Complex Traits and Candidate Genes: Estimation of Genetic Variance Components Across Modes of Inheritance

Mitchell J. Feldmann[1,*], Giovanny Covarrubias-Pazaran[2] and Hans-Peter Piepho[3]

[1]Department of Plant Sciences, University of California, Davis, One Shield Ave, Davis, 95616, California, United States of America
[2]International Maize and Wheat Improvement Center (CIMMYT), Km. 45, Carretera México-Veracruz, El Batán, 56130, Texcoco, Edo. de México, México
[3]Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Fruwirthstrasse 23, Stuttgart, 70599, Germany

*Corresponding author: mjfeldmann@ucdavis.edu

Abstract

Large-effect loci—those discovered by genome-wide association studies or linkage mapping—associated with key traits segregate amidst a background of minor, often undetectable genetic effects in both wild and domesticated plants and animals. Accurately attributing mean differences and variance explained to the correct components in the linear mixed model (LMM) analysis is important for both selecting superior progeny and parents in plant and animal breeding, but also for gene therapy and medical genetics in humans. Marker-assisted prediction (MAP) and its successor, genomic prediction (GP), have many advantages for selecting superior individuals and understanding disease risk. However, these two approaches are less often integrated to simultaneously study the modes of inheritance of complex traits. This simulation study demonstrates that the average semivariance can be applied to models incorporating Mendelian, oligogenic, and polygenic terms, simultaneously, and yields accurate estimates of the variance explained for all relevant terms. Our previous research focused on large-effect loci and polygenic variance exclusively, and in this work we want to synthesize and expand the average semivariance framework to a multitude of different genetic architectures and the corresponding mixed models. This framework independently accounts for the effects of large-effect loci and the polygenic genetic background and is universally applicable to genetics studies in humans, plants, animals, and microbes.

Keywords: Average semivariance, Linear mixed model, Variance component estimation, Polygenic inheritance, Oligogenic inheritance, Mendelian inheritance

Introduction

Today, LMMs are routinely applied in breeding and quantitative genetics research and are used for the prediction of genetic values in plants and animals (VanRaden 2008; Hayes *et al.* 2009; Albrecht *et al.* 2011; Endelman 2011; Crossa *et al.* 2014; Meuwissen *et al.* 2016), or polygenic risk scores (PRSs) in humans (de los Campos *et al.* 2010; Dudbridge 2013; Wray *et al.* 2019; Truong *et al.* 2020; de Los Campos *et al.* 2013; Lello *et al.* 2018, 2019), to estimate the heritability of traits in target populations (Visscher *et al.* 2006, 2008; de los Campos *et al.* 2015; Lehermeier *et al.* 2017; Legarra 2016), and to estimate ecological and evolutionary genetic parameters of behavioral traits (Walsh and Lynch 2018; Walsh *et al.* 2020; Oldroyd 2012; Hemani *et al.* 2013; Ariyomo *et al.* 2013). Genetic values are constructed from a combination of genetic effects; including Mendelian factors; which may have both additive effect and dominance deviations (Pincot *et al.* 2018, 2022), oligogenic factors consisting of few genetic factors and their epistatic interactions appropriate for marker-assisted prediction (MAP) (Tang *et al.* 2006), a polygenic term consisting of a dense genome-wide framework of markers assumed to have minor effects appropriate for genomic prediction (GP); which may also account of additive and dominance sources of variance (Pincot *et al.* 2020; Brandariz and Bernardo 2019), and a residual genetic term consisting of all genetic effects not accounted for by the previous genetic factors (Rutkoski *et al.* 2014; Rice and Lipka 2019; DeWitt *et al.* 2021). The ultimate objective in breeding applications is, typically, predicting the genotypic value, e.g., breeding value or genetic merit of a candidate individual (Knapp 1998; Piepho *et al.* 2008; Piepho 2009; VanRaden 2008; Luby and Shaw 2001; Collard and Mackill 2007). For loci to provide actionable gains or diagnoses, they must explain a significant proportion of phenotypic and genetic variation in a population with alleles in segregation at target loci.

Candidate gene discovery through genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping is prolific in plant and animal populations (Lander and Botstein 1989; Lander and Schork 1994; Visscher *et al.* 2012, 2017; Korte and Farlow 2013; Yu *et al.* 2006). Despite decades of directional selection in many plant populations, loci impacting traits of interest still segregate, even in advanced breeding materials, and these genome-wide analyses have succeeded in implicating numerous genes and genomic regions in the control of a wide variety of both simple and complex traits (Tang *et al.* 2006; Pincot *et al.* 2018; Wassom *et al.* 2008; Demmings *et al.* 2019a; Rutkoski *et al.* 2014; Rice and Lipka 2019; DeWitt *et al.* 2021; Han *et al.* 2018; Xin *et al.* 2020; Kim and Reinke 2019; Gage *et al.* 2020; Visscher *et al.* 2012, 2017; Andersson 2001; Hayes and Goddard 2001; Anderson *et al.* 2007; Septiningsih *et al.* 2009; Hayes *et al.* 2010; Saatchi *et al.* 2014; Seabury *et al.* 2017), although the utility of such marker-trait associations may not be fully realized (Bernardo 2004, 2016). Large-effect and statistically significant loci typically only explain a fraction of the genetic and phenotypic variance in a population (Feldmann *et al.*

2021), along with the polygenic fraction (Feldmann *et al.* 2022), except in extreme scenarios when Mendelian factors wholly control a trait.

Discovered loci rarely, if ever, explain 100% of the genetic variance, and understanding the multiple sources of variation and how they relate can help breeders and research prioritize targets and mitigate risk (Bernardo 2004, 2014). Genes with significant effects often dominate the 'non-missing heritability,' but they can also mask or obscure the effects of other quantitatively acting genes and pleiotropically affect multiple quantitative phenotypes (Mackay 2001; Mackay *et al.* 2009; Lorenz and Cohen 2012; De Villemereuil *et al.* 2018; Eichler *et al.* 2010). For example, mutations in the *BRCA2* gene can have large effects, but be incompletely penetrant, interact with other genes, and may be necessary but insufficient for predicting breast, ovarian, and other cancer risks in women (Gaudet *et al.* 2010). Accurately partitioning the Mendelian, oligogenic, and polygenic sources of variance allows researchers to assess how much value, or risk, specific loci confer.

Here, we use simulations to show that the ASV provides accurate variance component estimates (VCEs) and variance component ratios for all relevant genetic terms regardless study design or population type, e.g., outbred or inbred. We sought to marry the our previously published works (Piepho 2019; Feldmann *et al.* 2021, 2022) and to present a fully realized ASV approach for typical LMM analyses in human, plant, animal, and microbial genetics. We demonstrate how these models can be extended to handle more complex genetic structures, including adding multiple explanatory loci and marker-marker interactions, incorporating non-additive dominance and epistasis variance, and partitioning marker variance into additive and dominance components. We provide examples of expressing the different models and extensions in the freely available `sommer` R package (Covarrubias-Pazaran 2016). We believe that the average semivariance is a powerful tool for answering these questions regardless of the organism, population, or trait.

Linear mixed model analysis and the average semivariance

The average semivariance (ASV) estimator of total variance (Piepho 2019) and the variance of single markers and marker-marker interactions (Feldmann *et al.* 2021) is half the average total pairwise variance of a difference between entries and can be decomposed into independent sources of variance, e.g., genetic and residual. In this article, we assume that researchers are able to independently replicate entries—as in clonally propagated or inbred crop species—or can collect repeated measures on entries (e.g., individuals, families, or strains)—as in humans and animals—and then estimate the least square means (LSMs), best linear unbiased estimators (BLUEs), or other adjusted entry means in the first stage of a two-stage analysis (Piepho *et al.* 2012; Schulz-Streeck *et al.* 2013; Damesa *et al.* 2017, 2019).

The key idea here is that the adjust entry means, in general, are considered the "phenotype" since we assume independent replication. In animal breeding, "de-regressed" best linear unbiased predictors (BLUPs) are used in GBLUP and GWAS analysis (Strandén and Mäntysaari 2010; Ricard *et al.* 2013; Calus *et al.* 2016; Konstantinov and Goddard 2020). The two-stage approach is commonly applied for GWAS and GP studies in plants (Pincot *et al.* 2018, 2020; Damesa *et al.* 2017; Dias *et al.* 2020; Gogel *et al.* 2018). For simplicity in our demonstration, we assume that the error variance of the observation is $\mathbf{R} = \mathbf{I}_n\sigma^2_\epsilon$, where n is the number of entries (e.g., individuals, accessions, genotypes, lines, or animals). The more general approach is to assume a general variance-covariance matrix $\mathbf{R}$ and, importantly, the average semivariance can efficiently deal with more general forms of $\mathbf{R}$ and integrated directly into single-stage or multi-stage analyses. We explore ASV in a fully efficient two-stage analysis below in this article.

The form of the linear mixed model (LMM) for this analysis assuming only one explanatory marker is:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_m m + \mathbf{I}g + \mathbf{I}G_R + \boldsymbol{\epsilon} \tag{1}$$

where y is the vector of LSMs with $y \sim \mathcal{N}(\mu, \mathbf{V})$, μ is the population mean and the only fixed effect, m is the random effect of the main-effect locus with $m \sim \mathcal{N}(0, \mathbf{I}\sigma^2_m)$, g is the random additive genetic effect associated with the genome-wide framework of marker excluding m with $g \sim \mathcal{N}(0, \mathbf{K}_{ASV}\sigma^2_g)$, G_R is the random residual genetic term—the portion of the total genetic effect not accounted for by m or g—with $G_R \sim \mathcal{N}(0, \mathbf{I}\sigma^2_{G_R})$, and ϵ is the random residual term with $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{R})$. We then calculated $\mathbf{K}_{ASV}$ as:

$$\mathbf{K}_{ASV} = \frac{\bar{\mathbf{X}}\bar{\mathbf{X}}^T}{(n-1)^{-1}tr(\bar{\mathbf{X}}\bar{\mathbf{X}}^T)} \tag{2}$$

where $\bar{\mathbf{X}} = \mathbf{P}\mathbf{X}$ is the mean-centered marker matrix, $\bar{\mathbf{K}} = \bar{\mathbf{X}}\bar{\mathbf{X}}^T$ is the realized genomic relationship or kinship matrix, $\mathbf{P} = \mathbf{I} - n^{-1}\mathbf{1}_n\mathbf{1}_n^T$ is the idempotent mean-centering matrix, and $tr(\cdot)$ is the trace. $\mathbf{Z}_m$ is a $n \times n_m$ dimension design matrix linking levels of the explanatory locus to LSMs in y, where n_m is the number of marker genotypes.

The ASV definition of total variance from LMM (1) is:

$$\begin{aligned}\theta^{ASV}_y &= (n-1)^{-1}tr(\mathbf{V}\mathbf{P}) \\ &= \theta^{ASV}_m + \theta^{ASV}_g + \theta^{ASV}_{G_R} + \theta^{ASV}_{\bar{\epsilon}}\end{aligned} \tag{3}$$

where θ^{ASV}_y is the total phenotypic variance, $\mathbf{V}$ is the variance-covariance among observations, θ^{ASV}_m is the average semivariance of the simple genetic term, θ^{ASV}_g is the average semivariance of the polygenic term, $\theta^{ASV}_{G_R}$ is the average semivariance of the residual genetic term, and $\theta^{ASV}_{\bar{\epsilon}}$ is the average semivariance of the residuals.

The ASV definition of the genomic variance is:

$$\begin{aligned}\theta^{ASV}_g &= (n-1)^{-1}\sigma^2_g tr(\mathbf{X}\mathbf{X}^T\mathbf{P}) \\ &= \left[\frac{tr(\bar{\mathbf{K}})}{n-1}\right]\sigma^2_g\end{aligned} \tag{4}$$

In general, we replace the unknown parameter values (σ^2_g) with their REML estimates ($\hat{\sigma}^2_g$) to obtain the ASV estimates ($\hat{\theta}^{ASV}_g$). Following this form, it is possible to extend LMM (1) to include dominance and epistatic sources of variance (see below). The ASV definition of the marker associated genetic variance is:

$$\begin{aligned}\theta^{ASV}_m &= (n-1)^{-1}\sigma^2_m tr(\mathbf{Z}_m\mathbf{Z}_m^T\mathbf{P}_m) \\ &= \left[\frac{(n - n^{-1}\sum_h n^2_{G:m_h})}{n-1}\right]\sigma^2_m \\ &= k_m\hat{\sigma}^2_m\end{aligned} \tag{5}$$

It is possible to extend this using the approach for multi-locus models as in (8), with and without marker-marker interactions, described in (Feldmann *et al.* 2021). The ASV definition of the residual genetic variance is:

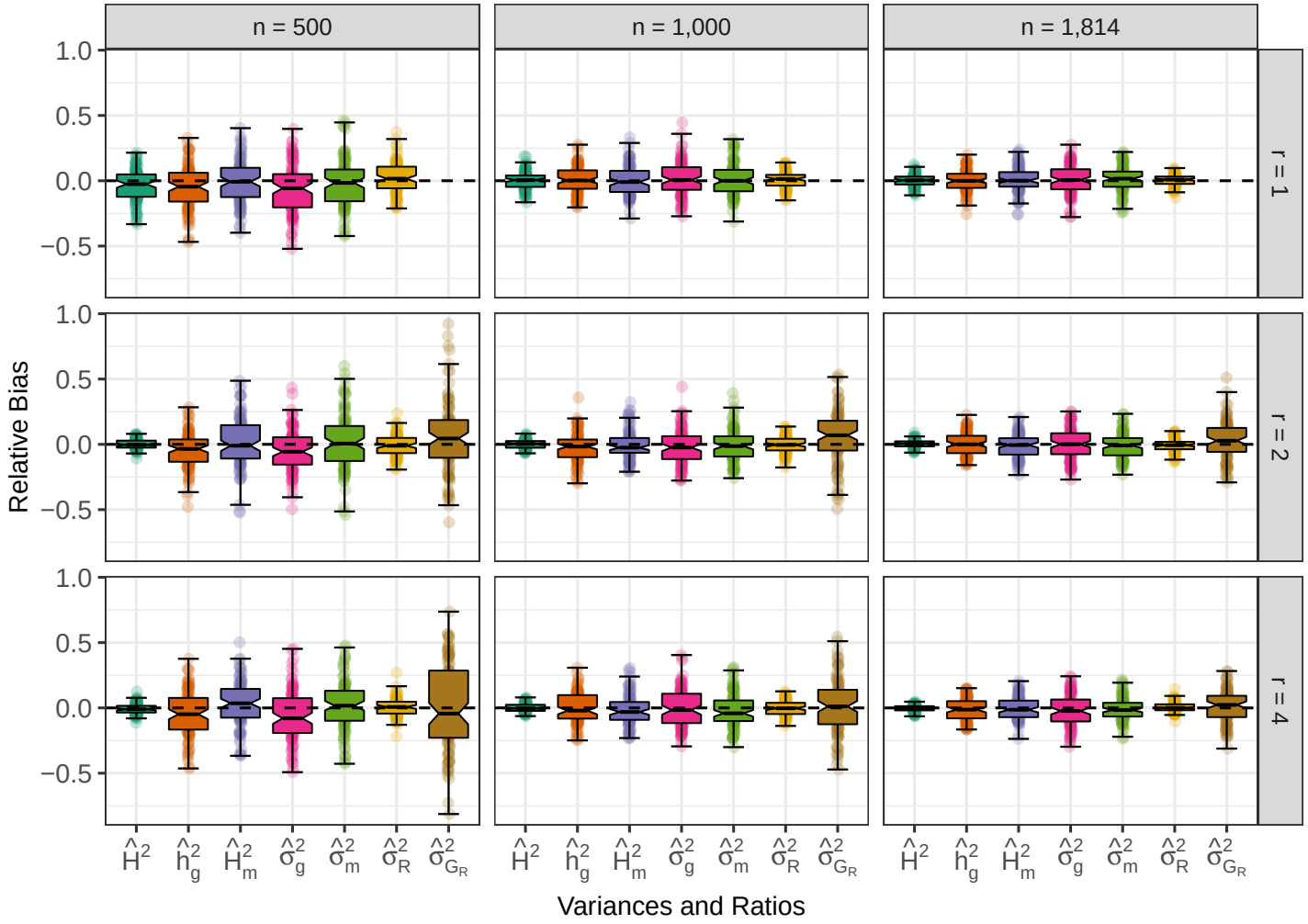

Figure 1 Effect of n and r on the relative bias of variance components and ratios in simulated outbred populations. Phenotypic observations were simulated for 100 samples with $n = 500, 1,000$, and $1,814$ (left to right) genotyped for $m = 5,000$ SNPs and the average heterozygosity $H = 0.38$. The relative bias of marker heritability, genomic heritability estimates ($\hat{h}^2_g$), broad sense heritability, genomic variance, marker variance, residual genetic variance, and residual variance heritability when the number of replicates of each entry (r) = 1 (upper panel), 2 (middle panel), and 4 (lower panel). The upper and lower halves of each box correspond to the first and third quartiles (the 25th and 75th percentiles). The notch corresponds to the median (the 50th percentile). The upper whisker extends from the box to the highest value that is within $1.5 \times IQR$ of the third quartile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the first quartile to the lowest value within $1.5 \times IQR$ of the quartile. The dashed line in each plot is the true value from simulations.

$$\begin{aligned}\theta^{ASV}_{G_R} &= (n-1)^{-1}\sigma^2_{G_R} tr(\mathbf{I}_n\mathbf{I}_n^T\mathbf{P}_n) \\ &= \sigma^2_{G_R}\end{aligned} \tag{6}$$

Importantly, all terms are estimated on the same scale as the residual variance θ^{ASV}_ϵ and are estimates on an entry-mean basis. The ASV definition of the residual variance is:

$$\begin{aligned}\theta^{ASV}_R &= (n-1)^{-1}\sigma^2_\epsilon tr(\mathbf{I}_n\mathbf{I}_n^T\mathbf{P}_n) \\ &= \sigma^2_\epsilon\end{aligned} \tag{7}$$

Linear mixed model extensions incorporating the average semivariance

While an important model, LMM (1) only covers a narrow scope of the possible genetic models and experiments that might exist, and we want to provide researchers with a clear strategy for expanding this approach to more complex systems. This section demonstrates how to partition the additive and dominance variance from a single marker, incorporate multiple explanatory loci, their interactions into the model, and non-additive polygenic terms, and achieve a fully efficient two-stage analysis. Depending on the population, trait, environment, etc. the unique components of the models demonstrated here can be hybridized and merged to accurately and holistically decompose the multitude of potential sources of genetic variation. The code to execute these models using the sommer v4.1.7 (Covarrubias-Pazaran 2016) is provided in the methods.

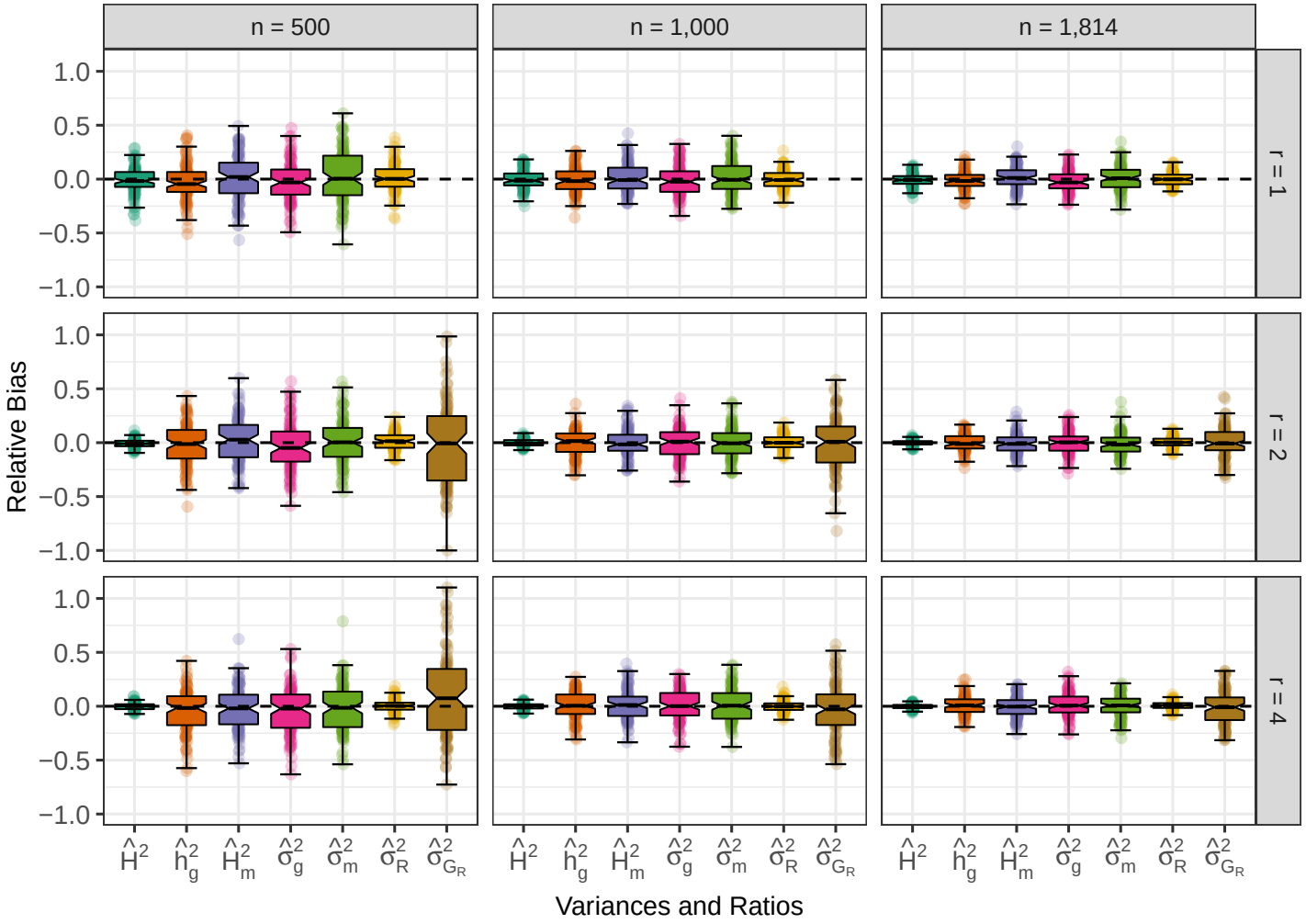

Figure 2 Effect of n and r on the relative bias of variance components and ratios in simulated inbred populations. Phenotypic observations were simulated for 100 samples with $n = 500, 1,000$, and $1,814$ (left to right) genotyped for $m = 5,000$ SNPs and the average heterozygosity $H = 0$. The relative bias of marker heritability, genomic heritability estimates ($\hat{h}^2_g$), broad sense heritability, genomic variance, marker variance, residual genetic variance, and residual variance heritability when the number of replicates of each entry (r) = 1 (upper panel), 2 (middle panel), and 4 (lower panel). The upper and lower halves of each box correspond to the first and third quartiles (the 25th and 75th percentiles). The notch corresponds to the median (the 50th percentile). The upper whisker extends from the box to the highest value that is within $1.5 \times IQR$ of the third quartile, where IQR is the inter-quartile range, or distance between the first and third quartiles. The lower whisker extends from the first quartile to the lowest value within $1.5 \times IQR$ of the quartile. The dashed line in each plot is the true value from simulations.

Extension #1: Incorporating multiple target loci and locus-locus interactions

It is common for multiple QTL to be implicated from genetic studies (Tang *et al.* 2006; Rutkoski *et al.* 2014; Vasconcellos *et al.* 2017; Lopdell *et al.* 2019; Legare *et al.* 2000; Cockerton *et al.* 2019; Rice and Lipka 2019; Demmings *et al.* 2019b), the utility of which is not always certain (Bernardo 2001, 2004). While the simulations in this paper rely exclusively on LMM (1), this model can be easily expanded to include multiple explanatory loci and their interactions or statistical epistasis (Moore and Williams 2005; Álvarez-Castro and Carlborg 2007), as demonstrated by (Feldmann *et al.* 2021). For example, the LMM with three main-effect loci, denoted m_1, m_2, and m_3, is:

$$\mathbf{y} = \mathbf{1}_n\mu + \sum_{i=1}^{3}\mathbf{Z}_{m_i}m_i + \sum_{\substack{i=1\\i<j}}^{2}\sum_{j=2}^{3}\mathbf{Z}_{m_{ij}}m_{ij} + \mathbf{Z}_{m_{123}}m_{123} + \mathbf{I}g + \mathbf{I}G_R + \boldsymbol{\epsilon} \quad (8)$$

where m_i is the random effect of the i-th main-effect marker, m_{ij} is the random effect of the two-way interaction between the i-th and j-th markers, and m_{123} is the random effect of the three-way interaction between the three main-effect loci. $\mathbf{Z}_{m_i}$, $\mathbf{Z}_{m_{ij}}$, and $\mathbf{Z}_{m_{123}}$ are design matrices that link levels of the explanatory marker and interactions to LSMs in y. The rest of the terms have the same definitions.

Extension #2: Partitioning θ_m^{ASV} into additive ($\theta_{m_\alpha}^{ASV}$) and dominance ($\theta_{m_\delta}^{ASV}$) components

The factor coding of the Mendelian and oligogenic markers is a different approach than is standard in GWAS (Korte and Farlow 2013; Visscher *et al.* 2012, 2017). In GWAS, markers are typically treated as fixed and coded numerically, e.g., the dosage model. Assuming that a researcher is working with an outbred species ($H \neq 0$), the dominance deviations can be significant, and partitioning the additive and dominance sources of variance from significant markers can be helpful in hybrid crop breeding and disease risk prognoses. Our goal is to partition θ_m^{ASV} into its additive ($\theta_{m_\alpha}^{ASV}$) and dominance ($\theta_{m_\delta}^{ASV}$) components.

Here, we demonstrate an LMM that can be used to partition the additive and dominance sources of variance of the main effect marker. The form of the linear mixed model (LMM) for this analysis assuming only one explanatory marker is:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_{m_\alpha} m_\alpha + \mathbf{Z}_{m_\delta} m_\delta + \mathbf{I}g + \mathbf{I}G_R + \boldsymbol{\epsilon} \quad (9)$$

where m_α is the random effect of the main-effect locus with $m_\alpha \sim \mathcal{N}(0, \mathbf{I}\sigma^2_{m_\alpha})$ and m_δ is the random effect of the main-effect locus with $m_\delta \sim \mathcal{N}(0, \mathbf{I}\sigma^2_{m_\delta})$. $\mathbf{Z}_{m_\alpha}$ is an $n \times 3$ design matrix linking marker genotypes to observations and $\mathbf{Z}_{m_\delta}$ is an $n \times 2$ design matrix linking genotypic state, either homozygous (AA and aa) or heterozygous (Aa), to observations. Other terms are as defined in LMM (1).

The average semivariance associated with m_α is obtained as in (5) by:

$$\begin{aligned} \hat{\theta}_{m_\alpha}^{ASV} &= (n-1)^{-1}\hat{\sigma}^2_{m_\alpha} tr(\mathbf{Z}_{m_\alpha}\mathbf{Z}^T_{m_\alpha}\mathbf{P}_{m_\alpha}) \\ &= \left[\frac{n - n^{-1}\sum_h n^2_{G:m_{\alpha_h}}}{n-1}\right]\hat{\sigma}^2_{m_\alpha} \end{aligned} \quad (10)$$

where $n_{G:m_{\alpha_h}}$ is the number of entries nested in the h-th marker genotype (Feldmann *et al.* 2021). The average semivariance associated with m_δ is obtained by :

$$\begin{aligned} \hat{\theta}_{m_\delta}^{ASV} &= (n-1)^{-1}\hat{\sigma}^2_{m_\delta} tr(\mathbf{Z}_{m_\delta}\mathbf{Z}^T_{m_\delta}\mathbf{P}_{m_\delta}) \\ &= \left[\frac{n - n^{-1}\sum_i n^2_{G:m_{\delta_i}}}{n-1}\right]\hat{\sigma}^2_{m_\delta} \end{aligned} \quad (11)$$

where $n_{G:m_{\delta_i}}$ is the number of entries nested in the i-th genetic state. The sum of $\left[k_{m_\alpha}\hat{\sigma}^2_{m_\alpha} + k_{m_\delta}\hat{\sigma}^2_{m_\delta}\right] = \left[\hat{\theta}_{m_\alpha}^{ASV} + \hat{\theta}_{m_\delta}^{ASV}\right] = \hat{\theta}_m^{ASV}$ and $\left[\hat{\theta}_{m_\alpha}^{ASV} + \hat{\theta}_{m_\delta}^{ASV}\right] - \hat{\theta}_m^{ASV} = 2.21 \times 10^{-5}$. $\hat{\theta}_m^{ASV}$ is an unbiased estimate of the variance explained by a marker (Feldmann *et al.* 2021). The likelihood ratio (LR) between LMM (1) and (9) was $LR \approx 0$ and was not significant in any simulated populations ($P_{LR} > 0.2$), suggesting that there is no appreciable difference between the model likelihood of (1) and (9). The same marker variance is estimated in both LMMs, (1) and (9), and the estimates are equal. Note that we were not able to fit LMM (9) in all software and had to use either `sommer::mmer()` or `asreml::asreml()`.

Extension #3: Incorporating additional polygenic terms for dominance (g_δ) deviations

LMM (1) can also be extended to include both additive (g_α) and dominance (g_δ) sources of genomic variance (Vitezica *et al.* 2013; Kumar *et al.* 2015; Vitezica *et al.* 2017; Xiang *et al.* 2018; Sun *et al.* 2014; Ali *et al.* 2020; Zhang *et al.* 2021; Martini *et al.* 2016). The form of the LMM for analysis with both g_α and g_δ assuming only one explanatory marker M is:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_m m + \mathbf{I}g_\alpha + \mathbf{I}g_\delta + \mathbf{I}G_R + \boldsymbol{\epsilon} \quad (12)$$

where g_α and g_δ are random effect vectors for the additive and dominance polygenic effects, respectively, with $g_\alpha \sim \mathcal{N}(0, \mathbf{K}_{ASV}\sigma^2_{g_\alpha})$ and $g_\delta \sim \mathcal{N}(0, \mathbf{K}^D_{ASV}\sigma^2_{g_\delta})$. The average semivariance dominance kernel is:

$$\mathbf{K}^D_{ASV} = \frac{\bar{\mathbf{W}}\bar{\mathbf{W}}^T}{(n-1)^{-1}tr(\bar{\mathbf{W}}\bar{\mathbf{W}}^T)} \quad (13)$$

where $\mathbf{W} = 1 - |\mathbf{X}|$, assuming $\mathbf{X}$ is coded [-1,0,1], and $\bar{\mathbf{W}} = \mathbf{PW}$. This is a feasible approach to improve genetic performance in crossbred populations with large dominance genetic variation (Nishio and Satoh 2014; Vitezica *et al.* 2017; Xiang *et al.* 2018; Wolfe *et al.* 2021). Both $\mathbf{K}_{ASV}$ and $\mathbf{K}^D_{ASV}$ have the matrix properties proposed by Speed and Balding (2015); i.e., $n^{-1}tr(\mathbf{K}) = 1$ and $n^{-2}\sum_i\sum_j K_{ij} = 0$. Not surprisingly, the dominance variance estimated with $\mathbf{K}^D_{ASV}$ were accurate and the relative bias from 100 simulated populations was -3.32%.

Further extensions for additive-by-additive $A \times A$ or additive-by-dominance $A \times D$ polygenic interactions are also possible (Nishio and Satoh 2014; Covarrubias-Pazaran 2016; Vitezica *et al.* 2017). These matrices are often calculated as the Hadamard product (element-wise multiplication, $\circ$) of $\mathbf{K}_{ASV}$ and/or $\mathbf{K}^D_{ASV}$, where the additive-by-additive epistasis GRM is $\mathbf{K}^I_{ASV} = \mathbf{K}_{ASV} \circ \mathbf{K}_{ASV}$. This matrix has the same essential properties as $\mathbf{K}_{ASV}$, and so we hypothesize that the ASV estimted variance components will be accurate for these terms as well.

Extension #4: Stage-wise LMM analysis for multi-environment trials (METs) and meta-analysis in plant breeding

Two-stage, or stage-wise, analyses are the *status quo* in plant breeding trials in both academic studies and seed industry (Piepho *et al.* 2012; Damesa *et al.* 2017, 2019; Endelman 2022). The reason for this is that plant breeders are often not interested in the performance *per se* of a line or hybrid *within* a specific location, unless the presence of cross-over (rank change) $G \times E$ is very large enough to make data from one target environment non-informative in another target environment. Instead, plant breeders are often more interested in the ranking and performance of entries averaged across all environments (Bernardo 2020). It is common then to fit a first model that accounts for the variation of random design elements, e.g., locations, years, blocks, and fixed genotype effects to obtain the estimated marginal means (EMMs) or best linear unbiased estimators (BLUEs) as adjusted entry means. These adjusted entry means are then used as the phenotype or response variable in GWAS and genomic prediction studies. However, the naive approach is not "fully efficient" (Piepho *et al.* 2012) and assumes that adjusted entry means are IID; i.e., $\mathbf{R} = \mathbf{I}\sigma^2_R$. However, due to incomplete block and augmented designs, missing data, and changes in experiment designs over time and location, IID entry means are rarely observed in practice. To fully utilize the data, however, the variance-covariance matrix of the estimates from Stage 1 must be included in Stage 2 (Piepho *et al.* 2012; Damesa *et al.* 2017), which is not possible with many software packages for genomics-assisted breeding.

The LMM for stage one is:

$$\mathbf{y} = \mathbf{X}G + \mathbf{Z}u + \boldsymbol{\epsilon}_e \quad (14)$$

where $\mathbf{X}$ is the fixed effect design matrix linking observations to entries, $\mathbf{Z}$ is the random effect design matrix for design (e.g., blocks) elements within each environment (e.g., years and locations), and ϵ_e are the residuals and $\epsilon_e \sim \mathcal{N}(0, \mathbf{R}_e)$, where $\mathbf{R}_e$ is the residual variance-covariance matrix estimated in the e-th environment. $\mathbf{R}_e$ can be estimated with or without spatial or autoregressive correlations (Farfan *et al.* 2015; Rodríguez-Álvarez *et al.* 2018; Anderson *et al.* 2018; Selle *et al.* 2020). This model is fitted for each environment independently. From these models, we obtain the adjusted entry means $\bar{y}$ and the residual variance covariance matrices $\mathbf{R}_e$ from each of $e = 1,...,n_e$ environments, where n_e are the number of environments. For CRD or experiments without design elements the obtained variance-covariance matrix will be diagonal. Assume that we have two environments, we will obtain $\mathbf{R}_1$ from environment 1 and $\mathbf{R}_2$ from environment 2. We can then construct the $2n \times 2n$ block-diagonal stage-one $\mathbf{\Omega}$ matrix as:

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 \end{bmatrix} \tag{15}$$

This block-diagonal form indicates that the residuals among entries are uncorrelated among environments. For simplification, $\mathbf{R}_e$ can be approximated by diagonal matrices in several different ways (Smith *et al.* 2001; Möhring and Piepho 2009; Welham *et al.* 2010; Piepho *et al.* 2012; Moehring *et al.* 2014), but here we use to the full variance-covariance matrix from each experiment e. Importantly, we need to carry $\mathbf{\Omega}$ over from the stage-one analyses to stage-two of the analysis.

The LMM for stage two is then:

$$\bar{\mathbf{y}} = \mathbf{1}\mu + \mathbf{X}E + \mathbf{Z}_m m + \mathbf{Z}_g g + \mathbf{Z}_{G_R} G_R + \boldsymbol{\epsilon}_2 \tag{16}$$

where $\bar{y}$ are the adjusted entry means from stage-one, μ is the population mean, $\mathbf{X}$ is the fixed effect design matrix linking environments to adjusted entry means, E are the fixed environmental effects, g is the random additive genetic effect associated with the genome-wide framework of marker excluding m with $g \sim \mathcal{N}(0, \mathbf{K}_{ASV}\sigma^2_g)$, G_R is the random residual genetic term—the portion of the total genetic effect not accounted for by m or g—with $G_R \sim \mathcal{N}(0, \mathbf{I}_n\sigma^2_{G_R})$, and ϵ_2 is the structured residual term from stage-one with $\epsilon_2 \sim \mathcal{N}(0, \mathbf{\Omega})$ This approach is accessible to researchers via the sommer, asreml, and StageWise packages in R (Covarrubias-Pazaran 2016; Butler 2021; Endelman 2022) and in SAS.

We created 100 simulated population ($n = 1,000$; $m = 5,000$) using a similar approach to the other simulations in this experiment. However, in this experiment we included Environmental and Block within Environment effects. We estimates the variance explained by the polygenic background, a large effect locus, the residual genetic variance, and non-genetic residual. The single stage analysis yielded relative biases of -0.67%, -0.33%, -0.67%, and 0.41% for the marker variance ($\hat{\sigma}^2_m$), genomic variance ($\hat{\sigma}^2_g$), residual genetic variance ($\hat{\sigma}^2_{G_R}$), and residual variance ($\hat{\sigma}^2_R$), respectively (Fig 3). The two stage analysis yielded relative biases of -0.83%, -4.08%, 0.15%, and 0.16% for the marker variance ($\hat{\sigma}^2_m$), genomic variance ($\hat{\sigma}^2_g$), residual genetic variance ($\hat{\sigma}^2_{G_R}$), and residual variance ($\hat{\sigma}^2_R$), respectively (Fig 3).

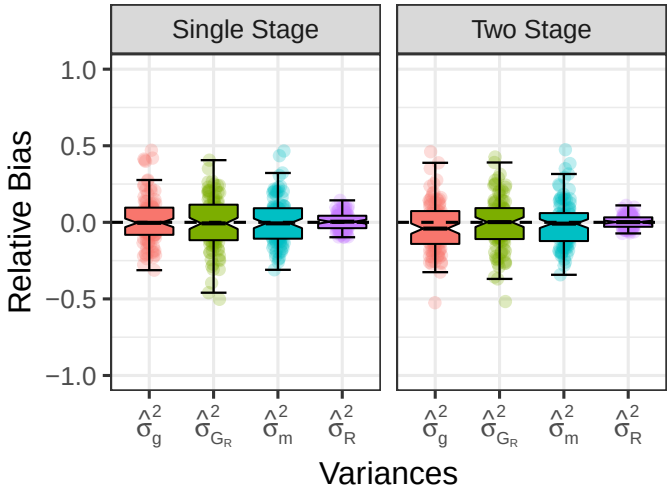

Figure 3 Single versus multi Stage analysis with two environments. The relative bias of genomic variance ($\hat{\sigma}^2_g$), marker variance ($\hat{\sigma}^2_m$), residual genetic variance ($\hat{\sigma}^2_{G_R}$), and residual variance ($\hat{\sigma}^2_R$) analysed in a single stage (left panel) or in two stages (right panel). The upper and lower halves of each box correspond to the first and third quartiles (the 25th and 75th percentiles). The notch corresponds to the median (the 50th percentile). The upper whisker extends from the box to the highest value that is within $1.5 \times IQR$ of the third quartile, where IQR is the interquartile range, or distance between the first and third quartiles. The lower whisker extends from the first quartile to the lowest value within $1.5 \times IQR$ of the quartile. The dashed line in each plot is the true value from simulations.

Extension #5: Incorporating k_M directly into LMM analyses

In (Feldmann *et al.* 2021), we introduced ASV into LMMs for individual markers in genetic analysis as a *post hoc* adjustment of the variance explained by a marker by k_M (5). This directly led to (Feldmann *et al.* 2022), in which we showed that ASV estimates of the genomic variance could be obtained by scaling the genomic relationship prior to the LMM analysis and introduced $\mathbf{K}_{ASV}$, eliminating the need for any *post hoc* adjustment. Using statistical packages such as sommer (Covarrubias-Pazaran 2016), we can directly apply k_M to the variance-covariance matrix for large effect loci M and their interaction in our model. Typically, the identity matrix is used as the variance-covariance matrix and levels of the random effect are assumed to have the same variance and no covariance. In (Feldmann *et al.* 2021) we multiplied the average marginal variance component by k_M to obtain the ASV component. Instead, if we define $\mathbf{K}_M = \mathbf{I}_{n_M} k_M^{-1}$, where $\mathbf{K}_M$ is $n_M \times n_M$ and n_M is the number of marker genotypes. We can essentially think of $\mathbf{K}_M$ in the same way that we think of genomic relationship matrices; e.g., $\mathbf{K}_{ASV}$, except that we apply $\mathbf{K}_M$ to the levels of the marker genotype instead of entries. With this approach, we maintain the levels of the factor come from the same variance and zero covariance, but our scaling factor embedded directly in the model eliminating the need for adjustment. Embedding k_M in the LMM analysis using $\mathbf{K}_M$ is equivalent to the *post hoc* adjustment that we proposed in (Feldmann *et al.* 2021), and so it is up to the user to determine which approach they prefer.

Results and Discussion

Candidate Genes and Complex Traits

Bernardo (2014) was the first to propose an integration of MAP and GP and since then empirical studies have validated the methodol-

ogy (Rutkoski *et al.* 2014; Zhang *et al.* 2014; Rice and Lipka 2019; Spindel *et al.* 2016) while others have shown little-to-no improvement over GP (Li *et al.* 2015; Galli *et al.* 2020), suggesting that modeling significant markers can improve prediction accuracy only when markers explain a *significant* portion of both genetic and phenotypic variance (Galli *et al.* 2020). With the high densities of genome-wide markers commonly assayed in gene finding studies, investigators often identify markers tightly linked to candidate or known causal genes as exemplified by diverse real world examples (Andersson 2001; Hayes and Goddard 2001; Anderson *et al.* 2007; Gaudet *et al.* 2010; Hayes *et al.* 2010; Jensen *et al.* 2012; Visscher *et al.* 2012; Septiningsih *et al.* 2009; Saatchi *et al.* 2014; Visscher *et al.* 2017; Freebern *et al.* 2020; Li *et al.* 2021; Korte and Farlow 2013). The candidate marker loci are nearly always initially identified by genome-wide searches using sequential (marker-by-marker) approaches such as GWAS and QTL analysis. Following the discovery of statistically significant marker-trait associations from a marker-by-marker genome-wide scan, the natural progression would be to analyze single- or multi-locus genetic models where the effects of the discovered loci are simultaneously corrected for the effects of other discovered loci, e.g., polygenic variation (Stroup *et al.* 2018; Gbur *et al.* 2020).

A marker will not explain a large portion of variance if that marker does not have a large, detectable effect and, thus, markers that explain a large portion of genetic variance will be the most useful for MAP. For example, consider Fusarium wilt resistance in strawberry which is conferred by a single dominant acting locus *Fw*1 (Pincot *et al.* 2018, 2022). This locus explains nearly 100% of both the phenotypic and genetic variance and the mean differences delineate resistant vs susceptible genotypes, and thus there is almost no added benefit of a genome-wide sample of markers over the single-marker assay (m) for product delivery and germplasm improvement. While variance explained is directly linked to the effect size, it is not a direct substitute. However, the random effect machinery allows for researchers to obtain variance component estimates and effect sizes (e.g., BLUPs) simultaneously (Searle *et al.* 1992) eliminating the need for multiple statistical models to assess the variance explained and the effect size of a target locus. The BLUP procedure is directly applied in this model, so it is natural to use the same statistical machinery to estimate GEBVs by GBLUP and the genetic effect of a locus.

As a point of contrast, yield in maize (*Zea mays*) is heritable but no single locus explains any appreciable amount of phenotypic or genotypic variance (Heffner *et al.* 2009, 2010; Yang *et al.* 2017; Brandariz and Bernardo 2019; Gage *et al.* 2020; Zhang *et al.* 2019). For improvement of yield in maize, GP is potentially a more valuable approach because the researcher, or breeder, can predict the polygenic value (g) without relying on any one particular locus, but instead capturing variation of a genome-wide sample of markers. The more challenging scenario is the intermediate case in which a trait is controlled by both loci that are discernible from the polygenic background and the polygenic background itself (Rutkoski *et al.* 2014; Rice and Lipka 2019; DeWitt *et al.* 2021).

The ratio between the variance explained by the oligogenic and polygenic terms with the total genetic or phenotypic variance is likely a a major factor determining the cost-benefit of incorporating MAP, GP, or both into a breeding or diagnostic program. Modeling a individual loci can be advantageous when the proportion of the phenotypic and genetic variance explained by the locus is reasonably large and not partially captured by other markers in linkage disequilibrium (LD) with the target (Bernardo 2014; Rutkoski *et al.* 2014; Rice and Lipka 2019; Pincot *et al.* 2018, 2022). Ideally, the targeted markers should not fit the marker effect size distribution assumptions, e.g., that all marker effects contribute equally to the genomic variance and are drawn from the same distribution (Piepho 2009; Endelman 2011; Habier *et al.* 2007) and should not be in high LD with a large number of other markers. With ASV, researchers can accurately estimate these parameters directly in LMM analyses.

Simulations confirm that ASV yields accurate estimates of all genetic variance components and ratios

As we show in our previous studies (Piepho 2019; Feldmann *et al.* 2021, 2022), ASV is ideal for estimating the variance explained by both single loci and GRMs. In our simulations, we included variation in population size, e.g., $n = 500, 1,000$, and $1,814$, and replication of entries, e.g., $r = 1, 2$, and 4 for both outbred (Fig 1) and inbred populations (Fig 2). We can see that the same pattern that has emerged as in previous studies; the ASV approach yields accurate, unbiased estimates of variance components and variance component ratios from LMM analyses regardless of the constitution of the population or the study design. Even when there is only one replicate per entry ($r = 1$) all of the explanatory genetic terms are still accurately partitioned from the total variance. As n increased from 500 to $1,814$, the precision of estimates increased dramatically (the sampling variance decreases). Increasing r from 1 to 4 did not affect precision or accuracy of genomic and marker associated variances. However, increased numbers of replicates did improve the precision of residual variance components. This is because entries are replicated among plots ($n \cdot r$), but markers and other genetic components are replicated among entries (n). Our simulations, in conjunction with our previous results (Piepho 2019; Feldmann *et al.* 2021, 2022), demonstrate that in most populations—human, animal, plant, or microbe—the average semivariance will yield accurate and easily interpreted estimates of different variance components.

Average semivariance in quantitative genetics and beyond

ASV is a strategy that can be used for estimating and partitioning the total variance into components (Piepho 2019), such as the variance explained by loci and locus-locus (Feldmann *et al.* 2021) and the genomic variance (Feldmann *et al.* 2022). The approach we are suggesting shares some common threads with the current thinking in quantitative genetics, particularly as it relates to genomic relatedness, genomic heritability, and genomic prediction (VanRaden 2008; Yang *et al.* 2010; Kang *et al.* 2010; Habier *et al.* 2013; Hayes *et al.* 2009; Meuwissen *et al.* 2001; Isik *et al.* 2017; Zas and Sampedro 2015; Potti and Canal 2011; Roff and Fairbairn 2015; Swarts *et al.* 2021; Nietlisbach *et al.* 2016; Ulrich *et al.* 2021; Fan *et al.* 2021) but it also deviates from the classic quantitative genetic model conceptually in that it assumes that marker effects are random variables (Falconer and Mackay 1996; Lynch and Walsh 1998; Bernardo 2001). We have demonstrated that these are a statistically valid set of assumptions, even though they deviate from the classic quantitative genetics perspective.

ASV has several beneficial elements that make ASV a viable option for quantitative genetics, but more importantly, it is appropriate for any quantitative discipline where variance components are of interest from plant and microbial biology to psychology and infant research. Namely:

1. **The definitions of the variance components using average semivariance are additive and sum to the phenotypic vari-**

ance. This means that the LMM can be extended to incorporate all explanatory components, e.g., dominance, epistasis, transcriptomic, and will yield accurate VCEs for all terms (Nishio and Satoh 2014; Vitezica *et al.* 2017; Xiang *et al.* 2018; Krause *et al.* 2019). This is not necessarily true for all definitions of variance components (Piepho 2019).

2. **ASV is well suited for mutli-stage analyses** At the center of ASV, is the idea that the "entry mean" is the phenotype *per se*, and not the observations (Piepho 2019; Feldmann *et al.* 2022). One interpretation is that individuals, not observations, are the primary source of variation. ASV yields accurate estimates of the genetic and genomic variance components in unreplicated, or partially replicated, designs common in humans and agricultural plants and animals (Cullis *et al.* 2006; Moehring *et al.* 2014; Cullis *et al.* 2020; Butler *et al.* 2014; González-Barrios *et al.* 2019). ASV also yields accurate estimates in the two-stage approaches to GP and GWAS in plants (Piepho *et al.* 2012; Damesa *et al.* 2017, 2019).
3. **ASV does not affect or impact the BLUPs or breeding value predictions.** ASV is only used to obtain accurate VCEs (Piepho 2019; Feldmann *et al.* 2022). It has been demonstrated that marker coding and different strategies for scaling and centering $\mathbf{Z}$ and $\mathbf{K}$ do not impact BLUPs or prediction accuracy (Strandén and Christensen 2011; Legarra 2016; Legarra *et al.* 2018), and, because ASV essentially works through a set of scalar coefficients determined by the experiment and population, this feature directly applies to this work.
4. **ASV works under many model assumptions in GLMM analyses** beyond the often-assumed variance-covariance structure in this study, e.g., $\mathbf{R} = \mathbf{I}\sigma^2_\epsilon$. ASV can be applied to designs accounting for spatial structure through auto-regressive correlations or spline-models (Rodríguez-Álvarez *et al.* 2018; De Resende *et al.* 2006; Selle *et al.* 2019, 2020; Burgueño *et al.* 2000; Borges *et al.* 2019; Hoefler *et al.* 2020). ASV can also be applied to data sets where the observational units lead to non-normality of residuals; i.e., ordinal disease scores and proportion scores (Piepho 2019).

As substantiated by our simulations in this study and in the context of our previous work, ASV with REML estimation of the underlying variance components yields accurate estimates for oligo and polygenic effect, both individually and collectively, and BLUPs of the the additive and dominance effects of marker loci (Piepho 2019; Feldmann *et al.* 2021, 2022). ASV directly yields accurate estimates of genomic heritability in the observed population and can be used to adjust deviations that arise from other commonly used methods for calculating genomic relationships regardless of the population constitution, such as inbred lines and F_1 hybrids, unstructured GWAS populations, or animal herds and flocks. We believe that $\mathbf{K}_{ASV}$ provides a powerful approach for directly estimating genomic heritability for the observed population regardless of study organism or experiment design (Visscher *et al.* 2006, 2007, 2008, 2010). In conclusion, our recommendation is that the average semi-variance approach be considered for general adoption by genetic researchers working in humans, microbes, or (un)domesticated plants and animals.

Methods and Materials

Computer Simulations

We generated 18 experiment designs with different population sizes of $n = 500, 1{,}000$, and $1{,}814$ and number of clonal replicates per entry $r = 1$, 2, and 4 for outbred $H = 0.38$ and inbred $H = 0.0$ populations. Clonal replicates are a special case common in plant genetics of hybrid (e.g., maize, rice, and sorghum) cropping systems and in clonally propagated species (e.g., strawberry, potato, and apple). In all examples, 100 populations genotyped at $m = 5{,}000$ loci. These $5{,}000$ SNPs were used to generate the purely additive polygenic background and one locus for the simple genetic effect. Marker genotypes, e.g., alleles, were drawn from a multivariate normal distribution with to replicate the population structure of the 1,814 mice from Valdar *et al.* (2006) using `MASS::mvrnorm()` and transformed such that the population was heterozygosity $H = 0.38$. We then estimated $\mathbf{K}_{ASV}$ and excluded the targeted locus from the calculation of $\mathbf{K}_{ASV}$. We also simulated residual genetic and residual effects each from a normal distribution with $\mu = 0$ and $\sigma_{G_R} = \sqrt{20}$ and $\sigma_R = \sqrt{30 \cdot r}$ using `stats::rnorm()`. A single explanatory locus was simulated with a segregation ratio of approximately $1:2:1$ for AA:Aa:aa marker genotypes was simulated with $\mu = 0$ and $\sigma_m = \sqrt{k_M \cdot 25}$ using `stats::rnorm()`. We did not control for the portion of additive vs dominance variance for the single marker. We simulated marker effects for all $m = 5{,}000$ loci following a normal distribution $\mu = 0$ and $\sigma_g = \sqrt{40/5000}$. When multiplied by the centered marker genotypes and summed, the score is taken as the true additive genetic value g of each individual. For each simulated population we expressed LMM (1) using `asreml::asreml()` Butler (2021). In the second set of simulations, we used the same approach and same mean and variance parameters. However, in this example we simulated full inbred lines in the background polygenic markers ($H = 0.0$) and in the foreground markers, e.g., $1:0:1$ for AA:Aa:aa. All plots are made with the `ggplot2` package Wickham (2016) in R 4.1.0 R Core Team (2020).

Model statements in `R/sommer` v4.1.7:

Incorporating One Target Locus into GBLUP LMM (1) is expressed as:

```
mmer(fixed  = Y ~ 1,
     random = ~ M +
                vsr(G, Gu = Kasv) +
                GR,
     rcov   = ~ units,
     data   = data)
```

where `data` is a $n \times 4$ matrix containing the phenotypic observations Y, a factor coding levels of M, a factor coding entries G, and a factor coding levels of G_R. The variable `units` is inferred by `sommer::mmer()` and can be considered as a column with as many levels as rows in the data (Covarrubias-Pazaran 2016). The factor levels of G and G_R are equivalent.

The version of this model with k_M embedded is expressed as:

```
mmer(fixed  = Y ~ 1,
     random = ~ vsr(M, Gu = KM) +
                vsr(G, Gu = Kasv) +
                GR,
     rcov   = ~ units,
     data   = data)
```

where `KM` is the matrix $\mathbf{K}_M = \mathbf{I}_{n_M} k_M^{-1}$. All other variables are the same as previously defined.

Incorporating Multiple Target Loci into GBLUP. LMM (8) is expressed as:

```
mmer(fixed  = Y ~ 1,
     random = ~ M1 + M2 + M3 +
```

```
                M12 + M13 + M23 + M123 +
                vsr(G, Gu = Kasv) +
                GR,
        rcov  = ~ units,
        data  = data)
```

where `data` is a $n \times 10$ matrix containing the phenotypic observations Y, seven columns corresponding to the marker effects and interactions, a factor coding entries G, and a factor coding levels of G_R. The factor coding of m_α has three levels corresponding to $AA : Aa : aa$ and a factor coding levels of m_δ has two levels corresponding to homozygous and heterozygous.

Partitioning Marker Variance into Additive and Dominance Components. LMM (9) is expressed as:

```
mmer(fixed  = Y ~ 1,
     random = ~ Ma + Md +
                vs(G, Gu = Kasv) +
                GR,
     rcov   = ~ units,
     data   = data)
```

where `data` is a $n \times 5$ matrix containing the phenotypic observations Y, a factor coding levels of m_α, a factor coding levels of m_δ, a factor coding entries G, and a factor coding levels of G_R. The factor coding of m_α has three levels corresponding to $AA : Aa : aa$ and a factor coding levels of m_δ has two levels corresponding to the genetic state—either homozygous or heterozygous.

Incorporating a Genomic Dominance Relationship Matrix into GBLUP. LMM (12) is expressed as:

```
mmer(fixed  = Y ~ 1,
     random = ~ M +
                vsr(Ga, Gu = Kasv) +
                vsr(Gd, Gu = Kasv_D) +
                GR,
     rcov   = ~ units,
     data   = data)
```

where `data` is a $n \times 5$ matrix containing the phenotypic observations Y, a factor coding levels of M, and three factors coding entries, e.g., G_α, G_δ, and G_R. The factor levels of G_α, G_δ, and G_R are equivalent.

Incorporating Stagewise Meta-analysis into GBLUP. LMM (14) is expressed as:

```
mmer(fixed  = Y ~ G,
     rcov   = ~ units,
     data   = data)
```

where `data` is a $n \times 2$ matrix containing the phenotypic observations Y and one factor coding G for the entry ID. Blocks and other within location design elements can be incorporated as random effects using the `random` = syntax. In sommer, $\mathbf{R}_e$s are obtained from each location as the 'VarBeta' matrix in the `sommer::mmer()` output. Specially, 'VarBeta' is the name of the model estimated variance covariance matrix among entry means in sommer. The $\mathbf{R}_e$s are then bound corner-to-corner, which is accomplished using `sommer::adiag1()` to obtain $\mathbf{\Omega}$. We then take the inverse of $\mathbf{\Omega}$ using `base::solve()`.

The LMM for stage 2 (16) is expressed as:

```
mmer(fixed    = Estimate ~ Env - 1,
     random   = ~ vsr(M, Gu = KM) +
                vsr(G, Gu = Kasv) +
                G:Env + GR,
     rcov     = ~ vsr(units,
                  Gti = matrix(invSigma2,1,1),
                  Gtc = matrix(3,1,1)),
     nIters   = 25,
     emWeight = rep(1,25),
     W        = invOmega,
     data     = data)
```

where where `data` is a $n \times 5$ matrix containing the adjusted entry means from stage 1 Y, a factor coding levels of M, two equivalent factors coding entries, e.g., G and G_R, and one factor coding environments Env. In this approach, we must fix the residual variance component equal to 1 so that the residual so that all the scaling of the `invOmega` $= \mathbf{\Omega}^{-1}$ is unaffected by the model estimation process. Within the `vs()` argument, the `Gti()` and `Gtc()` arguments are used to set the initial value of the variance component equal to the inverse of the variance among adjusted entry means (`invSigma2` $= \hat{\sigma}^{-2}$) and to constrain the variance component estimation to a fixed value by setting the first argument equal to 3 (Covarrubias-Pazaran 2022). In this example we use 25 iterations of 100% expectation-maximization algorithm; however, the EM and NR methods can be exchanged or averaged, by changing the `emWeight` argument.

Data Availability

Zenodo repository coming soon. For now, code is available by request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding Statement

This research was supported by grants to Steven J. Knapp from the United States 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). HPP was supported by the German Research Foundation (DFG) grant PI 377/24-1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: MJF, HPP **Data curation:** MJF **Formal Analysis:** MJF **Funding Acquisition:** HPP **Investigation:** MJF, HPP **Methodology:** MJF, GCP **Project administration:** MJF, HPP **Resources:** HPP **Software:** MJF, GCP **Supervision:** MJF, HPP **Validation:** MJF **Visualization:** MJF **Writing – original draft preparation:** MJF, HPP **Writing – review & editing:** MJF, HPP, GCP

Literature cited

Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC. 2011. Genome-based prediction of testcross values in maize. Theo. Appl. Genet.. 123:339.

Ali M, Zhang L, DeLacy I, Arief V, Dieters M, Pfeiffer WH, Wang J, Li H. 2020. Modeling and simulation of recurrent phenotypic and genomic selections in plant breeding under the presence of epistasis. The Crop Journal. 8:866–877.

Álvarez-Castro JM, Carlborg O. 2007. A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. Genetics. 176:1151–1167.

Anderson JA, Chao S, Liu S. 2007. Molecular breeding using a major qtl for fusarium head blight resistance in wheat. Crop Sci.. 47:S–112.

Anderson SL, Mahan AL, Murray SC, Klein PE. 2018. Four parent maize (fpm) population: Effects of mating designs on linkage disequilibrium and mapping quantitative traits. The plant genome. 11:170102.

Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet.. 2:130–138.

Ariyomo TO, Carter M, Watt PJ. 2013. Heritability of boldness and aggressiveness in the zebrafish. Behavior genetics. 43:161–167.

Bernardo R. 2001. What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci.. 41:1–4.

Bernardo R. 2004. What proportion of declared qtl in plants are false? Theor. Appl. Genet.. 109:419–424.

Bernardo R. 2014. Genomewide selection when major genes are known. Crop Sci.. 54:68–75.

Bernardo R. 2016. Bandwagons I, too, have known. Theor. Appl. Genet.. 129:2323–2332.

Bernardo R. 2020. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something blue. Heredity. 125:375–385.

Borges A, González-Reymundez A, Ernst O, Cadenazzi M, Terra J, Gutiérrez L. 2019. Can spatial modeling substitute for experimental design in agricultural experiments? Crop Science. 59:44–53.

Brandariz SP, Bernardo R. 2019. Small ad hoc versus large general training populations for genomewide selection in maize biparental crosses. Theo. Appl. Genet.. 132:347–353.

Burgueño J, Cadena A, Crossa J, Banziger M, Gilmour A, Cullis B. 2000. User's guide for spatial analysis of field variety trials using asreml.

Butler D. 2021. *asreml: Fits the Linear Mixed Model*. R package version 4.1.0.160.

Butler DG, Smith AB, Cullis BR. 2014. On the design of field experiments with correlated treatment effects. Journal of Agricultural, Biological, and Environmental Statistics. 19:539–555.

Calus M, Vandenplas J, Ten Napel J, Veerkamp R. 2016. Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. Journal of dairy science. 99:6403–6419.

Cockerton HM, Li B, Vickerstaff RJ, Eyre CA, Sargent DJ, Armitage AD, Marina-Montes C, Garcia A, Passey A, Simpson DW *et al*. 2019. Identifying verticillium dahliae resistance in strawberry through disease screening of multiple populations and image based phenotyping. Frontiers Plant Sci.. 10:924.

Collard BCY, Mackill DJ. 2007. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. London, Ser. B. 363:557–572.

Covarrubias-Pazaran G. 2016. Genome-assisted prediction of quantitative traits using the r package sommer. PloS One. 11:e0156744.

Covarrubias-Pazaran G. 2022. Changes and faqs for the sommer package.

Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y *et al*. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity. 112:48–60.

Cullis BR, Smith AB, Cocks NA, Butler DG. 2020. The design of early-stage plant breeding trials using genetic relatedness. Journal of Agricultural, Biological and Environmental Statistics. 25:553–578.

Cullis BR, Smith AB, Coombes NE. 2006. On the design of early generation variety trials with correlated data. Journal of agricultural, biological, and environmental statistics. 11:381–393.

Damesa TM, Hartung J, Gowda M, Beyene Y, Das B, Semagn K, Piepho HP. 2019. Comparison of weighted and unweighted stage-wise analysis for genome-wide association studies and genomic selection. Crop Sci.. 59:2572–2584.

Damesa TM, Möhring J, Worku M, Piepho HP. 2017. One step at a time: stage-wise analysis of a series of experiments. Agron. J.. 109:845–857.

de los Campos G, Gianola D, Allison DB. 2010. Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat. Rev. Genet.. 11:880–886.

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

de Los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D. 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS genetics. 9:e1003608.

De Resende MDV, Thompson R, Welham S. 2006. Multivariate spatial statistical analysis of longitudinal data in perennial crops. Revista de Matemática e Estatística, São Paulo,. 24:147–169.

De Villemereuil P, Morrissey MB, Nakagawa S, Schielzeth H. 2018. Fixed-effect variance and the estimation of repeatabilities and heritabilities: issues and solutions. J. Evol. Bio.. 31:621–632.

Demmings EM, Williams BR, Lee CR, Barba P, Yang S, Hwang CF, Reisch BI, Chitwood DH, Londo JP. 2019a. Quantitative trait locus analysis of leaf morphology indicates conserved shape loci in grapevine. Frontiers Plant Sci.. 10:1373.

Demmings EM, Williams BR, Lee CR, Barba P, Yang S, Hwang CF, Reisch BI, Chitwood DH, Londo JP. 2019b. Quantitative trait locus analysis of leaf morphology indicates conserved shape loci in grapevine. Frontiers Plant Sci.. 10:1373.

DeWitt N, Guedira M, Lauer E, Murphy JP, Marshall D, Mergoum M, Johnson J, Holland JB, Brown-Guedira G. 2021. Characterizing the oligogenic architecture of plant growth phenotypes informs genomic selection approaches in a common wheat population. BMC Genomics. 22:1–18.

Dias K, Piepho H, Guimarães L, Guimarães PdO, Parentoni S, Pinto MdO, Noda R, Magalhães J, Guimarães C, Garcia A *et al*. 2020. Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theoretical and Applied Genetics. 133:443–455.

Dudbridge F. 2013. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9:e1003348.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet.. 11:446–450.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrblup. The Plant Genome. 4:250–255.

Endelman JB. 2022. Stagewise: Two-stage analysis of multi-environment trials for genomic selection and gwas. R package version 0.20.

Falconer D, Mackay T. 1996. *Introduction to Quantitative Genetics*. Harlow, Essex, UK. Longmans Green.

Fan M, Hall ML, Roast M, Peters A, Delhey K. 2021. Variability,

heritability and condition-dependence of the multidimensional male colour phenotype in a passerine bird. Heredity. pp. 1–12.

Farfan IDB, De La Fuente GN, Murray SC, Isakeit T, Huang PC, Warburton M, Williams P, Windham GL, Kolomiets M. 2015. Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of maize hybrids in the subtropics. PloS one. 10:e0117737.

Feldmann MJ, Piepho HP, Bridges WC, Knapp SJ. 2021. Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. PLoS genetics. 17:e1009762.

Feldmann MJ, Piepho HP, Knapp SJ. 2022. Average semivariance directly yields accurate estimates of the genomic variance in complex trait analyses. G3. 12:jkac080.

Freebern E, Santos DJ, Fang L, Jiang J, Gaddis KLP, Liu GE, VanRaden PM, Maltecca C, Cole JB, Ma L. 2020. Gwas and fine-mapping of livability and six disease traits in holstein cattle. BMC Genomics. 21:1–11.

Gage JL, Monier B, Giri A, Buckler ES. 2020. Ten years of the maize nested association mapping population: impact, limitations, and future directions. The Plant Cell. 32:2083–2093.

Galli G, Alves FC, Morosini JS, Fritsche-Neto R. 2020. On the usefulness of parental lines gwas for predicting low heritability traits in tropical maize hybrids. PloS one. 15:e0228724.

Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, Segrè AV, McGee K, McGuffog L, Kartsonaki C *et al*. 2010. Common genetic variants and modification of penetrance of brca2-associated breast cancer. PLoS Genet. 6:e1001183.

Gbur EE, Stroup WW, McCarter KS, Durham S, Young LJ, Christman M, West M, Kramer M. 2020. *Analysis of generalized linear mixed models in the agricultural and natural resources sciences*. volume 156. John Wiley & Sons.

Gogel B, Smith A, Cullis B. 2018. Comparison of a one-and two-stage mixed model analysis of australia's national variety trial southern region wheat data. Euphytica. 214:1–21.

González-Barrios P, Díaz-García L, Gutiérrez L. 2019. Mega-environmental design: Using genotype× environment interaction to optimize resources for cultivar testing. Crop Science. 59:1899–1915.

Habier D, Fernando RL, Dekkers JC. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 177:2389–2397.

Habier D, Fernando RL, Garrick DJ. 2013. Genomic blup decoded: a look into the black box of genomic prediction. Genetics. 194:597–607.

Han K, Lee HY, Ro NY, Hur OS, Lee JH, Kwon JK, Kang BC. 2018. Qtl mapping and gwas reveal candidate genes controlling capsaicinoid content in capsicum. Plant Biotech. J.. 16:1546–1558.

Hayes B, Goddard ME. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol.. 33:1–21.

Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet.. 6:e1001139.

Hayes BJ, Visscher PM, Goddard ME. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res.. 91:47–60.

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci.. 50:1681–1690.

Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. Crop Sci.. 49:1–12.

Hemani G, Knott S, Haley C. 2013. An evolutionary perspective on epistasis and the missing heritability. PLoS genetics. 9:e1003295.

Hoefler R, González-Barrios P, Bhatta M, Nunes JA, Berro I, Nalin RS, Borges A, Covarrubias E, Diaz-Garcia L, Quincke M *et al*. 2020. Do spatial designs outperform classic experimental designs? Journal of Agricultural, Biological and Environmental Statistics. 25:523–552.

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

Jensen J, Su G, Madsen P. 2012. Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. BMC Genet.. 13:44.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong Sy, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet.. 42:348–354.

Kim SM, Reinke RF. 2019. A novel resistance gene for bacterial blight in rice, xa43 (t) identified by gwas, confirmed by qtl mapping using a bi-parental population. PLoS One. 14:e0211775.

Knapp SJ. 1998. Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci.. 38:1164–1174.

Konstantinov K, Goddard M. 2020. Application of multivariate single-step snp best linear unbiased predictor model and revised snp list for genomic evaluation of dairy cattle in australia. Journal of Dairy Science. 103:8305–8316.

Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 9:29.

Krause MR, González-Pérez L, Crossa J, Pérez-Rodríguez P, Montesinos-López O, Singh RP, Dreisigacker S, Poland J, Rutkoski J, Sorrells M *et al*. 2019. Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3: Genes, Genomes, Genetics. 9:1231–1247.

Kumar S, Molloy C, Muñoz P, Daetwyler H, Chagné D, Volz R. 2015. Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. G3: Genes, Genomes, Genetics. 5:2711–2718.

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

Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. Science. 265:2037–2048.

Legare ME, Bartlett FS, Frankel WN. 2000. A major effect qtl determined by multiple genes in epileptic el mice. Genome Research. 10:42–48.

Legarra A. 2016. Comparing estimates of genetic variance across different relationship models. Theoretical population biology. 107:26–30.

Legarra A, Lourenco DA, Vitezica ZG. 2018. Bases for genomic prediction. http://genoweb.toulouse.inra.fr/~alegarra/GSIP.pdf. Accessed: 2021-05-24.

Lehermeier C, De los Campos G, Wimmer V, Schön CC. 2017. Genomic variance estimates: With or without disequilibrium covariances? J. Anim. Breed. Genet.. 134:232–241.

Lello L, Avery SG, Tellier L, Vazquez AI, de los Campos G, Hsu SD. 2018. Accurate genomic prediction of human height. Genetics. 210:477–497.

Lello L, Raben TG, Yong SY, Tellier LC, Hsu SD. 2019. Genomic prediction of 16 complex disease risks including heart attack, diabetes, breast and prostate cancer. Scientific reports. 9:1–16.

Li B, VanRaden P, Null D, O'Connell J, Cole J. 2021. Major quanti-

tative trait loci influencing milk production and conformation traits in guernsey dairy cattle detected on bos taurus autosome 19. J. Dairy Sci.. 104:550–560.

Li H, Wang J, Bao Z. 2015. A novel genomic selection method combining gblup and lasso. Genetica. 143:299–304.

Lopdell TJ, Tiplady K, Couldrey C, Johnson TJ, Keehan M, Davis SR, Harris BL, Spelman RJ, Snell RG, Littlejohn MD. 2019. Multiple qtl underlie milk phenotypes at the csf2rb locus. Genet. Sel. Evol.. 51:3.

Lorenz K, Cohen BA. 2012. Small-and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. Genetics. 192:1123–1132.

Luby JJ, Shaw DV. 2001. Does marker-assisted selection make dollars and sense in a fruit breeding program? HortScience. 36:872–879.

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

Mackay TF. 2001. Quantitative trait loci in drosophila. Nat. Rev. Genet.. 2:11–20.

Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet.. 10:565.

Martini JW, Wimmer V, Erbe M, Simianer H. 2016. Epistasis and covariance: how gene interaction translates into genomic relationship. Theo. Appl. Genet.. 129:963–976.

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

Meuwissen T, Hayes B, Goddard M. 2016. Genomic selection: A paradigm shift in animal breeding. Animal Frontiers. 6:6–14.

Moehring J, Williams ER, Piepho HP. 2014. Efficiency of augmented p-rep designs in multi-environmental trials. Theoretical and applied genetics. 127:1049–1060.

Möhring J, Piepho H. 2009. Comparison of weighting in two-stage analyses of series of experiments. Crop Sci. 49:1988.

Moore JH, Williams SM. 2005. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays. 27:637–646.

Nietlisbach P, Keller LF, Postma E. 2016. Genetic variance components and heritability of multiallelic heterozygosity under inbreeding. Heredity. 116:1–11.

Nishio M, Satoh M. 2014. Including dominance effects in the genomic blup method for genomic evaluation. PloS one. 9:e85792.

Oldroyd BP. 2012. Domestication of honey bees was associated with expansion of genetic diversity. Molecular ecology. 21:4409–4411.

Piepho H, Möhring J, Melchinger A, Büchse A. 2008. Blup for phenotypic selection in plant breeding and variety testing. Euphytica. 161:209–228.

Piepho HP. 2009. Ridge regression and extensions for genomewide selection in maize. Crop Sci.. 49:1165–1176.

Piepho HP. 2019. A coefficient of determination (R^2) for generalized linear mixed models. Biom. J.. 61:860–872.

Piepho HP, Moehring J, Schulz-Streeck T, Ogutu JO. 2012. A stage-wise approach for the analysis of multi-environment trials. Biom. J.. 54:844–860.

Pincot DD, Feldmann MJ, Hardigan MA, Vachev MV, Henry PM, Gordon TR, Bjornson M, Rodriguez A, Cobo N, Famula RA *et al*. 2022. Novel fusarium wilt resistance genes uncovered in natural and cultivated strawberry populations are found on three non-homoeologous chromosomes. Theoretical and Applied Genetics. pp. 1–25.

Pincot DD, Hardigan MA, Cole GS, Famula RA, Henry PM, Gordon TR, Knapp SJ. 2020. Accuracy of genomic selection and long-term genetic gain for resistance to verticillium wilt in strawberry. The Plant Genome. 13:e20054.

Pincot DD, Poorten TJ, Hardigan MA, Harshman JM, Acharya CB, Cole GS, Gordon TR, Stueven M, Edger PP, Knapp SJ. 2018. Genome-wide association mapping uncovers fw1, a dominant gene conferring resistance to fusarium wilt in strawberry. G3: Genes, Genomes, Genet.. 8:1817–1828.

Potti J, Canal D. 2011. Heritability and genetic correlation between the sexes in a songbird sexual ornament. Heredity. 106:945–954.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ricard A, Danvy S, Legarra A. 2013. Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in french show-jumping horses. Journal of Animal Science. 91:1076–1085.

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

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PH. 2018. Correcting for spatial heterogeneity in plant breeding experiments with p-splines. Spatial Statistics. 23:52–71.

Roff D, Fairbairn D. 2015. Bias in the heritability of preference and its potential impact on the evolution of mate choice. Heredity. 114:404–412.

Rutkoski JE, Poland JA, Singh RP, Huerta-Espino J, Bhavani S, Barbier H, Rouse MN, Jannink JL, Sorrells ME. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. The Plant Genome. 7.

Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. 2014. Large-effect pleiotropic or closely linked qtl segregate within and across ten us cattle breeds. BMC Genomics. 15:1–17.

Schulz-Streeck T, Ogutu JO, Piepho HP. 2013. Comparisons of single-stage and two-stage approaches to genomic selection. Theoretical and applied genetics. 126:69–82.

Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, Bhattarai EK, Molaei M, Freetly HC, Hansen SL *et al*. 2017. Genome-wide association study for feed efficiency and growth traits in us beef cattle. BMC Genomics. 18:1–25.

Searle SR, Casella G, McCulloch C. 1992. *Variance components*. John Wiley & Sons.

Selle ML, Steinsland I, Hickey JM, Gorjanc G. 2019. Flexible modelling of spatial variation in agricultural field trials with the r package inla. Theoretical and Applied Genetics. 132:3277–3293.

Selle ML, Steinsland I, Powell O, Hickey JM, Gorjanc G. 2020. Spatial modelling improves genetic evaluation in smallholder breeding programs. Genetics Selection Evolution. 52:1–17.

Septiningsih EM, Pamplona AM, Sanchez DL, Neeraja CN, Vergara GV, Heuer S, Ismail AM, Mackill DJ. 2009. Development of submergence-tolerant rice cultivars: the sub1 locus and beyond. Annals of Botany. 103:151–160.

Smith A, Cullis B, Thompson R. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics. 57:1138–1147.

Speed D, Balding DJ. 2015. Relatedness in the post-genomic era: is it still useful? Nat. Rev. Genet.. 16:33–44.

Spindel J, Begum H, Akdemir D, Collard B, Redoña E, Jannink J, McCouch S. 2016. Genome-wide prediction models that incorporate de novo gwas are a powerful new tool for tropical rice improvement. Heredity. 116:395–408.

Strandén I, Christensen OF. 2011. Allele coding in genomic evalua-

tion. Genet. Sel. Evol.. 43:1–11.
Strandén I, Mäntysaari EA. 2010. A recipe for multiple trait deregression. Interbull Bulletin. pp. 21–21.
Stroup WW, Milliken GA, Claassen EA, Wolfinger RD. 2018. *SAS for mixed models: introduction and basic applications*. SAS Institute.
Sun C, VanRaden PM, Cole JB, O'Connell JR. 2014. Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. PloS one. 9:e103934.
Swarts K, Bauer E, Glaubitz JC, Ho T, Johnson L, Li Y, Li Y, Miller Z, Romay C, Schön CC *et al*. 2021. Joint analysis of days to flowering reveals independent temperate adaptations in maize. Heredity. 126:929–941.
Tang S, Leon A, Bridges WC, Knapp SJ. 2006. Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Sci.. 46:721–734.
Truong B, Zhou X, Shin J, Li J, van der Werf JH, Le TD, Lee SH. 2020. Efficient polygenic risk scores for biobank scale data by exploiting phenotypes from inferred relatives. Nat. Comms.. 11:1–11.
Ulrich GF, Zemp N, Vorburger C, Boulain H. 2021. Quantitative trait locus analysis of parasitoid counteradaptation to symbiont-conferred resistance. Heredity. pp. 1–14.
Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JNP, Mott R, Flint J. 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet.. 38:879–887.
VanRaden PM. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci.. 91:4414–4423.
Vasconcellos RC, Oraguzie OB, Soler A, Arkwazee H, Myers JR, Ferreira JJ, Song Q, McClean P, Miklas PN. 2017. Meta-qtl for resistance to white mold in common bean. PloS One. 12:e0171685.
Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. Am. J. Hum. Genet.. 90:7–24.
Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era—concepts and misconceptions. Nat. Rev. Genet.. 9:255–266.
Visscher PM, Macgregor S, Benyamin B, Zhu G, Gordon S, Medland S, Hill WG, Hottenga JJ, Willemsen G, Boomsma DI *et al*. 2007. Genome partitioning of genetic variation for height from 11,214 sibling pairs. The American Journal of Human Genetics. 81:1104–1110.
Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG. 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet.. 2.
Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 2017. 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet.. 101:5–22.
Visscher PM, Yang J, Goddard ME. 2010. A commentary on 'common snps explain a large proportion of the heritability for human height'by yang et al.(2010). Twin Research and Human Genetics. 13:517–524.
Vitezica ZG, Legarra A, Toro MA, Varona L. 2017. Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. Genetics. 206:1297–1307.
Vitezica ZG, Varona L, Legarra A. 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics. 195:1223–1230.
Walsh B, Lynch M. 2018. *Evolution and selection of quantitative traits*. Oxford University Press.
Walsh JT, Garnier S, Linksvayer TA. 2020. Ant collective behavior is heritable and shaped by selection. The American Naturalist. 196:541–554.
Wassom JJ, Mikkelineni V, Bohn MO, Rocheford TR. 2008. qtl for fatty acid composition of maize kernel oil in illinois high oil× b73 backcross-derived lines. Crop Sci.. 48:69–78.
Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR. 2010. A comparison of analysis methods for late-stage variety evaluation trials. Australian & New Zealand Journal of Statistics. 52:125–149.
Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
Wolfe MD, Chan AW, Kulakow P, Rabbi I *et al*. 2021. Genomic mating in outbred species: predicting cross usefulness with additive and total genetic covariance matrices. BioRxiv. .
Wray NR, Kemper KE, Hayes BJ, Goddard ME, Visscher PM. 2019. Complex trait prediction from genome data: contrasting ebv in livestock to prs in humans: genomic prediction. Genetics. 211:1131–1141.
Xiang T, Christensen OF, Vitezica ZG, Legarra A. 2018. Genomic model with correlation between additive and dominance effects. Genetics. 209:711–723.
Xin F, Zhu T, Wei S, Han Y, Zhao Y, Zhang D, Ma L, Ding Q. 2020. Qtl mapping of kernel traits and validation of a major qtl for kernel length-width ratio using snp and bulked segregant analysis in wheat. Scientific reports. 10:1–12.
Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW. 2010. Common snps explain a large proportion of the heritability for human height. Nat. Genet.. 42:565.
Yang J, Mezmouk S, Baumgarten A, Buckler ES, Guill KE, McMullen MD, Mumm RH, Ross-Ibarra J. 2017. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLoS Genet.. 13:e1007019.
Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet.. 38:203–208.
Zas R, Sampedro L. 2015. Heritability of seed weight in maritime pine, a relevant trait in the transmission of environmental maternal effects. Heredity. 114:116–124.
Zhang H, Yin L, Wang M, Yuan X, Liu X. 2019. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. Frontiers in genetics. 10:189.
Zhang J, Liu F, Reif JC, Jiang Y. 2021. On the use of gblup and its extension for gwas with additive and epistatic effects. G3 Genes| Genomes| Genetics. .
Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, Li J, Simianer H. 2014. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PloS one. 9:e93017.