File S3: MAGBLUP-RI and -RAE variance estimates

MAGBLUP-RI and MAGBLUP-RAE were both evaluated for their precision in variance component estimation using simulated genotypes and traits.

Simulated genotypes

In our simulations, genotypes were generated based on the genotypic data of the Flint-Dent dataset. The dent lines were simulated by randomly sampling each chromosome from all existing versions of the given chromosome within the real dent lines. The flint lines were simulated in an equivalent manner. The admixed lines were simulated by generating gametes from hybrids produced at random between simulated dent and flint lines. The SNPs were located on a genetic map for which the genetic distance between pairs of markers was calculated as being proportional to their physical distance, and the scale parameter was determined relative to chromosome 1 (200 cM for around 300 Mbp). For each chromosome, the recombination breakpoints were sampled in a Poisson distribution with parameter λ equal to the length of the chromosome in Morgan. As with the real dataset, the datasets simulated according to this procedure included 970 individuals (300 dent, 304 flint and 366 admixed lines).

Simulated traits

Phenotypes were simulated according to the procedure described in the Material and methods section of the manuscript (i.e. 3 genetic configurations, 1,000 QTLs and a heritability of 0.8).

Assessment of the precision of variance estimates

The precision of the genetic variance component estimates of MAGBLUP-RI and MAGBLUP-RAE was evaluated based on simulated phenotypes and genotypes, using all 482,013 SNPs including the 1,000 QTLs. The covariances matrices were computed using the genotypic and allele ancestry information of the simulated datasets.

The variance components of MAGBLUP-RI are defined conditionally to allele effects. For a given genetic configuration, the precision of variance estimates was evaluated using one simulated trait (i.e. one sample of allele effects), and 100 simulated population samples. The variance estimates were compared to the three reference variances: σ^2_S, $\sigma^2_{G_D}$ and $\sigma^2_{G_F}$. Each reference variance was computed using the simulated allele effects and the reference allele frequencies that were estimated on the real dataset (e.g. $\sigma^2_{G_D} = \sum_{m=1}^{M} \widehat{f}_{mD}(1-\widehat{f}_{mD})(\beta^1_{mD} - \beta^0_{mD})^2)$. For the segregation variance, the approximation $\sigma^2_S \approx \sum_{m=1}^{M} (\mu_{mD} - \mu_{mF})^2$ was used as a reference, as the number of loci was large and the deviations effects were 0-centered (see Supplementary File S1). Note that the difference between the exact and the approximate reference segregation variance was negligible in all genetic configurations.

The variance components of MAGBLUP-RAE are defined conditionally to genotypes and allele ancestries. For a given genetic configuration, the precision of variance estimates was evaluated using one simulated population sample and 100 simulated traits (i.e. 100 samples of allele effects). The estimates were compared to the three reference variances: σ^2_U, $\sigma^2_{U_D}$ and $\sigma^2_{U_F}$. Each variance was computed using the variances of allele effects and the number of QTLs (e.g. $\sigma^2_{U_D} = M\sigma^2_{\delta_D}$).

For each model, the procedure was replicated 100 times (i.e. over 100 simulated traits for MAGBLUP-RI and over 100 simulated population samples for MAGBLUP-RAE), in order to evaluate the mean bias, standard deviation (SD) and root mean square error (RMSE) of the estimates. For a given variance component σ^2, the mean bias is computed as: $\frac{1}{100}\sum_{i=1}^{100} B_i$ with $B_i = \frac{1}{100}\sum_{j=1}^{100}\left(\widehat{\sigma^2_{ij}} - \sigma^2_i\right)$ where i is the i^{th} simulated trait and j is the j^{th} population sample for

MAGBLUP-RI (and conversely for MAGBLUP-RAE), the mean SD is computed as: $= \frac{1}{100}\sum_{i=1}^{100} S_i$ with $S_i = \sqrt{\frac{1}{100}\sum_{j=1}^{100}\left(\widehat{\sigma^2_{ij}} - \sigma^2_i\right)^2}$ and the mean RMSE is computed as: $= \frac{1}{100}\sum_{i=1}^{100}\sqrt{B_i^2 + S_i^2}$.

MAGBLUP-RI variance estimates

As the variance components of MAGBLUP-RI are defined conditionally to allele effects, the precision of their estimation was evaluated by simulating a single trait (single sample of QTL allele effects) for each genetic configuration and by estimating variance components over 100 population samples. For a given simulated trait, the variances were generally well estimated in all genetic configurations (Table A). This procedure was replicated over 100 simulated traits to study the mean bias, SD and RMSE, also presented in Table A. Very limited biases were obtained for all components, but the segregation variance σ^2_S tended to be estimated less accurately than the group-specific genetic variances. Note that the segregation variance σ^2_S is null for none of the three genetic configurations as both group-specific allele effects and group-specific allele frequencies at QTLs lead to a non-zero reference value.

Table A. Mean and SD of variance components (Var.) estimated by MAGBLUP-RI for a given simulated trait (Simulated trait 1) over 100 simulated populations samples according to each type of genetic configuration (G. conf.), as well as mean bias, SD and RMSE of estimates over 100 replicates (Simulated traits 1 to 100)

		Simulated trait 1			**Simulated traits 1 to 100**		
G. conf.	**Var.**	**Reference**	**Mean**	**SD**	**Mean bias**	**Mean SD**	**Mean RMSE**
Main	σ^2_S	140.81	110.89	81.11	-1.51	87.05	96.44
Main	$\sigma^2_{G_D}$	644.88	547.73	57.59	-0.53	68.27	76.95
Main	$\sigma^2_{G_F}$	653.76	610.75	63.01	-1.87	70.63	76.83
Dev.	σ^2_S	2761.18	2642.91	464.96	-40.08	424.98	600.37
Dev.	$\sigma^2_{G_D}$	310.47	274.83	64.09	-0.49	59.55	63.67
Dev.	$\sigma^2_{G_F}$	1079.20	1068.47	126.37	-2.62	106.04	113.40
Main+Dev.	σ^2_S	3017.74	3026.54	580.02	-6.60	564.53	743.03
Main+Dev.	$\sigma^2_{G_D}$	909.51	959.02	125.60	3.53	139.68	154.09
Main+Dev.	$\sigma^2_{G_F}$	1510.76	1517.89	142.18	-0.37	184.99	202.12

MAGBLUP-RAE variance estimates

As the variance components of MAGBLUP-RAE are defined conditionally to genotypes and alleles ancestries, the precision of their estimation was evaluated by simulating 100 traits (100 samples of QTL allele effects) for each genetic configuration and by estimating variance components using a single population sample. For a given simulated population sample, the variances were generally well estimated in all genetic configurations (Table B). This procedure was replicated over 100 simulated population samples to study the mean bias, SD and RMSE, as presented in Table B. Limited positive biases were observed for variance components expected to be equal to zero along with little negative biaises for variance components expected to be non null, suggesting a minor trade-off between variance component estimates.

Table B. Mean and standard deviation (SD) of variance components (Var.) estimated by MAGBLUP-RAE for 100 simulated traits on a given simulated population sample (Simulated pop 1) according to each type of genetic configuration (G. conf.), as well as mean bias, SD and RMSE of estimates over 100 replicates (Simulated pop 1 to 100)

		Simulated population 1			**Simulated populations 1 to 100**		
G. conf.	**Var.**	**Reference**	**Mean**	**SD**	**Mean bias**	**Mean SD**	**Mean RMSE**
Main	σ^2_U	2000	1981.73	203.44	-45.11	211.03	216.37
Main	$\sigma^2_{U_D}$	0	28.55	56.30	26.94	58.11	64.12
Main	$\sigma^2_{U_F}$	0	32.70	51.74	27.64	58.60	65.87
Dev.	σ^2_U	0	151.40	199.90	166.23	221.11	276.87
Dev.	$\sigma^2_{U_D}$	1000	880.99	282.17	-148.94	285.96	323.22
Dev.	$\sigma^2_{U_F}$	3000	2842.24	335.35	-151.97	378.77	408.93
Main+Dev.	σ^2_U	2000	1960.12	629.25	-13.20	676.78	678.97
Main+Dev.	$\sigma^2_{U_D}$	1000	1069.56	653.31	44.69	613.64	617.27
Main+Dev.	$\sigma^2_{U_F}$	3000	3023.27	568.65	47.78	651.71	655.36