Heritability in Plant Breeding on a Genotype-Difference Basis

Paul Schmidt,* Jens Hartung,* Jörn Bennewitz,† and Hans-Peter Piepho*,1
*Biostatistics Unit, Institute of Crop Science and †Institute of Animal Science, University of Hohenheim, Stuttgart, 70599 ,Germany
ORCID ID: 0000-0003-1528-2082 (P.S.)

ABSTRACT In plant breeding, heritability is often calculated (i) as a measure of precision of trials and/or (ii) to compute the response to selection. It is usually estimated on an entry-mean basis, since *the phenotype* is usually an aggregated value, as genotypes are replicated in trials, which stands in contrast with animal breeding and human genetics. When this was first proposed, assumptions such as balanced data and independent genotypic effects were made that are often violated in modern plant breeding trials/analyses. Due to this, multiple alternative methods have been proposed, aiming to generalize heritability on an entry-mean basis. In this study, we propose an extension of the concept for heritability on an entry-mean to an entry-difference basis, which allows for more detailed insight and is more meaningful in the context of selection in plant breeding, because the correlation among entry means can be accounted for. We show that under certain circumstances our method reduces to other popular generalized methods for heritability estimation on an entry-mean basis. The approach is exemplified via four examples that show different levels of complexity, where we compare six methods for heritability estimation on an entry-mean basis to our approach (example codes: https://github.com/PaulSchmidtGit/Heritability). Results suggest that heritability on an entry-difference basis is a well-suited alternative for obtaining an overall heritability estimate, and in addition provides one heritability per genotype as well as one per difference between genotypes.

KEYWORDS heritability; plant breeding; mixed models

THE idea behind measures of heritability as used in plant breeding is relatively simple: they express the proportion of the total phenotypic variance that is attributable to the average effects of genes, which in turn determines the degree of resemblance between relatives (Falconer and Mackay 2005, chapter 10). Lourenço *et al.* (2017) phrase it as "the extent to which a phenotype is genetically determined." A phenotype is the composite of an organism's observable traits. It results from (i) the expression of the organism's genotype, (ii) the influence of environmental factors, and (iii) the interactions between both. Thus, heritability investigates the relationship between observed/phenotypic values with phenotypic variance σ_p^2 and their respective underlying true genotypic values ($\boldsymbol{g}$) with genotypic variance σ_g^2. We can furthermore dissect $\boldsymbol{g}$ and σ_g^2 into additive, dominance, and epistasis components to extract average effects of alleles and breeding values ($\boldsymbol{a}$), with variance σ_a^2. Depending on whether genotypic values or breeding values are considered, we refer to broad-sense heritability (H^2) or narrow-sense heritability (h^2), respectively (*e.g.*, Xu 2013). Accordingly, there is a clear distinction between H^2 and h^2. However, note that the approach presented in this article is relevant for both measures, which is why we will refer to *heritability* in general throughout this article, unless it is necessary to refer to one of the two specifically. The true genotypic/breeding values (and their variances) are of course unknown, but can be estimated/predicted from phenotypic data.

Piepho and Möhring (2007) pointed out that heritability was originally proposed in the context of animal breeding where the observations are collected from individuals, and are used in mixed models to estimate genetic parameters or breeding values by utilizing the additive genetic relationship of all individuals. It is important to note that in animal breeding, each individual usually has its own and unique genotype. Hence, an *animal* and a *genotype* refer to the same thing: a single individual. Naturally, this means that

doi: https://doi.org/10.1534/genetics.119.302134
Manuscript received March 18, 2019; accepted for publication June 17, 2019; published Early Online June 27, 2019.
Supplemental material available at FigShare: https://doi.org/10.25386/genetics.7370168.
[1]Corresponding author: Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Fruwirthstrasse 23, Stuttgart, 70599, Germany. E-mail: piepho@uni-hohenheim.de

genotypes/animals cannot be truly replicated in or across environments. Instead, the same genotype/animal may only be measured repeatedly over time (Mrode and Thompson 2014, chapter 1.3.2) and/or via its progeny/relatives (Mrode and Thompson 2014, chapter 3.4). Repeated observations of individuals are either modeled by adding a random uncorrelated permanent environment effect to the mixed model or by treating repeated observations as different traits.

However, in plant breeding, we usually have crop cultivars/lines/varieties that are represented by a large group of individual plants with exactly the same genotype. This is mainly because most crop cultivars are clones, inbred lines, or hybrids. Thus, in contrast to animal breeding, a *genotype* in plant breeding does not refer to just a single individual, but to all individual plants belonging to the same cultivar and thus sharing the same genotype. In other words, a genotype in animal breeding is a single, genetically unique individual, whereas a genotype in plant breeding is usually a group of genetically identical individuals (though there are exceptions; see the *Discussion*). This allows for true replication of a genotype in and across multiple environments, even at a single time point and thus without repeated measurements in time. Notice that plant breeders may also decide to have repeated measures over time and/or account for observations of related genotypes, but the salient feature of plant genotypes as opposed to animal genotypes is the availability of true replication.

As a consequence, the multiple observations of the same cultivar are usually aggregated to obtain a single phenotypic value. Levels of aggregation range from individual plants to means of many plants, with the same genotype tested across several locations and years in designed experiments. The need for this aggregation in plant breeding results is an additional step that is not necessary in animal breeding. Since we do not consider analyses of individual plants in this article, the phenotypic value for a genotype will always be assumed to be some sort of mean value. In this context, heritability is referred to as heritability on an entry-mean (*i.e.*, genotype-mean) basis, which stands in contrast with heritability in animal breeding and human genetics. This difference can also be deduced from the level of the heritability, which is usually much lower for complex traits in humans and animals compared to the entry-mean heritability of complex traits in plant breeding. Nowadays, the aggregation of the multiple observations into a mean phenotypic value per genotype is usually either done by best linear unbiased estimation (BLUE) or best linear unbiased prediction (BLUP). Note that the use of BLUPs here is not the same as in the framework of breeding value estimation typical for animal breeders (see *Materials and Methods*). The choice between BLUE and BLUP in plant breeding depends on the goal of the analysis and both are commonly used in practice [see Piepho *et al.* (2008), for more information and a review on the decision-making].

In early work on measures of heritability for plant breeding trials (*e.g.*, Hanson and Robinson 1963), it was assumed that (i) the trial design is completely balanced/orthogonal, (ii) genotypic effects are independent, and (iii) variances and covariances are constant, so that the arithmetic mean across all observations for a genotype is the natural choice to aggregate to a single phenotypic value. We will from now on refer to such a scenario as the *simple*, *balanced setting*. Only in a simple, balanced setting does heritability have multiple direct interpretations. Specifically, H^2 is (I) the ratio of genetic variance to phenotypic variance, (II) the slope coefficient of a linear regression of the genotypic values on the phenotypic values [γ_{ap}], (III) the squared correlation between genotypic values and phenotypic values [r^2_{ap}], and (IV) the ratio of response to selection [ΔG] to selection differential [S] [see Falconer and Mackay (2005), pages 160 and 186]. For the simple, balanced setting, these interpretations can be represented by

$$H^2 = \overbrace{\frac{\sigma^2_g}{\sigma^2_p}}^{I} = \overbrace{\gamma_{gp}}^{II} = \overbrace{r^2_{gp}}^{III} = \overbrace{\frac{\Delta G}{S}}^{IV}, \tag{1}$$

where ΔG is often also referred to as the genetic gain and S is the mean phenotypic value of the selected genotypes, expressed as a deviation from the population mean. Note that when heritability is estimated via (1)-IV, it is referred to as the realized heritability, as it requires ΔG to be known, which means that selection has already occurred and that the offspring have been observed. By rewriting (1)-IV, we obtain the more directly interpretable *breeder's equation*:

$$\Delta G = H^2 S, \tag{2}$$

which allows the prediction of the expected response to selection and is the key reason why heritability plays such an important role in plant (and animal) breeding. Hanson and Robinson (1963, page 128) get to the heart of it: "Heritability has value primarily as a method of quantifying the concept of whether progress from selection for a plant character is relatively easy or difficult to make in a breeding program. A plant breeder, through experience, can perhaps rate a series of characters on their response to selection. Heritability gives a numerical description of this concept."

The standard method of estimating H^2 for phenotypic mean values is by plugging estimates for σ^2_g and σ^2_p into (1)-I. Given the simple, balanced setting, these variances can easily be estimated from mean squares and their respective expected mean squares of a conventional ANOVA (Yan 2014, chapter 1). For a single environment, where n_g genotypes are tested in n_r replicates in a completely randomized design (CRD), the observed data may be modeled as

$$y_{ik} = \mu + g_i + \varepsilon_{ik}, \tag{3}$$

where y_{ik} is the kth observation of the ith genotype, μ is the intercept, g_i is the effect for the ith genotype, and ε_{ik} is the plot error effect corresponding to y_{ik}. The phenotypic value of

genotype i can be obtained as the arithmetic mean across replicates, $\bar{y}_{i\cdot}$, which is also the BLUE in this simple, balanced setting. Assuming that g_i and ε_{ik} are random, with independent distribution, zero mean, and variances σ^2_g and σ^2_ε, respectively, we can calculate the phenotypic variance of $\bar{y}_{i\cdot}$ as:

$$\sigma^2_p = \sigma^2_g + \frac{\sigma^2_\varepsilon}{n_r}. \tag{4}$$

When genotypes are tested in a multienvironment trial (MET), where an environment denotes a year-by-location combination and CRDs are used at all environments, the observed data may be modeled as

$$y_{ikt} = \mu + g_i + e_t + (ge)_{it} + \varepsilon_{ikt}, \tag{5}$$

where y_{ikt} is the kth observation of the ith genotype at the tth environment, μ is the intercept, g_i is the main effect for the ith genotype, e_t is the main effect for the tth environment, $(ge)_{it}$ is the itth genotype-by-environment interaction effect, and ε_{ikt} is the plot error effect corresponding to y_{ikt}. Again, since we have a simple, balanced setting, we can obtain phenotypic values as $\bar{y}_{i\cdot\cdot} = BLUE(\mu + g_i)$. Assuming that g_i, e_t, $(ge)_{it}$, and ε_{ikt} are random, with independent distributions, zero mean, and variances σ^2_g, σ^2_e, σ^2_{ge} and σ^2_ε, respectively, one may then calculate the phenotypic variance as

$$\sigma^2_p = \sigma^2_g + \frac{\sigma^2_{ge}}{n_e} + \frac{\sigma^2_\varepsilon}{n_e n_r}, \tag{6}$$

where n_e is the number of environments (Hallauer *et al.* 2010, page 59). This approach assumes a single variance for genotype-by-environment interactions $(g \times e)$, even when multiple locations were tested across multiple years. In the latter case, one may instead model the environmental effects via separate year, and location main and interaction effects as

$$y_{ikmq} = \mu + g_i + f_m + l_q + (fl)_{mq} + (gf)_{im} + (gl)_{iq} + (gfl)_{imq} + \varepsilon_{ikmq}, \tag{7}$$

where y_{ikmq} is the kth observation of the ith genotype in the mth year and the qth location, μ is the intercept, g_i is the main effect of the ith genotype, f_m is the main effect for the mth year, l_q is the main effect for the qth location, $(fl)_{mq}$ is the mqth year-by-location interaction effect, $(gf)_{im}$ is the imth genotype-by-year interaction effect, $(gl)_{iq}$ is the iqth genotype-by-location interaction effect, $(gfl)_{imq}$ is the imqth genotype-by-year-by-location interaction effect, and ε_{ikmq} is the plot error effect corresponding to y_{ikmq}. Once more, we have the phenotypic values for each genotype as $\bar{y}_{i\cdots} = BLUE(\mu + g_i)$. Assuming g_i, f_m, l_q, $(fl)_{mq}$, $(gf)_{im}$, $(gl)_{iq}$, $(gfl)_{imq}$, and ε_{ikmq} are random, with independent distribution, zero mean, and variances σ^2_g, σ^2_f, σ^2_l, σ^2_{fl}, σ^2_{gf}, σ^2_{gl}, σ^2_{gfl}, and σ^2_ε, respectively, we can calculate the phenotypic variance as:

$$\sigma^2_p = \sigma^2_g + \frac{\sigma^2_{gl}}{n_l} + \frac{\sigma^2_{gf}}{n_f} + \frac{\sigma^2_{gfl}}{n_l n_f} + \frac{\sigma^2_\varepsilon}{n_l n_f n_r}, \tag{8}$$

where n_f and n_l are the number of years and locations, respectively (Becker 2011; Yan 2014). Note that when a MET is conducted either at multiple locations within a single year or at a single location but across multiple years, (8) simplifies to (6). We will refer to heritability estimates involving (4), (6), or (8) and (1)-I as the *standard heritability* (H^2_{Std}).

By examining (4), (6), and (8), it can be seen that the calculation of σ^2_p always involves the genotypic variance and all $g \times e$ variances. However, it does not incorporate purely environmental variance components, such as σ^2_e in the context of (6), and σ^2_f, σ^2_l, and σ^2_{fl} in the context of (8) [but see Yan (2014, page 3)]. Hence, referring to σ^2_p as the phenotypic variance may be considered misleading, as it neglects the purely environmental effects and therefore is definitely not the variance of phenotypic mean values. However, a key feature of the perspective on heritability put forward in this paper is that σ^2_p, as defined in (3)–(8), actually coincides with half the variance of a difference between two phenotypic mean values in the simple, balanced setting. Taking (5) as an example, the variance of a genotype mean is

$$var(\bar{y}_{i\cdot\cdot}) = \sigma^2_g + \frac{\sigma^2_e}{n_e} + \frac{\sigma^2_{ge}}{n_e} + \frac{\sigma^2_\varepsilon}{n_e n_r} \neq \sigma^2_p, \tag{9}$$

whereas the variance of a difference between means of genotypes i and j is $var(\bar{y}_{i\cdot\cdot}) + var(\bar{y}_{j\cdot\cdot}) - 2cov(\bar{y}_{i\cdot\cdot}, \bar{y}_{j\cdot\cdot})$, which in this simple, balanced setting reduces to

$$var(\bar{y}_{i\cdot\cdot} - \bar{y}_{j\cdot\cdot}) = 2\left(\sigma^2_g + \frac{\sigma^2_{ge}}{n_e} + \frac{\sigma^2_\varepsilon}{n_e n_r}\right) = 2\sigma^2_p. \tag{10}$$

It can be argued that this focus on genotype "differences" makes more sense than a focus on genotype means or effects themselves, since the goal is to select the best-performing genotype(s) to maximize ΔG and the ranking of genotypes is uniquely determined by all pairwise differences, whereas the individual genotypic effects and the absolute (mean) performance level as such do not inform about the ranking (Searle *et al.* 1992; Piepho *et al.* 2008). Accordingly, the correct ranking of genotypes and thus the precision of estimating genotype differences is more relevant than the precision of the genotype effect estimates. Furthermore, we found that especially in some older publications, the definitions of heritability refer to differences: Knight (1948) defines heritability as "the portion of the observed variance for which differences in heredity are responsible," while Hanson and Robinson (1963, page 125) state that the "concept of heritability originated as an attempt to describe whether differences actually observed between individuals arose from the differences in genetic makeup between the individuals or

resulted from different environmental forces." Finally, it is no coincidence that two published alternative heritability estimation methods also involve the variance of a difference between genotypes: Cullis *et al.* (2006) proposed to calculate heritability as

$$H^2_{Cullis} = 1 - \frac{\bar{v}^{BLUP}_{\Delta\cdot\cdot}}{2\sigma^2_g}, \tag{11}$$

where $\bar{v}^{BLUP}_{\Delta\cdot\cdot}$ is the mean variance of a difference of two BLUPs for the genotypic effect and σ^2_g is the genotypic variance. The BLUE counterpart was suggested by Piepho and Möhring (2007) as

$$H^2_{Piepho} = \frac{\sigma^2_g}{\sigma^2_g + \bar{v}^{BLUE}_{\Delta\cdot\cdot}/2}, \tag{12}$$

where $\bar{v}^{BLUE}_{\Delta\cdot\cdot}$ is the mean variance of a difference of two genotypic BLUEs and σ^2_g is the genotypic variance. Both measures reduce to H^2_{Std} in the simple, balanced setting.

These considerations suggest that heritability is best defined in terms of pairwise comparisons among genotypes, and this is in fact the key perspective taken in this paper. We aim to provide two things. First, an elaboration of how a heritability on an entry-difference basis (H^2_Δ/h^2_Δ) can be defined. Second, a comparison of how the application of H^2_Δ/h^2_Δ compares to other published methods for estimating heritability on an entry-mean basis. To do so, we give a derivation of H^2_Δ and subsequently apply it alongside six other heritability measures in four example analyses with real data.

Materials and Methods

Mixed model theory

We assume a standard linear mixed model for the observed data vector $\boldsymbol{y}$, which is of the form

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}, \tag{13}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are vectors of fixed and random effects, respectively, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the associated design matrices, and $\boldsymbol{\varepsilon}$ is a vector of random residual errors. The random effects $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ are assumed to be independently distributed as $\boldsymbol{u} \sim MVN(0, \boldsymbol{G})$ and $\boldsymbol{\varepsilon} \sim MVN(0, \boldsymbol{R})$, such that $\boldsymbol{y} \sim MVN(\boldsymbol{X\beta}, \boldsymbol{V})$, where $MVN(.,.)$ denotes the multivariate normal distribution with a mean vector given as the first argument and a variance–covariance matrix as the second. To obtain estimates for $\boldsymbol{\beta}$ as well as predictions for $\boldsymbol{u}$, the mixed model equations (Henderson 1986; Searle *et al.* 1992) can be solved as

$$\begin{bmatrix} BLUE(\boldsymbol{\beta}) \\ BLUP(\boldsymbol{u}) \end{bmatrix} = \boldsymbol{C} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}, \tag{14}$$

where

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix}^{-}. \tag{15}$$

Note that we will explicitly use the words "estimate/estimator" and "predict/predictor" for BLUEs and BLUPs, respectively. In the context of variety trials, genotype effects $\boldsymbol{g}$ may be considered to be fixed or random. The choice between these two options depends on the goal of the analysis and both are commonly used in practice [see Piepho *et al.* (2008) for a review on decision-making as well as Appendix A]. Accordingly, we consider both as complementing cases in this article.

If $\boldsymbol{g}$ is assumed to be a fixed effect, we obtain genotypic BLUEs ($\hat{\boldsymbol{g}}^{BLUE}$) as a subset of $BLUE(\boldsymbol{\beta})$ with $var(\hat{\boldsymbol{g}}^{BLUE}|\boldsymbol{g}) = \boldsymbol{C}_{11(\boldsymbol{g})}$ as the submatrix of $\boldsymbol{C}_{11}$ associated with $\boldsymbol{g}$. If $\boldsymbol{g}$ is assumed to be a random effect, we obtain genotypic BLUPs ($\hat{\boldsymbol{g}}^{BLUP}$) as a subset of $BLUP(\boldsymbol{u})$ with $var(\hat{\boldsymbol{g}}^{BLUP}|\boldsymbol{g}) = \boldsymbol{C}_{22(\boldsymbol{g})}$ as the submatrix of $\boldsymbol{C}_{22}$ associated with $\boldsymbol{g}$. Note that in the case where $\boldsymbol{g}$ is the only fixed/random effect in the model, its estimators/predictors and their variance matrices are not subsets of, but equal to, $BLUE(\boldsymbol{\beta})/BLUP(\boldsymbol{u})$ and $\boldsymbol{C}_{11}/\boldsymbol{C}_{22}$, respectively. Notice further that BLUPs can always be computed from entry means (BLUEs) using a stage-wise approach (see Appendix A). Only in the simple, balanced setting is the correlation of genotypic BLUEs and BLUPs equal to unity, and thus the ranking of genotypes does not change. On a side note, it should be mentioned that one may want to use linear combinations of $BLUE(\boldsymbol{\beta})$, such as, *e.g.*, *least squares means* (SAS Institute Inc. 2013) or *estimated marginal means* (Searle *et al.* 2012) as phenotypic mean values instead. While this certainly results in different estimated values and (co)variances, it does not affect the heritability measures proposed in this article. This is because for all methods that rely on $\hat{\boldsymbol{g}}^{BLUE}$, the focus lies on differences, and any additional variance brought in via the linear combination with nongenetic effects cancels out when computing differences.

Example data sets

To exemplify the approach proposed in this article, we use four examples that show different levels of complexity. In examples 1 and 2, a single-environment data set (Wright 2017) is analyzed, while in example 3 a subset (*i.e.*, a single environment) of the MET data set in example 4 (Hadasch *et al.* 2016) is analyzed.

Example 1: The R-package *agridat* (Wright 2017) provides yield data from an oat trial at a single environment. The trial had 24 genotypes and was laid out as an α-design, with three complete replicates and six incomplete blocks of size four within each replicate. Yet, in this first example, we will analyze the data as if the trial were laid out as CRD, and thus ignore the information about the complete replicates and incomplete blocks. Therefore, the model is simply

$$y_{ik} = \mu + g_i + \varepsilon_{ik}, \tag{16}$$

where y_{ik} is the kth observation of the ith genotype, μ is the intercept, g_i is the effect for the ith genotype, and ε_{ik} is the plot error effect corresponding to y_{ik} with $var(\boldsymbol{\varepsilon}) = \boldsymbol{I}_{72}\sigma^2_\varepsilon$. Accordingly, this example is a simple, balanced setting.

Example 2: Here, we analyzed the same data set as in example 1, but included effects accounting for the trial's design in the model

$$y_{iko} = \mu + g_i + r_k + b_{ko} + \varepsilon_{iko}, \tag{17}$$

where y_{iko} is the observation of the ith genotype in the oth block of the kth replicate, μ is the intercept, g_i is the effect for the ith genotype, r_k is the effect for the kth replicate, b_{ko} is the effect for the oth block in the kth replicate, and e_{iko} is the plot error effect corresponding to y_{iko} with $var(\boldsymbol{\varepsilon}) = \boldsymbol{I}_{72}\sigma^2_{\varepsilon}$. To recover interblock information, b_{ko} was modeled as random with $var(\boldsymbol{b}) = \boldsymbol{I}_{18}\sigma^2_b$, whereas r_k was taken as fixed.

Example 3: This data set was taken from Hadasch *et al.* (2016). It comprises plot data of 89 lettuce varieties (*i.e.*, "geno1"–"geno89") tested at three environments (*i.e.*, "env1"–"env3"), each laid out as a randomized complete block design (RCBD). The measured trait was resistance to downy mildew scored on a scale ranging from 0 to 5. We note that despite the acknowledged discreteness of this response variable, residual plots did not indicate an appreciable deviation of residuals from the normal distribution. All 89 varieties were genotyped with a total of 300 markers [*i.e.*, 95 single nucleotide polymorphisms and 205 amplified fragment length polymorphism markers, see Hayes *et al.* (2014) for details] so that a marker matrix $\boldsymbol{M}_{89\times 300}$ was provided. The biallelic marker M_{iw} for the ith genotype, and the wth marker with alleles A_1 (*i.e.*, the reference allele) and A_2, was coded as 1 for A_1A_1, -1 for A_2A_2, and 0 for A_1A_2 and A_2A_1. The kinship matrix $\boldsymbol{K}$ was obtained as

$$\boldsymbol{K} = \boldsymbol{M}\boldsymbol{M}'. \tag{18}$$

We note that for our purposes, the scaling of $\boldsymbol{M}$ is largely immaterial, but see Appendix D for different scaling options of $\boldsymbol{M}$ and the associated interpretations of genetic variance. In this example, we analyzed only the second environment ("env2") and thus have 89 genotypes at a single environment laid out as RCBD with three replicates, which we modeled as

$$y_{ik} = \mu + a_i + r_k + \varepsilon_{ik}, \tag{19}$$

where y_{ik} is the observation of the ith genotype in the kth replicate, μ is the intercept, a_i is the additive effect for the ith genotype, r_k is the effect for the kth replicate, and ε_{ik} is the plot error effect corresponding to y_{ik} with $var(\boldsymbol{\varepsilon}) = \boldsymbol{I}_{267}\sigma^2_{\varepsilon}$. For simplicity, we assume here that there are no genetic effects for dominance or epistasis, but an extension of the model to cover this case is straightforward (Wellmann and Bennewitz 2011; Dias *et al.* 2018; Viana *et al.* 2018).

Example 4: In contrast to example 3, we here analyzed all three environments jointly. The data set is not completely balanced, as (i) environment env1 has two complete replicates, while environments env2 and env3 have three, (ii) there are no observations for genotypes geno38 and geno49 at environment env1, and (iii) there are no observations for genotype geno81 at either env1 or env3. The linear mixed model used for this analysis was

$$y_{ikt} = \mu + a_i + e_t + (ge)_{it} + r_{kt} + \varepsilon_{ikt}, \tag{20}$$

where y_{ikt} is the observation of the ith genotype in the kth replicate at the tth environment, μ is the intercept, a_i is the additive effect for the ith genotype, e_t is the effect for the tth environment, $(ge)_{it}$ is the itth genotype-by-environment interaction, r_{kt} is the effect for the kth replicate at the tth environment, and ε_{ikt} is the plot error effect corresponding to y_{ikt}. We allowed for heterogeneous error variances between environments, assuming $var(\boldsymbol{\varepsilon}) = \oplus_{t=1}^{n_e}\boldsymbol{I}_{\boldsymbol{n}_{obs(t)}}\sigma^2_{\varepsilon_t}$, where $\oplus$ is the direct sum operator, and $n_{obs(t)}$ and $\sigma^2_{\varepsilon_t}$ are the number of observations and plot error variance at the tth environment, respectively. Finally, e_t and r_k were taken as fixed effects, while $(ge)_{it}$ was taken as random with $var(\boldsymbol{ge}) = \boldsymbol{I}_{263}\sigma^2_{ge}$.

Modeling of genotypic main effect in all examples: Note that some of the heritability measures used in this article require estimates from both the model with a fixed genotype main effect as well as the model with a random genotype main effect. For example, H^2_{Piepho} in (12) needs an estimate for $\bar{v}^{BLUE}_{\Delta\cdot\cdot}$ as well as for σ^2_g. Thus, in this article we always fit both potential models in parallel, and accordingly obtain both $\hat{\boldsymbol{g}}^{BLUE}$ and $\hat{\boldsymbol{g}}^{BLUP}$ for the same example. For models where $\boldsymbol{g}$ is assumed to be a random effect, we assume $var(\boldsymbol{g}) = \boldsymbol{I}_{\boldsymbol{n_g}}\sigma^2_g$. Additionally, we make use of the marker information in examples 3 and 4, and thus estimate additive genotypic effects $\boldsymbol{a}$ by assuming $var(\boldsymbol{a}) = \boldsymbol{K}\sigma^2_a$.

Further, notice that for each example the variance components in the model with the fixed genotype main effect are fixed to the corresponding estimates in the model where $\boldsymbol{g}$ is assumed to be a random effect. This is done to rule out the potential influence of an estimation error on variance component estimation between the two models (Schmidt *et al.* 2019).

Pairwise heritability

As shown in (1), heritability can be interpreted and thus estimated in multiple ways. To derive a *pairwise heritability* or *heritability on an entry-difference basis* ($H^2_{\Delta}/h^2_{\Delta}$), we can exploit and slightly alter (1)-III so that it applies to phenotypic and genotypic differences instead of their respective mean values. We here first derive H^2_{Δ} under the assumption of independent genotypic effects with constant variance and afterward generalize H^2_{Δ} to correlated genotypic effects with nonconstant (co)variances. For both scenarios, we differentiate between $\hat{\boldsymbol{g}}^{BLUE}$ and $\hat{\boldsymbol{g}}^{BLUP}$.

Independent genotypic effects with constant variance: Given independent genotypic effects with constant variance σ^2_g (as in the simple, balanced setting), and irrespective of

whether $\hat{\boldsymbol{g}}^{BLUE}$ or $\hat{\boldsymbol{g}}^{BLUP}$ is used, the variance of the true difference between genotypes i and j is

$$var(g_i - g_j) = 2\sigma_g^2. \tag{21}$$

As stated before, the conditional variance, or prediction error variance, of the predicted BLUPs given the observed data is $var(\hat{\boldsymbol{g}}^{BLUP}|\boldsymbol{g}) = \boldsymbol{C}_{22(\boldsymbol{g})}$. However, notice that the marginal variance of $\hat{\boldsymbol{g}}^{BLUP}$is $var(\hat{\boldsymbol{g}}^{BLUP}) = \boldsymbol{G}_{(\boldsymbol{g})} - \boldsymbol{C}_{22(\boldsymbol{g})}$, where $\boldsymbol{G}_{(\boldsymbol{g})}$ is the submatrix of $\boldsymbol{G}$ associated with $\boldsymbol{g}$ (see Searle *et al.* 1992, chapter 7.4.d.). As a result, we find that the variance of a difference between two BLUPs of genotypes i and j is

$$var\left(\hat{g}_i^{BLUP} - \hat{g}_j^{BLUP}\right) = 2\sigma_g^2 - v_{\Delta ij}^{BLUP}, \tag{22}$$

where $v_{\Delta ij}^{BLUP}$ is the prediction error variance of a difference between BLUPs of genotypes i and j, and can be obtained via $\boldsymbol{C}_{22(\boldsymbol{g})}$. Furthermore, we find the covariance to be

$$cov\left(g_i - g_j, \hat{g}_i^{BLUP} - \hat{g}_j^{BLUP}\right) = 2\sigma_g^2 - v_{\Delta ij}^{BLUP}. \tag{23}$$

Hence, making use of (21), (22), and (23), the squared correlation between the true difference and its predictor, which is the heritability of the predictor of the difference $(g_i - g_j)$, is found to be

$$H^2_{\Delta ij_{BLUP}} = \left(\frac{2\sigma_g^2 - v_{\Delta ij}^{BLUP}}{\sqrt{2\sigma_g^2\left(2\sigma_g^2 - v_{\Delta ij}^{BLUP}\right)}}\right)^2 = \frac{2\sigma_g^2 - v_{\Delta ij}^{BLUP}}{2\sigma_g^2}. \tag{24}$$

This pairwise heritability is a quantity that explicitly accounts for unbalanced data, is analogous to the coefficient of determination (CD) (Piepho 2019), and can be reported in its own right. Yet, as the number of genotypes increases, the number of genotype differences n_Δ can quickly get very large ($n_\Delta = n_g[n_g - 1]/2$). Thus, it is not a convenient statistic to report. However, the statistic can be averaged per genotype, so that a genotype-specific average heritability is obtained, *i.e.*

$$\bar{H}^2_{\Delta i\cdot_{BLUP}} = \frac{1}{n_g - 1}\sum_{j\neq i} H^2_{\Delta ij_{BLUP}}. \tag{25}$$

Note that (25) is similar in spirit to the repeatability for an individual genotype's BLUP as is commonly reported in animal breeding (Laloë 1993; Mrode and Thompson 2014). Finally, if we go further and instead average (24) across all pairs of genotypes, we obtain a single average pairwise heritability as

$$\bar{H}^2_{\Delta\cdot\cdot_{BLUP}} = \frac{2}{n_g(n_g - 1)}\sum_i\sum_{j<i} H^2_{\Delta ij_{BLUP}}. \tag{26}$$

Hence, $\bar{H}^2_{\Delta\cdot\cdot_{BLUP}}$ has the interpretation of an arithmetic mean of all pairwise heritabilities $H^2_{\Delta ij_{BLUP}}$ and is (given independent genotypic effects with constant variance) the algebraically equivalent simplification of H^2_{Cullis}, since the pairwise heritability in (24) has a constant denominator.

If $\hat{\boldsymbol{g}}^{BLUE}$ is used, we take the marginal variance of a difference between two BLUEs of genotypes i and j as

$$var\left(\hat{g}_i^{BLUE} - \hat{g}_j^{BLUE}\right) = 2\sigma_g^2 + v_{\Delta ij}^{BLUE}, \tag{27}$$

where $v_{\Delta ij}^{BLUE}$ is the variance of a difference between BLUEs of genotypes i and j, and can be obtained via $\boldsymbol{C}_{11(\boldsymbol{g})}$. Moreover,

$$cov\left(g_i - g_j, \hat{g}_i^{BLUE} - \hat{g}_j^{BLUE}\right) = 2\sigma_g^2. \tag{28}$$

Analogous to (24), only this time applying (21), (27), and (28), the squared correlation between the true difference and its estimator, which is the heritability of the estimator of the difference $(g_i - g_j)$, is found to be

$$H^2_{\Delta ij_{BLUE}} = \left(\frac{2\sigma_g^2}{\sqrt{2\sigma_g^2\left(2\sigma_g^2 + v_{\Delta ij}^{BLUE}\right)}}\right)^2 = \frac{2\sigma_g^2}{2\sigma_g^2 + v_{\Delta ij}^{BLUE}}. \tag{29}$$

In accordance with (25) and (26), we could now take the arithmetic mean of $H^2_{\Delta ij_{BLUE}}$ per genotype and across all pairs. However, the inconvenient difference compared to $\bar{H}^2_{\Delta\cdot\cdot_{BLUP}}$ in (24) is that the pairwise heritabilities in (29) do not have a constant denominator. As a result, taking the arithmetic mean across all $H^2_{\Delta ij_{BLUE}}$ does not lead to the algebraically equivalent simplification of H^2_{Piepho}. However, if we take the harmonic mean instead of the arithmetic mean for all pairwise heritabilities as

$$\bar{H}^2_{\Delta\cdot\cdot_{BLUE}} = \frac{n_g(n_g - 1)}{2\sum_i\sum_{i<j}\frac{1}{H^2_{\Delta ij_{BLUE}}}}, \tag{30}$$

the resulting average does indeed coincide with H^2_{Piepho}. Correspondingly, we take the harmonic mean per genotype as:

$$\bar{H}^2_{\Delta i\cdot_{BLUE}} = \frac{n_g - 1}{\sum_{i<j}\frac{1}{H^2_{\Delta ij_{BLUE}}}}. \tag{31}$$

Notice that since we here assume independent genotypic effects with constant variance, both $\bar{H}^2_{\Delta\cdot\cdot_{BLUP}}$ and $\bar{H}^2_{\Delta\cdot\cdot_{BLUE}}$ can, in fact, also be obtained by averaging the numerator and denominator in (24) and (29) separately across pairs (see Appendix B).

Generalization to correlated genotypic effects with nonconstant variances: So far, we have assumed independent genotypic effects with constant variance σ_g^2. Yet, the approach outlined above naturally extends to the case where effects have nonconstant variance, (*i.e.*, $var(g_i) = \sigma^2_{g(i,i)}$) and/or are correlated (*i.e.*, $cov(g_i, g_j) = \sigma_{g(i,j)}$), *e.g.*, due to pedigree- or marker-based kinship. The variance of the true difference between genotypes i and j then becomes

$$var(g_i - g_j) = \sigma^2_{g(i,i)} + \sigma^2_{g(j,j)} - 2\sigma_{g(i,j)}. \tag{32}$$

When BLUP is used, the pairwise heritability in (24) accordingly generalizes to

$$\begin{aligned} H^2_{\Delta ij_{BLUP}} &= \frac{var(g_i - g_j) - v^{BLUP}_{\Delta ij}}{var(g_i - g_j)} \\ &= \frac{\sigma^2_{g(i,i)} + \sigma^2_{g(j,j)} - 2\sigma_{g(i,j)} - v^{BLUP}_{\Delta ij}}{\sigma^2_{g(i,i)} + \sigma^2_{g(j,j)} - 2\sigma_{g(i,j)}}. \end{aligned} \tag{33}$$

When BLUE is used, the pairwise heritability in (29) generalizes to

$$\begin{aligned} H^2_{\Delta ij_{BLUE}} &= \frac{var(g_i - g_j)}{var(g_i - g_j) + v^{BLUE}_{\Delta ij}} \\ &= \frac{\sigma^2_{g(i,i)} + \sigma^2_{g(j,j)} - 2\sigma_{g(i,j)}}{\sigma^2_{g(i,i)} + \sigma^2_{g(j,j)} - 2\sigma_{g(i,j)} + v^{BLUE}_{\Delta ij}}. \end{aligned} \tag{34}$$

These pairwise heritabilities can be averaged as before, *i.e.*, as per Equations (25) and (26) for BLUP, and Equations (31) and (30) for BLUE, to obtain $H^2_{\Delta i\cdot}$ and $H^2_{\Delta\cdot\cdot}$. Unfortunately, the average over all pairwise heritabilities cannot generally be simplified algebraically for either BLUE or BLUP, since in both cases neither the denominator nor the numerator of the pairwise heritability is constant. But a simplification is forthcoming if, as an approximation, we average the numerator and denominator of the pairwise heritability separately (see Appendix C).

Other proposals to compute heritability

Besides H^2_{Std}, H^2_{Cullis} (11), and H^2_{Piepho} (12), we calculated three additional generalized heritability measures that are in common usage.

H² Oakey: An alternative measure of heritability, which is sometimes referred to as *the generalized heritability*, can be traced back to Laloë (1993) and was proposed by Oakey *et al.* (2006) in the context of plant breeding. It has gained popularity due to its ability to account for heterozygosity/covariances in $\boldsymbol{G}_{(\boldsymbol{g})}$ (*e.g.*, Mathews *et al.* 2008; Rodríguez-Álvarez *et al.* 2018). In this approach, contrasts of the true and predicted genotypic effects are defined as $\boldsymbol{c}'\boldsymbol{g}$ and $\boldsymbol{c}'\hat{\boldsymbol{g}}^{BLUP}$, respectively. Note that, unlike for H^2_{Δ} where all differences/contrasts between genotype pairs are considered, the contrast vector $\boldsymbol{c}$ can be any linear combination of genotypic effects where the elements of $\boldsymbol{c}$ sum to 0. This allows different genotypic effects to have different heritabilities, while at the same time reducing to a single scalar quantity. Similar to our derivation, they make use of (1)-III, with the difference that it applies to $\boldsymbol{c}'\boldsymbol{g}$ and $\boldsymbol{c}'\hat{\boldsymbol{g}}^{BLUP}$, and thus represents the squared correlation between the true and predicted genotypic contrasts defined by $\boldsymbol{c}$. The vector $\boldsymbol{c}$ is then chosen so that this squared correlation (*i.e.*, the heritability) is maximized. *Components of the full heritability* are defined as

$$\begin{aligned} \lambda &= \left(\frac{cov(\boldsymbol{c}'\boldsymbol{g},\ \boldsymbol{c}'\hat{\boldsymbol{g}}^{BLUP})}{\sqrt{var(\boldsymbol{c}'\boldsymbol{g})var(\boldsymbol{c}'\hat{\boldsymbol{g}}^{BLUP})}} \right)^2 \\ &= \frac{\boldsymbol{c}'\boldsymbol{G}_{(\boldsymbol{g})}\left(\boldsymbol{I}_{n_g} - \boldsymbol{G}^{-1}_{(\boldsymbol{g})}\boldsymbol{C}_{22(\boldsymbol{g})}\right)\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_{(\boldsymbol{g})}\boldsymbol{c}} = \frac{\boldsymbol{c}'\boldsymbol{G}_{(\boldsymbol{g})}\boldsymbol{D}\boldsymbol{c}}{\boldsymbol{c}'\boldsymbol{G}_{(\boldsymbol{g})}\boldsymbol{c}}, \end{aligned} \tag{35}$$

where $\boldsymbol{c}$ is an eigenvector of the matrix $\boldsymbol{D} = \boldsymbol{I}_{n_g} - \boldsymbol{G}^{-1}_{(\boldsymbol{g})}\boldsymbol{C}_{22(\boldsymbol{g})}$ with constraint $\boldsymbol{c}'\boldsymbol{G}_{(\boldsymbol{g})}\boldsymbol{c} = 1$ and λ is the associated eigenvalue. The largest among the eigenvalues, λ_ζ ($\zeta = \{1, \ldots, n_\lambda\}$), equals the maximized squared correlation of $\boldsymbol{c}'\boldsymbol{g}$ and $\boldsymbol{c}'\hat{\boldsymbol{g}}^{BLUP}$. Note that $n_\lambda = n_g$ and due to constraints on $\hat{\boldsymbol{g}}^{BLUP}$, $n_z < n_g$ eigenvalues will be zero. The generalized heritability is then defined as the mean of all nonzero eigenvalues:

$$H^2_{Oakey} = \frac{\sum_{\zeta=1}^{n_g}\lambda_\zeta}{n_g - n_z} = \frac{\sum_{\zeta=n_z+1}^{n_g}\lambda_\zeta}{n_g - n_z}. \tag{36}$$

In more recent work, Rodríguez-Álvarez *et al.* (2018) show how random spatial variation in plant breeding experiments can be analyzed using tensor product P-splines, which is a two-dimensional smoothing technique. In this context, they denote the trace of $\boldsymbol{D}$ as the *effective dimension* of the genotypic effects and reexpress H^2_{Oakey} as:

$$H^2_{Oakey} = \frac{trace(\boldsymbol{D})}{n_g - n_z}. \tag{37}$$

They point out that the notion of effective dimension is well known in the smoothing context, where it can be interpreted as a complexity measure for a given model and its component effects.

Simulated: This method was proposed by Piepho and Möhring (2007), and is also based on the squared correlation between $\boldsymbol{g}$ and $\hat{\boldsymbol{g}}^{BLUP}$ in (1)-III. It first defines the variance–covariance matrix of all random effects $\boldsymbol{u}$ and the target genotypic effects $\boldsymbol{g}$ as

$$var\begin{pmatrix}\boldsymbol{g}\\ \boldsymbol{u}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G}_{(\boldsymbol{g})} & \boldsymbol{U}\\ \boldsymbol{U}' & \boldsymbol{G}\end{pmatrix}, \tag{38}$$

so that $\boldsymbol{U} = cov(\boldsymbol{g}, \boldsymbol{u})$. They then define $\boldsymbol{\Omega}$ as the variance–covariance matrix of the joint distribution of the true and predicted genotypic effects:

$$\begin{aligned} \boldsymbol{\Omega} &= var\begin{pmatrix}\boldsymbol{g}\\ \hat{\boldsymbol{g}}^{BLUP}\end{pmatrix} \\ &= \begin{pmatrix}\boldsymbol{G}_{(\boldsymbol{g})} & \boldsymbol{F}\boldsymbol{G}^{-1}(\boldsymbol{G} - \boldsymbol{C}_{22})\boldsymbol{G}^{-1}\boldsymbol{F}'\\ \boldsymbol{F}\boldsymbol{G}^{-1}(\boldsymbol{G} - \boldsymbol{C}_{22})\boldsymbol{G}^{-1}\boldsymbol{F}' & \boldsymbol{F}\boldsymbol{G}^{-1}(\boldsymbol{G} - \boldsymbol{C}_{22})\boldsymbol{G}^{-1}\boldsymbol{F}'\end{pmatrix}, \end{aligned} \tag{39}$$

Table 1 Summary of heritability measures

Heritability measure			Accounts for heteroscedasticity/covariances		
Overall	Per genotype	Per pair	In $\boldsymbol{G}_{(\boldsymbol{g})}$	In $\boldsymbol{C}_{11(\boldsymbol{g})}$	In $\boldsymbol{C}_{22(\boldsymbol{g})}$
H^2_{Std}	—	—	No/no	No/no	—
$\bar{H}^2_{\Delta\cdot\cdot_{BLUE}}$	$\bar{H}^2_{\Delta i\cdot_{BLUE}}$	$H^2_{\Delta ij_{BLUE}}$	Yes/yes	Yes/yes	—
H^2_{Piepho}	—	—	No/no	Yes/yes	—
$\bar{H}^2_{\Delta\cdot\cdot_{BLUP}}$	$\bar{H}^2_{\Delta i\cdot_{BLUP}}$	$H^2_{\Delta ij_{BLUP}}$	Yes/yes	—	Yes/yes
H^2_{Cullis}	—	—	No/no	—	Yes/yes
H^2_{Oakey}	—	—	Yes/yes	—	Yes/yes
$\bar{r}^2_{\cdot}$	r^2_i	—	Yes/no	—	Yes/no
H^2_{sim}	—	—	Yes/yes	—	Yes/yes

where $\boldsymbol{F} = \begin{pmatrix} \boldsymbol{G}_{(\boldsymbol{g})} & \boldsymbol{U} \end{pmatrix}$. Subsequently, $\boldsymbol{\Omega}$ can be decomposed as $\boldsymbol{\Omega} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$ by singular value decomposition or Cholesky decomposition. Finally, values for $\boldsymbol{g}$ and $\hat{\boldsymbol{g}}^{BLUP}$ can be simulated (*i.e.*, $\boldsymbol{g}_{\boldsymbol{sim}}$ and $\hat{\boldsymbol{g}}^{\boldsymbol{BLUP}}_{\boldsymbol{sim}}$) for an experiment, with the same design and genotypic relationships as those underlying the actual data $\boldsymbol{y}$, as

$$\boldsymbol{d}_{\boldsymbol{sim}} = \begin{pmatrix} \boldsymbol{g}_{\boldsymbol{sim}} \\ \hat{\boldsymbol{g}}^{\boldsymbol{BLUP}}_{\boldsymbol{sim}} \end{pmatrix} = \boldsymbol{\Gamma}\boldsymbol{z}_{\boldsymbol{sim}}, \tag{40}$$

where $\boldsymbol{z}_{\boldsymbol{sim}}$ is a vector with $2n_g$ simulated random independent standard normal deviates. Thus, for each of the $n_{sim} = \{1, \ \ldots, s\}$ simulation runs, a new vector $\boldsymbol{z}_{\boldsymbol{sim}}$ of independent standard normal deviates is randomly generated, so that we can obtain $\boldsymbol{d}_{\boldsymbol{sim}}$, and compute the squared sample correlation (r^2_s) of $\boldsymbol{g}_{\boldsymbol{sim}}$ and $\hat{\boldsymbol{g}}^{\boldsymbol{BLUP}}_{\boldsymbol{sim}}$. By running a large number of simulations, we can obtain the simulated expected squared correlation of predicted and true genotypic effects as

$$H^2_{sim} = \frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} r^2_s. \tag{41}$$

Notice that even in the *simple*, *balanced setting*, this method is not exactly identical to H^2_{Std}, but it is asymptotically equivalent for an increasing number of genotypes:

$$H^2_{sim} = E\left(r^2_{g,\hat{g}}\right) \approx \left(\frac{cov(\boldsymbol{g}, \hat{\boldsymbol{g}}^{BLUP})}{\sqrt{var(\boldsymbol{g}) var(\hat{\boldsymbol{g}}^{BLUP})}} \right)^2 = H^2_{Std}. \tag{42}$$

It can be argued that H^2_{sim} is preferable over, *e.g.*, H^2_{Piepho} or H^2_{Cullis}, as it captures the entire variance–covariance structure in the data in a more comprehensive manner (Piepho and Möhring 2007; Schmidt *et al.* 2019). Further notice that this approach also allows direct simulation of the response to selection (see Appendix E).

Reliability: In animal breeding and thus mostly in the context of a single observation per genotype, the reliability (r^2), also known as the CD, is a popular statistic expressing the squared correlation between predicted and true breeding value, and is closely related to heritability (Laloë *et al.* 1996; Laloë and Phocas 2003; Kuehn *et al.* 2007; Piepho 2019). We can estimate the reliability for the (genotypic/breeding value of the) ith entry as

$$r^2_i = 1 - \frac{var\left(\hat{g}^{BLUP}_i\right)}{var(g_i)}, \tag{43}$$

where $var(\hat{g}^{BLUP}_i)$ is the ith diagonal element of the $\boldsymbol{C}_{22(\boldsymbol{g})}$ matrix and $var(g_i)$ is ith diagonal element of the $\boldsymbol{G}_{(\boldsymbol{g})}$ matrix (Mrode and Thompson 2014, chapter 9.3.4.). Accordingly, r^2_i does not account for the off-diagonal elements of either $\boldsymbol{C}_{22(\boldsymbol{g})}$ or $\boldsymbol{G}_{(\boldsymbol{g})}$ [but see Mrode and Thompson (2014), Appendix D]. We can further obtain the mean reliability as

$$\bar{r}^2_{\cdot} = \frac{1}{n_g} \sum_{i=1}^{n_g} r^2_i. \tag{44}$$

Computation of heritability measures: As summarized in Table 1, we computed eight overall heritability measures (H^2_{Std},$\bar{H}^2_{\Delta\cdot\cdot_{BLUE}}$, H^2_{Piepho}, $\bar{H}^2_{\Delta\cdot\cdot_{BLUP}}$, H^2_{Cullis}, H^2_{Oakey}, $\bar{r}^2_{\cdot}$, and H^2_{Sim}), three genotype-wise heritability measures ($\bar{H}^2_{\Delta i\cdot_{BLUE}}$, $\bar{H}^2_{\Delta i\cdot_{BLUP}}$, and r^2_i), and two pair-wise heritability measures ($H^2_{\Delta ij_{BLUE}}$ and $H^2_{\Delta ij_{BLUP}}$) for each of the four examples. When assuming $var(\boldsymbol{a}) = \boldsymbol{K}\sigma^2_a$ for examples 3 and 4, we are estimating narrow-sense heritabilities, otherwise broad-sense heritabilities were computed. The former is not possible for H^2_{Std}, H^2_{Cullis}, and H^2_{Piepho}, however, as these measures implicitly assume a constant genotypic variance and no covariances. We therefore based estimates of these heritability measures on the alternative model that assumes $var(\boldsymbol{g}) = \boldsymbol{I}_{\boldsymbol{n_g}}\sigma^2_g$ and accordingly completely ignores marker information. Finally, we used an *ad hoc* solution to compute H^2_{Std} via (6) in example 4, where we had slightly unbalanced data and environment-specific error variance estimates. Although only 86 out of 89 genotypes were present at all three environments, and one environment only had two replicates instead of three, we considered $n_e = n_r = 3$. Thus, we took the respective maximum value, which is often done in practice. Furthermore, we computed the mean error variance as $\sigma^2_{\bar{\varepsilon}} = \sum \sigma^2_{\bar{\varepsilon}_t}/n_e$.

Software: All statistical analyses were done by REML with the mixed model package ASReml-R version 3.0 (Gilmour *et al.*

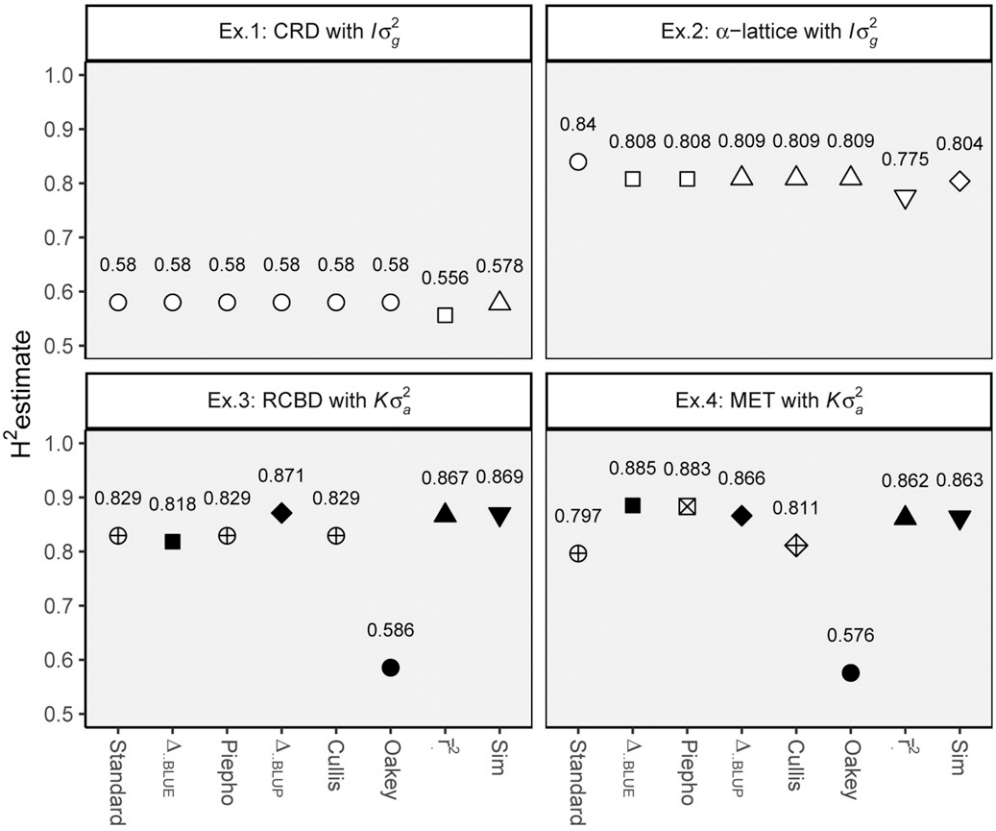

Figure 1 Overall heritability estimates for each method and example. For clearer comparison, (i) estimated values are shown above each symbol and (ii) within each experiment identical values show the same symbol. Black symbols indicate that $var(\boldsymbol{a}) = \boldsymbol{K}\sigma_a^2$ was assumed for the estimation, whereas for white symbols $var(\boldsymbol{g}) = \boldsymbol{I}\sigma_g^2$ was assumed. BLUE, best linear unbiased estimation; BLUP, best linear unbiased prediction; CRD, completely randomized design; Ex, example; RCBD, randomized complete block design.

2009) in the R Statistical Computing Environment (R Core Team 2015) and SAS 9.4 (SAS Institute Inc. 2013).

Data availability

The data used in examples 1 and 2 are the "john.alpha" data set provided in the R package agridat (Wright 2017), which is available at https://cran.r-project.org/web/packages/agridat/index.html. Supplemental material and the data used in examples 3 and 4 are available at https://figshare.com/articles/Lettuce_trial_phenotypic_and_marker_data_/8299493. R code is available at https://github.com/PaulSchmidtGit/Heritability.

Results

In this section, we present overall heritability estimates first, followed by genotype-wise and finally pairwise heritability estimates. The results are accompanied by minimal interpretations, whereas an elaboration on general reasons for differences between estimates of the different methods is provided in the *Discussion*.

Overall heritability

Figure 1 shows estimates for overall heritability obtained via all eight methods for all examples. As expected for the simple, balanced setting of example 1, all estimates are identical, except for the slightly smaller estimates obtained for $\bar{r}_{\cdot}^2$ and H_{sim}^2. A similar picture is found for example 2, with the exception of H_{Std}^2 displaying a notably larger estimate than the rest, which is due to the method's inability to account for the variance of the random incomplete block effect. Examples 3 and 4 show a more heterogeneous picture, with h_{Oakey}^2 displaying estimates that are strikingly lower than the rest.

Genotype-wise heritability

In example 1, estimates for the pairwise heritability measures $H_{\Delta ij_{BLUE}}^2$ and $H_{\Delta ij_{BLUP}}^2$ were constant across all pairs (≈0.580), and hence equal to $\bar{H}_{\Delta i\cdot_{BLUE}}^2$, $\bar{H}_{\Delta i\cdot_{BLUP}}^2$, $\bar{H}_{\Delta \cdot\cdot_{BLUE}}^2$, and $\bar{H}_{\Delta \cdot\cdot_{BLUP}}^2$, which in turn were equal to H_{Std}^2, H_{Cullis}^2, and H_{Oakey}^2. Analogously, the slightly lower r_i^2 estimates were also constant (≈0.556) across genotypes and, therefore, equal to $\bar{r}_{\cdot}^2$.

In example 2, estimates for genotype-wise heritability measures were no longer constant, neither between nor within each method. Instead, each method found two unique, but relatively similar, estimates and each one for half of the genotypes, respectively, *i.e.*, 0.80792 and 0.80801 for $\bar{H}_{\Delta i\cdot_{BLUE}}^2$, 0.80911 and 0.80916 for $\bar{H}_{\Delta i\cdot_{BLUP}}^2$, and 0.77537 and 0.77547 for r_i^2.

Figure 2 details the heritability estimates per genotype for examples 3 and 4. It can be seen that the estimates are much more heterogeneous compared to examples 1 and 2. In fact, all of the 89 genotype-wise estimates per method were different from each other in both example 3 and example 4. In both of the latter, and irrespective of the heritability method, genotypes geno49, geno82, and geno88 showed noticeably lower estimates than the rest. This is related to the fact that the additive variances of these three genotypes stand out with a size of only ≈30% of the average (results not shown).

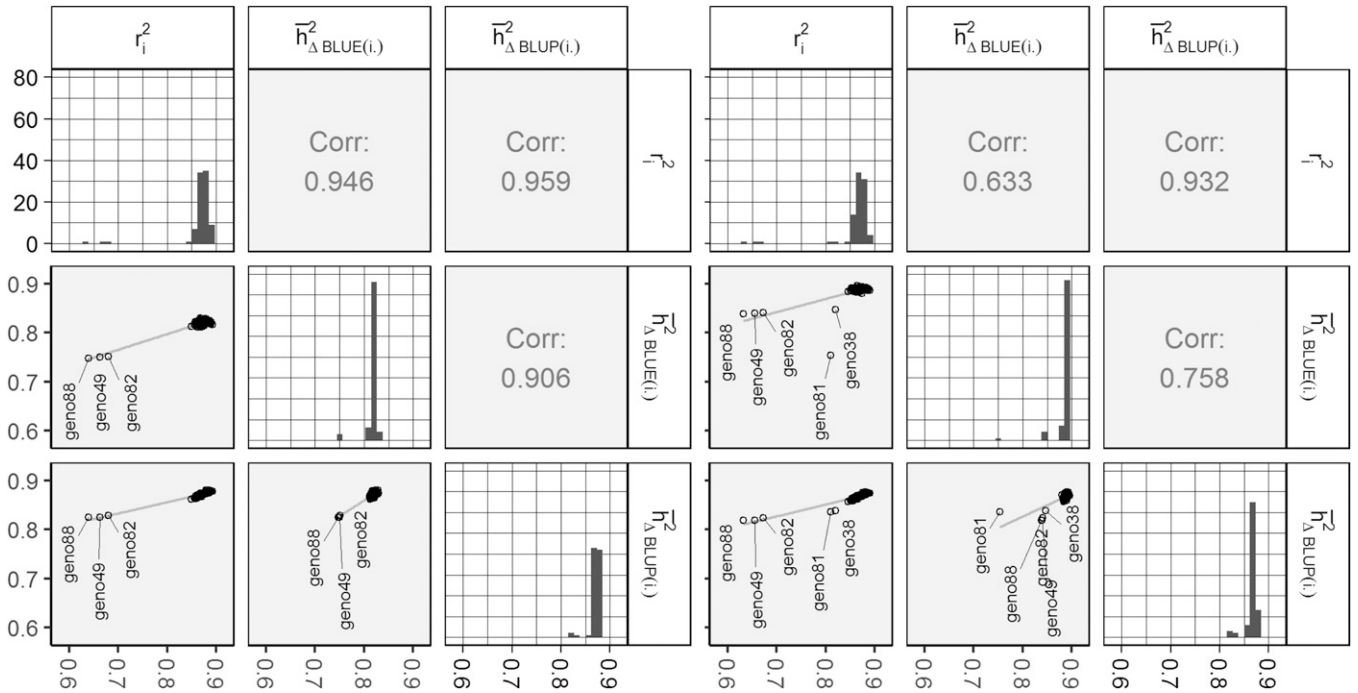

Figure 2 Summary plot for genotype-wise heritability estimates for example 3 (left) and example 4 (right). Plots on the bottom left show scatter plots with simple linear regression lines and genotype-labels for outlying estimates. The plots on the diagonal show histograms for the respective column. The upper right corner displays the Pearson correlation estimate. BLUE, best linear unbiased estimation; BLUP, best linear unbiased prediction; Corr, correlation.

We additionally found relatively low estimates for geno49 and geno81 (*i.e.*, two of the genotypes with missing observations) in example 4 (Figure 2).

In example 3, estimates between the three methods all showed correlations larger than 0.9, with $\bar{h}^2_{\Delta i \cdot BLUP}$ and r_i^2 displaying the highest correlation of ∼0.959. In example 4, the correlation between these two was still high (≈0.932), while the other two correlation estimates dropped below 0.76 (Figure 2).

Pairwise heritability

Figure 3 details the estimated pairwise heritability measures for all examples. As mentioned above, estimates for $H^2_{\Delta ij_{BLUE}}$ and $H^2_{\Delta ij_{BLUP}}$ in example 1 were constant across all $n_\Delta = 276$ genotype pairs (≈0.580). In example 2, estimates for each method split into two clusters, respectively, yet even the overall estimate ranges were relatively small (0.802–0.817 for $H^2_{\Delta ij_{BLUE}}$ and 0.803–0.818 for $H^2_{\Delta ij_{BLUP}}$). The heterogeneous estimates in example 2 are due to the recovery of interblock information via the random incomplete block effect in (17). It results from the varying number of times two genotypes (or even their neighboring genotypes) appear together in the same incomplete block and thus can be compared directly (John and Williams 1995).

Estimates in example 3 mostly form a single cluster for both methods, respectively, ranging from ∼0.70–0.88 for $H^2_{\Delta ij_{BLUE}}$ and 0.80–0.92 for $H^2_{\Delta ij_{BLUP}}$, so that the cluster is located left of/above the diagonal in Figure 3. There are three exceptionally low estimates for both methods, which are those of the differences geno49–geno82, geno49–geno88, and geno82–geno88, *i.e.*, the three genotypes with relatively small additive variances as mentioned before. Generally speaking, genotype pairs with a lower additive genotypic covariance $cov(g_i, g_j)$ tend to show higher estimates for both $H^2_{\Delta ij_{BLUE}}$ and $H^2_{\Delta ij_{BLUP}}$ (Figure 3).

In most aspects, results from example 4 are similar to those in example 3. The most striking difference is the second cluster of lower pairwise heritability estimates. As highlighted in Figure 3, this cluster consists only of the 88 differences from/to genotype geno81, *i.e.*, the genotype present only at a single of the three environments. Notice that while not highlighted, the differences from/to the two genotypes present at only two of the environments (*i.e.*, geno38 and geno49) also form similar clusters, but they are merely located at the lower left border of the main cluster and thus do not stand out as drastically. Furthermore, the main cluster here lies more or less on the diagonal. Finally, the range of estimates for $cov(g_i, g_j)$ is smaller than in example 3. This is mostly due to the $g \times e$ variance that for example 4 in (20) could be estimated separately via the genotype-by-environment interaction effect $(ge)_{it}$ as $var(\boldsymbol{ge}) = \boldsymbol{I}_{263}\sigma^2_{ge}$, whereas fo example 3 the $g \times e$ variance is confounded within the variance of the genotype main effect, since there is no genotype-by-environment interaction effect

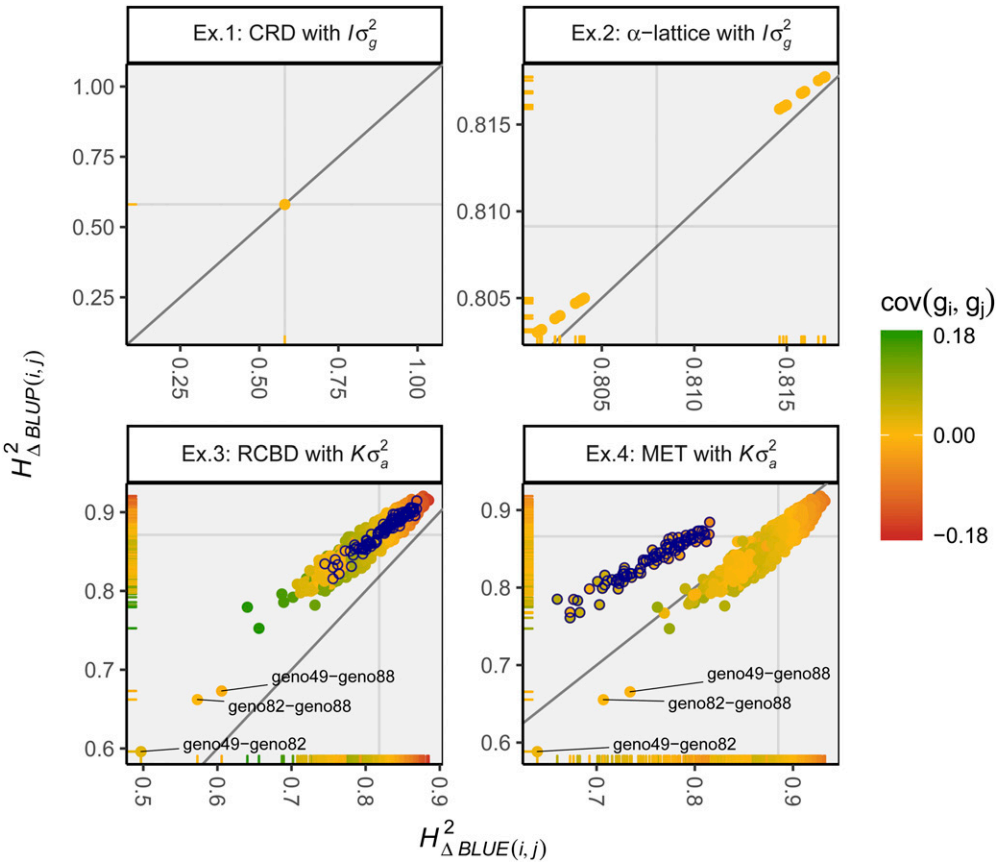

Figure 3 All pairwise heritability estimates. Color indicates covariance in $\boldsymbol{G}_{(\boldsymbol{g})}$ for the respective pair. Circles with blue border are pairs with genotype "geno81," which is only present at a single environment. Vertical and horizontal gray lines represent $\bar{H}^2_{\Delta\cdot\cdot BLUE}$ and $\bar{H}^2_{\Delta\cdot\cdot BLUP}$ for examples 1 and 2, and $\bar{h}^2_{\Delta\cdot\cdot BLUE}$ and $\bar{h}^2_{\Delta\cdot\cdot BLUP}$ for examples 3 and 4. BLUE, best linear unbiased estimation; BLUP, best linear unbiased prediction; CRD, completely randomized design; Ex, example; MET, multienvironment trial; RCBD, randomized complete block design.

in (19), as data from only a single environment are present.

Discussion

Variations of the mixed model for replicated plant data

A consequence of having multiple individuals represent the same genotype is that the definition of the genetic variance σ_g^2 requires careful consideration; rather unproblematic are genetically completely homogeneous cultivars such as inbred lines, doubled-haploid (DH) lines, clones, or hybrids. Yet, it should be noted that when it comes to genetically heterogeneous population cultivars in rye, for example, the matter is no longer straightforward (Bernal-Vasquez *et al.* 2017). Note that estimates for $\bar{h}^2_{\Delta\cdot\cdot BLUE}$, $\bar{h}^2_{\Delta\cdot\cdot BLUP}$ increased with decreasing covariance $cov(g_i, g_j)$ as seen in the *Results* section for pairwise heritability. Note further that when we assumed $var(\boldsymbol{a}) = \boldsymbol{K}\sigma_a^2$ for the (additive) genotype main effect in the analysis of example 4, for simplicity we did not correspondingly make use of $\boldsymbol{K}$ for modeling the genotype–environment interaction effect, but instead assumed $var(\boldsymbol{ge}) = \boldsymbol{I}_{263}\sigma_{ge}^2$. Yet, to account for marker-by-environment interaction, one may instead assume *e.g.*, $var(\boldsymbol{ge}) = \boldsymbol{K}\otimes\boldsymbol{I}_3\sigma_{ge}^2$, where $\otimes$ denotes the Kronecker product. Although Schulz-Streeck *et al.* (2012) found that the improvement of accounting for marker-by-environment interaction can be negligible, the reader may easily decide to include $\boldsymbol{K}$ for random genotype-interaction effects and it may actually be advisable if the variances for those effects are expected to be relatively large (Bernal-Vasquez *et al.* 2017).

Relatedness of heritability methods

When both $\boldsymbol{G}_{(\boldsymbol{g})}$ and $\boldsymbol{C}_{22(\boldsymbol{g})}$ are proportional to identity matrices, all heritability measures give identical estimates. However, this only holds for a linear random model, since even a single fixed effect [*e.g.*, μ in (16)] results in nonzero off-diagonal elements in $\boldsymbol{C}_{22(\boldsymbol{g})}$. Yet, it should be noted that such a purely random model is not practically relevant in plant (and animal) breeding.

When $\boldsymbol{G}_{(\boldsymbol{g})} = \boldsymbol{I}_{n_g}\sigma_g^2$ and $\boldsymbol{C}_{22(\boldsymbol{g})}$ have compound symmetry structure (*i.e.*, simple, balanced setting), all methods except H^2_{Sim} and $\bar{r}^2_{\cdot}$ give identical estimates, which we confirmed in the simple, balanced setting of example 1. Furthermore, both H^2_{Δ} methods display that same value across all genotypes and pairs. Since we always had to assume $\boldsymbol{G}_{(\boldsymbol{g})} = \boldsymbol{I}_{n_g}\sigma_g^2$ for H^2_{Std}, H^2_{Piepho}, and H^2_{Cullis}, example 3 technically reduces to a simple, balanced setting for these measures as well, since it involves a single environment laid out as RCBD with balanced data. Accordingly, these three methods yielded identical estimates in that scenario as well (Figure 1). The estimates for $\bar{r}^2_{\cdot}$ differ, because off-diagonal elements are ignored in its estimation. However, notice that with increasing n_g, the off-diagonal

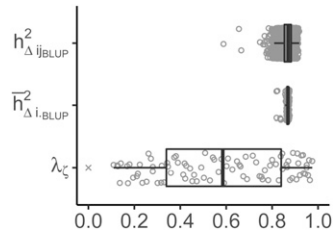

Figure 4 Individual estimates and boxplots for $h^2_{\Delta ij_{BLUP}}$, $\bar{h}^2_{\Delta i._{BLUP}}$, and λ_ζ (*i.e.*, eigenvalues needed for h^2_{Oakey}) from example 4. The zero eigenvalue is denoted by the symbol ×. BLUP, best linear unbiased prediction.

elements of $\boldsymbol{C}_{22(\boldsymbol{g})}$ approach 0, and thus $\bar{r}^2_{\cdot}$ approaches the other measures and, as stated before, so does H^2_{Sim}.

In the case where $\boldsymbol{G}_{(\boldsymbol{g})} = \boldsymbol{I}_{n_g}\sigma^2_g$ and $\boldsymbol{C}_{22(\boldsymbol{g})}$ is an unstructured matrix, $\bar{H}^2_{\Delta \cdot\cdot_{BLUP}}$ and H^2_{Cullis} give identical estimates. This is analogously true for $\bar{H}^2_{\Delta \cdot\cdot_{BLUE}}$ and H^2_{Piepho}, and for $\boldsymbol{C}_{11}$ (rather than $\boldsymbol{C}_{22(\boldsymbol{g})}$) being an unstructured matrix. Rodríguez-Álvarez *et al.* (2018) show a connection between $\bar{r}^2_{\cdot}$ and H^2_{Oakey}. They point out that given $\boldsymbol{G}_{(\boldsymbol{g})_{h_g}} = \boldsymbol{I}_{n_g}\sigma^2_g$, the mean reliability can be expressed as $\bar{r}^2_{\cdot} = \frac{1}{n_g}\sum_{i=1} r^2_i = trace(\boldsymbol{I}_{n_g} - \boldsymbol{G}^{-1}_{(\boldsymbol{g})}\boldsymbol{C}_{22(\boldsymbol{g})})/n_g = trace(\boldsymbol{D})/n_g$ and thus as a special case of H^2_{Oakey}, ignoring the number of zero eigenvalues. This is in agreement with Laloë (1993), and our results confirm this relationship for examples 1 and 2 (results not shown). However, we did not find it to be true for examples 3 and 4, where $\boldsymbol{G}_{(\boldsymbol{g})} \neq \boldsymbol{I}_{n_g}\sigma^2_g$. Note that in their article Rodríguez-Álvarez *et al.* (2018) refer to the mean reliability ($\bar{r}^2_{\cdot}$) as H^2_C in Cullis *et al.* (2006), even though the latter method is based on differences [see (11)], whereas the former is based on individual genotypes [see (43)].

Interestingly, though not practically very relevant, r^2_i and $\bar{H}^2_{\Delta i._{BLUP}}$ give identical estimates per genotype for diagonal $\boldsymbol{G}_{(\boldsymbol{g})}$ and $\boldsymbol{C}_{22(\boldsymbol{g})}$.

Finally, when $\boldsymbol{G}_{(\boldsymbol{g})}$ displays covariances (*e.g.*, as $\boldsymbol{G}_{(\boldsymbol{g})} = \boldsymbol{K}\sigma^2_g$) and $\boldsymbol{C}_{22(\boldsymbol{g})}$ is an unstructured matrix, none of the five measures that are appropriate for this case (*i.e.*, $\bar{h}^2_{\Delta \cdot\cdot_{BLUE}}$, $\bar{h}^2_{\Delta \cdot\cdot_{BLUP}}$, h^2_{Oakey}, $\bar{r}^2_{\cdot}$, and h^2_{Sim}) give identical results. Nevertheless, with the exception of h^2_{Oakey}, heritability estimates within examples 3 and 4 were relatively similar (Figure 1). This is true even for the methods that did not include kinship information (*i.e.*, H^2_{Std}, H^2_{Piepho}, and H^2_{Cullis}). Yet, when Ould Estaghvirou *et al.* (2013) simulated phenotypic and marker data for a single location, they found that these three methods gave lower estimates than the rest.

The most striking results are those for h^2_{Oakey} in examples 3 and 4. Notice that Lourenço *et al.* (2017) also found outstandingly low estimates for h^2_{Oakey} in models with $\boldsymbol{G}_{(\boldsymbol{g})} = \boldsymbol{K}\sigma^2_g$ for two real data sets and one simulated data set. This raises the question whether H^2_{Oakey} is suited for cases where $\boldsymbol{G}_{(\boldsymbol{g})} = \boldsymbol{K}\sigma^2_g$. While it is true that H^2_{Oakey} estimates are equal to H^2_{Std} in the simple, balanced setting (of example 1), this is only a necessary, but not a sufficient, condition for H^2_{Oakey} to be generally applicable also with correlated genotypes. It can be argued that H^2_{Oakey} and $\bar{H}^2_{\Delta \cdot\cdot}$ are similar in the sense that both are based on contrasts between genotypes. However, the important difference is that for $\bar{H}^2_{\Delta \cdot\cdot}$ the contrasts of interest (*i.e.*, all genotype pairs) are chosen according to the breeders' goals, while the contrasts in H^2_{Oakey} are determined by the data set. Laloë (1993) shows that the smallest and largest nonzero eigenvalues, λ_ζ, are the lower and upper limit for the heritability of all possible contrasts of genotypes. Accordingly, they are also the lower and upper limit for all $H^2_{\Delta ij}/h^2_{\Delta ij}$, which is confirmed by our results (Figure 4). It may be noted that H^2_{Oakey} is a simple average of "canonical" heritabilities (eigenvalues) corresponding to canonical contrasts and derived from the canonical decomposition of covariance matrices. As such, low heritabilities indicate a low design efficiency, as the smallest eigenvalue appears to be a good indicator of the robustness of the design and a measure of the part of the genetic trend that can be predicted (Laloë and Phocas 2003). In contrast, a pairwise difference is a combination of low and high heritability contrasts that corresponds to the weighted mean of canonical heritabilities. This may go some way toward explaining the observed discrepancies in estimates of h^2_{Oakey} and the alternatives considered here.

Genotype-specific information gained from H^2_Δ

Figure 2 and Figure 3 are good examples of how H^2_Δ/h^2_Δ can give a more in-depth insight into the outcome of a plant breeding trial. More precisely, the results show that investigating genotype-specific, or even genotype-contrast-specific, heritability measures can (i) summarize how variable/ambiguous an overall heritability estimate is and (ii) point out individual genotypes that stand out.

The first point may seem obvious by now, but it must be realized that whenever only a single (overall) heritability estimate is obtained, there is no information whatsoever about the dispersion of that estimate, only about its central tendency. We argue that investigating $\bar{H}^2_{\Delta i.}/\bar{h}^2_{\Delta i.}$ and/or $H^2_{\Delta ij}/h^2_{\Delta ij}$ gives more comprehensive insight, since they explicitly show the variability across genotypes and thus, bringing us to the second point, point out exceptional genotypes such as geno38, geno49, geno81, geno82, and geno88 in example 4. Notice that we identified two separate reasons for why these five genotypes were exceptional (low additive variance and missing observations; see *Results* section). Moreover, looking at the cluster of estimates for geno81 in example 4 in Figure 3, it becomes clear how a single genotype can display notably lower pairwise heritability estimates than the rest. These estimates are obviously decreasing the overall heritability estimate for this example, which would be especially unfavorable in a scenario where the breeder does not have great interest in this particular genotype. In the end, none of this would become apparent by only looking at overall heritability estimates (Figure 1).

What do we really want heritability for and how does H^2_Δ help?

There are two main reasons why heritability on an entry-mean basis is of interest in plant breeding. On one hand, it is plugged

into the breeder's Equation (2) to predict the response to selection. On the other hand, it is a descriptive measure used to assess the usefulness and precision of results from cultivar evaluation trials. In the simple, balanced setting, heritability is suited for both purposes. However, whenever we depart from the simple, balanced setting, H^2_{Std} is no longer suited for either of the two purposes. Furthermore, any alternative measure can ultimately only aim to generalize heritability as a descriptive measure. Regarding the purpose of heritability as a mean to estimate the response to selection, we reiterate the view of Piepho and Möhring (2007): instead of trying to approximate heritability using some *ad hoc* measure, one should simulate the response to selection directly. As stated before, this can be done via the same simulation-based approach used to obtain H^2_{Sim} and is exemplified in their work (also see Appendix E).

In terms of the descriptive function of heritability, we believe that H^2_{Δ} is a valuable extension of heritability on an entry-mean basis, because its genotype-wise and pair-wise estimates give more detailed insight, and ultimately allow for better decision-making by the breeder. Furthermore, we would like to argue that its derivation, calculation, and interpretation is rather intuitive, which makes us optimistic about its acceptance in practice. Our view is further supported by Laloë (1993), who also suggested extending reliability to pairwise contrasts in the animal breeding framework.

Conclusions

Irrespective of whether a breeder ultimately decides to apply H^2_{Δ}, we think that understanding the idea behind it alone raises awareness for the mentioned problems with heritability outside the scope of simple, balanced settings and thus can potentially improve the use of heritability in breeding programs. In the end, all heritability measures aim to inform about the same underlying subject matter, and, albeit via different methodologies, they all do. What should therefore be held above all, especially given the mentioned ambiguity problems in a nonsimple, unbalanced setting, is to report how heritability was estimated. Hanson and Robinson (1963, page 612) put it in this crucial context of reproducible science by pointing out that "One must extrapolate in the real world, and one must use estimates of heritability derived by someone else, especially with new crops in new environments. Therefore, it is important to specify exactly how published estimates were obtained in order that others may extrapolate."

Acknowledgments

We thank Ivan Simko and Ryan J. Hayes (US Department of Agriculture-Agricultural Research Service, Crop Improvement and Protection Research Unit, Salinas, CA) for providing the lettuce data. We also wish to thank two anonymous reviewers for their very helpful comments. This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft grant PI 377/18-1).

Literature Cited

Becker, H., 2011 *Pflanzenzüchtung*. Ed. 2 Ulmer, Stuttgart, Germany.

Bernal-Vasquez, A.-M., A. Gordillo, M. Schmidt, and H.-P. Piepho, 2017 Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. BMC Genet. 18: 51. https://doi.org/10.1186/s12863-017-0512-8

Cullis, B. R., A. B. Smith, and N. E. Coombes, 2006 On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11: 381–393. https://doi.org/10.1198/108571106X154443

Dias, K. O. D. G., S. A. Gezan, C. T. Guimarães, A. Nazarian, L. da Costa *et al.*, 2018 Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity 121: 24–37. https://doi.org/10.1038/s41437-018-0053-6

Falconer, D. S., and T. F. C. Mackay, 2005 *Introduction to Quantitative Genetics*, Ed. 4. Pearson Prentice Hall, Harlow.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson, 2009 *ASReml user guide release 3.0*. VSN International Ltd., Hemel Hempstead.

Hadasch, S., I. Simko, R. J. Hayes, J. O. Ogutu, and H.-P. Piepho, 2016 Comparing the predictive abilities of phenotypic and marker-assisted selection methods in a biparental lettuce population. Plant Genome 9. Available at: https://dl.sciencesocieties.org/publications/tpg/abstracts/9/1/plantgenome2015.03.0014. https://doi.org/10.3835/plantgenome2015.03.0014

Hallauer, A. R., M. J. Carena, and J. B. Miranda Filho, 2010 *Quantitative Genetics in Maize Breeding*, Vol. 3. Springer, New York.

Hanson, W., and H. Robinson, 1963 *Statistical Genetics and Plant Breeding*. National Academy of Sciences - National Research Council, Washington, DC.

Hayes, R. J., C. H. Galeano, Y. Luo, R. Antonise, and I. Simko, 2014 Inheritance of decay of fresh-cut lettuce in a recombinant inbred line population from Salinas 88 × La Brillante. J. Am. Soc. Hortic. Sci. 139: 388–398. https://doi.org/10.21273/JASHS.139.4.388

Henderson, C. R., 1986 Statistical methods in animal improvement. Historical overview, pp. 2–14 in *Advances in Statistical Methods for Genetic Improvement of Livestock*, Vol. 18, edited by D. Gianola, and K. Hammond. Springer, Berlin. https://doi.org/10.1007/978-3-642-74487-7_1

John, N. A., and E. R. Williams, 1995 *Cyclic and Computer Generated Designs*, Ed. 2. Chapman and Hall, London. https://doi.org/10.1007/978-1-4899-7220-0

Knight, R. L., 1948 Dictionary of Genetics, Chronica Botánica Co., Waltham, Massachusetts.

Kuehn, L. A., R. M. Lewis, and D. R. Notter, 2007 Managing the risk of comparing estimated breeding values across flocks or herds through connectedness: a review and application. Genet. Sel. Evol. 39: 225–247. https://doi.org/10.1051/gse:2007001

Laloë, D., 1993 Precision and information in linear models of genetic evaluation. Genet. Sel. Evol. 25: 557. https://doi.org/10.1186/1297-9686-25-6-557

Laloë, D., and F. Phocas, 2003 A proposal of criteria of robustness analysis in genetic evaluation. Livest. Prod. Sci. 80: 241–256. https://doi.org/10.1016/S0301-6226(02)00092-1

Laloë, D., F. Phocas, and F. Menissier, 1996 Considerations on measures of precision and connectedness in mixed linear

models of genetic evaluation. Genet. Sel. Evol. 28: 359. https://doi.org/10.1186/1297-9686-28-4-359

Lourenço, V. M., P. C. Rodrigues, A. M. Pires, and H.-P. Piepho, 2017 A robust DF-REML framework for variance components estimation in genetic studies. Bioinformatics 33: 3584–3594. https://doi.org/10.1093/bioinformatics/btx457

Mathews, K. L., M. Malosetti, S. Chapman, L. McIntyre, M. Reynolds *et al.*, 2008 Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor. Appl. Genet. 117: 1077–1091. https://doi.org/10.1007/s00122-008-0846-8

Mrode, R. A., and R. Thompson, 2014 *Linear Models for the Prediction of Animal Breeding Values*, Ed. 3. Centre for Agriculture and Bioscience International, Wallingford. https://doi.org/10.1079/9781780643915.0000

Oakey, H., A. Verbyla, W. Pitchford, B. Cullis, and H. Kuchel, 2006 Joint modeling of additive and non-additive genetic line effects in single field trials. Theor. Appl. Genet. 113: 809–819. https://doi.org/10.1007/s00122-006-0333-z

Ould Estaghvirou, S. B., J. O. Ogutu, T. Schulz-Streeck, C. Knaak, M. Ouzunova *et al.*, 2013 Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. BMC Genomics 14: 860. https://doi.org/10.1186/1471-2164-14-860

Piepho, H.-P., 2019 A coefficient of determination (R^2) for generalized linear mixed models. Biom. J. 61: 860–872. https://doi.org/10.1002/bimj.201800270

Piepho, H.-P., and J. Möhring, 2007 Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177: 1881–1888. https://doi.org/10.1534/genetics.107.074229

Piepho, H.-P., J. Möhring, A. E. Melchinger, and A. Büchse, 2008 BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161: 209–228. https://doi.org/10.1007/s10681-007-9449-8

Piepho, H.-P., J. Möhring, T. Schulz-Streeck, and J. O. Ogutu, 2012a A stage-wise approach for the analysis of multi-environment trials. Biom. J. 54: 844–860. https://doi.org/10.1002/bimj.201100219

Piepho, H.-P., J. O. Ogutu, T. Schulz-Streeck, B. Estaghvirou, A. Gordillo *et al.*, 2012b Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. Crop Sci. 52: 1093–1104. https://doi.org/10.2135/cropsci2011.11.0592

R Core Team, 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Rodríguez-Álvarez, M. X., M. P. Boer, F. A. van Eeuwijk, and P. H. C. Eilers, 2018 Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spat. Stat. 23: 52–71. https://doi.org/10.1016/j.spasta.2017.10.003

SAS Institute Inc, 2013 *Base SAS 9.4 Procedures Guide: Statistical Procedures*, *2*. SAS Institute Inc., Cary, NC.

SAS Institute Inc., 2017 *SAS/STAT® 14.3 User's Guide*. SAS Institute Inc., Cary, NC.

Schmidt, P., J. Hartung, J. Rath, and H.-P. Piepho, 2019 Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci. 59: 525–536. https://doi.org/10.2135/cropsci2018.06.0376

Schulz-Streeck, T., J. O. Ogutu, Z. Karaman, C. Knaak, and H.-P. Piepho, 2012 Genomic selection using multiple populations. Crop Sci. 52: 2453–2461. https://doi.org/10.2135/cropsci2012.03.0160

Searle, S. R., G. Casella, and C. E. McCulloch, 1992 *Variance Components*. Wiley, New York. https://doi.org/10.1002/9780470316856

Searle, S. R., F. M. Speed, and G. A. Milliken, 2012 Population marginal means in the linear model. An alternative to least squares means. Am. Stat. 34: 216–221. https://doi.org/10.1080/00031305.1980.10483031

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–4423. https://doi.org/10.3168/jds.2007-0980

Viana, J. M. S., H. D. Pereira, G. B. Mundim, H.-P. Piepho, and F. F. E. Silva, 2018 Efficiency of genomic prediction of non-assessed single crosses. Heredity 120: 283–295. https://doi.org/10.1038/s41437-017-0027-0

Wellmann, R., and J. Bennewitz, 2011 The contribution of dominance to the understanding of quantitative genetic variation. Genet. Res. 93: 139–154. https://doi.org/10.1017/S0016672310000649

Wright, K., 2017 agridat: Agricultural Datasets. R package version 1.13. dataset: john.alpha. https://CRAN.R-project.org/package=agridat.

Xu, S., 2013 *Principles of Statistical Genomics*. Springer, New York. https://doi.org/10.1007/978-0-387-70807-2

Yan, W., 2014 *Crop Variety Trials: Data Management and Analysis*. John Wiley & Sons Inc., New York. https://doi.org/10.1002/9781118688571

Communicating editor: A. Charcosset

Appendix

Appendix A: Computing $\hat{g}^{BLUP}$ From $\hat{g}^{BLUE}$ in a Stage-Wise Approach

Consider the analysis of a series of trials where the aim is to obtain $\hat{\boldsymbol{g}}^{BLUP}$. One may use a single-stage analysis with a random genotype main effect. Alternatively, it is also possible to use a stage-wise analysis, where the first stage obtains $\hat{\boldsymbol{g}}^{BLUE}$ from which $\hat{\boldsymbol{g}}^{BLUP}$ are predicted in the second stage. Both approaches are equivalent, if variance components are known. In practice, small differences are encountered, as variances need to be estimated (Piepho *et al.* 2012a).

Single-stage analysis

Assume that we can write the general single-stage model in (13) as

$$\boldsymbol{y} = \boldsymbol{X_A}\boldsymbol{\beta_A} + \boldsymbol{Z_A}\boldsymbol{u_A} + \boldsymbol{Z_B}\boldsymbol{u_B} + \boldsymbol{Z_W}\boldsymbol{u_W} + \boldsymbol{\varepsilon}, \tag{45}$$

where $\boldsymbol{\beta_A}$ is a vector of fixed across-environment effects (subscript A), $\boldsymbol{u_A}$ is a vector of random across-environment effects with $\boldsymbol{u_A} \sim (0, \boldsymbol{G_A})$, $\boldsymbol{u_B}$ is a vector of random between-environment effects (subscript B) with $\boldsymbol{u_B} \sim (0, \boldsymbol{G_B})$, $\boldsymbol{u_W}$ is a vector of random within-environment effects (subscript W) with $\boldsymbol{u_W} \sim (0, \boldsymbol{G_W})$, $\boldsymbol{\varepsilon}$ is a vector of plot errors with $\boldsymbol{\varepsilon} \sim (0, \boldsymbol{R})$, and $\boldsymbol{y}$ is the observed data vector with $\boldsymbol{y} \sim (0, \boldsymbol{V})$, where $\boldsymbol{V} = \boldsymbol{Z_A}\boldsymbol{G_A}\boldsymbol{Z_A^T} + \boldsymbol{Z_B}\boldsymbol{G_B}\boldsymbol{Z_B^T} + \boldsymbol{Z_W}\boldsymbol{G_W}\boldsymbol{Z_W^T} + \boldsymbol{R}$ [see Piepho *et al.* (2012a)]. Thus, $\boldsymbol{G}$ here is a block diagonal matrix with blocks $\boldsymbol{G_A}$, $\boldsymbol{G_B}$, and $\boldsymbol{G_W}$ on its diagonal. Given a MET similar to example 4, $\boldsymbol{X_A}\boldsymbol{\beta_A}$ would be a general intercept, $\boldsymbol{u_A}$ the random genotype main effect (*i.e.* $\boldsymbol{g}$), $\boldsymbol{u_B}$ the random environment main effect and random genotype-by-environment interaction effect, $\boldsymbol{u_W}$ the random block effect at each environment, and $\boldsymbol{\varepsilon}$ the plot error. Accordingly, $\boldsymbol{G_A}$ would correspond to the kinship matrix $\boldsymbol{K}$, whereas $\boldsymbol{G_B}$ and $\boldsymbol{G_W}$ would be proportional to identity matrices.

Two-stage analysis

Model (45) has a two-stage representation, with the first stage model given by

$$\boldsymbol{y} = \boldsymbol{X_{A1}}\boldsymbol{\beta_{A1}} + \boldsymbol{Z_B}\boldsymbol{u_B} + \boldsymbol{Z_W}\boldsymbol{u_W} + \boldsymbol{\varepsilon}_1, \tag{46}$$

where effects are defined as in (45), while the subscript 1 denotes that the corresponding parameters are estimated in this first stage and thus $\boldsymbol{V}_1 = \boldsymbol{Z_B}\boldsymbol{G_B}\boldsymbol{Z_B^T} + \boldsymbol{Z_W}\boldsymbol{G_W}\boldsymbol{Z_W^T} + \boldsymbol{R}_1$. Notice that in contrast to (45), $\boldsymbol{Z_A}\boldsymbol{u_A}$ is missing here, since the genotype main effect needs to be taken as fixed at this stage (Piepho *et al.* 2012a) and is therefore comprised in $\boldsymbol{X_{A1}}\boldsymbol{\beta_{A1}}$. For simplicity, we can assume that here the genotypic main effects are the only fixed effects so that $\boldsymbol{\beta_{A1}}$ are the genotype means. Accordingly, we can obtain genotypic BLUEs as

$$BLUE(\boldsymbol{\beta_{A1}}) = \hat{\boldsymbol{g}}^{BLUE} = \left(\boldsymbol{X}_{A1}^{'}\ \boldsymbol{V}_1^{-1}\boldsymbol{X_{A1}}\right)^{-1}\boldsymbol{X}_{A1}^{'}\boldsymbol{V}_1^{-1}\boldsymbol{y}, \tag{47}$$

with variance

$$var\left(\hat{\boldsymbol{g}}^{BLUE}\right) = \left(\boldsymbol{X}_{A1}^{'}\boldsymbol{V}_1^{-1}\boldsymbol{X_{A1}}\right)^{-1} = \boldsymbol{R}_2. \tag{48}$$

In the second stage we then have

$$\hat{\boldsymbol{g}}^{BLUE} = \boldsymbol{X_{A2}}\boldsymbol{\beta_{A2}} + \boldsymbol{Z_{A2}}\boldsymbol{u_A} + \boldsymbol{\varepsilon}_2, \tag{49}$$

where $\boldsymbol{\varepsilon}_2 \sim (0, \boldsymbol{R}_2)$ and just like in (45), the random genotype main effect $\boldsymbol{g} = \boldsymbol{u_A} \sim (0, \boldsymbol{G_A})$ where $\boldsymbol{G_A}$ corresponds to the kinship matrix $\boldsymbol{K}$. Notice that $\boldsymbol{X_{A2}}\boldsymbol{\beta_{A2}} = 1_{n_g}\mu$ and $\boldsymbol{Z_{A2}} = \boldsymbol{I}_{n_g}$. Accordingly, we have $\boldsymbol{V}_2 = \boldsymbol{Z_{A2}}\boldsymbol{G_A}\boldsymbol{Z_{A2}^T} + \boldsymbol{R}_2$ and for the mixed model equations we find

$$\begin{bmatrix} \boldsymbol{X}_{A2}^{'}\boldsymbol{R}_2^{-1}\boldsymbol{X}_{A2}^{'} & \boldsymbol{X}_{A2}^{'}\boldsymbol{R}_2^{-1}\boldsymbol{Z_{A2}} \\ \boldsymbol{Z}_{A2}^{'}\boldsymbol{R}_2^{-1}\boldsymbol{X_{A2}} & \boldsymbol{Z}_{A2}^{'}\boldsymbol{R}_2^{-1}\boldsymbol{Z_{A2}} + \boldsymbol{G_A}^{-1} \end{bmatrix}^{-} \begin{bmatrix} \hat{\boldsymbol{\beta}}_{A2} \\ \hat{\boldsymbol{g}}^{BLUP} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_{A2}^{'}\boldsymbol{R}_2^{-1}\hat{\boldsymbol{g}}^{BLUE} \\ \boldsymbol{Z}_{A2}^{'}\boldsymbol{R}_2^{-1}\hat{\boldsymbol{g}}^{BLUE} \end{bmatrix}. \tag{50}$$

It is important to note that we can write

$$\boldsymbol{X_A} = \boldsymbol{X_{A1}}\boldsymbol{X_{A2}} \tag{51}$$

and

$$\boldsymbol{Z_A} = \boldsymbol{X_{A1}}\boldsymbol{Z_{A2}}. \tag{52}$$

Finally, plugging in (48) and (49) into (50), while making use of (51) and (52), it can be seen that the mixed model equation solutions for the second stage are equivalent to those for the single-stage analysis.

Appendix B: An Alternative Estimation for $\bar{H}^2_{\Delta\cdot\cdot}$ in the Case of Independent Genotypic Effects with Constant Variance by Averaging Separately Across Numerator and Denominator

Given independent genotypic effects with constant variance, averaging numerator and denominator of $H^2_{\Delta ij_{BLUP}}$ in (24) separately first leads to the algebraically equivalent simplification

$$\bar{H}^2_{\Delta\cdot\cdot_{BLUP}} = \frac{\sum_i\sum_{j<i}\mathrm{var}(g_i - g_j) - v^{BLUP}_{\Delta ij}}{\sum_i\sum_{j<i}\mathrm{var}(g_i - g_j)} = \frac{2\sigma^2_g - \frac{2}{n_g(n_g-1)}\sum_i\sum_{j<i}v^{BLUP}_{\Delta ij}}{2\sigma^2_g} = \frac{2\sigma^2_g - \bar{v}^{BLUP}_{\Delta\cdot\cdot}}{2\sigma^2_g} = 1 - \frac{\bar{v}^{BLUP}_{\Delta\cdot\cdot}}{2\sigma^2_g} = H^2_{Cullis}. \tag{53}$$

Analogously, we find for $H^2_{\Delta ij_{BLUE}}$ in (29) that

$$\bar{H}^2_{\Delta\cdot\cdot_{BLUE}} = \frac{\sum_i\sum_{j<i}\mathrm{var}(g_i - g_j)}{\sum_i\sum_{j<i}\mathrm{var}(g_i - g_j) + v^{BLUE}_{\Delta ij}} = \frac{2\sigma^2_g}{2\sigma^2_g + \frac{2}{n_g(n_g-1)}\sum_i\sum_{j<i}v^{BLUE}_{\Delta ij}} = \frac{2\sigma^2_g}{2\sigma^2_g + \bar{v}^{BLUE}_{\Delta\cdot\cdot}} = \frac{\sigma^2_g}{\sigma^2_g + \bar{v}^{BLUE}_{\Delta\cdot\cdot}/2} = H^2_{Piepho}. \tag{54}$$

Appendix C: An Approximation for $\bar{H}^2_{\Delta\cdot\cdot}$ in the General Case

It can be shown (Ould Estaghvirou *et al.* 2013, Additional Files 1) that

$$\sum_i\sum_{j<i}v^{BLUE}_{\Delta ij} = trace\Big(n_g\boldsymbol{P_\mu}\boldsymbol{C}_{11(\boldsymbol{g})}\Big), \tag{55}$$

where $\boldsymbol{P_\mu} = \boldsymbol{I_{ng}} - n_g^{-1}1_{n_g}1^T_{n_g}$ with 1_{n_g} being a vector of ones. Thus, $\boldsymbol{P_\mu}$ is a matrix that centers for the overall mean, such that $\bar{v}^{BLUE}_{\Delta\cdot\cdot} = \frac{2}{n_g(n_g-1)}trace(n_g\boldsymbol{P_\mu}\boldsymbol{C}_{11(\boldsymbol{g})})$. Analogously, we have

$$\sum_i\sum_{j<i}v^{BLUP}_{\Delta ij} = trace\Big(n_g\boldsymbol{P_\mu}\boldsymbol{C}_{22(\boldsymbol{g})}\Big), \tag{56}$$

such that $\bar{v}^{BLUP}_{\Delta\cdot\cdot} = \frac{2}{n_g(n_g-1)}trace(n_g\boldsymbol{P_\mu}\boldsymbol{C}_{22(\boldsymbol{g})})$ and finally

$$\sum_i\sum_{j<i}var(g_i - g_j) = trace\Big(n_g\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})}\Big), \tag{57}$$

such that $\bar{v}^{g}_{\Delta\cdot\cdot} = \frac{2}{n_g(n_g-1)}trace(n_g\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})})$. By inserting (55), (56), and (57) into (34), we get

$$\overline{\bar{H}}^2_{\Delta\cdot\cdot_{BLUE}} = \frac{trace\Big(\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})}\Big)}{trace\Big(\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})}\Big) + trace\Big(\boldsymbol{P_\mu}\boldsymbol{C}_{11(\boldsymbol{g})}\Big)}, \tag{58}$$

where, loosely speaking, $trace(\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})})$ captures the genotypic variance, $trace(\boldsymbol{P_\mu}\boldsymbol{C}_{11(\boldsymbol{g})})$ captures the environmental variance, and hence $trace(\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})}) + trace(\boldsymbol{P_\mu}\boldsymbol{C}_{11(\boldsymbol{g})})$ captures the phenotypic variance. Notice that (58) coincides with Method 4 of Ould Estaghvirou *et al.* (2013). Based on (33), we find correspondingly

$$\overline{\bar{H}}^2_{\Delta\cdot\cdot_{BLUP}} = \frac{trace\Big(\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})}\Big) - trace\Big(\boldsymbol{P_\mu}\boldsymbol{C}_{22(\boldsymbol{g})}\Big)}{trace\Big(\boldsymbol{P_\mu}\boldsymbol{G}_{(\boldsymbol{g})}\Big)}\ . \tag{59}$$

Appendix D: A Standardization of *M* To Directly Capture the Total Genotypic Variance

As shown in Appendix C, $\bar{v}^{g}_{\Delta\cdot\cdot}$ captures the average genotypic variance of a difference between two genotypes. In accordance with the notion of this article, we could obtain half the average variance of a difference as an estimate for *the total genotypic variance*. Notice that this approach is in line with the average semi variance $(\sigma^2_{g(asv)})$ used by Piepho (2019) and can be defined as

$$\sigma^2_{g(asv)} = 0.5\bar{v}^{g}_{\Delta\cdot\cdot} = \frac{trace\left(\boldsymbol{P_\mu G_{(g)}}\right)}{n_g - 1}. \tag{60}$$

In the case of $\boldsymbol{G_{(g)}} = \boldsymbol{MM'}\sigma^2_a = \boldsymbol{K}\sigma^2_a$, we can do some useful rearrangements:

$$\begin{aligned}\sigma^2_{g(asv)} &= \frac{trace\left(\boldsymbol{P_\mu MM'}\right)\sigma^2_g}{n_g - 1} = \frac{trace\left(\boldsymbol{P_\mu P_\mu MM'}\right)\sigma^2_g}{n_g - 1} \\ &= \frac{trace\left(\boldsymbol{P_\mu MM'P_\mu}\right)\sigma^2_g}{n_g - 1} = \frac{trace\left(\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}'\right)\sigma^2_g}{n_g - 1},\end{aligned} \tag{61}$$

where $\tilde{\boldsymbol{M}} = \boldsymbol{P_\mu M}$ is column mean-centered. Notice that $\tilde{\boldsymbol{M}}$ is equal to what is referred to as $\boldsymbol{Z}$ in VanRaden (2008) (page 4416). Notice further that the standardization presented here works for an arbitrary coding of markers and arbitrary scale. As a result, the trace of $\tilde{\boldsymbol{G}}_{(g)} = \tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}'\widehat{\sigma^2_a} = \tilde{\boldsymbol{K}}\widehat{\sigma^2_a}$ directly returns $\hat{\sigma}^2_{g(asv)}$. We can standardize further:

$$\overline{\tilde{M}} = \frac{\tilde{\boldsymbol{M}}}{\sqrt{\frac{trace\left(\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}'\right)}{n_g - 1}}} \tag{62}$$

and define $\overline{\boldsymbol{G}}_{(g)} = \overline{\boldsymbol{M}}\overline{\boldsymbol{M}}'\overline{\overline{\sigma^2_a}} = \overline{\boldsymbol{K}}\overline{\overline{\sigma^2_a}}$. Now the total genotypic variance simplifies to

$$\sigma^2_{g(asv)} = trace(\boldsymbol{P_\mu}\frac{\overline{\boldsymbol{M}}\overline{\boldsymbol{M}}'\overline{\overline{\sigma^2_a}})}{n_g - 1} = \overline{\overline{\sigma^2_a}}, \tag{63}$$

which simplifies (58) and (59) to

$$\bar{h}^2_{\Delta BLUE} = \frac{\overline{\overline{\sigma^2_a}}}{\overline{\overline{\sigma^2_a}} + trace\left(n_g\boldsymbol{P_\mu C}_{11(\boldsymbol{g})}\right)} \tag{64}$$

and

$$\bar{h}^2_{\Delta BLUP} = \frac{\overline{\overline{\sigma^2_a}} - trace\left(n_g\boldsymbol{P_\mu C}_{22(\boldsymbol{g})}\right)}{\overline{\overline{\sigma^2_a}}}. \tag{65}$$

Thus, if $\overline{\tilde{\boldsymbol{M}}}$ is standardized as shown above, the genotypic variance we are trying to capture is conveniently estimated directly as the variance for $\boldsymbol{a}$. However, it should be noted that both $\tilde{\boldsymbol{K}}$ and $\overline{\tilde{\boldsymbol{K}}}$ are singular, and thus may lead to problems depending on the statistical software being used: ASReml-R 3.0 (Gilmour *et al.* 2009) will show an error message, while PROC MIXED in SAS 9.4 (SAS Institute Inc. 2013) will effectively compute a modified version of the mixed model equations (Piepho *et al.* 2012b; SAS Institute Inc. 2017).

We would like to point out that neither $\tilde{\boldsymbol{K}}$ nor $\overline{\tilde{\boldsymbol{K}}}$ are equal to the genomic relationship matrices proposed in VanRaden (2008). Their first method obtains the general relationship matrix as:

$$\boldsymbol{K_{VanRaden1}} = \frac{\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}'}{\sum_{w=1}^{n_m} 2\boldsymbol{p_w}(1 - \boldsymbol{p_w})}, \tag{66}$$

where $\boldsymbol{p_w}$ are the mean frequencies of the second allele across all genotypes at marker $\boldsymbol{w} = \{1, \ldots, n_m\}$ and can also be expressed as $\boldsymbol{p_w} = \frac{1'_{n_g}\boldsymbol{M}}{2n_g}$. Their second method obtains the matrix as $\boldsymbol{K_{VanRaden2}} = \tilde{\boldsymbol{M}}\boldsymbol{L}\tilde{\boldsymbol{M}}'$ where $\boldsymbol{L}$ is a diagonal with elements

$$L_{w,w} = \frac{1}{n_m[2\boldsymbol{p_w}(1-\boldsymbol{p_w})]}. \tag{67}$$

Since we have $\tilde{\boldsymbol{K}} = \tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}'$, we find that opposed to $\tilde{\boldsymbol{K}}$ and therefore $\overline{\overline{\boldsymbol{K}}}$, (66) and (67) additionally require a division involving allele frequencies. It may also be pointed out that in a population with Hardy–Weinberg equilibrium, $trace(\boldsymbol{MM}')$ corresponds to the heterozygosity of the markers. However, it should be kept in mind that completely genetically homogeneous cultivars—such as inbred lines, DH-lines, clones, or hybrids—are definitely not in Hardy–Weinberg equilibrium. Finally, note that if we approximate $\bar{h}^2_{\Delta\cdot\cdot BLUE}$ via (59) and implement $\hat{\sigma}^2_{g(asv)}$ into h^2_{Piepho}, we find that $\bar{\bar{h}}^2_{\Delta\cdot\cdot BLUE} = h^2_{Piepho}$ even in the general case. Analogously, we have $\bar{\bar{h}}^2_{\Delta\cdot\cdot BLUP} = h^2_{Cullis}$.

Appendix E: Directly Simulate Response to Selection

As proposed by Piepho and Möhring (2007), one may directly simulate the response to selection via the same approach used to obtain H^2_{Sim} in this article. Since in (40) we simulate n_g pairs of true ($\boldsymbol{g_{sim}}$) and predicted ($\hat{\boldsymbol{g}}^{\boldsymbol{BLUP}}_{\boldsymbol{sim}}$) genotypic values, we can select a subset with the best values for $\hat{\boldsymbol{g}}^{\boldsymbol{BLUP}}_{\boldsymbol{sim}}$ and compute the response to selection for the sth simulation run as

$$R_s = \frac{\sum_{i\in\varphi_s}\boldsymbol{g_{sim}}}{n_{sel}}, \tag{68}$$

where φ_s is the selected subset of genotypes in the sth simulation run and n_{sel} is the number of selected genotypes (*i.e.* $n_{sel} \leq n_g$). Analogous to (41), we can then obtain the simulated expected response to selection as

$$R_{sim} = n^{-1}_{sim}\sum_{s=1}^{n_{sim}} R_s. \tag{69}$$

Note that, in principle, the simulation-based approach could also be adapted to simulate the joint distribution of true ($\boldsymbol{g_{sim}}$) and estimated ($\hat{\boldsymbol{g}}^{\boldsymbol{BLUE}}_{\boldsymbol{sim}}$) genotypic values. Furthermore, one could then compare the simulated expected response to selection values from both simulation approaches to determine whether to ultimately model the genotypic main effect as fixed or random. Yet, not least because of the ability to account for kinship information, we generally expect an analysis with random genotypic effects to be more useful than one with fixed genotypic effects [see Piepho *et al.* (2008)].