Accurate Estimation of Marker-Associated Genetic Variance and Heritability in Complex Trait Analyses

Mitchell J. Feldmann[1,*], **Hans-Peter Piepho**[2], **William C. Bridges**[3] **and Steven J. Knapp**[1,*]

[1]Department of Plant Sciences, University of California–Davis, Davis, California, United States of America, [2]Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany, [3]Department of Mathematical Sciences, Clemson University, Clemson, South Carolina, United States of America

ABSTRACT The emergence of high-throughput, genome-scale approaches for identifying and genotyping DNA variants has been a catalyst for the development of increasingly sophisticated whole-genome association and genomic prediction approaches, which together have revolutionized the study of complex traits in human, animal, and plant populations. These approaches have uncovered a broad spectrum of genetic complexity across traits and organisms, from a small number of detectable loci to an unknown number of undetectable loci. The heritable variation observed in a population is often partly caused by the segregation of one or more large-effect (statistically detectable) loci. Our study focused on the accurate estimation of the proportion of the genetic variance explained by such loci (p), a parameter estimated to quantify and predict the importance of causative loci or markers in linkage disequilibrium with causative loci. Here, we show that marker-associated genetic variances are systematically overestimated by standard statistical methods. The upward bias is purely mathematical in nature, unrelated to selection bias, and caused by the inequality between the genetic variance among progeny and sums of partitioned marker-associated genetic variances. We discovered a straightforward mathematical correction factor (k_M) that depends only on degrees of freedom and the number of entries, is constant for a given experiment design, expands to higher-order genetic models in a predictable pattern, and yields bias-corrected estimates of marker-associated genetic variance and heritability.

KEYWORDS heritability; linear mixed model; quantitative trait locus; single nucleotide polymorphism; average marginal variance; average semi-variance

Introduction

The introduction of chromosome-scale approaches for identifying and mapping quantitative trait loci (QTL) was an important milestone in quantitative genetics (Lander and Botstein 1989; Walsh 2001; Hill 2012; Churchill 2016). Through "interval mapping" Lander and Botstein (1989) showed that a genome could be searched to identify QTL predicted by Mendelian genetics to underlie the quantitative genetic variation observed in a population. Their approach relied on genotyping individuals with a genome-wide framework of DNA markers. The emergence of increasingly powerful methods for identifying and genotyping DNA variants has facilitated genome-wide searches for markers linked to QTL by exploiting the historic recombination in populations through genome-wide association studies (GWAS) (Visscher *et al.* 2012, 2017). Because the genetic factors underlying quantitative traits are often exceedingly complex and difficult to resolve (Mackay 2001; Huang and Mackay 2016; Visscher and Yang 2016), the QTL discovered with these approaches often only explain a fraction of the heritable variation for a trait (Walsh 2001; Bernardo 2008; Hill 2012; Visscher *et al.* 2012; Bernardo 2016).

Building on the foundation of genome-wide QTL discovery, Lande and Thompson (1990) described how genotypes for DNA markers in linkage disequilibrium (LD) with statistically significant QTL could be combined with phenotypes in an optimum index for artificial selection in animals and plants. They showed that the index weights and predicted genetic gains (ΔG) from marker-assisted index selection were functions of narrow-sense heritability ($h^2 = \sigma^2_A/\sigma^2_P$) and the proportion of the additive genetic variance explained by marker loci in LD with the underlying QTL ($p_A = \sigma^2_{M_A}/\sigma^2_A$), where σ^2_A is the additive genetic variance, $\sigma^2_{M_A}$ is the proportion of σ^2_A associated with statistically significant marker loci incorporated into the marker-assisted selection (MAS) index, and σ^2_P is the phenotypic variance (Neimann-Sorensen and Robertson 1961; Lande and Thompson 1990; Visscher *et al.* 2006). More generally, $p = \sigma^2_M/\sigma^2_G$, where σ^2_G is the genetic variance among entries (individuals or families) and

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Saturday 11th April, 2020
*Corresponding author: sjknapp@ucdavis.edu

σ^2_M is the proportion of σ^2_G explained by the intra- and inter-locus effects of marker loci. The interpretation of σ^2_G depends on the genetic or pedigree relationships among entries (Lynch and Walsh 1998), whereas the interpretation of σ^2_M depends on the marker effects specified in the model (Lande and Thompson 1990). Technically, $H^2_M = \sigma^2_M/\sigma^2_P$ is the fraction of the phenotypic variance, similar to broad-sense heritability ($H^2 = \sigma^2_G/\sigma^2_P$), associated with a statistically significant subset of markers in LD with the underlying QTL.

The marker-assisted selection (MAS) approach envisioned by Lande and Thompson (1990) was ultimately supplanted by genomic selection, where breeding values are estimated by whole-genome regression (WGR) methods in populations genotyped with a dense genome-wide framework of DNA markers (Meuwissen *et al.* 2001; VanRaden 2008; Hill 2012, 2014; Visscher *et al.* 2019). When applying WGR methods, the additive effects for statistically significant and non-significant markers (the entire framework) are summed across the genome (Meuwissen *et al.* 2001), as opposed to identifying, selecting, and summing the additive effects of statistically significant markers only, as in the Lande and Thompson (1990) MAS index. WGR has, in addition, facilitated the development of the theory and methods for assumption-free estimation of 'genomic' heritability and other genetic parameters (Visscher *et al.* 2006; VanRaden 2008; Visscher *et al.* 2008; de los Campos *et al.* 2015).

With the emergence of genomic prediction methods, the focus in quantitative genetic studies has heavily shifted away from QTL discovery *per se* towards genomic prediction and the estimation of genomic-estimated breeding values (GEBVs) (Meuwissen *et al.* 2001; Heffner *et al.* 2009; Piepho 2009; Daetwyler *et al.* 2010; de los Campos *et al.* 2013, 2015; Crossa *et al.* 2017; Lello *et al.* 2018; Damesa *et al.* 2019). Nevertheless, quantitative genetic studies typically entail a genome-wide search for large-effect (statistically significant) QTL, in addition to genomic prediction using WGR methods (Mackay 2001; Walsh 2001; Rutkoski *et al.* 2014; de los Campos *et al.* 2015; Lello *et al.* 2018; Wang and Xu 2019; Lello *et al.* 2019; Rice and Lipka 2019).

Although the concepts of marker-associated genetic variance and marker heritability (H^2_M) are integral to the application genome-informed approaches in quantitative genetics, estimates of these parameters have rarely been reported (Lander and Botstein 1989; Lande and Thompson 1990; Lynch and Walsh 1998; Walsh 2001; Hill 2012; de los Campos *et al.* 2015). We discovered that p and H^2_M are systematically overestimated by standard statistical approaches. To explore this problem, we investigated the accuracy of two methods for estimating p and H^2_M: average marginal variance (AMV) and average semi-variance (ASV) (Henderson 1953; Searle and Gruber 1971; Piepho 2019). As shown below, AMV estimators are identical to the well known ANOVA or method-of-moments estimators (Henderson 1953; Searle and Gruber 1971). On the surface, the estimation of p and H^2_M should be straightforward; however, when applying the Lande and Thompson (1990) index to a selection problem in sunflower (Tang *et al.* 2006), we observed AMV estimates of p and H^2_M that were greater than 1.0, the theoretical limit for both parameters. Here, we describe the root cause of the overestimation problem, identify the source of the bias, compare the accuracy of AMV and ASV estimators of p and H^2_M, and show that ASV yields accurate estimates, whereas AMV yields systematically upwardly biased estimates of both ratios. Finally, we present a straightforward approach for bias-correcting AMV estimates of p and H^2_M.

Materials and Methods

Linear Mixed Models for ANOVA Estimation

The linear mixed models needed for ANOVA estimation of σ^2_G and σ^2_M are developed here. Consider an experiment where n_G entries (e.g., individuals or families) are phenotyped for a normally distributed quantitative trait, genotyped for one or more markers, and tested in a balanced completely randomized experiment design with r_G replications/entry, n_M marker genotypes/locus, and r_M replications/marker genotype. Here, we study balanced designs with a single explanatory marker locus ($M1$) for simplicity. Solutions for more complex genetic models and unbalanced data are shown in Appendices A, B, and C.

The LMM for estimating σ^2_G is:

$$y_{jk} = \mu + G_j + \epsilon_{jk} \tag{1}$$

where y_{jk} is the jk^{th} phenotypic observation, μ is the population mean, G_j is the random effect of the jth entry, ϵ_{jk} is the random effect of the jk^{th} residual, $G_j \sim N(0, \sigma^2_G)$, $\epsilon_{jk} \sim N(0, \sigma^2_\epsilon)$, $j = 1, 2, ..., n_G$, and $k = 1, 2, ..., r_G$. The is the classic LMM for estimating σ^2_G using ANOVA or REML methods (Table 1).

Suppose entries were genotyped for a single marker locus ($M1$) in LD with a QTL affecting the trait. The LMM for estimating the genetic variance associated with $M1$ (σ^2_{M1}) is:

$$y_{ijk} = \mu + M1_i + G : M1_{ij} + \epsilon_{ijk} \tag{2}$$

where y_{ijk} is the ijk^{th} phenotypic observation, $M1_i$ is the random effect of the ith marker genotype at locus $M1$, $G : M1_{ij}$ is the random effect of the jth entry nested in the ith marker genotype, ϵ_{ijk} is the random effect of the ijk^{th} residual, $i = 1, 2, 3$, $M1_i \sim N(0, \sigma^2_{M1})$, $G : M1_{ij} \sim N(0, \sigma^2_{G:M1})$, and $\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$. This is the classic LMM for estimating σ^2_{M1} using ANOVA or REML methods (Table 1).

ANOVA Estimators of the Variance Components

Statistical analyses of LMMs (1) and (2) are needed to obtain estimates of σ^2_M, σ^2_G, p, and H^2_M. ANOVA estimators of the between-entry (σ^2_G) and residual (σ^2_ϵ) variance components for LMM (1) with balanced data are:

$$\hat{\sigma}^2_G = \frac{MS_G - MS_\epsilon}{r_G} \tag{3}$$

and

$$\hat{\sigma}^2_\epsilon = MS_\epsilon \tag{4}$$

where the mean sum of squares (MSs) are defined in Table (1), $E(\hat{\sigma}^2_G) = \sigma^2_G$, and $E(\hat{\sigma}^2_\epsilon) = \sigma^2_\epsilon$. Similarly, for a single marker locus ($M1$), ANOVA estimators of the genetic variance associated with $M1$ (σ^2_{M1}), the entries nested in the marker locus ($\sigma^2_{G:M1}$), and residual variance components for LMM (2) with balanced data are:

$$\hat{\sigma}^2_{M1} = \frac{MS_{M1} - MS_{G:M1}}{r_G n_{G:M1}} \tag{5}$$

and

$$\hat{\sigma}^2_{G:M1} = \frac{MS_{G:M1} - MS_\epsilon}{r_G} \tag{6}$$

and

$$\hat{\sigma}^2_\epsilon = MS_\epsilon$$

where $n_{G:M1}$ is the number of entries nested in $M1$ and the MSs are defined in Table 1, $E(\hat{\sigma}^2_{M1}) = \sigma^2_{M1}$, and $E(\hat{\sigma}^2_{G:M1}) = \sigma^2_{G:M1}$. The residuals in LMMs (1) and (2) are identical when the data are balanced. Hence, for a single marker locus, the classic estimator of marker-associated genetic variance is:

$$\hat{p}_{M1} = \frac{\hat{\sigma}^2_{M1}}{\hat{\sigma}^2_G} \tag{7}$$

and the estimator of broad-sense marker heritability on an entry-mean basis is:

$$\hat{H}^2_{M1} = \frac{\hat{\sigma}^2_{M1}}{\hat{\sigma}^2_G + \hat{\sigma}^2_{\bar{\epsilon}}/r_G} = \frac{\hat{\sigma}^2_{M1}}{\hat{\sigma}^2_{\bar{p}}} \tag{8}$$

where $\hat{\sigma}^2_{\bar{p}}$ is the phenotypic variance on an entry-mean basis.

Linear Mixed Models for AMV and ASV Estimation

The phenotypic observations for AMV and ASV estimation of the variance components are the adjusted entry-level means ($\bar{y}_{ij\bullet}$) from LMM (1), which are stored in an n_G-element vector ($\bar{y}$). These are best linear unbiased estimates (BLUEs) of the entry means (Piepho *et al.* 2012; Piepho 2019). The LMM equivalent to (2) for the entry-level means analysis of the effect of a single marker locus ($M1$) is:

$$\bar{y}_{ij\bullet} = \mu + M1_i + G : M1_{ij} + \bar{\epsilon}_{ij\bullet} \tag{9}$$

where $\bar{y}_{ij\bullet}$ is the phenotypic mean for the ij^{th} entry, μ is the population mean, $M1_i$ is the random effect of the ith marker genotype, $var(M1_i) = \sigma^2_{M1}$, $G : M1_{ij}$ is the random effect of entries nested in $M1$, $var(G : M1_{ij}) = \sigma^2_{G:M1}$, $\bar{\epsilon}_{ij\bullet}$ is the residual error, and $var(\bar{\epsilon}_{ij\bullet}) = r_G^{-1}\sigma^2_\epsilon$.

With the entry-mean vector ($\bar{y}$) as input, the LMM equivalent to (9) is:

$$\bar{y} = 1_n\mu + g + \bar{\epsilon} \tag{10}$$

where 1_{n_G} is an n_G-vector of ones, μ is the population mean, g is an n_G-vector of entry effects, $g \sim N(0, G)$, $\bar{\epsilon}$ is n_G-vector of residual effects, and $\bar{\epsilon} \sim N(0, R)$. The residual variance-covariance matrix (R) is estimated in the first stage of the two-stage ASV analysis (Piepho *et al.* 2012; Damesa *et al.* 2017, 2019). The structure of the entry variance-covariance matrix (G) depends on the model for marker locus effects (g). The g effect can be decomposed into the random effects of marker loci and entries nested in marker loci:

$$g = g_1 + g_2 + ... + g_c = \mathbf{Z}_1u_1 + \mathbf{Z}_2u_2 + ... + \mathbf{Z}_cu_c \tag{11}$$

where the $\mathbf{Z}_c$ are design matrices and u_c are independent random effect vectors for c genetic factors in the model, e.g., marker loci, interactions between marker loci, and entries nested in marker loci with individual variance-covariance matrices G_c. Hence, for a single marker locus $M1$, the c terms in (10) and (11) account for the genetic variation associated with $M1$ and the residual genetic variation associated with entries nested in $M1$ ($G : M1$). For LMM (10), the variance-covariance matrix on an entry-means basis is:

$$var(\bar{y}) = V = G + R = \sum_c (Z_cG_cZ_c^T) + R \tag{12}$$

where $\bar{y}$ is a vector of adjusted entry-level means. AMV and ASV estimators of the variance components (g_c) in LMMs (10) and (11) are shown in the results section, along with algebraic expressions of the biases of ASV estimators of p and H^2_M for different decompositions of g in (11).

Biases of AMV and ASV Estimators of p and H^2_M

True variance components in LMMs (1) and (2) are needed to estimate the biases of AMV and ASV estimators of p and H^2_M. Here, we will use sample variances of realized effects to define true variances. Following Estaghvirou *et al.* (2013), the true variance for the between-entry (total genetic) effect in LMM (1), is:

$$s^2_G = (n_y - 1)^{-1}\sum_{i=1}^{n_y}(\bar{g}_i - \bar{g}_\bullet)^2$$

where $\bar{g} = \mathbf{Z}_{M1}u_{M1} + \mathbf{Z}_{G:M1}u_{G:M1}$, $\bar{g}_i$ is the ith element of $\bar{g}$, $i = 1, 2, 3, ..., n$, and $\bar{g}_\bullet$ is the mean of $\bar{g}$. We note that $E(s^2_G) = E(\hat{\sigma}^2_G) = \sigma^2_G$. The true variance for the residual effect in LMM (1) is:

$$s^2_{\bar{\epsilon}} = (n_y - 1)^{-1}\sum_{i=1}^{n_y}(\bar{\epsilon}_i - \bar{\epsilon}_\bullet)^2$$

where $\bar{\epsilon}_i$ is the ith element of $\bar{\epsilon}$, $i = 1, 2, 3, ..., n$, and $\bar{\epsilon}_\bullet$ is the mean of $\bar{\epsilon}$. We note that $E(s^2_{\bar{\epsilon}}) = E(\hat{\sigma}^2_{\bar{\epsilon}}) = \sigma^2_{\bar{\epsilon}}$. The true variance for the effect of entries nested in $M1$ in LMM (2) is:

$$s^2_{G:M1} = (n_y - 1)^{-1}\sum_{i=1}^{n_y}(\bar{q}_i - \bar{q}_\bullet)^2$$

where $\bar{q} = \mathbf{Z}_{G:M1}u_{G:M1}$, $\bar{q}_i$ is the effect of the ith entry nested in marker locus 1, $i = 1, 2, 3, ..., n$), and $\bar{q}_\bullet$ is the mean of $\bar{q}$ across entries. We note that $E(s^2_{G:M1}) = E(\hat{\sigma}^2_{G:M1}) = \sigma^2_{G:M1}$. Finally, The true variance for the effect of a single marker locus ($M1$) in LMM (2) is:

$$s^2_{M1} = (n_y - 1)^{-1}\sum_{i=1}^{n_y}(m_i - \bar{m})^2$$

where $m = \mathbf{Z}_{M1}u_{M1}$, m_i is the effect of the ith genotype for marker locus 1, $i = 1, 2$, or 3, and $\bar{m}$ is the mean of m across entries. We demonstrate in the results that $E(s^2_{M1}) \leq E(\hat{\sigma}^2_{M1}) = \sigma^2_{M1}$. This inequality is at the heart of the problem we are studying and drives the overestimation of p and H^2_M.

The bias of the AMV estimator of σ^2_{M1} is:

$$bias(\hat{\theta}^{AMV}_{M1}) = E[\hat{\theta}^{AMV}_{M1}] - s^2_{M1}$$

where $\hat{\theta}^{AMV}_{M1}$ is the AMV estimate of σ^2_{M1} and $E[\hat{\theta}^{AMV}_{M1}]$ is the expected value of $\hat{\theta}^{AMV}_{M1}$. Similarly, the bias of the ASV estimator is:

$$bias(\hat{\theta}^{ASV}_{M1}) = E[\hat{\theta}^{ASV}_{M1}] - s^2_{M1}$$

where $\hat{\theta}^{ASV}_{M1}$ is the ASV estimator of σ^2_M and $E[\hat{\theta}^{ASV}_{M1}]$ is the expected value of $\hat{\theta}^{ASV}_{M1}$.

As shown in the results, the bias-correction factor (k_M) is a constant determined by the independent variable structure. The relative bias, which is expressed as a proportion of the true value (s^2_{M1}), is also constant for a given true value. The relative bias of the AMV estimator is:

$$RB(\hat{\theta}_{M1}^{AMV}) = \frac{E[\hat{\theta}_{M1}^{AMV}] - s_{M1}^2}{s_{M1}^2}$$

and the relative bias of the ASV estimator is:

$$RB(\hat{\theta}_{M1}^{ASV}) = \frac{E[\hat{\theta}_{M1}^{ASV}] - s_{M1}^2}{s_{M1}^2}$$

The biases and relative biases are determined through computer simulations.

Simulation Experiments

We used simulation to estimate the accuracy of AMV and ASV estimators of p and H_M^2. Plot-level phenotypic observations (y_{ijk}) for LMMs (1) and (2) were simulated for $n_M = 3$ genotypes/marker locus and different n_G, r_G, r_M, and H^2. The phenotypic observations were simulated for 21 combinations of these variables with balanced or unbalanced data structures as shown in Table 2. As shown below, the bias is a mathematical constant for a given experiment design (fixed n_G, r_G, n_M, and r_M). Simulations were performed to verify that result and estimate the biases associated with AMV and ASV estimates of σ_M^2 for different independent variable structures (study designs) (Table 2). We simulated 1,000 samples (iterations) for every study design using standard statistical methods (Burton *et al.* 2006; Morris *et al.* 2019). The different effects (random normal variates) were summed according to LMMs (1) and (2) to obtain $n = n_G r_G$ phenotypic observations in each sample for every study design.

The phenotypic observations for each sample were obtained by generating random normal variates for entries, marker loci, and residuals using the R function 'rnorm()' with known means and variances at the plot-level (R Core Team 2019). These plot-level observations were used to estimate the variance components in LMMs (1) and (2) and assess the accuracy of AMV and ASV estimators of p and H_M^2. In study designs 1-6, the true variances of marker loci (s_{M1}^2) and entries nested in marker loci ($s_{G:M1}^2$) were allowed to vary between samples such that the true marker heritability was between 0 and 1. Study designs 1-6 demonstrate how different experimental conditions, number of explanatory marker loci (m) and design balance, affect estimates of the variance explained by marker loci (Fig. 1; Table 2). In study designs 7-21, the true variances of the independent variables were fixed for all samples, which allowed us to estimate the bias and relative bias associated with the different estimators. Study designs 7-21 demonstrate how different independent variables, r_G and n_G, and H_M^2 affect the estimates of bias and relative bias (Figs. 2-4; Table 2). The variance components were estimated using REML in the 'lme4::lmer()' v1.1-21 (Bates *et al.* 2015) package in R v3.6.0 (R Core Team 2019).

Estimation Examples

We reanalyzed data from previously published studies on sunflower (Tang *et al.* 2006) and strawberry (Pincot *et al.* 2018) to compare AMV and bias-corrected AMV estimates of p and H_M^2 and illustrate the application of the bias-correction methods presented herein. For the sunflower study, two replications (r_G = 2) of $n_G = 146$ recombinant inbred lines (RILs) were phenotyped for seed oil concentration (g/kg) and genotyped for three marker loci (B, P, and HYP) with two homozygous marker genotypes/locus (Tang *et al.* 2006). The LMM for the analysis of (2), expanded to three marker loci with interactions, is:

$$\begin{aligned} y_{hijkl} =& \mu + B_h + P_i + HYP_j \\ &+ B_h \times P_i + B_h \times HYP_j + P_i \times HYP_j \\ &+ B_h \times P_i \times HYP_j + G_k : (B_h \times P_i \times HYP_j) \\ &+ \epsilon_{hijkl} \end{aligned} \tag{13}$$

where B_h is the hth effect of the B locus, P_i is the ith effect of the P locus, HYP_j is the jth effect of the HYP locus, $G_k : (B \times P \times HYP)$ is the kth effect of entries nested in $B \times P \times HYP$ marker loci, and ϵ_{hijkl} is the $hijkl^{th}$ residual effect. The data for RILs were balanced, whereas the data for marker genotypes were slightly unbalanced. Each of the eight $B \times P \times HYP$ homozygotes were observed in the RIL population; however, the number of entries nested in each marker genotype ($n_{G:M}$) varied from $n_{G:B} = 81 : 65$, $n_{G:P} = 60 : 86$, and $n_{G:HYP} = 70 : 76$. Variance components for LMMs (1) and (13) were estimated using the REML method in 'lme4::lmer()' (Bates *et al.* 2015). The marker-associated genetic variances for individual marker loci and two- and three-way interactions among marker loci were bias-corrected using approaches described in Appendices B-D.

For the strawberry study, four replications (r_G) of 565 individuals (n_G) from a genome-wide association study (GWAS) were phenotyped for resistance to Fusarium wilt and genotyped for single nucleotide polymorphism (SNP) markers in LD with $Fw1$, a dominant gene conferring resistance to the pathogen (Pincot *et al.* 2018). The replications were asexually propagated clones of individuals; hence, the expected causal variance among individuals was equal to the total genetic variation in the population, analogous to that expected among monozygotic twins (Lynch and Walsh 1998). Genetic parameters were estimated for two SNP markers ($AX493$ and $AX396$) that were tightly linked to $Fw1$ (Pincot *et al.* 2018). The genotypic data for both markers were highly unbalanced. Genotype numbers were $141AA : 282Aa : 141aa$ for $AX493$ and $16AA : 177Aa : 371aa$ for $AX396$, where A and a are alternate SNP alleles. The variance components for the strawberry study were estimated for LMMs (1) and (2) using the REML method in 'lme4::lmer()' (Bates *et al.* 2015). The marker-associated genetic variances for both loci were bias-corrected using the approach described in Appendix B.

Data Availability

The custom R scripts for reproducing our simulations have been deposited in a public GitHub repository (https://github.com/mjfeldmann/VarCompSim). The simulated data shown in Figures 1-4 have been deposited in a public Zenodo repository (http://dx.doi.org/10.5281/zenodo.3742421).

Results

ANOVA Estimators

Through genetic analyses of quantitative phenotypes in plant populations, we discovered that the sum of REML estimates of σ_M^2 and $\sigma_{G:M}^2$ from LMM (2) were greater than the REML estimate of σ_G^2 from LMM (1) (Table 4). Intuitively, we expected $\hat{\sigma}_M^2$ and $\hat{\sigma}_{G:M}^2$ to sum to $\hat{\sigma}_G^2$; however, our analyses show that estimates of variance components from the individual LMMs are not corrected for differences in the number of observations for different factors in the models. While the estimation of $p = \sigma_M^2/\sigma_G^2$ seems straightforward, we found that this parameter is

systematically overestimated using standard ANOVA methods. ANOVA estimators of the individual variance components from the two LMMs are unbiased (Searle and Gruber 1971); however, the ratio of $\hat{\sigma}^2_M$ from LMM (2) to $\hat{\sigma}^2_G$ from LMM (1) is upwardly biased because mathematical constants (k_M) are hidden in the expected values of the parameters estimated in the two models.

We discovered and show below that ANOVA estimators of σ^2_G from LMM (1) are equal to:

$$\hat{\sigma}^2_G = k_{M1}\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1}$$

where $\hat{\sigma}^2_G$ is the ANOVA estimator of σ^2_G from LMM (1), $\hat{\sigma}^2_{M1}$ and $\hat{\sigma}^2_{G:M1}$ are ANOVA estimators of σ^2_{M1} and $\sigma^2_{G:M1}$ from LMM (2), and k_{M1} is the coefficient hidden in the expected mean squares of the two LMMs. While the sums of squares are additive ($SS_G = SS_{M1} + SS_{G:M1}$), the mean squares are not. This is where the problem arises with estimating p and H^2_M as functions of the ANOVA estimators from LMMs (1) and (2). The 'hidden' coefficients (k_M) are the source of the bias in ANOVA estimators of σ^2_M (Tables 1-3).

To define the k_M algebraically, we equated ANOVA estimators of the variance components from LMM (1) with those from LMM (2). Here we show the solution for a model with a single marker locus ($M1$) with balanced data. Solutions for more complex genetic models and unbalanced data are shown in Table 1 and Appendices A, B, and C. From LMM (1) and equation (3), the ANOVA estimator of the between entry variance component is:

$$\begin{aligned}\hat{\sigma}^2_G &= \frac{MS_G - MS_\epsilon}{r_G} \\ &= \frac{SS_G/df_G - SS_\epsilon/df_\epsilon}{r_G} \\ &= \frac{1}{df_G r_G}SS_G - \frac{1}{df_\epsilon r_G}SS_\epsilon \qquad (14)\end{aligned}$$

where SS_G and SS_ϵ are sums of squares for the between entry and residual effect in LMM (1) and $SS_G = SS_{M1} + SS_{G:M1}$ (Table 1). After substituting $SS_{M1} + SS_{G:M1}$ for SS_G in (14), we obtained:

$$\begin{aligned}(14)\text{ cont.} &= \frac{1}{df_G r_G}(SS_{M1} + SS_{G:M1}) - \frac{1}{df_\epsilon r_G}SS_\epsilon \\ &= \frac{1}{df_G r_G}[df_{M1}(\hat{\sigma}^2_\epsilon + r_{G:M1}\hat{\sigma}^2_{G:M1} + r_{M1}\hat{\sigma}^2_{M1}) \\ &+ df_{G:M1}(\hat{\sigma}^2_\epsilon + r_{G:M1}\hat{\sigma}^2_{G:M1})] - \frac{1}{df_\epsilon r_G}df_\epsilon\hat{\sigma}^2_\epsilon \\ &= \frac{df_{M1}r_{M1}}{df_G r_G}\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1} \\ &= \frac{df_{M1}n_{G:M1}}{df_G}\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1} \\ &= k_{M1}\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1} \qquad (15)\end{aligned}$$

From (15), the sum of ANOVA estimators of σ^2_{M1} and $\sigma^2_{G:M1}$ from LMM (1) is greater than the ANOVA estimator of σ^2_G from LMM (2):

$$\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1} > \hat{\sigma}^2_G = k_{M1}\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1} \qquad (16)$$

Hence, $\hat{\sigma}^2_{M1} + \hat{\sigma}^2_{G:M1}$ overestimates $\hat{\sigma}^2_G$ by $(1 - k_M)\hat{\sigma}^2_{M1}$.

The bias-correction coefficients (k_M) range from zero to one and are mathematical constants for a particular experiment design and data structure (Tables 1-4). For a single marker locus ($M1$), $r_{M1} = n_G/n_{M1}$; hence, $r_{M1} < n_G$ and $0 < k_{M1} < 1$ (Table 1). Similar inequalities exist for the bias correction coefficients (k_M) in more complicated genetic models and differ for different levels of effects in the genetic model (Table 3; Appendices A, B, and C). For a two locus model with balanced data, $k_{M1} = k_{M2} < k_{M1\times M2}$ (Tables 1-3). The coefficient for the two locus interaction ($k_{M1\times M2}$) is larger than the coefficients for either locus (k_{M1} or k_{M2}) because the denominator ($df_G r_G$) is constant, whereas the numerators increase and approach the denominator as the degrees of freedom for marker effects increases (Tables 1-4). We predicted that multiplying ANOVA estimates of σ^2_M by k_M would correct the bias and yield unbiased estimates of σ^2_M, p, and H^2_M. We tested this through computer simulations, the results of which are presented after we show the mathematical relationships among ANOVA, AMV, and ASV estimators of the variance components and demonstrate that the ASV estimators of p and H^2_M are unbiased.

Average Marginal Variance Estimators

Although the formula and notation for AMV estimators of the variance components are less familiar than the well known and widely used formula and notations for Henderson (1953) method-of-moments or ANOVA estimators (Searle and Gruber 1971), they are identical. AMV estimation uses LMM (9) with the entry means ($\bar{y}_{ij\bullet}$) as input, whereas ANOVA uses LMMs (1 and 2) and the original phenotypic observations (y_{ijk}) as input. The AMV defintion of the total variance among observations for LMM (9) is:

$$\theta^{AMV}_{\bar{y}} = n_G{}^{-1}tr(V) = \sum_c \theta^{AMV}_{g_c} + \theta^{AMV}_{\bar{\epsilon}}$$

where V is the variance-covariance matrix of observations, n_G is the number of entries, $tr(V)$ is the trace of V, $\theta^{AMV}_{g_c} = n_G{}^{-1}tr(Z_c G_c Z_c^T)$ is the marginal variance explained by the cth genetic factor in the model (e.g., marker loci, interactions among marker loci, and entries nested in marker loci), Z_c are design matrices for the c genetic factors, $\theta^{AMV}_{\bar{\epsilon}} = n_G{}^{-1}tr(R)$ is the AMV estimator of the residual variance, and R is the residual variance-covariance matrix.

From LMM (9), the AMV estimator of of the variance explained by $M1$ is:

$$\begin{aligned}\hat{\theta}^{AMV}_{M1} &= (n_G)^{-1}\hat{\sigma}^2_{M1}tr(Z_{u_{M1}}Z^T_{u_{M1}}) \\ &= \frac{(n_{M1})n_{G:M1}}{n_G}\hat{\sigma}^2_{M1} \\ &= \hat{\sigma}^2_{M1}\end{aligned}$$

where $Z_{u_{M1}} = I_{n_{M1}} \otimes 1_{n_{G:M1}}$ and u_{M1} is a vector of random effects for $M1$. We note that $E(\hat{\sigma}^2_{M1}) = \theta^{AMV}_{M1} = \sigma^2_{M1}$. The AMV estimator of of the variance explained by $G : M1$ is:

$$\begin{aligned}\theta^{AMV}_{G:M1} &= (n_G)^{-1}\sigma^2_{G:M1}tr(Z_{u_{G:M1}}Z^T_{u_{G:M1}}) \\ &= \frac{n_G}{n_G}\sigma^2_{G:M1} \\ &= \sigma^2_{G:M1}\end{aligned}$$

where $u_{G:M1}$ is a vector of random entry nested in $M1$ effects. Hence, from (14) and (15), the sum of AMV estimators of σ^2_{M1} and $\sigma^2_{G:M1}$ is greater than the ANOVA estimator of σ^2_G:

$$\hat{\theta}^{AMV}_{M1} + \hat{\theta}^{AMV}_{G:M1} \geq \hat{\sigma}^2_G$$

Average Semi-Variance Estimators

The average semi-variance or average variance of differences among observations leads to a definition of the total variance that provides a natural way to account for heterogeneity of variance and covariance among observations (Piepho 2019; Schmidt *et al.* 2019a). ASV can be defined for any variance-covariance structure in a generalized LMM and allows for missing and unbalanced data. The ASV measure of total variance is half the average total pairwise variance of a difference between entries:

$$\theta^{ASV}_{\bar{y}} = (n_G - 1)^{-1} tr(VP_{n_G})$$

where $P_{n_G} = I_{n_G} - {n_G}^{-1} J_{n_G}$, I_{n_G} is an $n_G \times n_G$ identity matrix, and J_{n_G} is an $n_G \times n_G$ unit matrix (Piepho 2019). $\theta^{ASV}_{\bar{y}}$ accounts for the variances and covariances of the phenotypes. By comparison, the sample variance of the adjusted entry-level means ($\bar{y}$) is:

$$s^2_{\bar{y}} = (n_G - 1)^{-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{y}_\bullet)^2$$

where $\bar{y}_\bullet = {n_G}^{-1} \sum_{i=1}^{n_G} \bar{y}_i = {n_G}^{-1} 1^T_{n_G} \bar{y}$. Interestingly, the expected value of $s^2_{\bar{y}}$ is:

$$\begin{aligned} E(s^2_{\bar{y}}) &= (n_G - 1)^{-1} E(\bar{y}^T P_{n_G} \bar{y}) \\ &= (n_G - 1)^{-1} tr(VP_{n_G}) \\ &= \theta^{ASV}_{\bar{y}} \end{aligned} \tag{17}$$

(Estaghvirou *et al.* 2013). The ASV measure of total variance can be decomposed into independent sources of variance (e.g., genetic and residual) according to LMM (12):

$$\begin{aligned} tr(VP_{n_G}) &= tr(GP_{n_G}) + tr(RP_{n_G}) \\ &= \sum_c tr(Z_c G Z_c^T P_{n_G}) + tr(RP_{n_G}) \end{aligned}$$

Thus,

$$\theta^{ASV}_{\bar{y}} = \sum_c \theta^{ASV}_{g_c} + \theta^{ASV}_{\bar{\epsilon}}$$

where $\theta^{ASV}_{g_c} = (n_G - 1)^{-1} tr(Z_c G_c Z_c^T P_{n_G})$ is the variance explained by the cth genetic factor (u_c), genetic factors are marker loci, interactions between marker loci, and entries nested in marker loci, and $\theta^{ASV}_{\bar{\epsilon}} = (n_G - 1)^{-1} tr(RP_{n_G})$ is the residual variance.

Equation (17) can be decomposed into the expected values of the sample variances of the random effects:

$$\begin{aligned} E(s^2_{\bar{y}}) &= (n_G - 1)^{-1} E(\bar{y}^T P_{n_G} \bar{y}) \\ &= (n_G - 1)^{-1} \left[\sum_c E(u_c^T Z_c^T P_{n_G} Z_c u_c) + E(\bar{\epsilon}^T P_{n_G} \bar{\epsilon}) \right] \\ &= \sum_c E(s^2_{g_c}) + E(s^2_{\bar{\epsilon}}) \end{aligned}$$

where $s^2_{g_c} = (n_G - 1)^{-1} \sum_{i=1}^{n_G} (g_{i(c)} - \bar{g}_{\bullet(c)})^2$, $g_{i(c)}$ is the ith element of $g_c = Z_c u_c$, $\bar{g}_{\bullet(c)} = {n_G}^{-1} \sum_{i=1}^{n_G} g_{i(c)}$, and $s^2_{\bar{\epsilon}} = (n_G - 1)^{-1} \sum_{i=1}^{n_G} (\bar{\epsilon}_i - \bar{\epsilon}_\bullet)^2$. Hence, the variance explained by the cth genetic factor is:

$$\theta^{ASV}_{g_c} = E(s^2_{g_c})$$

where c indexes marker loci, interactions between marker loci, and entries nested in marker loci.

For a single marker locus $M1$, $\theta^{ASV}_{M1} = E(s^2_{M1})$. The ASV estimator of the variance explained by $M1$ is:

$$\begin{aligned} \hat{\theta}^{ASV}_{M1} &= (n_G - 1)^{-1} \hat{\sigma}^2_{M1} tr(Z_{u_{M1}} Z^T_{u_{M1}} P) \\ &= \frac{(n_{M1} - 1) n_{G:M1}}{n_G - 1} \hat{\sigma}^2_{M1} \\ &= \frac{df_{M1} n_{G:M1}}{df_G} \hat{\sigma}^2_{M1} \\ &= k_{M1} \hat{\sigma}^2_{M1} \\ &= \hat{\sigma}^2_{M1*} \end{aligned} \tag{18}$$

where $Z_{u_{M1}} = I_{n_{M1}} \otimes 1_{n_{G:M1}}$, $df_G = n_G - 1$, $df_{M1} = n_{M1} - 1$, and $df_{G:M1} = df_G - df_{M1}$ (Table 1). The asterisk subscript ($*$) is used to identify ASV (bias-corrected) estimates of σ^2_M, p, and H^2_M. Consistent with (16), $\hat{\theta}^{ASV}_{M1} < \hat{\theta}^{AMV}_{M1}$ by the factor k_{M1}. As shown in Tables (1-3), k_M varies across genetic models and experiment designs.

The ASV estimator of the variance explained by $G : M1$ is:

$$\begin{aligned} \hat{\theta}^{ASV}_{G:M1} &= (n_G - 1)^{-1} \hat{\sigma}^2_{G:M1} tr(Z_{u_{G:M1}} Z^T_{u_{G:M1}} P) \\ &= \frac{n_G - 1}{n_G - 1} \hat{\sigma}^2_{G:M1} \\ &= \hat{\sigma}^2_{G:M1} \end{aligned} \tag{19}$$

From (18), the coefficient for bias-correcting AMV estimates of σ^2_{M1} is:

$$k_{M1} = \frac{df_{M1} n_{G:M1}}{df_G}$$

Note that this definition of k_{M1} is equivalent to the definitions in Table 3 with r_G factored out. Hence, from (14) and (15), the sum of ASV estimators is equal to the ANOVA estimator of $\hat{\sigma}^2_G$:

$$\hat{\theta}^{ASV}_{M1} + \hat{\theta}^{ASV}_{G:M1} = \hat{\sigma}^2_G$$

Thus, in contrast to (7) and (8), the bias-corrected estimator of p for a single marker locus ($M1$) is:

$$\hat{p}_{M1*} = \frac{\hat{\sigma}^2_{M1*}}{\hat{\sigma}^2_G} = \frac{k_{M1} \hat{\sigma}^2_{M1}}{\hat{\sigma}^2_G}$$

Similarly, the bias-corrected estimator of H^2_M for a single marker locus is:

$$\hat{H}^2_{M1*} = \frac{\hat{\sigma}^2_{M1*}}{\hat{\sigma}^2_G + \hat{\sigma}^2_\epsilon / r_G} = \frac{k_{M1} \hat{\sigma}^2_{M1}}{\hat{\sigma}^2_G + \hat{\sigma}^2_\epsilon / r_G} = \frac{k_{M1} \hat{\sigma}^2_{M1}}{\hat{\sigma}^2_{\bar{P}}}$$

where $\hat{\sigma}^2_G + \hat{\sigma}^2_\epsilon / r_G = \hat{\sigma}^2_{\bar{P}}$ is the phenotypic variance on an entry-mean basis (Lynch and Walsh 1998).

Simulations

The systematic biases of AMV estimators of p and H^2_M were verified by analyses of simulated samples for 21 different study designs and parameter combinations (Figures 1-4; Supplemental Table 2). As predicted by (15) and (18), ASV estimates were unbiased, whereas AMV estimates were upwardly biased across simulation studies. The magnitude of the bias increased as H^2_M increased but was proportionally constant (Figure 1). For a single locus with balanced data, for example, H^2_M and bias increased as the effect of the QTL increased (Figure 1A). For experiments with multiple loci, the bias of the ANOVA estimators of marker-marker interactions was less than that for individual marker loci as a function of the underlying bias-correction coefficients (k_M). This suggests that as the number of observed levels of marker-marker interaction approaches the number of entries (n_G), the value of k_M approaches 1.0.

The bias was greater for unbalanced than for balanced data (Fig. 1). As shown for an F_2 population segregating 1 AA : 2 Aa : 1 aa for a single marker locus, AMV and ASV estimates of H^2_M were significantly more variable for unbalanced than balanced data (Fig. 1). The effect of unbalanced data was more extreme for the F_2 simulation (Fig. 1D) than for simulations where 10-33% of the observations were missing for markers with roughly equal numbers of replications/marker genotype (Fig. 1E-F). The increased dispersion of AMV and ASV estimates for the F_2 simulation (Figure 1D) relative to estimates for the other simulations (Figure 1A-C and E-F) resulted from lower precision of the variance component estimates. Even though the ASV estimates are unbiased, sampling variances among the simulated F_2 samples were larger and yielded a small percentage of H^2_M estimates slightly outside the theoretical range (Figure 1D). For the other simulations (Fig. 1A-C and E-F), none of the ASV estimates exceeded 1.0.

The relative biases were not affected by the number of replications of entries or the number of entries, although the precision of σ^2_M estimates increased as n_G increased (Figures 2-3). Similarly, the relative biases were not affected by H^2_M; however, estimate precision was strongly affected by H^2_M (Figure 4). These simulations show that the relative biases of the AMV estimators of σ^2_M and H^2_M are mathematically constant for a given experiment and study design and that the bias-correction we applied yielded unbiased estimates of σ^2_M, p, and H^2_M (Figures 1-4).

Case Study #1: Three Marker Loci With Slightly Unbalanced Genotypic Data

We estimated the marker-associated genetic variances for three loci (B, P, and HYP) segregating in a sunflower RIL mapping population (Tang *et al.* 2006) using LMM (13) (Table 4). Uncorrected (AMV) and bias-corrected (ASV) REML estimates of the marker-associated genetic variances for HYP, $B \times P$, $P \times HYP$, $B \times P \times HYP$ were zero (Table 4). HYP was retained in the model because the marker-associated variance for $B \times HYP$ was non-zero. The AMV estimate of σ^2_M (17.85) was greater than the ASV estimate of σ^2_M (9.24) for the full three-locus model, as predicted by (15) and (18). Similarly, the AMV estimate of marker heritability (0.79) was nearly two-fold greater than the ASV estimate of marker heritability (0.41) for the full three-locus model. The AMV estimate of the contribution of the B locus to marker heritability (0.51) was double that of the ASV estimate (0.25). Hence, AMV yielded inflated estimates of the contributions of these marker loci to the heritable variation for the trait under study in this population (seed oil concentration). The ASV (bias-corrected AMV) estimate for the full three-locus model suggests that the intra- and interlocus effects of these loci explained 43.0% of the genetic variation ($\hat{p} = 9.24/21.61 = 0.43$), as opposed to 79.0% for the upwardly biased AMV estimate ($\hat{p} = 17.85/21.61 = 0.82$) (Table 4).

Case Study #2: A Single Marker Locus With Highly Unbalanced Genotypic Data

The application of the bias-correction is illustrated here for highly unbalanced genotypic data. Variance components were estimated for two SNP markers ($AX493$ and $AX396$) in LD with a gene ($Fw1$) affecting resistance to Fusarium wilt in a strawberry GWAS population (n_G = 564) that was genotyped with a genome-wide framework of SNP markers (Pincot *et al.* 2018). Both SNP markers were highly significant with $-log_{10}p$-values of 6.61×10^{-31} for $AX493$ and 2.95×10^{-222} for $AX396$. Genotype frequencies were highly unbalanced for both markers with a scarcity of AA homozygotes (2.8%) for $AX396$ ($16AA : 177Aa : 371aa$) and 1 : 2: 1 genotypic ratio for $AX493$ ($141AA : 282Aa : 141aa$). The k_M for these data ($k_{AX493} = 0.62$ and $k_{AX396} = 0.47$) were calculated as shown in Appendix B.

The AMV estimate of H^2_M for $AX396$ exceeded 1.0, a telltale sign of k_M-bias (Table 4). AMV estimates of σ^2_M and H^2_M for both SNP markers were double or nearly double their ASV estimates (Table 4). The bias-corrected ASV estimate for $AX396$ suggested that the causal locus ($Fw1$) explained approximately 62% of the genetic variance for resistance to Fusarium wilt ($\hat{p}_{M*} = 0.62$), versus a non-sensical 133% for the biased AMV estimate ($\hat{p}_M = 1.33$). Even with bias correction, the sum of ASV estimates of σ^2_M and $\sigma^2_{G:M}$ for $AX493$ was slightly greater than the ASV estimate of σ^2_G. This result was consistent with the pattern we observed for highly unbalanced marker genotypic data in our simulation studies where a certain fraction of bias-corrected estimates exceeded the theoretical limit for heritability because of decreased precision (Figure 1). The k_M-bias problem would not necessarily have been detected in the analysis of $AX396$ because the p and H^2_M estimates fell within the acceptable range (0,1), e.g., $(\hat{\theta}^{ASV}_{AX396} + \hat{\theta}^{ASV}_{G:AX396})/\hat{\theta}^{ASV}_{\bar{P}} = 0.71$. This estimate was nevertheless still upwardly biased (Table 4).

Discussion

We originally suspected that selection bias was the culprit behind the overestimation of p and H^2_M (Beavis *et al.* 1994; Xu 2003; Bernardo 2004); however, the k_M-bias problem is purely statistical in nature and operates in combination with selection bias to inflate estimates of the importance (statistical contributions) of QTL to the heritable variation in a population (Table 3; Appendices A, B, and C). Although the concept of marker heritability was introduced early in the application of genome-informed approaches in quantitative genetics, σ^2_M, p and H^2_M have not been widely estimated (Lande and Thompson 1990; Utz *et al.* 2000; Walsh 2001; Bernardo 2004; Visscher *et al.* 2006; Bernardo 2008; Visscher *et al.* 2008; Mackay *et al.* 2009; Manolio *et al.* 2009; Cockram and Mackay 2018; Schmidt *et al.* 2019b; Wang and Xu 2019). Our results show that the overestimation of p and H^2_M is greatest for large-effect QTL, which were the initial focus of our study (Figure 1). As shown in the empirical examples, the overestimation problem would only be obvious in practice for cases where estimates of p and H^2_M exceeded 1.0, as was the case in the two empirical examples we examined (Table 4). The two-fold difference between ASV estimates of H^2_M for the two

SNP markers in the strawberry example illustrates the challenge with accurately estimating the proportion of the heritable variation associated with a marker. Because neither SNP appears to be a causal variant, the difference in the H_M^2 estimates stems from historic recombination between the causal locus ($Fw1$) and the SNP marker loci ($AX396$ was significantly more predictive than $AX493$). Regardless, the contributions of both loci were overestimated by AMV estimates of H_M^2.

Genomic heritability is routinely estimated in GWAS and genomic prediction studies where one of the goals is to accurately estimate the proportion of the genetic variance associated with thousands of DNA markers, in addition to estimating the 'missing' heritability (Visscher *et al.* 2006, 2008; VanRaden 2008; de los Campos *et al.* 2015). As shown by de los Campos *et al.* (2015), the mathematical relationship between heritability (estimated using family or pedigree relationships) and genomic heritability (estimated using genotypic relationships) is complex and not completely clear (Legarra *et al.* 2009). However, the proposed definition of marker heritability can be linked to genomic heritability, as described in Schmidt *et al.* (2019a), through the VanRaden (2008) definition of genomic relatedness. We used different terms to distinguish these parameters, one of which is directly estimated from statistically significant marker genotypes (marker heritability) and another of which is estimated from genetic relationship matrices (genomic heritability). For practical purposes, H_M^2 is the proportion of the heritability associated with one or more statistically significant markers, as opposed to the proportion associated with a genome-wide sample of both non-significant and statistically significant markers (Visscher *et al.* 2006, 2008; de los Campos *et al.* 2015).

Variance component ratios similar to p and H_M^2 are used in diverse fields of study, most commonly to estimate the fraction of the total variance of a dependent variable associated with a specific factor or independent variable (Searle 1995; Sun *et al.* 2010; Piepho 2019; Schmidt *et al.* 2019a,b). These parameters include repeatability, intra-class correlation, and coefficient of determination (Hill and Nicholas 1974; Utz *et al.* 2000; Schmidt *et al.* 2019a,b; Piepho 2019). The mathematical definitions and interpretations of these parameters are analogous to heritability (Robertson 1959; Piepho and Möhring 2007; Benfey and Mitchell-Olds 2008; Sun *et al.* 2010; Isik *et al.* 2017; Schmidt *et al.* 2019a,b). We suspect that biases similar to those described for marker heritability could be hidden in the estimators of these parameters, a problem that warrants further study. The approach we described appears to have applications in fields of study other than genetics to obtain unbiased estimates of the relevant parameters.

Abbreviations

AMV: average marginal variance; ANOVA: analysis of variance; ASV: average semi-variance; GEBV: genome estimates breeding value GWAS: genome-wide association study; LD: linkage disequilibirum; MAS: marker assisted selection; QTL: quantitative trait locus; REML: restricted maximum likelihood; SNP: single nucleotide polymorphism; WGR: whole-genome regression.

Author's Contributions

MJF, WCB, and SJK conceived the overall project. MJF performed the analyses. MJF, HPP, WCB, and SJK wrote the paper.

Competing Interests

The authors declare no competing interests.

Acknowledgements

We thank Duke Pauli for comments and advice and for reviewing this manuscript.

Funding

This research was supported by grants to SJK from the United Stated Department of Agriculture (http://dx.doi.org/10.13039/100000199) National Institute of Food and Agriculture (NIFA) Specialty Crops Research Initiative (# 2017-51181-26833) and California Strawberry Commission (http://dx.doi.org/10.13039/100006760), in addition to funding from the University of California, Davis (http://dx.doi.org/10.13039/100007707).

Literature Cited

Bates, D., M. Mächler, B. Bolker, and S. Walker, 2015 Fitting linear mixed-effects models using lme4. Journal of Statistical Software **67**: 1–48.

Beavis, W., O. Smith, D. Grant, and R. Fincher, 1994 Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Science **34**: 882–896.

Benfey, P. N. and T. Mitchell-Olds, 2008 From genotype to phenotype: systems biology meets natural variation. Science **320**: 495–497.

Bernardo, R., 2004 What proportion of declared QTL in plants are false? Theoretical and Applied Genetics **109**: 419–424.

Bernardo, R., 2008 Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Science **48**: 1649–1664.

Bernardo, R., 2016 Bandwagons I, too, have known. Theoretical and Applied Genetics **129**: 2323–2332.

Burton, A., D. G. Altman, P. Royston, and R. L. Holder, 2006 The design of simulation studies in medical statistics. Statistics in Medicine **25**: 4279–4292.

Churchill, G. A., 2016 Eric Lander and David Botstein on mapping quantitative traits. Genetics **203**: 1.

Cockram, J. and I. Mackay, 2018 Genetic mapping populations for conducting high-resolution trait mapping in plants. In *Plant Genetics and Molecular Biology*, pp. 109–138, Springer.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, *et al.*, 2017 Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science **22**: 961–975.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams, 2010 The impact of genetic architecture on genome-wide evaluation methods. Genetics **185**: 1021–1031.

Damesa, T. M., J. Hartung, M. Gowda, Y. Beyene, B. Das, *et al.*, 2019 Comparison of weighted and unweighted stage-wise analysis for genome-wide association studies and genomic selection. Crop Science **59**: 2572–2584.

Damesa, T. M., J. Möhring, M. Worku, and H.-P. Piepho, 2017 One step at a time: stage-wise analysis of a series of experiments. Agronomy Journal **109**: 845–857.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. Calus, 2013 Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics **193**: 327–345.

de los Campos, G., D. Sorensen, and D. Gianola, 2015 Genomic heritability: what is it? PLoS Genetics **11**: e1005048.

Estaghvirou, S. B. O., J. O. Ogutu, T. Schulz-Streeck, C. Knaak, M. Ouzunova, *et al.*, 2013 Evaluation of approaches for esti-

mating the accuracy of genomic prediction in plant breeding. BMC Genomics **14**: 860.

Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009 Genomic selection for crop improvement. Crop Science **49**: 1–12.

Henderson, C. R., 1953 Estimation of variance and covariance components. Biometrics **9**: 226–252.

Hill, W. G., 2012 Quantitative genetics in the genomics era. Current Genomics **13**: 196–206.

Hill, W. G., 2014 Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. Genetics **196**: 1–16.

Hill, W. G. and F. Nicholas, 1974 Estimation of heritability by both regression of offspring on parent and intra-class correlation of sibs in one experiment. Biometrics **3**: 447–468.

Huang, W. and T. F. C. Mackay, 2016 The genetic architecture of quantitative traits cannot be inferred from variance component analysis. PLoS Genetics **12**: e1006421.

Isik, F., J. Holland, and C. Maltecca, 2017 *Genetic data analysis for plant and animal breeding*. Springer.

Lande, R. and R. Thompson, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124**: 743–756.

Lander, E. S. and D. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185–199.

Legarra, A., I. Aguilar, and I. Misztal, 2009 A relationship matrix including full pedigree and genomic information. Journal of Dairy Science **92**: 4656–4663.

Lello, L., S. G. Avery, L. Tellier, A. I. Vazquez, G. de los Campos, *et al.*, 2018 Accurate genomic prediction of human height. Genetics **210**: 477–497.

Lello, L., T. G. Raben, S. Y. Yong, L. C. Tellier, and S. D. Hsu, 2019 Genomic prediction of complex disease risk. bioRxiv p. 506600.

Lynch, M. and B. Walsh, 1998 *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA.

Mackay, T. F. C., 2001 The genetic architecture of quantitative traits. Annual Review of Genetics **35**: 303–339.

Mackay, T. F. C., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10**: 565.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, *et al.*, 2009 Finding the missing heritability of complex diseases. Nature **461**: 747.

Meuwissen, T., B. Hayes, and M. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics **157**: 1819–1829.

Morris, T. P., I. R. White, and M. J. Crowther, 2019 Using simulation studies to evaluate statistical methods. Statistics in Medicine **38**: 2074–2102.

Neimann-Sorensen, A. and A. Robertson, 1961 The association between blood groups and several production characteristics in three danish cattle breeds. Acta Agriculturae Scandinavica **11**: 163–196.

Piepho, H.-P., 2009 Ridge regression and extensions for genomewide selection in maize. Crop Science **49**: 1165–1176.

Piepho, H.-P., 2019 A coefficient of determination (R^2) for generalized linear mixed models. Biometrical Journal **61**: 860–872.

Piepho, H.-P., J. Moehring, T. Schulz-Streeck, and J. O. Ogutu, 2012 A stage-wise approach for the analysis of multi-environment trials. Biometrical Journal **54**: 844–860.

Piepho, H.-P. and J. Möhring, 2007 Computing heritability and selection response from unbalanced plant breeding trials. Genetics **177**: 1881–1888.

Pincot, D. D., T. J. Poorten, M. A. Hardigan, J. M. Harshman, C. B. Acharya, *et al.*, 2018 Genome-wide association mapping uncovers fw1, a dominant gene conferring resistance to fusarium wilt in strawberry. G3: Genes, Genomes, Genetics **8**: 1817–1828.

R Core Team, 2019 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rice, B. and A. E. Lipka, 2019 Evaluation of rr-blup genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. The Plant Genome **12**.

Robertson, A., 1959 The sampling variance of the genetic correlation coefficient. Biometrics **15**: 469–485.

Rutkoski, J. E., J. A. Poland, R. P. Singh, J. Huerta-Espino, S. Bhavani, *et al.*, 2014 Genomic selection for quantitative adult plant stem rust resistance in wheat. The Plant Genome **7**.

Schmidt, P., J. Hartung, J. Bennewitz, and H.-P. Piepho, 2019a Heritability in plant breeding on a genotype-difference basis. Genetics **212**: 991–1008.

Schmidt, P., J. Hartung, J. Rath, and H.-P. Piepho, 2019b Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Science **59**: 525–536.

Searle, S. R., 1995 An overview of variance component estimation. Metrika **42**: 215–230.

Searle, S. R. and M. H. Gruber, 1971 *Linear models*. Wiley Online Library.

Sun, G., C. Zhu, M. Kramer, S. Yang, W. Song, *et al.*, 2010 Variation explained in mixed-model association mapping. Heredity **105**: 333–340.

Tang, S., A. Leon, W. C. Bridges, and S. J. Knapp, 2006 Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Science **46**: 721–734.

Utz, H. F., A. E. Melchinger, and C. C. Schön, 2000 Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics **154**: 1839–1849.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. Journal of Dairy Science **91**: 4414–4423.

Visscher, P. M., M. A. Brown, M. I. McCarthy, and J. Yang, 2012 Five years of GWAS discovery. The American Journal of Human Genetics **90**: 7–24.

Visscher, P. M., W. G. Hill, and N. R. Wray, 2008 Heritability in the genomics era—concepts and misconceptions. Nature Reviews Genetics **9**: 255–266.

Visscher, P. M., S. E. Medland, M. A. Ferreira, K. I. Morley, G. Zhu, *et al.*, 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genetics **2**.

Visscher, P. M., N. R. Wray, and C. S. Haley, 2019 "arte et labore"—a blackburn rovers fan's legacy in human complex trait genetics. Journal of Animal Breeding and Genetics **136**: 273–278.

Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, *et al.*, 2017 10 years of GWAS discovery: biology, function, and translation. The American Journal of Human Genetics **101**: 5–22.

Visscher, P. M. and J. Yang, 2016 A plethora of pleiotropy across

complex traits. Nature Genetics **48**: 707–708.
Walsh, B., 2001 Quantitative genetics in the age of genomics. Theoretical Population Biology **59**: 175–184.
Wang, M. and S. Xu, 2019 Statistical power in genome-wide association studies and quantitative trait locus mapping. Heredity **123**: 287–306.
Xu, S., 2003 Theoretical basis of the Beavis effect. Genetics **165**: 2259–2268.

Appendices

Appendix A. ASV Estimators of Marker-Associated Genetic Variances for Two Loci With Balanced Data

Here we develop the ASV estimator of the variance explained by markers for an experiment where n_G entries are phenotyped for a normally distributed quantitative trait in a completely randomized experiment design, the data are balanced, and the variance components for the random effects of two marker loci ($M1$ and $M2$) are estimated using LMM (2):

$$\begin{aligned} y_{hijk} =& \mu + M1_h + M2_i + (M1 \times M2)_{hi} \\ &+ G : (M1 \times M2)_{hij} + \epsilon_{hijk} \end{aligned} \tag{20}$$

where y_{hijk} is the *hijk-th* phenotypic observation, μ is the population mean, h = 1, 2, or 3, i = 1, 2, or 3, j = 1, 2, ..., n_G, k = 1, 2, ..., r_G, $M1_h$ is the random effect of marker locus 1 with $var(M1_h) = \sigma^2_{M1}$, $M2_i$ is the random effect of marker locus 2 with $var(M2_i) = \sigma^2_{M2}$, $(M1 \times M2)_{hi}$ is the random effect of the interaction between marker loci 1 and 2 with $var(M1 \times M2_{hi}) = \sigma^2_{M1\times M2}$, $G : (M1 \times M2_{hij})$ is the random effect of entries nested in marker loci with $var(G : M1 \times M2_{hij}) = \sigma^2_{G:M1\times M2}$ and ϵ_{hijk} is the residual effect with $var(\epsilon_{hijk}) = \sigma^2_\epsilon$.

The entry-mean or ASV model for the analysis of two loci is:

$$\begin{aligned} \bar{y}_{hij\bullet} =& \mu + M1_h + M2_i + (M1 \times M2)_{hi} \\ &+ G : (M1 \times M2)_{hij} + \bar{\epsilon}_{hij\bullet} \end{aligned}$$

where $\bar{y}_{hij\bullet}$ is the entry-mean, $\bar{\epsilon}_{hij\bullet}$ is the residual with $var(\bar{\epsilon}_{hij\bullet}) = r_G^{-1}\sigma^2_\epsilon$, and the other variables are as defined for (20).

The variance explained by the marker effect (u_{M1}) using the ASV approach (Piepho 2019) is:

$$\begin{aligned} \hat{\theta}^{ASV}_{M1} &= (n_G - 1)^{-1}\hat{\sigma}^2_{M1} tr(Z_{u_{M1}} Z^T_{u_{M1}} P) \\ &= \frac{(n_{M1} - 1)n_{G:M1\times M2} n_{M1}}{n_G - 1}\hat{\sigma}^2_{M1} \\ &= \frac{df_{M1} n_{G:M1\times M2} n_{M1}}{df_G}\hat{\sigma}^2_{M1} \end{aligned} \tag{21}$$

where $P = I_{n_G} - n_G^{-1} J_{n_G}$, $Z_{u_{M1}} = (I_{n_{M1}} \otimes 1_{n_{M1}}) \otimes 1_{n_{G:M1\times M2}}$. Substituting the appropriate values for $M2$ into (21) yields $\hat{\theta}^{ASV}_{M2}$. Similarly, the variance explained by the interaction between marker loci 1 and 2 ($u_{M1\times M2}$) using the ASV approach is:

$$\begin{aligned} \hat{\theta}^{ASV}_{M1\times M2} &= (n_G - 1)^{-1}\hat{\sigma}^2_{M1\times M2} tr(Z_{u_{M1\times M2}} Z^T_{u_{M1\times M2}} P) \\ &= \frac{(n_{M1\times M2} - 1)n_{G:M1\times M2}}{n_G - 1}\hat{\sigma}^2_{M1\times M2} \\ &= \frac{(df_{M1\times M2} + df_{M1} + df_{M2})n_{G:M1\times M2}}{df_G}\hat{\sigma}^2_{M1\times M2} \end{aligned} \tag{22}$$

where $Z_{u_{M1\times M2}} = I_{n_{M1\times M2}} \otimes 1_{n_{G:M1\times M2}}$.

Finally, the variance explained by the random effect of entries nested in marker loci ($u_{G:M1\times M2}$) is:

$$\begin{aligned} \hat{\theta}^{ASV}_{G:M1\times M2} &= (n_G - 1)^{-1}\hat{\sigma}^2_{G:M1\times M2} tr(Z_{u_{G:M1\times M2}} Z^T_{u_{G:M1\times M2}} P) \\ &= \frac{n_G - 1}{n_G - 1}\hat{\sigma}^2_{G:M1\times M2} \\ &= \hat{\sigma}^2_{G:M1\times M2} \end{aligned}$$

From (21), the coefficient for bias-correcting AMV estimates of σ^2_{M1} (the genetic variance explained by $M1$) is:

$$k_{M1} = \frac{df_{M1} n_{G:M1\times M2} n_{M1}}{df_G}$$

Note that this definition of k_{M1} is equivalent to the definitions in Table 3 with r_G factored out. Similarly, from (22), the coefficient for bias-correcting AMV estimates of $\sigma^2_{M1\times M2}$ (the genetic variance explained by the interaction between $M1$ and $M2$) is:

$$k_{M1\times M2} = \frac{(df_{M1\times M2} + df_{M1} + df_{M2})n_{G:M1\times M2}}{df_G}$$

This definition of $k_{M1\times M2}$ is equivalent to the definitions in Table 3 with r_G factored out.

The ASV definition of the total genetic variance is:

$$\theta^{ASV}_G = \theta^{ASV}_{M1} + \theta^{ASV}_{M2} + \theta^{ASV}_{M1\times M2} + \theta^{ASV}_{G:M1\times M2} \tag{23}$$

The ANOVA estimators of the variance components for LMM (1) with balanced data are previously defined in (3) and (4). The ANOVA estimators of σ^2_{M1}, σ^2_{M2}, $\sigma^2_{M1\times M2}$, and $\sigma^2_{G:M1\times M2}$, for LMM (20) are:

$$\hat{\sigma}^2_{M1} = \frac{MS_{M1} - MS_{M1\times M2}}{r_G n_{G:M1\times M2} n_{M1}} \tag{24}$$

and

$$\hat{\sigma}^2_{M2} = \frac{MS_{M2} - MS_{M1\times M2}}{r_G n_{G:M1\times M2} n_{M2}} \tag{25}$$

and

$$\hat{\sigma}^2_{M1\times M2} = \frac{MS_{M1\times M2} - MS_{G:M1\times M2}}{r_G n_{G:M1\times M2}} \tag{26}$$

and

$$\hat{\sigma}^2_{G:M1\times M2} = \frac{MS_{G:M1\times M2} - MS_\epsilon}{r_G} \tag{27}$$

where MS_{M1} is the expected mean sum of squares for marker locus 1, MS_{M2} is the expected mean sum of squares for marker locus 2, and so on for the other expected mean sum of squares from the Table 1.

Thus, the total genetic variance can be estimated by substituting the bias-corrected ANOVA estimators (24 - 27) into (23):

$$
\begin{aligned}
\hat{\theta}_G^{ASV} &= \hat{\theta}_{M1}^{ASV} + \hat{\theta}_{M2}^{ASV} + \hat{\theta}_{M1\times M2}^{ASV} + \hat{\theta}_{G:M1\times M2}^{ASV} \\
&= k_{M1}\hat{\sigma}_{M1}^2 + k_{M2}\hat{\sigma}_{M2}^2 + k_{M1\times M2}\hat{\sigma}_{M1\times M2}^2 \\
&\quad + \hat{\sigma}_{G:M1\times M2}^2
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{df_{M1}}{df_G r_G}[MS_{M1} - MS_{M1\times M2}) \\
&\quad + \frac{df_{M2}}{df_G r_G}(MS_{M2} - MS_{M1\times M2}) \\
&\quad + \frac{df_{M1\times M2} + df_{M1} + df_{M2}}{df_G r_G} \\
&\quad \times (MS_{M1\times M2} - MS_{G:M1\times M2}) \\
&\quad + \frac{1}{r_G}(MS_{G:M1\times M2} - MS_\epsilon)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{r_G}MS_G - \frac{df_G}{df_G r_G}MS_\epsilon \\
&= \frac{MS_G - MS_\epsilon}{r_G} \\
&= \hat{\sigma}_G^2
\end{aligned}
$$

When the entries are adjusted means (BLUEs), the ASV definition of marker heritability is:

$$
\hat{H}_{M1*}^2 = \frac{(\frac{df_{M1} n_{G:M1\times M2} n_{M1}}{df_G})\hat{\sigma}_{M1}^2}{\hat{\sigma}_G^2 + \hat{\sigma}_\epsilon^2/r_G} \tag{28}
$$

Substituting the appropriate values for $M2$ into (28) yields the bias-corrected estimates (H_{M2*}^2).

Appendix B. ASV Estimator of the Marker-Associated Genetic Variance for One Locus with Unbalanced Data

ASV estimators of σ_{M1}^2 and $\sigma_{G:M1}^2$ are developed here for unbalanced data. As before, the phenotypic observations are entry-means ($\bar{y}_{hi\bullet}$) and the LMM for the entry-mean analysis is (9). In vector notation, (9) can be written as:

$$
\bar{y} = 1_n\mu + (\oplus_h 1_{n_{G:M1_h}})u_{M1} + I_{n_G}u_{G:M1} + \bar{\epsilon}
$$

where $n_{G:M1_h}$ is the number of entries nested in the hth marker genotype, $n_G = \sum_h n_{G:M1_h}$, and observations are sorted by markers and entries within markers. The vectors u_{M1} and $u_{G:M1}$ hold the random effects $M1_h$ and $G : M1_{hi}$ with $var(u_{M1}) = I_{n_{M1}}\sigma_{M1}^2$ and $var(u_{G:M1}) = I_{n_G}\sigma_{G:M1}^2$, respectively. The proportion of the genetic variance explained by the marker locus is:

$$
\begin{aligned}
\hat{\theta}_{M1}^{ASV} &= (n_G - 1)^{-1}\hat{\sigma}_{M1}^2 tr(Z_{u_{M1}}Z_{u_{M1}}^T P) \\
&= \frac{n_G - n_G^{-1}\sum_h n_{G:M1_h}^2}{df_G}\hat{\sigma}_{M1}^2
\end{aligned}
$$

where $P = I_{n_G} - n_G^{-1}J_{n_G}$, $Z_{u_{M1}} = \oplus_h 1_{n_{G:M1_h}}$ and $SS_{n_{G:M1}} = \sum_h (n_{G:M1_h} - \bar{n}_{G:M1_\bullet})^2$. For the balanced case, $SS_{n_{G:M1}} = 0$ and $n_G/n_{M1} = n_{G:M1}$. The residual genetic variance associated with the entry nested nested in $M1$ genotypes is:

$$
\begin{aligned}
\hat{\theta}_{G:M1}^{ASV} &= (n_G - 1)^{-1}\hat{\sigma}_{G:M1}^2 tr(Z_{u_{G:M^1}}Z_{u_{G:M1}}^T P) \\
&= \frac{n_G - 1}{n_G - 1}\hat{\sigma}_{G:M1}^2 \\
&= \hat{\sigma}_{G:M1}^2
\end{aligned}
$$

Hence, the general form for k_{M1} is:

$$
k_{M1} = \frac{n_G - n_G^{-1}\sum_h n_{G:M1_h}^2}{df_G}
$$

Appendix C. ASV Estimator of the Marker-Associated Genetic Variance for Two Loci with Unbalanced Data

ASV estimators for an experiment with two loci are developed here for unbalanced data. As before, the phenotypic observations are entry-means ($\bar{y}_{hij\bullet}$) and the LMM for the entry-mean analysis is (20). In vector notation, (20) can be written as:

$$
\begin{aligned}
\bar{y} &= 1_n\mu + (\oplus_h 1_{n_{G:M1_h}})u_{M1} + (\oplus_i 1_{n_{G:M2_i}})u_{M2} \\
&\quad + (\oplus_{hi} 1_{n_{M1_h\times M2_i}})u_{M1\times M2} + I_{n_G}u_{G:M1\times M2} + \bar{\epsilon}
\end{aligned}
$$

where $n_{G:M1_h\times M2_i}$ is the number of entries nested in the hi^{th} marker genotype, $n_G = \sum_{hi} n_{G:M1_h\times M2_i}$, and observations are sorted by markers and entries within markers. The vectors u_{M1}, u_{M2}, $u_{M1\times M2}$ and $u_{G:M1\times M2}$ hold the random effects $M1_h$, $M2_i$, $M1_h \times M2_i$ and $G : M1_h \times M2_{ij}$ with $var(u_{M1}) = I_{n_{M1}}\sigma_{M1}^2$, $var(u_{M2}) = I_{n_{M2}}\sigma_{M2}^2$, $var(u_{M1\times M2}) = I_{n_{M1\times M2}}\sigma_{M1\times M2}^2$, and $var(u_{G:M1\times M2}) = I_{n_G}\sigma_{G:M1\times M2}^2$, respectively. The proportion of the genetic variance explained by the marker locus is:

$$
\begin{aligned}
\hat{\theta}_{M1}^{ASV} &= (n_G - 1)^{-1}\hat{\sigma}_{M1}^2 tr(Z_{u_{M1}}Z_{u_{M1}}^T P) \\
&= \frac{n_G - n_G^{-1}\sum_h n_{G:M1_h}^2}{df_G}\hat{\sigma}_{M1}^2
\end{aligned} \tag{29}
$$

where $P = I_{n_G} - n_G^{-1}J_{n_G}$, $Z_{u_{M1}} = (I_{n_{M1}} \otimes 1_{n_{M1}}) \otimes 1_{n_{G:M1\times M2}}$. Substituting the appropriate values for $M2$ into (29) yields θ_{M2}^{ASV}. Similarly, the variance explained by the interaction between marker loci 1 and 2 ($u_{M1\times M2}$) using the ASV approach is:

$$
\begin{aligned}
\hat{\theta}_{M1\times M2}^{ASV} &= (n_G - 1)^{-1}\hat{\sigma}_{M1\times M2}^2 tr(Z_{u_{M1\times M2}}Z_{u_{M1\times M2}}^T P) \\
&= \frac{n_G - n_G^{-1}\sum_{hi} n_{G:M1_h\times M2_i}^2}{df_G}\hat{\sigma}_{M1\times M2}^2
\end{aligned} \tag{30}
$$

where $Z_{u_{M1\times M2}} = I_{n_{M1\times M2}} \otimes 1_{n_{G:M1\times M2}}$.

Finally, the variance explained by the random effect of entries nested in marker loci ($u_{G:M1\times M2}$) is:

$$
\begin{aligned}
\hat{\theta}_{G:M1\times M2}^{ASV} &= (n_G - 1)^{-1}\hat{\sigma}_{G:M1\times M2}^2 tr(Z_{u_{G:M1\times M2}}Z_{u_{G:M1\times M2}}^T P) \\
&= \frac{n_G - 1}{n_G - 1}\hat{\sigma}_{G:M1\times M2}^2 \\
&= \hat{\sigma}_{G:M1\times M2}^2
\end{aligned}
$$

From (29), the coefficient for bias-correcting AMV estimates of σ_{M1}^2 (the genetic variance explained by $M1$) is:

$$k_{M1} = \frac{n_G - n_G^{-1}\sum_h n^2_{G:M1_h}}{df_G}$$

Similarly, from (30), the coefficient for bias-correcting AMV estimates of $\sigma^2_{M1\times M2}$ (the genetic variance explained by the interaction between $M1$ and $M2$) is:

$$k_{M1\times M2} = \frac{n_G - n_G^{-1}\sum_{hi} n^2_{G:M1_h\times M2_i}}{df_G}$$

Appendix D. ASV Estimator of the Marker-Associated Genetic Variance for Three Loci with Unbalanced Data

ASV estimators for an experiment with three loci are developed here for unbalanced data. The phenotypic observations are entry-means ($\bar{y}_{hij\bullet}$). In vector notation, the LMM for the entry-mean analysis is:

$$\begin{aligned}\bar{y} &= 1_n\mu + (\oplus_h 1_{n_{G:M1_h}})u_{M1}\\ &+ (\oplus_i 1_{n_{G:M2_i}})u_{M2} + (\oplus_j 1_{n_{G:M3_j}})u_{M3}\\ &+ (\oplus_{hi} 1_{n_{M1_h\times M2_i}})u_{M1\times M2} + (\oplus_{hj} 1_{n_{M1_h\times M3_j}})u_{M1\times M3}\\ &+ (\oplus_{ij} 1_{n_{M2_i\times M3_j}})u_{M2\times M3} + (\oplus_{hij} 1_{n_{M1_h\times M2_i\times M3_j}})u_{M1\times M2\times M3}\\ &+ I_{n_G} u_{G:M1\times M2\times M3} + \bar{\epsilon}\end{aligned}$$

where $n_{G:M1_h\times M2_i\times M3_j}$ is the number of entries nested in the hij^{th} marker genotype, $n_G = \sum_{hij} n_{G:M1_h\times M2_i\times M3_j}$, and observations are sorted by markers and entries within markers. Vectors u hold the random effects for the appropriate random effects of marker loci, marker-marker interations, and entries nested in marker loci as in Appendix B and C. The proportion of the genetic variance explained by the marker locus is:

$$\begin{aligned}\hat{\theta}^{ASV}_{M1} &= (n_G-1)^{-1}\hat{\sigma}^2_{M1} tr(Z_{u_{M1}} Z^T_{u_{M1}} P) \qquad (31)\\ &= \frac{n_G - n_G^{-1}\sum_h n^2_{G:M1_h}}{df_G}\hat{\sigma}^2_{M1}\end{aligned}$$

where $P = I_{n_G} - n_G^{-1} J_{n_G}$, $Z_{u_{M1}} = (I_{n_{M1}} \otimes 1_{n_{M1}}) \otimes 1_{n_{G:M1\times M2}}$. Substituting the appropriate values for $M2$ into (31) yields θ^{ASV}_{M2}. Similarly, the variance explained by the interaction between marker loci 1 and 2 ($u_{M1\times M2}$) using the ASV approach is:

$$\begin{aligned}\hat{\theta}^{ASV}_{M1\times M2} &= (n_G-1)^{-1}\hat{\sigma}^2_{M1\times M2} tr(Z_{u_{M1\times M2}} Z^T_{u_{M1\times M2}} P) \qquad (32)\\ &= \frac{n_G - n_G^{-1}\sum_{hi} n^2_{G:M1_h\times M2_i}}{df_G}\hat{\sigma}^2_{M1\times M2}\end{aligned}$$

where $Z_{u_{M1\times M2}} = I_{n_{M1\times M2}} \otimes 1_{n_{G:M1\times M2}}$. The variance explained by the three-way marker interaction ($u_{M1\times M2\times M3}$) using the ASV approach is:

$$\begin{aligned}&\hat{\theta}^{ASV}_{M1\times M2\times M3} = \qquad (33)\\ &\quad (n_G-1)^{-1}\hat{\sigma}^2_{M1\times M2\times M3} tr(Z_{u_{M1\times M2\times M3}} Z^T_{u_{M1\times M2\times M3}} P)\\ &\quad = \frac{n_G - n_G^{-1}\sum_{hij} n^2_{G:M1_h\times M2_i\times M3_j}}{df_G}\hat{\sigma}^2_{M1\times M2\times M3}\end{aligned}$$

where $Z_{u_{M1\times M2\times M3}} = I_{n_{M1\times M2\times M3}} \otimes 1_{n_{G:M1\times M2\times M3}}$.

Finally, the variance explained by the random effect of entries nested in marker loci ($u_{G:M1\times M2\times M3}$) is:

$$\begin{aligned}&\hat{\theta}^{ASV}_{G:M1\times M2\times M3} =\\ &\quad (n_G-1)^{-1}\hat{\sigma}^2_{G:M1\times M2\times M3} tr(Z_{u_{G:M1\times M2\times M3}} Z^T_{u_{G:M1\times M2\times M3}} P)\\ &\quad = \frac{n_G-1}{n_G-1}\hat{\sigma}^2_{G:M1\times M2\times M3}\\ &\quad = \hat{\sigma}^2_{G:M1\times M2\times M3}\end{aligned}$$

From (31), the coefficient for bias-correcting AMV estimates of σ^2_{M1} (the genetic variance explained by $M1$) is:

$$k_{M1} = \frac{n_G - n_G^{-1}\sum_h n^2_{G:M1_h}}{df_G}$$

Similarly, from (32), the coefficient for bias-correcting AMV estimates of $\sigma^2_{M1\times M2}$ (the genetic variance explained by the interaction between $M1$ and $M2$) is:

$$k_{M1\times M2} = \frac{n_G - n_G^{-1}\sum_{hi} n^2_{G:M1_h\times M2_i}}{df_G}$$

From (33), the coefficient for bias-correcting AMV estimates of $\sigma^2_{M1\times M2\times M3}$ (the genetic variance explained by the three-way marker loci interaction) is:

$$k_{M1\times M2\times M3} = \frac{n_G - n_G^{-1}\sum_{hij} n^2_{G:M1_h\times M2_i\times M3_j}}{df_G}\hat{\sigma}^2_{M1\times M2\times M3}$$

Table 1 Analyses of variance for LMMs (1) and (2) with balanced data where r_G replicates of n_G entries (individuals or families) are phenotyped for a normally distributed quantitative trait, replicates are arranged in a completely randomized experiment design, and parameters are estimated for one marker locus ($M1$) or two marker loci ($M1$ and $M2$). Sources of variation, degrees of freedom (df), mean squares (MS), and expected mean squares ($E[MS]$) are shown for ANOVA estimators of σ^2_G for LMM (1) and ANOVA estimators of σ^2_{M1} for marker locus $M1$, σ^2_{M2} for marker locus 2, and $\sigma^2_{M1\times M2}$ for the interaction between marker loci $M1$ and $M2$ using LMM (2). ANOVA breakdowns are shown for a single marker locus ($M1$) or two marker loci ($M1$ and $M2$). The residuals are identical for the three analyses.

Source	df^a	MS	E[MS]^b
Entry (G)	$df_G = n_G - 1$	$MS_G = \frac{SS_G}{df_G}$	$\sigma^2_\epsilon + r_G\sigma^2_G$
$M1$	$df_{M1} = n_{M1} - 1$	$MS_{M1} = \frac{SS_{M1}}{df_{M1}}$	$\sigma^2_\epsilon + r_G\sigma^2_{G:M1} + r_G n_{G:M1}\sigma^2_{M1}$
$G:M1$	$df_{G:M1} = df_G - df_{M1}$	$MS_{G:M1} = \frac{SS_{G:M1}}{df_{G:M1}}$	$\sigma^2_\epsilon + r_G\sigma^2_{G:M1}$
$M1$	$df_{M1} = n_{M1} - 1$	$MS_{M1} = \frac{SS_{M1}}{df_{M1}}$	$\sigma^2_\epsilon + r_G\sigma^2_{G:M1\times M2} + r_G n_{G:M1\times M2}\sigma^2_{M1\times M2}$ $+r_G n_{G:M1\times M2} n_{M1}\sigma^2_{M1}$
$M2$	$df_{M2} = n_{M2} - 1$	$MS_{M2} = \frac{SS_{M2}}{df_{M2}}$	$\sigma^2_\epsilon + r_G\sigma^2_{G:M1\times M2} + r_G n_{G:M1\times M2}\sigma^2_{M1\times M2}$ $+r_G n_{G:M1\times M2} n_{M2}\sigma^2_{M2}$
$M1\times M2$	$df_{M1\times M2} = (n_{M1}-1)(n_{M2}-1)$	$MS_{M1\times M2} = \frac{SS_{M1\times M2}}{df_{M1\times M2}}$	$\sigma^2_\epsilon + r_G\sigma^2_{G:M1\times M2} + r_G n_{G:M1\times M2}\sigma^2_{M1\times M2}$
$G:M1\times M2$	$df_{G:M1\times M2}$ $= df_G - (df_{M1} + df_{M2} + df_{M1\times M2})$	$MS_{G:M1\times M2} = \frac{SS_{G:M1\times M2}}{df_{G:M1\times M2}}$	$\sigma^2_\epsilon + r_G\sigma^2_{G:M1\times M2}$
Residual	$df_\epsilon = n_G(r_G - 1)$	$MS_\epsilon = \frac{SS_\epsilon}{df_\epsilon}$	σ^2_ϵ

[a] n_G = number of entries, n_{M1} = number of genotypes for marker locus 1, n_{M2} = number of genotypes for marker locus 2, $n_{G:M1}$ = number of entries nested in genotypes for marker locus 1 and 2, $n_{G:M2}$ = number of entries nested in genotypes for marker locus 1 and 2, $n_{G:M1\times M2}$ = number of entries nested in genotypes for marker locus 1 and 2, and n_y = number of independent observations.
[b] σ^2_G = genetic variance among entries, σ^2_{M1} = genetic variance explained by marker locus 1, σ^2_{M2} = genetic variance explained by marker locus 2, $\sigma^2_{M1\times M2}$ = genetic variance explained by the interaction between marker locus 1 and 2, $\sigma^2_{G:M1}$ = residual genetic variance among entries nested in marker locus 1, and $\sigma^2_{G:M1\times M2}$ = residual genetic variance among entries nested in marker locus 1 and 2.

Table 2 Simulation Study Designs and Variables. One thousand samples were simulated for each of 21 study designs with different marker heritabilities (H^2_M), n_M = 3 genotypes/marker locus, one to three marker loci (m), r_M replicates/marker genotype, n_G entries, r_G replicates/entry, and n total observations. The number of replications/marker genotype for study design 4 was equivalent to the expected number for the segregation of a co-dominant DNA marker in an F_2 population (1 AA : 2 Aa : 1 aa). The r_M, n_G, and r_G for study designs 5 and 6 varied because 10% and 33% of the data were randomly deleted (treated as missing).

Study Design	m	r_M	n_G	r_G	n	H^2_M Range
1	1	900	540	5	2,700	(0.0, 1.0)
2	2	900	540	5	2,700	(0.0, 1.0)
3	3	900	540	5	2,700	(0.0, 1.0)
4	1	675:1,350:675	540	5	2,700	(0.0, 1.0)
5	1	≤900	≤540	≤5	2,430	(0.0, 1.0)
6	1	≤900	≤540	≤5	1,800	(0.0, 1.0)
7	1	600	900	2	1,800	0.50
8	1	1,500	900	5	4,500	0.50
9	1	3,000	900	10	9,000	0.50
10	1	6,000	900	20	18,000	0.50
11	1	15,000	900	50	45,000	0.50
12	1	750	450	5	2,250	0.50
13	1	1,500	900	5	4,500	0.50
14	1	3,000	1,800	5	9,000	0.50
15	1	6,000	3,600	5	18,000	0.50
16	1	12,000	7,200	5	36,000	0.50
17	1	750	450	5	2,250	0.05
18	1	750	450	5	2,250	0.25
19	1	750	450	5	2,250	0.50
20	1	750	450	5	2,250	0.75
21	1	750	450	5	2,250	0.95

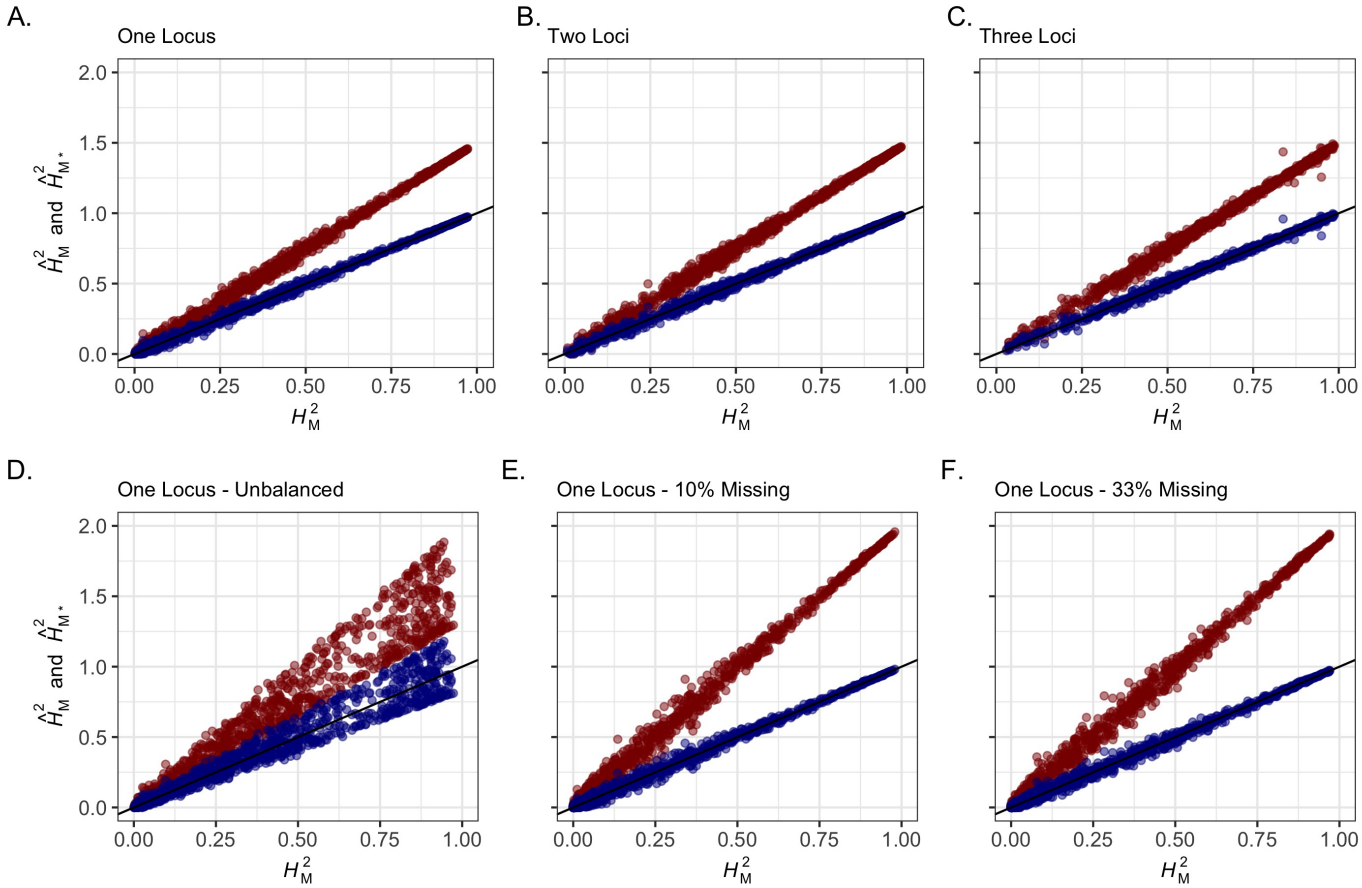

Figure 1 Accuracy of AMV and ASV Estimates of Marker Heritability. AMV and ASV estimates of H^2_M are shown for 1,000 segregating populations simulated for different numbers of entries (n_G individuals or families), five replications/entry (r_G = 5), true marker heritabilities ranging from 0 to 1, and one to three marker loci with three genotypes/marker locus (n_{M1} = 3). True marker heritabilities are shown on the x-axis. Uncorrected and bias-corrected estimates of H^2_M are shown on the y-axis, where $\hat{H}^2_M$ are the uncorrected AMV estimates (red highlighted observations), and $\hat{H}^2_{M*}$ are the bias-corrected ASV estimates (blue highlighted observations). AMV and ASV estimates of H^2_M from simulations are shown for: (A) one locus with balanced data for $n_G = 540$ entries (study design 1); (B) two marker loci ($M1$, $M2$, and $M1 \times M2$) with balanced data for $n_G = 540$ (study design 2); (C) three marker loci ($M1$, $M2$, $M3$, $M1 \times M2$, $M1 \times M3$, $M2 \times M3$, and $M1 \times M2 \times M3$) with balanced data for $n_G = 540$ (study design 3); (D) an F_2 population segregating 1:2:1 for one marker locus with $r_M = 675$ for both homozygotes, $r_M = 1,350$ for the heterozygote, and $n_G = 540$ (study design 4); (E) one locus for $n_G \leq 540$ entries where 10% of the phenotypic observations are randomly missing (the data are unbalanced) (study design 5); and (F) one locus for $n_G \leq 540$ entries where 33% of the phenotypic observations are randomly missing (the data are unbalanced) (study design 6).

Table 3 Coefficients (k_M) for calculating bias-corrected AMV estimates of marker-associated genetic variances (σ^2_M) for one to three marker loci ($M1$, $M2$, and $M3$), where r_G replicates of n_G entries are phenotyped, entries are genotyped for three marker loci ($M1$, $M2$, and $M3$), data are balanced, and parameters are estimated using ANOVA estimators for LMM (2). The k_M are functions of degrees of freedom (df) for marker effects, r_G, and the number of replications of marker genotypes (r_M).

Source	r_M [a]	df	Uncorrected Variance ($\hat{\sigma}^2$) [b]	Bias-Corrected Variance ($\hat{\sigma}^2_*$) [c]
$M1$	r_{M1}	df_{M1}	$\hat{\sigma}^2_{M1}$	$k_{M1}\hat{\sigma}^2_{M1} = \frac{df_{M1}r_{M1}}{df_G r_G}\hat{\sigma}^2_{M1}$
$M2$	r_{M2}	df_{M2}	$\hat{\sigma}^2_{M2}$	$k_{M2}\hat{\sigma}^2_{M2} = \frac{df_{M2}r_{M2}}{df_G r_G}\hat{\sigma}^2_{M2}$
$M3$	r_{M3}	df_{M3}	$\hat{\sigma}^2_{M3}$	$k_{M3}\hat{\sigma}^2_{M3} = \frac{df_{M3}r_{M3}}{df_G r_G}\hat{\sigma}^2_{M3}$
$M1 \times M2$	$r_{M1\times M2}$	$df_{M1\times M2}$	$\hat{\sigma}^2_{M1\times M2}$	$k_{M1\times M2}\hat{\sigma}^2_{M1\times M2} = \frac{(df_{M1}+df_{M2}+df_{M1\times M2})r_{M1\times M2}}{df_G r_G}\hat{\sigma}^2_{M1\times M2}$
$M1 \times M3$	$r_{M1\times M3}$	$df_{M1\times M3}$	$\hat{\sigma}^2_{M1\times M3}$	$k_{M1\times M3}\hat{\sigma}^2_{M1\times M3} = \frac{(df_{M1}+df_{M3}+df_{M1\times M3})r_{M1\times M3}}{df_G r_G}\hat{\sigma}^2_{M1\times M3}$
$M2 \times M3$	$r_{M2\times M3}$	$df_{M2\times M3}$	$\hat{\sigma}^2_{M2\times M3}$	$k_{M2\times M3}\hat{\sigma}^2_{M2\times M3} = \frac{(df_{M2}+df_{M3}+df_{M2\times M3})r_{M2\times M3}}{df_G r_G}\hat{\sigma}^2_{M2\times M3}$
$M1 \times M2 \times M3$	$r_{M1\times M2\times M3}$	$df_{M1\times M2\times M3}$	$\hat{\sigma}^2_{M1\times M2\times M3}$	$k_{M1\times M2\times M3}\hat{\sigma}^2_{M1\times M2\times M3} = \frac{(df_G - df_{G:M1\times M2\times M3})r_{M1\times M2\times M3}}{df_G r_G}\hat{\sigma}^2_{M1\times M2\times M3}$

[a] r_{M1}, r_{M2}, r_{M3}, $r_{M1\times M2}$, ..., $r_{M1\times M2\times M3}$ are the respective number of replicates of marker genotypes.
[b] $\hat{\sigma}^2_{M1}$, $\hat{\sigma}^2_{M2}$, $\hat{\sigma}^2_{M3}$, $\hat{\sigma}^2_{M1\times M2}$, ..., and $\hat{\sigma}^2_{M1\times M2\times M3}$ are the respective genetic variances for marker loci $M1$, $M2$, and $M3$ and interactions among them ($M1 \times M2$, $M1 \times M3$, $M2 \times M3$, and $M1 \times M2 \times M3$).
[c] df_G = degrees of freedom for entry.

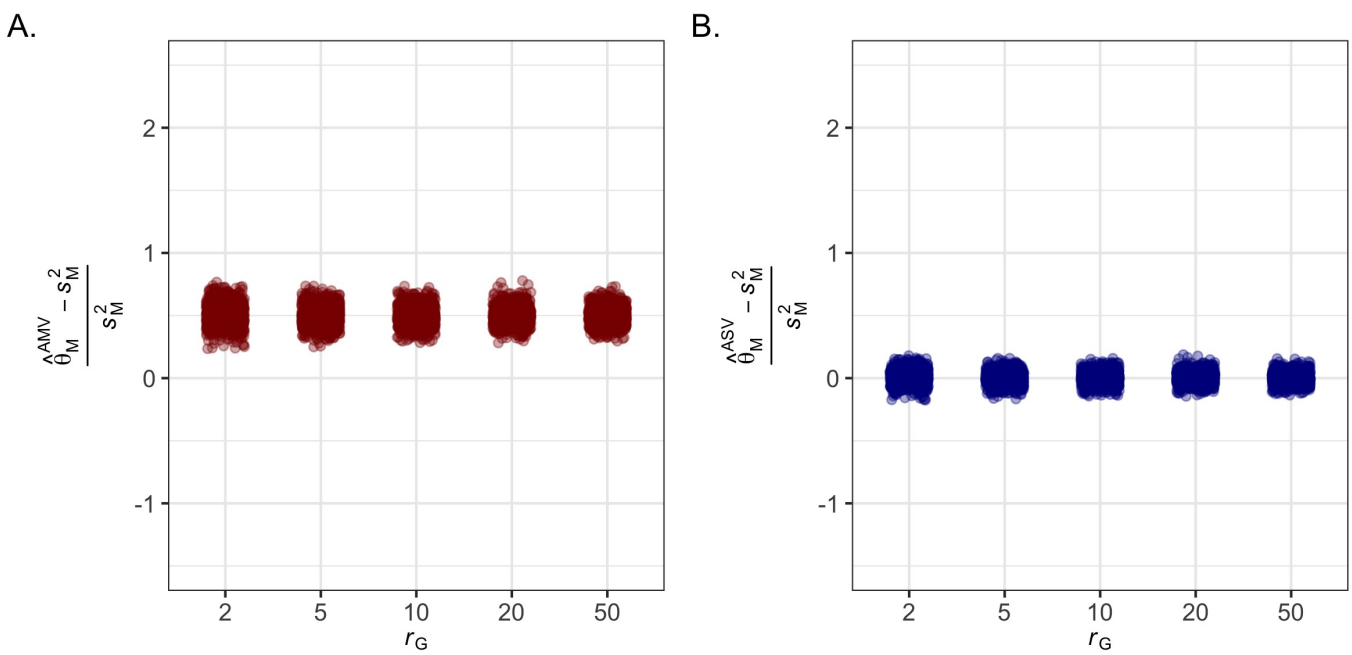

Figure 2 Effect of r_G on the Bias of AMV and ASV Estimates of σ^2_{M1}. Phenotypic observations were simulated for 1,000 populations segregating for a single marker locus with three genotypes ($n_{M1} = 3$), $n_G = 900$ progeny, and r_G = 2, 5, 10, 20, or 50 (study designs 7-11). The marker locus was assumed to be in complete linkage disequilibrium with a single QTL with marker heritabilities (H^2_{M1}) equal to 0.50. (A) Distribution of the relative biases of AMV estimates of σ^2_{M1} for different r_G. The relative bias ($RB[\theta^{AMV}_M] = 0.498$) was identical for different r_G. (B) Distribution of the relative biases of ASV estimates of σ^2_{M1} for different r_G. The relative bias ($RB[\theta^{ASV}_M] = 0.00$) was identical for different r_G.

Table 4 Uncorrected and bias-corrected estimates of marker-associated genetic variances (σ^2_M) and heritabilities (H^2_M) for three marker loci (B, P, and HYP) segregating in a sunflower recombinant inbred line (RIL) population ($n_G = 146$) with balanced data (Tang *et al.* 2006) and two marker loci (AX396 & AX493) segregating in a strawberry genome-wide association study (GWAS) population ($n_G = 565$) with unbalanced data (Pincot *et al.* 2018). The marker loci in the sunflower study were associated with QTL affecting seed oil content. The genetic variances associated with those loci were estimating using LMM (13). The marker loci in the strawberry study were associated with a locus ($Fw1$) affecting resistance to Fusarium wilt. The genetic variances associated with those loci were estimated using LMM (2). The coefficients (k_M) for bias-correcting AMV estimates of σ^2_M were found as described in Table 3 for balanced data and Appendix B for unbalanced data.

				Uncorrected		Bias-Corrected	
Case Study	Source	k_M	Variance Component	$\hat{\sigma}^2$	$\hat{H}^2$	$\hat{\sigma}^2_*$	$\hat{H}^2_*$
Tang *et al.* (2006)	Entry (G)	–	σ^2_G	21.61	0.95	21.61	0.95
	$M + G : M$	–	$\sigma^2_B + ... + \sigma^2_{G:B\times P\times HYP}$	30.76	1.35	22.15	0.98
	M	–	$\sigma^2_B + ... + \sigma^2_{B\times P\times HYP}$	17.85	0.79	9.24	0.41
	B	0.48	σ^2_B	11.57	0.51	5.59	0.25
	P	0.47	σ^2_P	1.26	0.06	0.60	0.03
	HYP	0.49	σ^2_{HYP}	2.9	0.13	1.41	0.06
	$B \times P$	0.77	$\sigma^2_{B\times P}$	0.21	0.01	0.17	0.01
	$B \times HYP$	0.78	$\sigma^2_{B\times HYP}$	1.89	0.08	1.46	0.06
	$P \times HYP$	0.77	$\sigma^2_{P\times HYP}$	0.00	0.00	0.00	0.00
	$B \times P \times HYP$	0.88	$\sigma^2_{B\times P\times HYP}$	0.00	0.00	0.00	0.00
	$G : B \times P \times HYP$	–	$\sigma^2_{G:B\times P\times HYP}$	12.91	0.57	12.91	0.57
	Residual (ϵ)	–	σ^2_ϵ	2.07	–	2.07	–
Pincot *et al.* (2018)	Entry (G)	–	σ^2_G	3.30	0.98	3.30	0.98
	$M + G : M$	–	$\sigma^2_{AX493} + \sigma^2_{G:AX493}$	4.01	1.20	3.45	1.03
	$AX493$ (M)	0.62	σ^2_{AX493}	1.48	0.44	0.93	0.27
	$G : AX493$	–	$\sigma^2_{G:AX493}$	2.53	0.75	2.53	0.75
	Residual (ϵ)	–	σ^2_ϵ	0.48	–	0.48	–
Pincot *et al.* (2018)	Entry (G)	–	σ^2_G	3.30	0.98	3.30	0.98
	$M + G : M$	–	$\sigma^2_{AX396} + \sigma^2_{G:AX396}$	4.77	1.42	2.39	0.71
	$AX396$ (M)	0.47	σ^2_{AX396}	4.47	1.33	2.09	0.62
	$G : AX396$	–	$\sigma^2_{G:AX396}$	0.30	0.09	0.30	0.09
	Residual (ϵ)	–	σ^2_ϵ	0.48	–	0.48	–

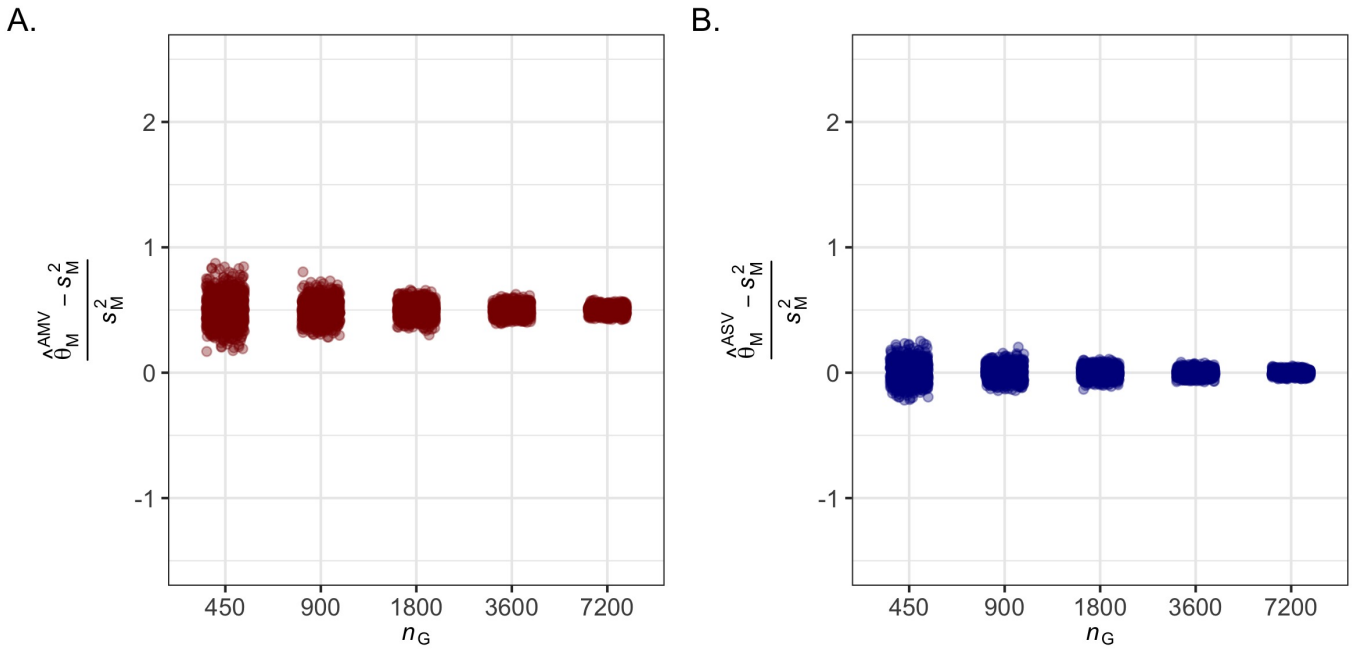

Figure 3 Effect of n_G on the Bias of AMV and ASV Estimates of σ^2_{M1}. Phenotypic observations were simulated for 1,000 populations segregating for a single marker locus with three genotypes ($n_{M1} = 3$), five replications/entry ($n_{M1} = 5$), and n_G = 450, 900, 1,800, 3,600, or 7,200 entries/population (study designs 12-16). The marker locus was assumed to be in complete linkage disequilibrium with a single QTL with marker heritabilities (H^2_{M1}) equat to 0.50. (A) Distribution of the relative biases of AMV estimates of σ^2_{M1} for different n_G. The relative bias ($RB[\theta_M^{AMV}] = 0.499$) was identical across the variables tested. (B) Distribution of the relative biases of ASV estimates of σ^2_{M1} for different n_G. The relative bias ($RB[\theta_M^{ASV}] = 0.00$) was identical across the variables tested.

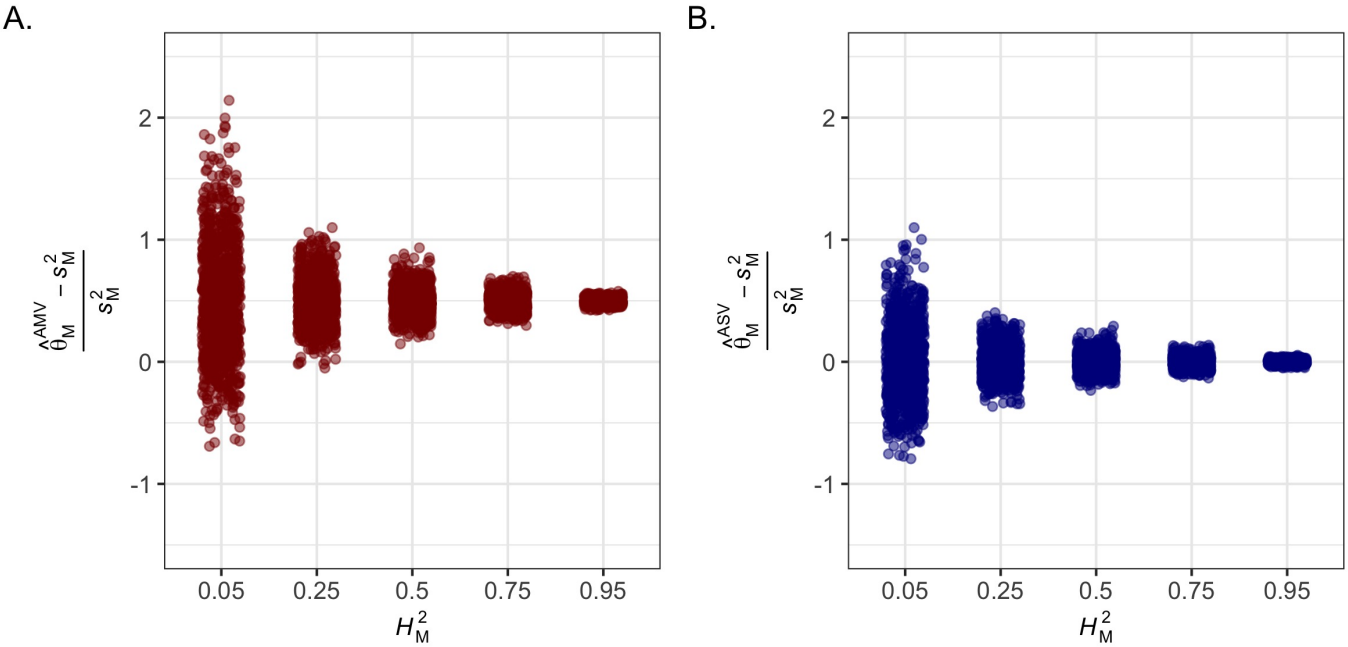

Figure 4 Effect of H^2_{M1} on the Bias of AMV and ASV Estimates of σ^2_{M1}. Phenotypic observations were simulated for 1,000 populations segregating for a single marker locus with three genotypes ($n_{M1} = 3$), five replications/entry ($n_{M1} = 5$), and $n_G = 450$ entries/population. The marker locus was assumed to be in complete linkage disequilibrium with a single QTL with marker heritabilities (H^2_{M1}) ranging from 0.05 to 0.95 (study designs 17-21). (A) Distribution of the relative biases of AMV estimates of σ^2_{M1} for different H^2_{M1}. The relative bias ($RB[\theta^{AMV}_M] = 0.496$) was identical across the variables tested. (B) Distribution of the relative biases of ASV estimates of σ^2_{M1} for different H^2_{M1}. The relative bias ($RB[\theta^{ASV}_M] = 0.0$) was identical across the variables tested.