FF) using GBLUP-RI, MAGBLUP-RI and MAGBLUP-RAE

	GBLUP-RI	MAGBLUP-RI	MAGBLUP-RAE
DFA_DFA	0.75 (0.04)	0.73 (0.04)	0.75 (0.04)
A_D	0.61 (0.05)	0.60 (0.05)	0.61 (0.05)
DFA_D	0.65 (0.05)	0.64 (0.05)	0.65 (0.05)
DF_D	0.68 (0.05)	0.67 (0.05)	0.68 (0.05)
D_D	0.72 (0.04)	-	-
F_D	0.43 (0.11)	-	-
A_F	0.68 (0.05)	0.64 (0.05)	0.68 (0.05)
DFA_F	0.70 (0.05)	0.67 (0.05)	0.69 (0.05)
DF_F	0.69 (0.05)	0.67 (0.05)	0.68 (0.05)
D_F	0.58 (0.08)	-	-
F_F	0.68 (0.05)	-	-
A_A	0.56 (0.07)	0.55 (0.07)	0.56 (0.07)
DFA_A	0.57 (0.08)	0.55 (0.08)	0.57 (0.07)
DF_A	0.58 (0.07)	0.56 (0.07)	0.58 (0.07)
D_A	0.54 (0.07)	-	-
F_A	0.55 (0.08)	-	-

Standard deviations over the predictive abilities of the 100 CV replicates are shown between brackets
“-” indicates that a model could not be applied for the given configuration

Supplementary Table S5. Mean predictive abilities over 100 CV replicates (SHO method) for ear leaf number (ELN) using GBLUP, MAGBLUP-RI and MAGBLUP-RAE

	GBLUP-RI	MAGBLUP-RI	MAGBLUP-RAE
DFA_DFA	0.70 (0.05)	0.69 (0.05)	0.70 (0.05)
A_D	0.44 (0.07)	0.42 (0.085)	0.44 (0.08)
DFA_D	0.49 (0.06)	0.45 (0.06)	0.49 (0.06)
DF_D	0.52 (0.06)	0.48 (0.06)	0.52 (0.06)
D_D	0.54 (0.05)	-	-
F_D	0.31 (0.11)	-	-
A_F	0.61 (0.06)	0.59 (0.07)	0.61 (0.06)
DFA_F	0.66 (0.05)	0.65 (0.06)	0.66 (0.05)
DF_F	0.67 (0.05)	0.65 (0.05)	0.67 (0.05)
D_F	0.58 (0.07)	-	-
F_F	0.67 (0.05)	-	-
A_A	0.48 (0.07)	0.45 (0.08)	0.47 (0.07)
DFA_A	0.48 (0.08)	0.45 (0.08)	0.47 (0.08)
DF_A	0.48 (0.07)	0.46 (0.08)	0.48 (0.07)
D_A	0.46 (0.07)	-	-
F_A	0.44 (0.08)	-	-

Standard deviations over the predictive abilities of the 100 CV replicates are shown between brackets
“-” indicates that a model could not be applied for the given configuration

Supplementary Table S6. Mean predictive abilities over 100 CV replicates (SHO method) for total number of leaves (TNL) using GBLUP, MAGBLUP-RI and MAGBLUP-RAE

	GBLUP-RI	MAGBLUP-RI	MAGBLUP-RAE
DFA_DFA	0.76 (0.04)	0.75 (0.04)	0.76 (0.04)
DFA_D	0.56 (0.06)	0.54 (0.06)	0.56 (0.06)
D_D	0.59 (0.05)	-	-
DF_D	0.57 (0.06)	0.56 (0.06)	0.58 (0.06)
F_D	0.39 (0.10)	-	-
A_D	0.53 (0.07)	0.51 (0.07)	0.54 (0.07)
DFA_	0.67 (0.05)	0.65 (0.05)	0.67 (0.05)
F_F	0.67 (0.05)	-	-
DF_F	0.68 (0.05)	0.66 (0.05)	0.67 (0.05)
D_F	0.56 (0.07)	-	-
A_F	0.64 (0.05)	0.62 (0.05)	0.64 (0.05)
DFA_A	0.52 (0.07)	0.51 (0.07)	0.52 (0.07)
A_A	0.53 (0.07)	0.51 (0.07)	0.52 (0.07)
DF_A	0.53 (0.08)	0.51 (0.08)	0.52 (0.08)
D_A	0.49 (0.07)	-	-
F_A	0.45 (0.07)	-	-

Standard deviations over the predictive abilities of the 100 CV replicates are shown between brackets
“-” indicates that a model could not be applied for the given configuration

Supplementary Table S7. Additive, epistatic and residual variance components of 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), with the p-value (pval) of the epistatic component using a likelihood-ratio LR test

	MF	FF	PH	ELN	TNL
σ_G^2	12.50	15.55	510.68	1.09	1.60
$\sigma_{(G \times G)}^2$	2.91	4.66	164.35	0.25	0.24
σ_E^2	2.13	0.92	45.68	0.19	0.34
pval	0.0075457	0.0000680	0.0278121	0.1650159	0.4571475

The existence of epistasis can be investigated using a test based on variance components. The epistatic variance component between pairs of loci was estimated on the joint dent, flint and admixed dataset using a model neglecting genetic structure called GBLUP-RI-epistatic: $\mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{g}_e + \mathbf{e}$, where \mathbf{y} is the vector of phenotypes, $\mathbf{1}$ is a vector of 1, μ is the global intercept, \mathbf{g} is the vector of additive genetic values with $\mathbf{g} \sim \mathcal{N}(0, \mathbf{K}\sigma_G^2)$, \mathbf{K} is the kinship matrix computed following Eq (3), σ_G^2 is the additive genetic variance, \mathbf{g}_e is the vector of epistatic deviations with $\mathbf{g}_e \sim \mathcal{N}(0, \mathbf{K} \circ \mathbf{K}\sigma_{(G \times G)}^2)$, $\sigma_{(G \times G)}^2$ is the epistatic genetic variance between pairs of loci, \mathbf{e} is the vector of errors with $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$, \mathbf{I} is the identity matrix, σ_E^2 is the error variance. Note that $\mathbf{K} \circ \mathbf{K}$ is the Hadamard product of the kinship matrix with itself. GBLUP-RI-epistatic can be seen as a simplified version of the one proposed by Vitezica *et al.* (2017), as purely homozygous lines were used. The epistatic variance component was tested using a LR test between this model and the same model without the term \mathbf{g}_e .

Vitezica, Z. G., A. Legarra, M. A. Toro, and L. Varona, 2017 Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics* 206: 1297–1307.

Supplementary Table S8. Variance components of real traits : male flowering (MF), female flowering (FF), plant height (PH), ear leaf number (ELN) and total number of leaves (TNL), estimated by GBLUP, MAGBLUP-RI and MAGBLUP-RAE using the 300 dent and 304 flint lines (and no admixed lines)

Model	Variance	MF	FF	PH	ELN	TNL
GBLUP-RI	σ_G^2	20.90 (2.33)	21.83 (2.84)	641.07 (85.59)	1.64 (0.30)	2.75 (0.25)
	σ_E^2	0.34 (0.39)	1.12 (0.65)	63.52 (25.27)	0.09 (0.05)	0.02 (0.05)
MAGBLUP-RI	$\sigma_{G_D}^2$	17.86 (2.08)	21.61 (2.33)	576.96 (75.08)	1.44 (0.17)	1.84 (0.24)
	$\sigma_{G_F}^2$	13.81 (1.69)	18.10 (1.99)	486.98 (68.49)	1.36 (0.17)	2.02 (0.25)
	σ_E^2	2.05 (0.71)	1.48 (0.74)	94.23 (30.67)	0.18 (0.07)	0.32 (0.10)
GBLUP-RAE	σ_G^2	48.64 (2.14)	60.14 (4.63)	1730.04 (75.08)	4.35 (0.40)	6.21 (0.66)
	σ_E^2	2.14 (0.69)	1.68 (0.74)	87.00 (28.90)	0.16 (0.06)	0.28 (0.09)
MAGBLUP-RAE	σ_U^2	40.14 (10.95)	53.11 (12.05)	1528.25 (296.71)	4.19 (0.49)	5.53 (0.74)
	$\sigma_{U_D}^2$	13.42 (10.93)	12.32 (11.73)	311.04 (261.00)	0.27 (0.35)	0.25 (0.56)
	$\sigma_{U_F}^2$	1.37 (10.98)	1.72 (11.96)	0.04 (265.11)	0.00 (0.37)	0.44 (0.61)
	σ_E^2	2.15 (0.71)	1.51 (0.73)	88.24 (29.11)	0.16 (0.06)	0.31 (0.09)

Standard errors are shown between brackets