Evaluating potential bias of REML estimators of heritability

Beatriz C. D. Cuyabano, Anders Christian Sørensen, Peter Sorensen

Evaluating potential bias of REML estimators of heritability

B.C.D. Cuyabano[1],A.C. Sørensen[1] & P. Sørensen[1]

[1] Aarhus University, Center for Quantitative Genetics and Genomics, Dept. Of Molecular Biology and Genetics, Blichers Allé 20, Postboks 50, 8830 Tjele, Denmark
beatriz.cuyabano@mbg.au.dk (Corresponding Author)

Summary

Genomic models that incorporate dense SNP genotypes are increasingly being used and studied for inference of variance parameters and narrow-sense heritability. The variance parameters of a linear mixed model linking a phenotype to SNP genotypes can be inferred using restricted maximum likelihood, which produces consistent, asymptotically normal estimates of variance components, when the SNP genotypes are those of the causal loci. Such properties are not guaranteed to hold when the covariance structure of the data specified by the genomic and the true models differs substantially. Since in practice we do not have knowledge of the true genetic relationship matrix among individuals, genomic models that incorporate SNP genotypes are used instead to compute a genomic relationship matrix. The patterns of realized relationships at different sets of loci (e.g., markers and causal loci) vary across the genome, and therefore a genomic relationship matrix may provide a poor description of true genetic relationships at causal loci, potentially leading to incorrect inferences. This work offers a theoretical analysis based on splitting the likelihood equations into components, isolating those that contribute to incorrect inferences, and providing an informative measure to compare the covariance structure of the data specified by the genomic and the true models. The theory presented is also used to evaluate and explain the success of a number of recently reported approaches in removing sources of bias of heritability estimates.

Keywords: heritability, genomic models, REML, likelihood misspecification

Introduction

Genomic models that incorporate dense marker information are increasingly being used and studied for inference of variance parameters and narrow-sense heritability (Yang et al., 2010; Golan and Rosset, 2011; Speed et al., 2012). We define a genomic model as any linear mixed model (LMM) linking a phenotype to SNP genotypes without knowledge of the causal quantitative trait loci (QTL) associated with the phenotype. The variance parameters of the LMM can be inferred using restricted maximum likelihood (REML) (Patterson and Thompson, 1971), which produces consistent, asymptotically normal estimates of variance components. These asymptotic properties of REML estimators are not guaranteed to hold when the likelihood of the genomic model used for inference differs from that of the true model that generated the data. In such a situation, the likelihood is misspecified. In a Gaussian setup this will be the case when the covariance structures of the data specified by the genomic and the true models differ.

The correct covariance structure requires knowledge of the true genetic relationship matrix ($\mathbf{G}_0$) among individuals at causal loci. Since these are typically unknown, in practice the genomic model makes use of marker genotypes instead in order to compute a genomic

relationship matrix ($\mathbf{G}$), which may provide a poor description of realized relationships at causal loci and this can lead to misspecification of the likelihood. In the setting that we explore in this work, the likelihood misspecification of a genomic model is exclusively due to the use of $\mathbf{G}$ instead of $\mathbf{G}_Q$.

REML was first implemented for populations of nominally unrelated individuals with a genomic model in Yang *et al.* (2010), where the focus of inference was the proportion of the variance of the quantitative trait (in that case, height in humans) explained by the LMM including all genotyped SNPs simultaneously. In recent years, concerns have been raised about the quality of inferences of variance parameters when genomic models have been used, without directly addressing the problem of likelihood misspecification.

The problem of misspecification of the likelihood of the genomic model was first raised by de los Campos *et al.*, 2013 and was studied using simulation by de los Campos *et al.*, 2015. In the latter study, a distinction was made between two genetic parameters: the additive genetic variance and the genomic variance or amount of additive variance that can be captured by regression on SNPs, not necessarily including the QTL.

In this work we look into the problem of misspecification of the likelihood to evaluate the bias of REML estimators of heritability. The objective was to provide a theoretical analysis based on the splitting of the likelihood equations into components, isolating those that contribute to incorrect inferences. The theory presented is also used to evaluate and explain the success of a number of recently reported approaches in removing sources of bias of heritability estimates.

Materials and methods

Consider that we have genomic data containing s SNP genotypes, which may or may not contain the QTL. SNPs that are not QTL are referred to as markers. We assume the following additive mixed model to relate SNPs to phenotypes:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{W}\mathbf{b} + \mathbf{e}, \qquad (1)$$

where μ is the overall mean, $\mathbf{W}$ is the $n \times s$ standardized SNP genotypes matrix (with $W_{ij} = (Z_{ij} - 2\theta_j)/\sqrt{2\theta_j(1-\theta_j)}$: $E(W_{ij}) = 0$, and $Var(W_{ij}) = 1$; $Z_{ij} \in \{0,1,2\}$ is the count of the minor allele at the *j-th* SNP with minor allele frequency (MAF) θ_j, of the *i-th* individual, for all *i=1,...,n* and *j=1,...,s*), $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}_s\sigma_b^2)$ is a $s \times 1$ vector of random SNP effects and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n\sigma_e^2)$ is a $n \times 1$ vector of the model's residuals. If $\mathbf{W}$ consists of QTL only ($\mathbf{W}_Q$), then equation (1) describes the true model that generates the phenotypes $\mathbf{y}$.

The covariance structure of the phenotype data that is specified by the genomic model ($Var(\mathbf{y} \mid \mathbf{W}) = s\sigma_b^2\mathbf{G} + \sigma_e^2\mathbf{I}_n = \sigma_e^2\mathbf{V}_\gamma$: $\mathbf{G} = s^{-1}\mathbf{W}\mathbf{W}'$, $\mathbf{V}_\gamma = \gamma\mathbf{G} + \mathbf{I}_n$, and $\gamma = s\sigma_b^2\sigma_e^{-2}$) differs from that specified by the true model ($Var(\mathbf{y} \mid \mathbf{W}_Q) = q\sigma_{bQ}^2\mathbf{G}_Q + \sigma_{eQ}^2\mathbf{I}_n = \sigma_{eQ}^2\mathbf{V}_{Q\gamma}$: $\mathbf{G}_Q = q^{-1}\mathbf{W}_Q\mathbf{W}_Q'$, q is the number of QTL, $\mathbf{V}_{Q\gamma} = \gamma_Q\mathbf{G}_Q + \mathbf{I}_n$, and $\gamma_Q = q\sigma_{bQ}^2\sigma_{eQ}^{-2}$), leading to a likelihood misspecification.

To evaluate the potential asymptotic bias of the REML estimators of the variance components under the genomic models, we evaluated the REML equation, comparing its behaviour to that of the REML equation under the true model, which is known to yield asymptotically unbiased estimators of the variance components, even if normality does not

hold and the number of QTL increases dramatically tending to infinity (Jiang, 1996). Differentiating the REML log-likelihood function (Harville, 1977) with respect to γ, we obtain the REML equation (Jiang, 2007):

$$\mathbf{y}'\left[\frac{\mathbf{P}_\gamma\mathbf{G}\mathbf{P}_\gamma}{tr(\mathbf{P}_\gamma\mathbf{G})}-\frac{\mathbf{P}_\gamma^2}{tr(\mathbf{P}_\gamma)}\right]\mathbf{y}=0 \quad , \tag{2}$$

where $\mathbf{P}_\gamma=\mathbf{V}_\gamma^{-1}-\mathbf{V}_\gamma^{-1}\mathbf{1}_n(\mathbf{1}_n'\mathbf{V}_\gamma^{-1}\mathbf{1}_n)^{-1}\mathbf{1}_n'\mathbf{V}_\gamma^{-1}$, and under the true model (QTL only), $\mathbf{P}_\gamma$ and $\mathbf{G}$ are sub-indexed with Q. $\hat{\gamma}$ is the solution of equation (2), and can be directly used to obtain the REML estimator of heritability, $\hat{h}^2=\hat{\gamma}/(1+\hat{\gamma})$, due to invariance property of maximum likelihood estimators.

Using the eigen-decomposition $\mathbf{G}=\mathbf{U}\Lambda\mathbf{U}'$, the REML equation (2) can be written as the non-observable REML equation, defined as:

$$\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\frac{\kappa_i}{(1+\gamma\lambda_i)^2}\left(\frac{\lambda_i-\lambda_j}{1+\gamma\lambda_j}\right)=0 \quad , \tag{3}$$

where $\kappa_i=\sum_{k=1}^{n}(\mathbf{U}_i'\mathbf{U}_{Qk})^2\lambda_{Qk}$ measures the correlations between eigen-vectors of the genomic and true models, weighted by the eigen-values of the true model (remembering that each eigen-value represents the amount of variance due to its respective component). In fact, κ_i is a measure that evaluates whether $\mathbf{G}$ is providing a poor description of $\mathbf{G}_Q$. We refer to equation (3) as non-observable because it is written as a function of $\mathbf{U}_Q$ and Λ_Q, which cannot be observed directly when only phenotype and genomic data are available, and we have no knowledge about the QTL.

Results and discussion

The term $\kappa_i=\sum_{k=1}^{n}(\mathbf{U}_i'\mathbf{U}_{Qk})^2\lambda_{Qk}$ in equation (3) is the key for the evaluation of the potential bias in REML estimators of heritability. Note that under the true model, $\kappa_{Qi}=\sum_{k=1}^{n}(\mathbf{U}_{Qi}'\mathbf{U}_{Qk})^2\lambda_{Qk}=\lambda_{Qi}$, and because the true models is known to yield asymptotically unbiased estimators of the variance components, the relationship between κ_i and λ_i is what determines (asymptotically) the presence and direction of bias of the REML estimators of heritability ($\hat{h}^2$) to the true heritability (h^2), as we present next:

1. $\kappa_i=\lambda_i$: $\hat{h}^2$ is asymptotically unbiased to h^2
2. $\kappa_i=c$, constant: $\hat{h}^2=0$, meaning that h^2 cannot be estimated by the model
3. $\kappa_i<\lambda_i$: $\hat{h}^2$ is asymptotically biased downwards to h^2
4. $\kappa_i>\lambda_i$: $\hat{h}^2$ is asymptotically biased upwards to h^2

We evaluated theoretically the relationship between κ_i and λ_i for eight scenarios that

are relevant in quantitative genetics studies. These scenarios differed in their population structure (unrelated or related individuals) and genetic architecture (regarding MAF and LD between QTL and markers). Table 1 presents the description of the scenarios, as well as the theoretical relationship between κ_i and λ_i in each scenario, for genomic models that comprised either the QTL and markers in the SNP genotypes, or the markers only. The relationships between κ_i and λ_i were obtained by equating $\kappa_i = \sum_{i=1}^{n} (\mathbf{U}_i'\mathbf{U}_{Qi})^2 \lambda_{Qi}$ under the assumptions of each scenario. Table 2 presents the description of the scenarios, as well as the asymptotic expectation of $\hat{h}^2$, denoted as $E_\infty(\hat{h}^2) = \lim_{n\to\infty} E(\hat{h}^2)$.

Using simulations, we studied the sampling properties of $\hat{h}^2$ for all the evaluated scenarios, and the results were used to support the theoretical expectations. Each scenario was replicated 1,000 times, in a population of 2,000 individuals (population size was considered large enough to guarantee the asymptotic properties, based on a preliminary study), with 20,000 SNPs, of which 3,000 were assigned as QTL. Figures 1-8 show that results obtained with simulations agree with the theory derived for all the scenarios. When the QTL are in complete LE with the markers, the addition of markers to the QTL genotypes does not induce bias to $\hat{h}^2$. When the QTL are in LD with the markers, the addition of markers to the QTL genotypes may induce bias to $\hat{h}^2$, if the structure of $\mathbf{G}$ differs substantially from the structure of $\mathbf{G}_Q$, as observed in scenarios in which the MAF distribution of the QTL and markers differ. The presence of related individuals in the population can alleviate the structural difference between $\mathbf{G}$ and $\mathbf{G}_Q$, attenuating the bias of $\hat{h}^2$.

The scenarios in which QTL and markers are in complete LE have been explored theoretical and empirically in other studies, with particular emphasis on the effect of the eigen-values of $\mathbf{G}$ on the likelihood of the (misspecified) genomic model (Jiang *et al*, 2014: Kumar *et al.*, 2015). Although Kumar *et al.* (2015) discussed the relevance of the difference between the eigen-vectors of $\mathbf{G}$ and $\mathbf{G}_Q$, they did not discuss this difference as correlations between the eigen-vectors, which are implied in the term κ_i that we evaluate. Moreover, the authors asses the performance of genomic models in estimating heritability mainly by describing the sensitivity of the likelihood to changes in the eigen-values. Indeed, the likelihood depends sensitively on all the eigen-values, but evaluation of the likelihood given a change in each eigen-value separately is not as informative as evaluation of the REML equation given a change in the distribution of all eigen-values simultaneously. Jiang et al. (2014) assesses the performance of genomic models in estimating heritability by studying the limiting behavior of the central term in equation (2), assuming the infinitesimal model. Although the analysis of scenarios in which QTL are in complete LE with the markers is very important to help us understand the mechanism of REML equations, we believe that scenarios in which QTL are in LD with the markers are more realistic. Moreover, in the presence of LD, the distribution of the MAF of QTL and markers may alter the correlations between phenotypes and genotypes, that are implied in equation (2), and can be measured by κ_i.

REML estimators of heritability using the method proposed by Yang *et al.* (2010), for scenarios in which QTL are in LD with the markers, can be biased to the true parameters, depending on the genetic architecture of the trait. This is the case when the distributions of MAF differ for QTL and markers, in populations with mostly unrelated individuals. Several approaches have been proposed, addressing the problem of biased heritability estimators, and we relate our theoretical evaluation to their results.

First, addressing the different MAF of the SNPs, (Speed *et al.*, 2012) suggests a

weighting of the SNPs by their MAF, which would give the same weighting to terms involving γ in equation (3), owing to the change in the definition of the heritability. $\mathbf{G}$ obtained using the SNPs suitably weighted according to the scenario will improve the relationship between κ_i and λ_i reducing the bias of $\hat{h}^2$. The definition of a suitable weight must be explored further, and the theory provided in this study provides a tool that can be used for such investigations.

Yang et al. (2015) suggested a method analogous to that proposed in Yang et al. (2010), by fitting the model with several genomic variance components, each of them relative to groups of SNPs with MAF within the same range. Non-observable REML equations can be obtained for each genomic variance component, and their analysis is analogous to that presented for one single component. Indeed the method is capable of removing the bias of $\hat{h}^2$. However, as observed by Yang *et al.* (2015), the increase in the number of variance components will increase the variance of $\hat{h}^2$, and, depending on the scenario evaluated, the estimates may be less reliable than those obtained by fitting a single genomic variance component. Edwards *et al.* (2016) also suggests the fitting of the model with several genomic variance components, however, grouping them to genomic feature (*i.e.* genes and their gene ontology), which requires the use of prior information about the genomic data. The results of their study showed that a relevant amount of variance was attributed to the significant features. The major advantage of grouping SNPs to genomic features instead of MAF, for the estimation of variance components, is that the prior genomic information used for grouping may lower the number of components, reducing the variance of $\hat{h}^2$. The use of prior genomic information to fit genomic models with multiple variance components was previously suggested by Speed and Balding (2014), who included a dynamic procedure to define a suitable partition of SNPs.

Considering the situation where prior genomic feature information is absent, Bayesian mixture models, such as BayesB (Meuwissen *et al.*, 2009) or BayesR (Erbe *et al.*, 2012), are reasonable solutions for assigning different distributions to groups of SNP effects (Edwards *et al.*, 2016). Again, non-observable REML equations can be used to evaluate $\hat{h}^2$, and the Bayesian models can be tuned using the information from our theoretical analysis.

Last but not least, a fifth approach includes related individuals to study populations, which can greatly reduce the bias of $\hat{h}^2$, when is exists. This is because the rare QTL induce genetic relationships between individuals. Whereas in populations of nominally unrelated individuals common markers mask those induced genetic relationships ($\mathbf{G} \to \mathbf{A} = \mathbf{I}_n \neq \mathbf{G}_Q$), drastically reducing the correlations between eigen-vectors of $\mathbf{G}$ and $\mathbf{G}_Q$, resulting in $\kappa_i < \lambda_i$, in populations of related individuals, the induced genetic relationships will better reflect the kinship matrix ($\mathbf{G} \approx \mathbf{A}$), improving the correlation between eigen-vectors of $\mathbf{G}$ and $\mathbf{G}_Q$, resulting in $\kappa_i \approx \lambda_i$ and less biased $\hat{h}^2$.

Conclusions

In a Gaussian setup, the likelihood of a genomic model is misspecified with respect to that of the true model that conceptually generated the data, due to the difference between the covariance structures of the data specified by these models. When used for inference of variance parameters and narrow-sense heritability, the misspecified likelihood may yield biased estimators of those parameters, and inferences must be interpreted with caution. Our study has shown that the bias of REML estimators of heritability is linked to the relationship

between the eigen-values and eigen-vectors of $\mathbf{G}$ and $\mathbf{G}_Q$, implied in the measure κ_i, and bias occurs when $\kappa_i \neq \lambda_i$. Moreover, the comparison between κ_i and λ_i not only identifies the potential bias of $\hat{h}^2$, but is also a very informative method for comparing $\mathbf{G}$ and $\mathbf{G}_Q$, which can be extended to a number of different approaches of obtaining $\hat{h}^2$ using REML on genomic models.

List of References

Campos, G. de los and D. Sorensen, 2013 Comments on pitfalls of predicting complex traits from SNP's. Nature Rev. Genet. 894 doi: 10.1038/nrg3457-c1

Campos, G. de los, D. Sorensen and D. Gianola, 2015 Genomic heritability: What is it? PLoS Genet. 11(5):e1005048

Edwards, S. M., I. F. Sørensen, P. Sarup, T. F. C. Mackay and P. Sørensen, 2016 Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *drosophila melanogaster*. Genetics 203(4) 1871-83

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman *et al.*, 2012 Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-29

Golan, D. and S. Rosset, 2011 Accurate estimation of heritability in genome wide studies using random effects models. Bioinformatics 27(13):i317-23

Harville, D. A., 1977 Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72(358):320-38

Jiang, J., 1996. REML estimation: asymptotic behavior and related topics. Annals Stat. 24(1):255-86

Jiang, J., 2007. Linear and generalized linear mixed models and their applications. Springer, New York

Jiang, J., C. Li, D. Paul, C. Yang and H. Zhao, 2014 High-dimensional genome-wide association study and misspecified mixed model analysis. ARXIV. eprint arXiv:1404.2355

Kumar, S. K., M. W. Feldman, D. H. Rehkopf and S. Tuljapurkar, 2015 Limitations of GCTA as a solution to the missing heritability problem. Proc. Natl. Acad. Sci. USA 113:E61–70

Meuwissen, T. H. E., T. R. Solberg, R. Shepherd and J. A. Woolliams, 2009 A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet. Sel. Evol. 41(2)

Patterson, H. D. and R. Thompson, 1971 Recovery of inter-block information when block sizes are unequal. Biometrika 58(3):545-54

Speed, D., G. Hemani, M. R. Johnson and D. J. Balding, 2012 Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. 91:1011-21

Speed, D. and D. J. Balding, 2014 MultiBLUP: improved SNP-based prediction for complex traits. Genome Res. 24:1550-7

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nature Genet. 42(7):565-9

Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen *et al.*, 2015 Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nature Genet. 47(10):1114-20

Table 1. Relationship between values of κ_i and λ_i for genomic models including QTL and markers (κ_{QM} and λ_{QM}) and including markers only (κ_M and λ_M), for eight different scenarios.

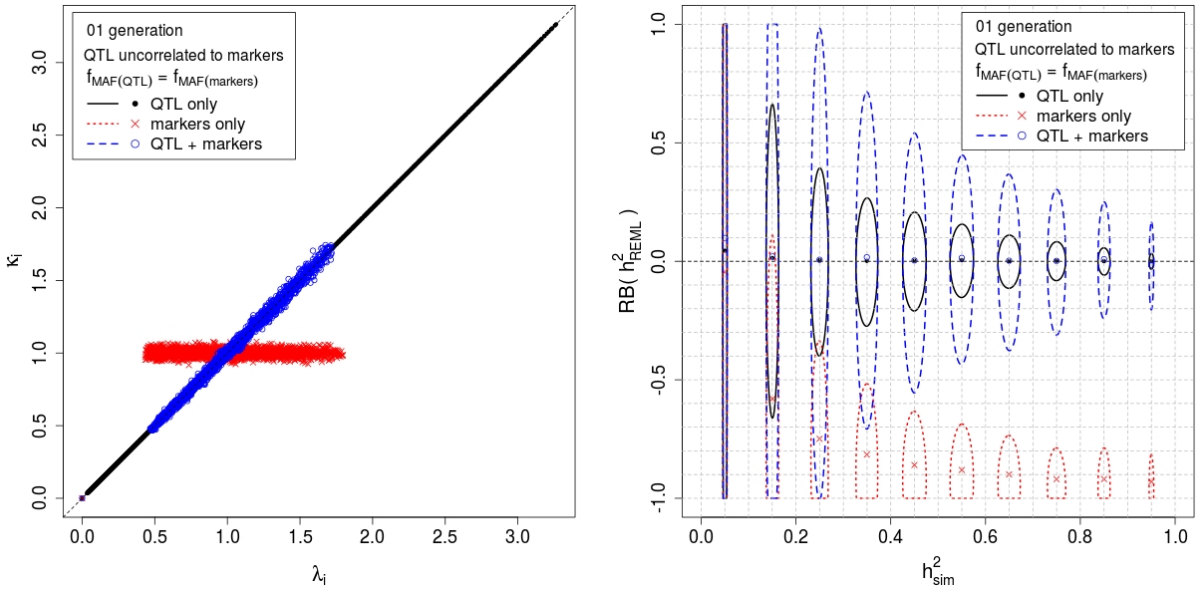
Scenario	Population	MAF	QTL/markers	κ_{QM}	κ_M
1	1 generation[1]	$MAF_{QTL} = MAF_{markers}$	complete LE	λ_{QM}	1
2	1 generation[1]	$MAF_{QTL} \neq MAF_{markers}$	complete LE	λ_{QM}	1
3	1 generation[1]	$MAF_{QTL} = MAF_{markers}$	LD	λ_{QM}	λ_M
4	2 generations	$MAF_{QTL} = MAF_{markers}$	LD	λ_{QM}	λ_M
5	10 generations	$MAF_{QTL} = MAF_{markers}$	LD	λ_{QM}	λ_M
6	1 generation[1]	$MAF_{QTL} \neq MAF_{markers}$	LD	λ_{QM}	λ_M
7	2 generations	$MAF_{QTL} \neq MAF_{markers}$	LD	λ_{QM}	λ_M
8	10 generations	$MAF_{QTL} \neq MAF_{markers}$	LD	λ_{QM}	λ_M

[1] completely unrelated individuals

Table 2. Asymptotic expectation of $\hat{h}^2$ for genomic models including QTL and markers ($\hat{h}^2_{QM}$) and including markers only ($\hat{h}^2_M$), based on the relationships between κ_i and λ_i, for eight different scenarios.

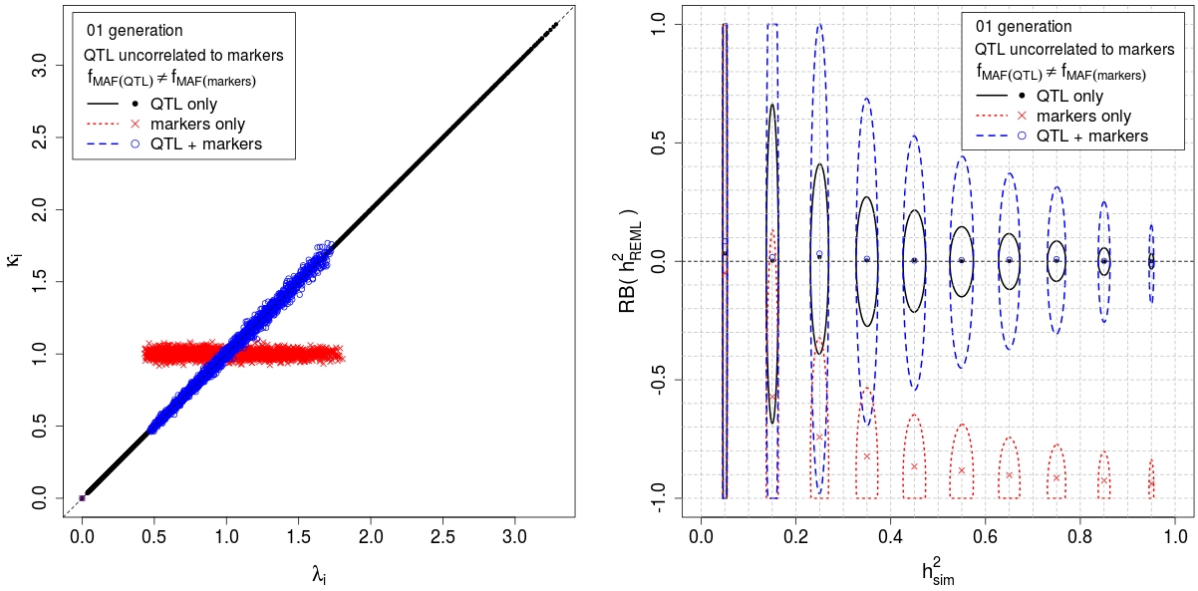
Scenario	Population	MAF	QTL/markers	$E_\infty(\hat{h}^2_{QM})$	$E_\infty(\hat{h}^2_M)$
1	1 generation[1]	$MAF_{QTL} = MAF_{markers}$	complete LE	h^2	0
2	1 generation[1]	$MAF_{QTL} \neq MAF_{markers}$	complete LE	h^2	0
3	1 generation[1]	$MAF_{QTL} = MAF_{markers}$	LD	h^2	$E_\infty(\hat{h}^2_{QM})$
4	2 generations	$MAF_{QTL} = MAF_{markers}$	LD	h^2	$E_\infty(\hat{h}^2_{QM})$
5	10 generations	$MAF_{QTL} = MAF_{markers}$	LD	h^2	$E_\infty(\hat{h}^2_{QM})$
6	1 generation[1]	$MAF_{QTL} \neq MAF_{markers}$	LD	h^2	$E_\infty(\hat{h}^2_{QM})$
7	2 generations	$MAF_{QTL} \neq MAF_{markers}$	LD	h^2	$E_\infty(\hat{h}^2_{QM})$
8	10 generations	$MAF_{QTL} \neq MAF_{markers}$	LD	h^2	$E_\infty(\hat{h}^2_{QM})$

[1] completely unrelated individuals



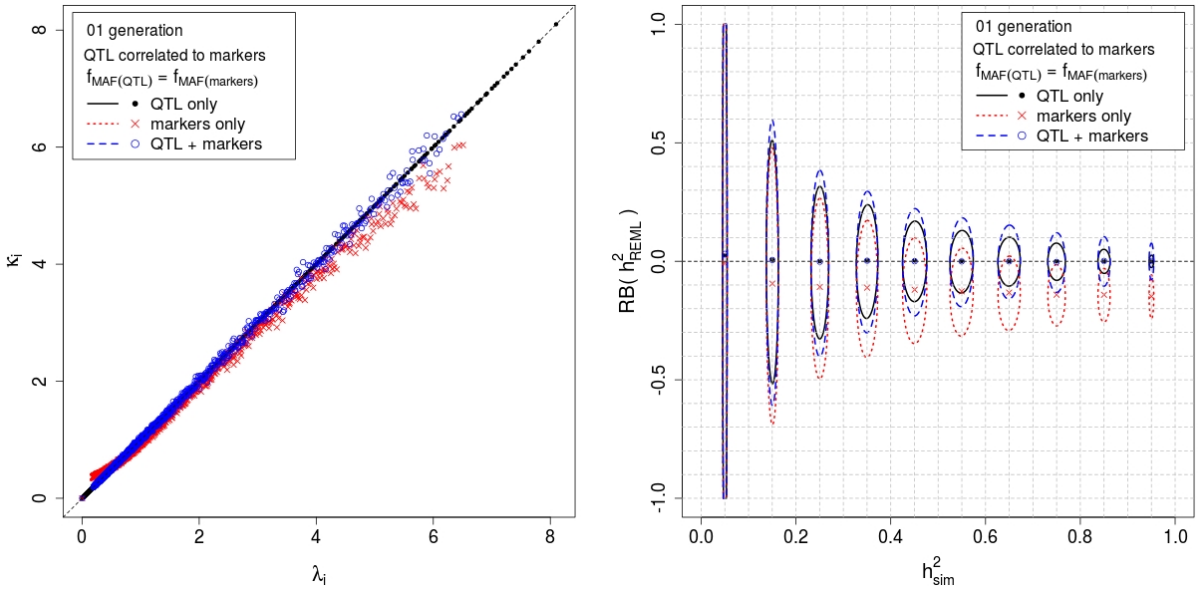
(a) (b)

Figure 1. Simulation results for scenario 1, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i and λ_i (b) relative bias of $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



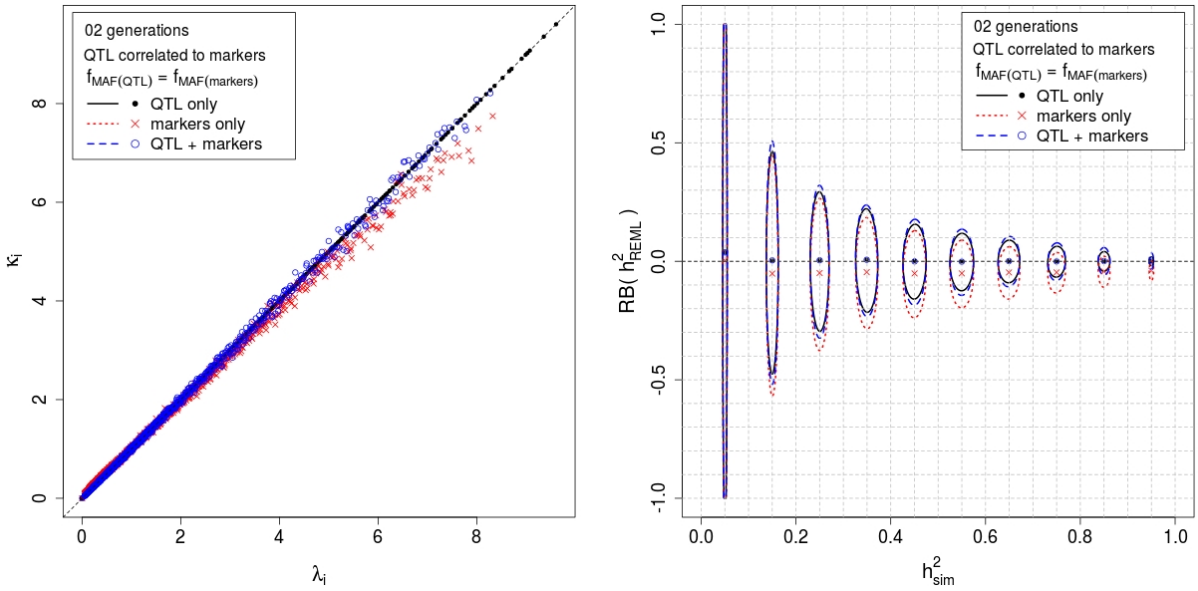
(a) (b)

Figure 2. Simulation results for scenario 2, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i and λ_i (b) relative bias of $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



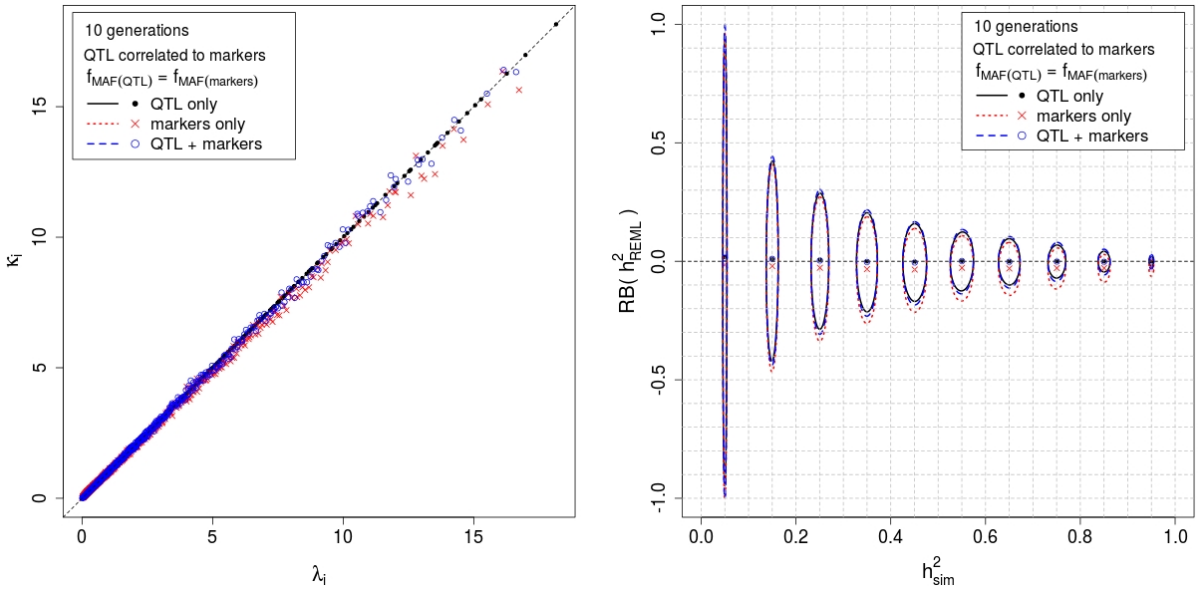
(a) (b)

Figure 3. Simulation results for scenario 3, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i and λ_i (b) relative bias of $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



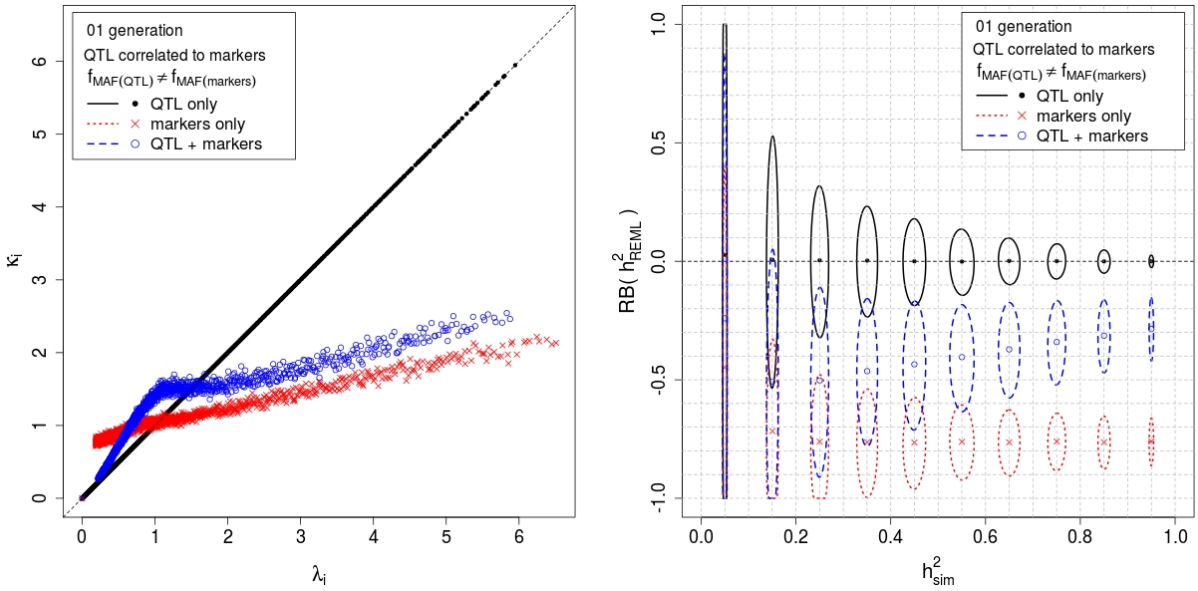
(a) (b)

Figure 4. Simulation results for scenario 4, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i and λ_i (b) relative bias of $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



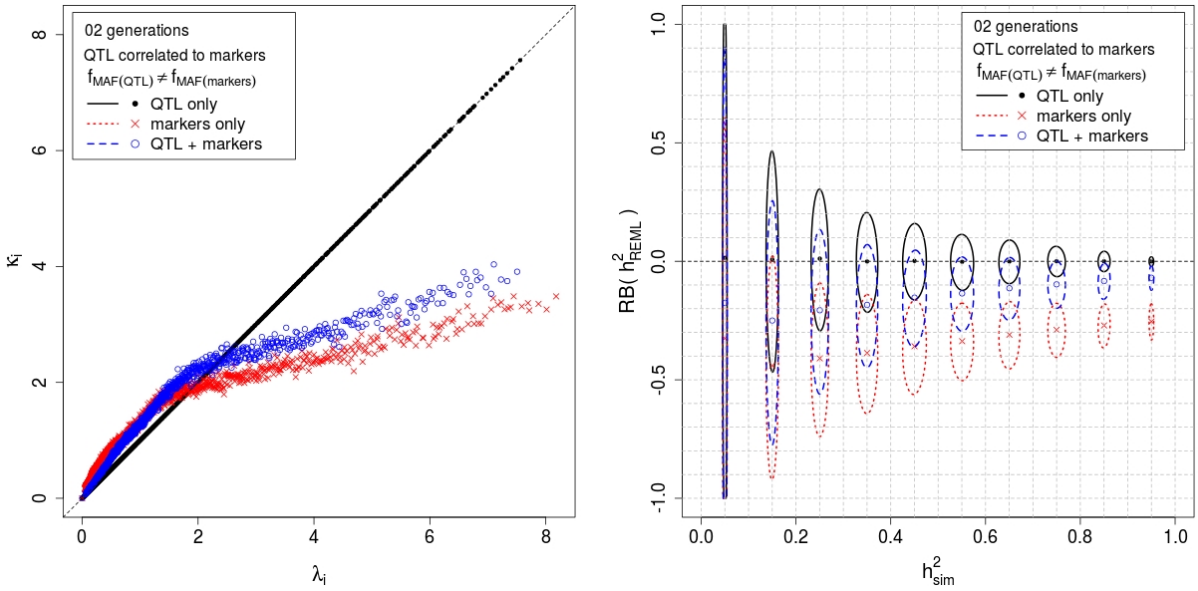
(a) (b)

Figure 5. Simulation results for scenario 5, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i and λ_i (b) relative bias of $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



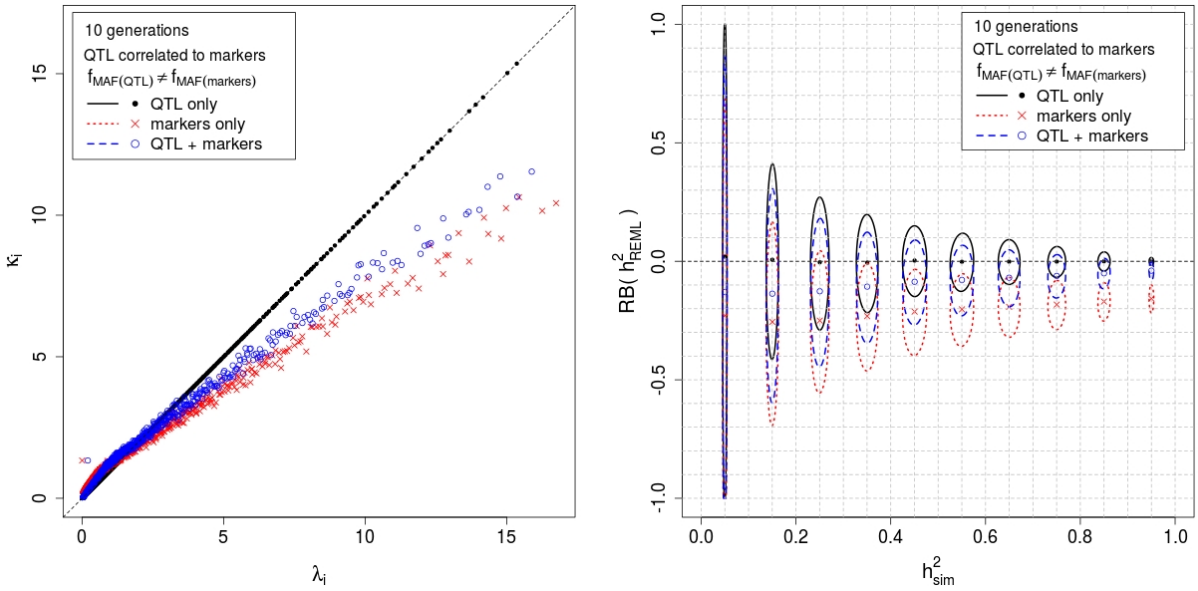
(a) (b)

Figure 6. Simulation results for scenario 6, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i and λ_i (b) relative bias of $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



(a) (b)

Figure 7. Simulation results for scenario 7, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i *and* λ_i *(b) relative bias of* $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.



(a) (b)

Figure 8. Simulation results for scenario 8, for genomic models including QTL only, including QTL and markers and including markers only (a) relationship between κ_i *and* λ_i *(b) relative bias of* $\hat{h}^2$, $RB(\hat{h}^2) = (\hat{h}^2 - h^2)/h^2$.