Marker-Based Estimation of Heritability in Immortal Populations

Willem Kruijer,*[,1] Martin P. Boer,* Marcos Malosetti,* Pádraic J. Flood,[†,‡] Bas Engel,* Rik Kooke,[§,†]
Joost J. B. Keurentjes,[†,] and Fred A. van Eeuwijk***

*Biometris, Wageningen University and Research Centre, 6700AC Wageningen, The Netherlands, [†]Laboratory of Genetics, Wageningen University and Research Centre, 6700AH Wageningen, The Netherlands, [‡]Horticulture and Product Physiology, Wageningen University and Research Centre, 6700AP Wageningen, The Netherlands, and [§]Laboratory of Plant Physiology, Wageningen University and Research Centre, 6700AR Wageningen, The Netherlands, and [**]Swammerdam Institute for Life Sciences, University of Amsterdam, 1090GE Amsterdam, The Netherlands

ABSTRACT Heritability is a central parameter in quantitative genetics, from both an evolutionary and a breeding perspective. For plant traits heritability is traditionally estimated by comparing within- and between-genotype variability. This approach estimates broad-sense heritability and does not account for different genetic relatedness. With the availability of high-density markers there is growing interest in marker-based estimates of narrow-sense heritability, using mixed models in which genetic relatedness is estimated from genetic markers. Such estimates have received much attention in human genetics but are rarely reported for plant traits. A major obstacle is that current methodology and software assume a single phenotypic value per genotype, hence requiring genotypic means. An alternative that we propose here is to use mixed models at the individual plant or plot level. Using statistical arguments, simulations, and real data we investigate the feasibility of both approaches and how these affect genomic prediction with the best linear unbiased predictor and genome-wide association studies. Heritability estimates obtained from genotypic means had very large standard errors and were sometimes biologically unrealistic. Mixed models at the individual plant or plot level produced more realistic estimates, and for simulated traits standard errors were up to 13 times smaller. Genomic prediction was also improved by using these mixed models, with up to a 49% increase in accuracy. For genome-wide association studies on simulated traits, the use of individual plant data gave almost no increase in power. The new methodology is applicable to any complex trait where multiple replicates of individual genotypes can be scored. This includes important agronomic crops, as well as bacteria and fungi.

KEYWORDS marker-based estimation of heritability; GWAS; genomic prediction; *Arabidopsis thaliana*; one- *vs.* two-stage approaches

NARROW-SENSE heritability is an important parameter in quantitative genetics, determining the response to selection and representing the proportion of phenotypic variance that is due to additive genetic effects (Jacquard 1983; Nyquist and Baker 1991; Ritland 1996; Visscher *et al.* 2006, 2008; Holland *et al.* 2010; Sillanpaa 2011). This definition of heritability goes back to Fisher (1918) and Wright (1920) almost a century ago. In plant species for which replicates of the same genotype are available (inbred lines, doubled haploids, clones), a different form of heritability, broad-sense heritability, is traditionally estimated by the intraclass correlation coefficient for genotypic effects, using estimates for within- and between-genotype variance. Broad-sense heritability is also referred to as repeatability and gives the proportion of phenotypic variance explained by heritable (additive) and nonheritable (dominance, epistasis) genetic variance.

With the arrival of high-density genotyping there is growing interest in marker-based estimation of narrow-sense heritability (WTCCC 2007; Yang *et al.* 2010, 2011; Speed *et al.* 2012; Vattikuti *et al.* 2012; Visscher and Goddard 2014). These estimates are obtained from mixed models containing random additive genetic effects, whose covariance structure is estimated from genetic markers. While marker-based heritability estimates have received much attention in human genetics, these are rarely reported for plant traits, despite their relevance in evolutionary genetics, in the dissection of complex traits, and in the ongoing debate on missing heritability (Manolio *et al.* 2009; Eichler

et al. 2010; Brachi et al. 2011; Lee et al. 2011; Zuk et al. 2012). Heritability estimates are also of great relevance to plant breeders, as they give a measure for the breeding potential of a trait. In addition, state-of-the-art phenotyping platforms are making experiments more reproducible, increasing the relevance of marker-based estimation of heritability, as well as comparison of estimates from different experiments.

Although marker-based heritability estimation for plant traits can in principle be achieved within the same analytic framework that has been developed for human traits, there are important differences. First, heritability of human traits is usually estimated using panels of unrelated individuals, to avoid confounding with geographical or environmental effects (Browning and Browning 2011). For most plant species panels of unrelated individuals are not available, as plant genotypes will often share ancestry or adaptation, inducing dependence between genotypes. At the same time, however, plant genotypes can be evaluated under the same experimental conditions (e.g., common garden or growth chamber), and the usual assumption is that this eliminates genotype–environment correlations. Second, plant genotypes are often phenotyped in several genetically identical replicates. This is the case for all so-called immortal populations (Keurentjes et al. 2007; Wijnen and Keurentjes 2014). Mixed-model analysis can then be performed either on the individual plant (or plot) data or on genotypic means. In the literature on multi-environment trials (Smith et al. 2001, 2005; Oakey et al. 2006; Piepho and Williams 2006; Piepho et al. 2006, 2012; Boer et al. 2007; Verbyla et al. 2007; Stich et al. 2008; Möhring and Piepho 2009; Van Eeuwijk et al. 2010; Welham et al. 2010; Malosetti et al. 2013) these approaches are referred to as respectively one-stage and two-stage. These works consider mostly populations for which a pedigree is available, typically experimental populations. In the context of genomic prediction and genome-wide association studies (GWAS) for natural populations, mixed-model analysis is usually performed using a two-stage approach. The (usually tacit) assumption is that the genotypic means and kinship coefficients contain all the relevant information for estimating the genetic and residual variance. Here we investigate the feasibility of marker-based estimation of heritability with one- and two-stage approaches and look at how heritability estimates affect genomic prediction with the best linear unbiased predictor (G-BLUP) and GWAS. Although our analysis of observed phenotypes focuses on the model plant *Arabidopsis thaliana*, the asymptotic variances of different heritability estimators were also computed for diverse panels of *Zea mays* (Riedelsheimer et al. 2012; Van Heerwaarden et al. 2012) and *Oryza sativa* (Zhao et al. 2011).

In both published data (Atwell et al. 2010) and new experiments, we found very large standard errors and sometimes unrealistically high estimates of heritability, which could not be explained by varying linkage disequilibrium (Speed et al. 2012). Much better heritability estimates were obtained when mixed-model analysis was performed at the individual plant level. These estimates were based on kinship information as well as additional information on within-genotype variability, and in simulations they were found to be up to 13 times more accurate than heritability estimates based on genotypic means. In genomic prediction, correlation between simulated and predicted genetic effects increased in some cases by as much as 49%. This is a substantial improvement that shows the importance of accurate heritability estimates in plant breeding programs.

All reported heritability estimates can be obtained using our R-package heritability, which is freely available from CRAN (http://cran.r-project.org/web/packages/heritability/index.html). In contrast to existing packages such as emma (Kang et al. 2008), rrblup (Endelman 2011), and synbreed (Wimmer et al. 2012), it provides confidence intervals for heritability estimates. We also present software for GWAS: our program scan_GLS can efficiently perform GWAS directly on the individual plant or plot-level data, as well as on the means, incorporating a nondiagonal error covariance structure.

Materials and Methods

We assume a natural population of n genotypes, where for each genotype a quantitative trait is measured on a number of genetically identical replicates of immortal lines. These replicates can refer to either individual plants (e.g., *Arabidopsis*) or plots in a field trial (e.g., maize). For convenience we use the terms "replicate" and "individual plant data" as synonyms throughout. In either case, the observations on replicates should not be confused with having multiple observations on the same individual. We focus on inbred lines, but apart from a few different constants, all expressions are valid for any diploid species. We do not partition the environmental variance into different contributing factors (see, e.g., Visscher et al. 2008); hence the environmental variance just equals the error variance. For simplicity we first present marker-based estimation of heritability for a completely randomized design, in the absence of additional covariates. The expressions for more general designs are given in *Appendix A*, which also provides a brief overview of the existing methodology for marker-based heritability estimation. Details on G-BLUP, GWAS, and the phenotypic data are given in *Appendixes C*, *D*, and *E*.

Genotypic data

We analyze simulated and real traits for subpopulations of the RegMap, which contains 1307 worldwide accessions of *A. thaliana* that have been genotyped at 214,051 SNP markers (Horton et al. 2012). We considered four subpopulations: 298 Swedish accessions [Swedish RegMap (Horton et al. 2012; Long et al. 2013)], 204 French accessions [French RegMap (Horton et al. 2012; Brachi et al. 2013)], 350 accessions from the HapMap population (Li et al. 2010), and a subset of 250 accessions that we refer to as the

structured RegMap (accession IDs are given in Supporting Information, Table S1). The structured RegMap accessions were chosen to have large differences in genetic relatedness (*i.e.*, variation in kinship coefficients, across pairs of accessions). These differences were hence largest in the structured RegMap and smallest in the HapMap (Figure S1 and Figure S2). For the asymptotic distributions of heritability estimators given below, we also considered marker-based kinship matrices for three populations of crop plants: the panel described in Van Heerwaarden *et al.* (2012) (*Z. mays*, 400 accessions), the panel described in Riedelsheimer *et al.* (2012) (*Z. mays*, 280 accessions), and the panel from Zhao *et al.* (2011) (*O. sativa*, 413 accessions).

Phenotypic data

We estimated heritabilities for two traits from the literature and four traits from new experiments. Two flowering traits from Atwell *et al.* (2010) were analyzed: days to flowering time under long day and vernalization (LDV) and days to flowering time under long day (LD). In two new experiments (*Appendix E*), leaf area (LA) 13 days after sowing was measured for the Swedish RegMap and the HapMap, using the same automated phenotyping platform. In the final experiment, the HapMap was phenotyped for bolting time (BT) and leaf width (LW). Trait descriptions and abbreviations are given in Table 1. In all experiments the individual plants were grouped into complete blocks, but due to nongerminating seeds and dead plants, there is some unbalance for all of the traits.

Marker-based estimation of heritability

Let K denote the genetic relatedness or kinship matrix of the genotypes, with elements

$$K_{i,j} = \frac{1}{p} \sum_{l=1}^{p} \frac{(x_{i,l} - f_l)(x_{j,l} - f_l)}{4f_l(1 - f_l)} \qquad (1)$$

(see, *e.g.*, Patterson *et al.* 2006 and Goddard *et al.* 2010). The numbers $x_{i,l} \in \{0, 2\}$ denote the minor allele count [because we focus on inbred lines, there is a small difference with the standard expression for outbreeders under Hardy–Weinberg equilibrium, where $x_{i,l} \in \{0, 1, 2\}$, and the constant 4 in (1) is replaced by 2] at marker l for genotype i, and $f_l \in [0, 1]$ is the minor allele frequency at marker l. The phenotypic response of replicate j of genotype i is modeled as

$$Y_{i,j} = \mu + G_i + E_{i,j} \quad (i = 1, \ldots, n, \quad j = 1, \ldots, r), \qquad (2)$$

where $G = (G_1, \ldots, G_n)$ has a $N(0, \sigma_A^2 K)$ distribution, and the errors $E_{i,j}$ have independent normal distributions with variance σ_E^2. We aim to estimate the heritability

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2},$$

which has been referred to as the "chip" heritability (Speed *et al.* 2012) and under certain assumptions equals narrow-

Table 1 Abbreviations of the trait names

Abbreviation	Trait
LDV	Days to flowering time under long day and vernalization
LD	Days to flowering time under long day
LA(S)	Leaf area 13 days after sowing (Swedish RegMap)
LA(H)	Leaf area 13 days after sowing (HapMap)
BT	Bolting time (HapMap)
LW	Leaf width (HapMap)

The first two traits were taken from Atwell *et al.* (2010).

sense heritability (*Appendix A*). In contrast to the *line* heritability $\sigma_A^2/(\sigma_A^2 + r^{-1}\sigma_E^2)$, the denominator contains the residual variance for a single individual (σ_E^2). Given restricted maximum-likelihood (REML) estimates $\hat{\sigma}_{A,r}^2$ and $\hat{\sigma}_{E,r}^2$, h^2 can be estimated by

$$\hat{h}_r^2 = \frac{\hat{\sigma}_{A,r}^2}{\hat{\sigma}_{A,r}^2 + \hat{\sigma}_{E,r}^2}, \qquad (3)$$

where the subscripts r stress the fact that $\hat{\sigma}_{A,r}^2$ and $\hat{\sigma}_{E,r}^2$ are estimates directly obtained from the replicates. Alternatively, h^2 can be estimated using the genotypic means

$$\overline{Y}_i = \frac{1}{r} \sum_{j=1}^{r} = \mu + G_i + \overline{E}_i, \quad G \sim N(0, \sigma_A^2 K),$$

$$\overline{E}_i \sim N(0, r^{-1}\sigma_E^2). \qquad (4)$$

Given REML estimates $\hat{\sigma}_{A,m}^2$ and $\hat{\sigma}_{E,m}^2$ for model (4), we have the heritability estimate

$$\hat{h}_m^2 = \frac{\hat{\sigma}_{A,m}^2}{\hat{\sigma}_{A,m}^2 + \hat{\sigma}_{E,m}^2}, \qquad (5)$$

where the subscripts m indicate that the estimates are based on genotypic means. We omit the letters r and m either if these are clear from the context or when a statement holds for both models (2) and (4). In human association studies, r is usually 1, and $\hat{h}_m^2$ and $\hat{h}_r^2$ are identical.

Repeatability and broad-sense heritability

Given a completely randomized design with r replicates, repeatability, or intraclass correlation can be estimated by

$$\hat{H}^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_G^2 + \hat{\sigma}_{\text{Env}}^2}, \text{with } \hat{\sigma}_G^2 = \frac{\text{MS}(G) - \text{MS}(\text{Env})}{r},$$

$$\hat{\sigma}_{\text{Env}}^2 = \text{MS}(\text{Env}), \qquad (6)$$

where MS(G) and MS(Env) are the mean sums of squares for genotype and residual error, obtained from analysis of variance (ANOVA). In the case that MS(G) < MS(Env), $\hat{\sigma}_G^2$ is set to zero. See Singh *et al.* (1993) or (in the context of sib analysis) Lynch and Walsh (1998, p. 563). Here we stick to the widely used notation $\hat{H}^2$, although repeatability equals broad-sense heritability (H^2) only under the assumption that all differences between genotypes are indeed genetic and

not due to, *e.g.*, genotype–environment correlation. In any case, no additional information on the genetic structure is used here, and hence the mean sums of squares for genotype contain all genetic effects, not only the additive ones. Consequently, $\hat{H}^2$ is an estimator of (an upper bound on) broad-sense heritability.

It is known to have a small bias (Nyquist and Baker 1991; Singh *et al.* 1993), which tends to zero when the number of genotypes increases. Additional fixed effects can be included in the ANOVA, which may reduce MS(Env) and give a higher heritability estimate. In the case that the genotypes have differing numbers of replicates, genetic variance is estimated with $\hat{\sigma}_G^2 = (MS(G) - MS(Env))/\bar{r}$, where

$$\bar{r} = (n-1)^{-1}\left[\sum_{i=1}^{n} r_i - \frac{\left(\sum_{i=1}^{n} r_i^2\right)}{\left(\sum_{i=1}^{n} r_i\right)}\right] \quad (7)$$

is the effective number of replicates (Lynch and Walsh 1998, p. 559). For a balanced design, $\bar{r} = r$.

Confidence intervals

Confidence intervals for broad-sense heritability H^2 were obtained from classical theory (*e.g.*, Lynch and Walsh 1998, p. 563) (for easy reference see File S1). Confidence intervals for h^2 were constructed using $\hat{\sigma}_A^2$ and $\hat{\sigma}_E^2$ and the inverse average-information (AI) matrix, obtained from the AI-REML algorithm (Gilmour *et al.* 1995). This 2×2 matrix provides an estimate of the covariance matrix of $(\hat{\sigma}_A^2, \hat{\sigma}_E^2)$. The delta method (see, *e.g.*, Van der Vaart 2000) applied to the function $(\sigma_A^2, \sigma_E^2) \rightarrow \sigma_A^2/(\sigma_A^2 + \sigma_E^2)$ then gives the asymptotic distribution of $\hat{h}^2$, from which a confidence interval for h^2 is calculated. In the case that the heritability is low or high, confidence intervals may be partly outside [0, 1]. We set all negative values to zero and all values larger than one equal to one. An alternative is application of the delta method to the function $(\hat{\sigma}_A^2, \hat{\sigma}_E^2) \rightarrow \log(\hat{\sigma}_A^2/\hat{\sigma}_E^2)$. This gives confidence intervals for $\log(\sigma_A^2/\sigma_E^2)$, which are then back transformed to confidence intervals for heritability. The latter intervals are referred to as "log-transformed" and the other intervals as "standard."

When the likelihood as a function of h^2 is monotonically increasing, the confidence intervals can become numerically unstable, alternating between [0, 1] and [1, 1]. However, we found that using the AI matrix obtained by setting $h^2 = 1 - \varepsilon$, the interval was always very close to [0, 1], for any $\varepsilon > 0$. We therefore defined the interval to be [0, 1] in the case of a monotonically increasing likelihood.

Simulations

Each simulated trait consists of $r = 3$ replicates for a subset of 200 accessions randomly drawn from a subpopulation of the *Arabidopsis* RegMap (Horton *et al.* 2012). We simulate according to model (2), except that the genetic effects G_i are not purely polygenic, but a mixture of QTL effects at a small

number of markers and a polygenic signal with genetic structure given by the kinship matrix defined in (1). More details are given in *Appendix B*. Different genetic architectures were considered, but all simulated genetic effects were additive. Hence, broad- and narrow-sense heritabilities were equal, and $\hat{H}^2$, $\hat{h}_r^2$, and $\hat{h}_m^2$ were in this case different estimators of the (same) simulated heritability.

For the comparison of heritability estimates we simulated traits for the structured RegMap, the HapMap, and the Swedish and French RegMap. For comparing one- and two-stage association mapping and genomic prediction only the structured RegMap and the HapMap were considered. For genomic prediction and the comparison of heritability estimates we simulated three levels of heritability, for each subpopulation: low (0.2), medium (0.5), and high (0.8). For each heritability level, 5000 traits were simulated. For the comparison of one- and two-stage association mapping, 1000 traits were simulated with heritability 0.5. In all simulations, a completely randomized design was assumed.

Results

Asymptotic variance

Applying general mixed-model theory (Casella and McCulloch 2006, p. 387) to model (2), it follows that the REML estimators $(\hat{\sigma}_{A,r}^2, \hat{\sigma}_{E,r}^2)$ are asymptotically Gaussian with covariance

$$\text{Var}\begin{pmatrix} \hat{\sigma}_{A,r}^2 \\ \hat{\sigma}_{E,r}^2 \end{pmatrix} \simeq 2\begin{bmatrix} \text{tr}(P(ZKZ^t)P(ZKZ^t)) & \text{tr}(P(ZKZ^t)P) \\ \text{tr}(P(ZKZ^t)P) & \text{tr}(PP) \end{bmatrix}^{-1}$$
$$=: \Sigma_{\hat{\sigma}_{A,r}^2, \hat{\sigma}_{E,r}^2},$$
$$(8)$$

where $P = V^{-1} - V^{-1}X(X^tV^{-1}X)^{-1}X^tV^{-1}$ for $V = \sigma_A^2 ZKZ^t + \sigma_E^2 I_{nr}$, $X = 1_{nr}$, and Z is the $nr \times n$ incidence matrix assigning plants to genotypes. The number of replicates r is considered fixed here. The asymptotic arguments are all with respect to n, the number of genotypes. The behavior of $\text{Var}(\hat{h}_m^2)$ as a function of n, h^2, and K can also be derived from approximations recently proposed in Visscher and Goddard 2014, independently from the present work. Although this is not made explicit in the notation, $\Sigma_{\hat{\sigma}_{A,r}^2, \hat{\sigma}_{E,r}^2}$, depends on the true (and usually unknown) values σ_A^2 and σ_E^2 through P and V. If (σ_A^2, σ_E^2) are estimated based on genotypic means following model (4), $(\hat{\sigma}_{A,m}^2, \hat{\sigma}_{E,m}^2)$ has asymptotic covariance $\Sigma_{\hat{\sigma}_{A,m}^2, \hat{\sigma}_{E,m}^2}$, which is obtained if we replace in (8) ZKZ^t by K and $X = 1_{nr}$ by $X = 1_n$ and substitute $V = \sigma_A^2 K + \sigma_E^2 r^{-1}I_n$ in the definition of P. The asymptotic variance of the heritability estimators $\hat{h}_r^2$ and $\hat{h}_m^2$ can now be obtained by application of the delta method (Van der Vaart 2000) to the function $(\sigma_A^2, \sigma_E^2) \rightarrow \sigma_A^2/(\sigma_A^2 + \sigma_E^2)$, with gradient

$$b_{\sigma_A^2, \sigma_E^2} = \frac{1}{(\sigma_A^2 + \sigma_E^2)^2}(\sigma_E^2, -\sigma_A^2)^t.$$

It then follows that given the true σ_A^2 and σ_E^2,

$$\text{Var}\left(\hat{h}_r^2\right) \simeq b_{\sigma_A^2,\sigma_E^2} \Sigma_{\hat{\sigma}_{A,r}^2,\hat{\sigma}_{E,r}^2} b_{\sigma_A^2,\sigma_E^2}^t,$$

$$\text{Var}\left(\hat{h}_m^2\right) \simeq b_{\sigma_A^2,\sigma_E^2} \Sigma_{\hat{\sigma}_{A,m}^2,\hat{\sigma}_{E,m}^2} b_{\sigma_A^2,\sigma_E^2}^t. \qquad (9)$$

It is easily verified that both variances do not depend on the absolute values of σ_A^2 and σ_E^2, but only on the ratio $h^2 = \sigma_A^2/(\sigma_A^2 + \sigma_E^2)$. To explore the potential gains in accuracy due to the use of individual plant data, we computed $\sqrt{\text{Var}(\hat{h}_r^2)}$ and the ratio $\sqrt{\text{Var}(\hat{h}_r^2)}\big/\sqrt{\text{Var}(\hat{h}_m^2)}$ of the two standard deviations, for several populations in A. thaliana, Z. mays, and O. sativa (Table 2) and three heritability levels ($h^2 = 0.2, 0.5, 0.8$).

The use of individual plant data gave a substantial improvement in accuracy for all populations, which was largest for the Arabidopsis HapMap [81% reduction with respect to $\sqrt{\text{Var}(\hat{h}_m^2)}$, for $r = 4$ and $h^2 = 0.8$] and smallest for the structured RegMap, the rice population from Zhao et al. (2011), and the maize population from Riedelsheimer et al. (2012) (13–15% reduction when $r = 2$ and $h^2 = 0.2$). For the maize population from Van Heerwaarden et al. (2012) and the other Arabidopsis populations the improvements were similar. The standard deviation of $\hat{h}_r^2$ decreased with increasing heritability and increasing numbers of replicates. The ratio of the standard deviations of $\hat{h}_r^2$ over those of $\hat{h}_m^2$ also decreased substantially with increasing heritability, but remained similar when increasing the number of replicates beyond 2.

It should be emphasized that the ratios in Table 2 are asymptotic: although the expressions in (9) depend on the kinship matrix K of a finite population, they may be good approximations only for (very) large populations with population structure similar to that defined by K. Such populations could be modeled as a function of population genetic processes such as drift and selection (Rousset 2002; Fu et al. 2003). However, for real plant populations marker information is available for at most a few thousand genotypes and more often several hundred. To assess variance and bias of $\hat{h}_r^2$ and $\hat{h}_m^2$ in such populations we used simulated traits.

Heritability estimates for simulated data

We analyzed simulated traits for the four populations of A. thaliana, for different genetic architectures. We also used the simulations to compare these marker-based estimators with the broad-sense heritability estimator $\hat{H}^2$, which ignores genetic relatedness, and to investigate the quality of heritability confidence intervals.

Heritability estimates for the structured RegMap and HapMap are given in Figure 1 and Table 3. Apart from $\hat{h}_m^2$ in the HapMap, none of the estimators showed marked bias. In particular, $\hat{H}^2$ had negligible bias, despite the fact that it does not account for genetic relatedness. As in the asymptotic results, $\hat{h}_r^2$ had lower variance than $\hat{h}_m^2$, the difference being largest for

**Table 2 Asymptotic distribution of marker-based heritability estimators**

Population	r	$h^2 = 0.2$	$h^2 = 0.5$	$h^2 = 0.8$
Structured RegMap	1	0.0702 (1.00)	0.0762 (1.00)	0.0497 (1.00)
Structured RegMap	2	0.0428 (0.85)	0.0406 (0.71)	0.0187 (0.49)
Structured RegMap	3	0.0337 (0.80)	0.0323 (0.66)	0.0153 (0.46)
Structured RegMap	4	0.0289 (0.77)	0.0284 (0.64)	0.0139 (0.46)
HapMap	1	0.1137 (1.00)	0.1310 (1.00)	0.1040 (1.00)
HapMap	2	0.0356 (0.46)	0.0292 (0.31)	0.0139 (0.19)
HapMap	3	0.0250 (0.40)	0.0226 (0.29)	0.0115 (0.18)
HapMap	4	0.0204 (0.37)	0.0199 (0.29)	0.0106 (0.19)
Swedish RegMap	1	0.0781 (1.00)	0.0827 (1.00)	0.0574 (1.00)
Swedish RegMap	2	0.0391 (0.73)	0.0344 (0.58)	0.0160 (0.36)
Swedish RegMap	3	0.0291 (0.67)	0.0269 (0.54)	0.0132 (0.34)
Swedish RegMap	4	0.0244 (0.64)	0.0237 (0.52)	0.0121 (0.33)
French RegMap	1	0.1141 (1.00)	0.1135 (1.00)	0.0768 (1.00)
French RegMap	2	0.0497 (0.66)	0.0416 (0.52)	0.0192 (0.33)
French RegMap	3	0.0360 (0.60)	0.0322 (0.49)	0.0158 (0.31)
French RegMap	4	0.0297 (0.58)	0.0283 (0.48)	0.0145 (0.31)
VH et al. (2012)	1	0.0680 (1.00)	0.0725 (1.00)	0.0502 (1.00)
VH et al. (2012)	2	0.0329 (0.71)	0.0290 (0.56)	0.0136 (0.37)
VH et al. (2012)	3	0.0244 (0.65)	0.0227 (0.53)	0.0113 (0.35)
VH et al. (2012)	4	0.0204 (0.62)	0.0200 (0.52)	0.0103 (0.36)
RH et al. (2012)	1	0.0698 (1.00)	0.0679 (1.00)	0.0399 (1.00)
RH et al. (2012)	2	0.0374 (0.81)	0.0346 (0.74)	0.0168 (0.60)
RH et al. (2012)	3	0.0286 (0.77)	0.0280 (0.72)	0.0141 (0.59)
RH et al. (2012)	4	0.0244 (0.76)	0.0249 (0.72)	0.0129 (0.60)
Zhao et al. (2011)	1	0.0642 (1.00)	0.0625 (1.00)	0.0329 (1.00)
Zhao et al. (2011)	2	0.0399 (0.87)	0.0351 (0.79)	0.0156 (0.65)
Zhao et al. (2011)	3	0.0311 (0.83)	0.0280 (0.76)	0.0128 (0.62)
Zhao et al. (2011)	4	0.0264 (0.81)	0.0246 (0.74)	0.0116 (0.62)

Shown are standard deviations of the estimators based on replicates and in parentheses the ratios of the standard deviations of the estimators based on replicates over those based on means. Asymptotic variances were obtained using the expressions in Equation 9. Four populations of A. thaliana (containing respectively 250, 350, 305, and 204 accessions), two populations of Z. mays [400 accessions (Van Heerwaarden et al. 2012) and 280 accessions (Riedelsheimer et al. 2012)] and one population of O. sativa [413 accessions (Zhao et al. 2011)] were considered. The second column gives the number of replicates (r). VH, Van Heerwaarden et al. (2012). RH, Riedelsheimer et al. (2012).

the HapMap. The magnitude of the differences was larger than expected from the asymptotic variances in Table 2, and in the HapMap standard errors were up to 13 times larger. As in the asymptotic framework, differences were smallest for the structured RegMap and largest for the HapMap.

Remarkably, also $\hat{H}^2$ was considerably more accurate than $\hat{h}_m^2$. The additional marker information appeared to be less important here than the loss of information on within-genotype variability. The estimator $\hat{h}_r^2$, which includes both sources of information, outperformed $\hat{H}^2$. Differences were largest for the structured RegMap (12.7% reduction when $h^2 = 0.2$, 42.6% reduction when $h^2 = 0.8$), where the large differences in relatedness provide more information than in the case of the HapMap (almost no difference when $h^2 = 0.2$ and 7% reduction when $h^2 = 0.8$). Simulation results for the Swedish and French RegMap are given in Figure S3 and Table S2. In terms of standard deviations of $\hat{h}_r^2$, $\hat{h}_m^2$, and $\hat{H}^2$, these populations were somewhere in between the structured RegMap and HapMap.

For a substantial proportion of the simulated traits in both the HapMap and the structured RegMap, we observed
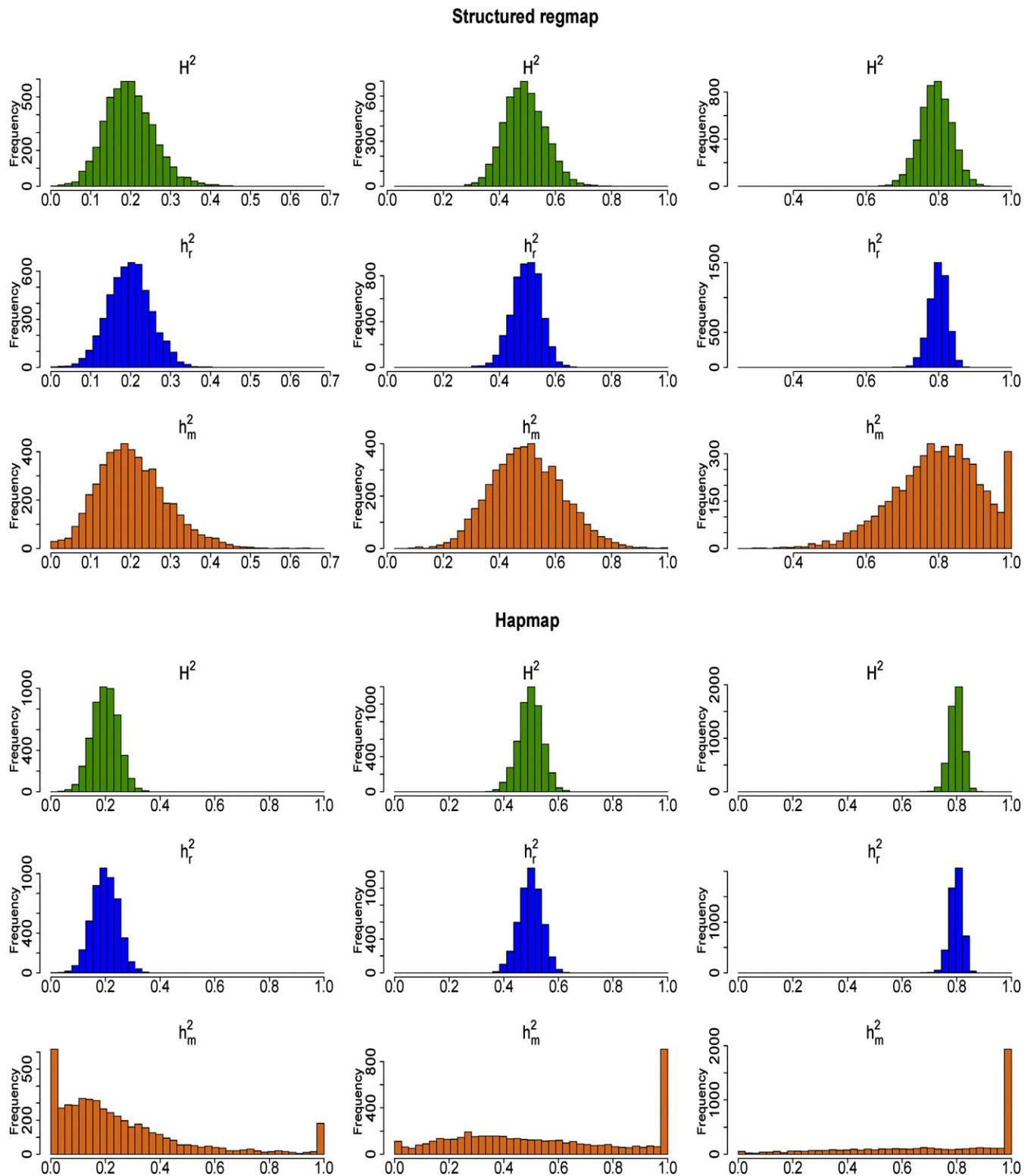
Structured regmap

Hapmap

Figure 1 Heritability estimates for 5000 simulated traits for random samples of 200 accessions drawn from the structured RegMap (top) and the HapMap (bottom). Twenty QTL were simulated, which explained half of the genetic variance. The simulated heritability was 0.2 (left), 0.5 (center), and 0.8 (right). Within each panel, the first row shows the ANOVA-based estimates of broad-sense heritability, the second row the mixed-model-based estimates based on the individual plant data, and the third row the mixed-model-based estimates based on genotypic means.

Table 3 Comparison of the marker-based heritability estimators $\hat{h}_r^2$ and $\hat{h}_m^2$ for simulated data

Estimator	Bias	Standard error	Relative standard error
Structured RegMap			
$h^2 = 0.2$			
Broad sense (H^2)	−0.00314	0.06079	0.70120
Individual level (h_r^2)	−0.00268	0.05306	0.61204
Means (h_m^2)	0.00475	0.08670	1.00000
$h^2 = 0.5$			
Broad sense (H^2)	−0.00938	0.07193	0.56365
Individual level (h_r^2)	−0.00400	0.05093	0.39915
Means (h_m^2)	−0.00164	0.12761	1.00000
$h^2 = 0.8$			
Broad sense (H^2)	−0.00825	0.04309	0.36087
Individual level (h_r^2)	−0.00138	0.02473	0.20712
Means (h_m^2)	−0.00388	0.11940	1.00000
HapMap			
$h^2 = 0.2$			
Broad sense (H^2)	−0.00128	0.04663	0.19404
Individual level (h_r^2)	−0.00154	0.04642	0.19317
Means (h_m^2)	0.05328	0.24033	1.00000
$h^2 = 0.5$			
Broad sense (H^2)	−0.00288	0.04280	0.13754
Individual level (h_r^2)	−0.00295	0.04120	0.13242
Means (h_m^2)	0.04400	0.31115	1.00000
$h^2 = 0.8$			
Broad sense (H^2)	−0.00282	0.02385	0.08263
Individual level (h_r^2)	−0.00231	0.02214	0.07672
Means (h_m^2)	−0.06137	0.28863	1.00000

We simulated 5000 traits, for random samples of 200 accessions drawn from the structured RegMap and HapMap. Twenty unlinked QTL were simulated, which explained 50% of the genetic variance. The simulated heritability was 0.2, 0.5, and 0.8. Standard errors are given relative to those of $\hat{h}_m^2$.

Table 4 Marker-based estimation of heritability

Interval	Coverage	Interval width
Structured RegMap		
$h^2 = 0.2$		
Broad sense	0.864	0.178
Individual level (standard)	0.926	0.201
Individual level (log-transformed)	0.957	0.203
Means (standard)	0.912	0.314
Means (log-transformed)	0.965	0.322
$h^2 = 0.5$		
Broad sense	0.735	0.160
Individual level (standard)	0.948	0.196
Individual level (log-transformed)	0.956	0.194
Means (standard)	0.909	0.460
Means (log-transformed)	0.961	0.433
$h^2 = 0.8$		
Broad sense	0.671	0.086
Individual level (standard)	0.941	0.094
Individual level (log-transformed)	0.941	0.094
Means (standard)	0.923	0.422
Means (log-transformed)	0.946	0.492
HapMap		
$h^2 = 0.2$		
Broad sense	0.948	0.178
Individual level (standard)	0.945	0.181
Individual level (log-transformed)	0.969	0.182
Means (standard)	0.835	0.538
Means (log-transformed)	0.916	0.678
$h^2 = 0.5$		
Broad sense	0.935	0.160
Individual level (standard)	0.948	0.164
Individual level (log-transformed)	0.952	0.163
Means (standard)	0.861	0.806
Means (log-transformed)	0.984	0.832
$h^2 = 0.8$		
Broad sense	0.924	0.084
Individual level (standard)	0.948	0.085
Individual level (log-transformed)	0.950	0.085
Means (standard)	0.892	0.882
Means (log-transformed)	0.914	0.915

Shown are width and coverage confidence intervals obtained from the individual plant data and the genotypic means. Results for broad-sense heritability intervals are reported for comparison. We simulated 5000 traits, for random samples of 200 accessions drawn from the structured RegMap (top) and HapMap (bottom). Twenty unlinked QTL were simulated, which explained 50% of the genetic variance. The simulated heritability was 0.2, 0.5, and 0.8.

a phenomenon that was not apparent from the asymptotic arguments above: $\hat{h}_m^2$ was close to or equal to 1, in which case the likelihood as a function of h^2 was monotonically increasing. This behavior is most likely to occur when the genetic relatedness matrix is close to compound symmetry (*i.e.*, all genotypes being equally related), in which case the likelihood is constant in h^2 (File S3). The small sample size can then lead to monotone likelihood profiles (Figure S4). Indeed $\hat{h}_m^2 = 1$ occurred most often for the HapMap: even for a simulated heritability of 0.5, $\hat{h}_m^2 > 0.99$ for 882 of the 5000 traits (Figure 1). For the HapMap and simulated $h^2 = 0.2$, we also observed $\hat{h}_m^2 = 0$.

The simulations were repeated for a genetic architecture where only 10% of the genetic variance consists of polygenic effects and the remaining 90% is determined by a single QTL (File S4). Even in this extreme scenario, $\hat{h}_r^2$ still performed very well, having almost negligible bias. Also the improvement in standard error relative to $\hat{h}_m^2$ remained similar.

Confidence intervals

For all simulated traits we calculated both standard and log-transformed confidence intervals for narrow-sense heritability (h^2), using genotypic means as well as individual plant data. Also confidence intervals for broad-sense heritability (H^2) were calculated. (We recall that only additive genetic effects were simulated, and hence $h^2 = H^2$.) Coverage and width of confidence intervals are given in Table 4 and Table S3 (Swedish and French RegMap). The standard 95% intervals obtained from the individual observations (*i.e.*, those associated with $\hat{h}_r^2$) had around 95% coverage in the simulations, while coverage was below 95% for the intervals based on genotypic means. The latter intervals also had larger width, which relates to the higher variance of $\hat{h}_m^2$ in simulations. For the HapMap average width was larger than 0.5, even at a simulated heritability of 0.2. For the simulated traits with $\hat{h}_m^2 = 1$, the corresponding confidence interval was [0, 1].

Table 5 Heritability estimates and confidence intervals for two flowering traits from Atwell *et al.* (2010) (LDV and LD) and four traits from new experiments

| Trait	$\hat{h}_r^2$	$\hat{h}_m^2$	$\hat{H}^2$
LDV	0.801 (0.756,0.840)	0.510 (0.303,0.714)	0.858 (0.826,0.886)
LD	0.933 (0.917,0.947)	1.000 (0.000,1.000)	0.966 (0.957,0.973)
LA(S)	0.209 (0.141,0.298)	0.088 (0.028,0.244)	0.235 (0.167,0.306)
LA(H)	0.378 (0.315,0.444)	0.339 (0.134,0.631)	0.388 (0.327,0.451)
BT	0.941 (0.929,0.950)	1.000 (0.000,1.000)	0.956 (0.947,0.963)
LW	0.553 (0.491,0.614)	0.155 (0.028,0.538)	0.530 (0.468,0.589)

Three estimators were used: the marker-based estimator using individual plant data ($\hat{h}_r^2$), the marker-based estimator using genotypic means ($\hat{h}_m^2$), and the ANOVA-based estimator of broad-sense heritability ($\hat{H}^2$). Horizontal lines separate traits measured in different experiments. Trait abbreviations are given in Table 1.

Differences between the log-transformed and the standard intervals were small for $\hat{h}_r^2$. For $\hat{h}_m^2$ the use of the log-transformed intervals gave bigger improvements, although coverage was still below 95%. Also the confidence intervals for broad-sense heritability had insufficient coverage. This is due to model misspecification: the analysis of variance does not incorporate the genetic structure, and uncertainty in the estimates is therefore underestimated. Coverage was still around 93% for the HapMap, whereas for the structured RegMap it decreased to 67.6% for $h^2 = 0.8$.

Heritability estimates for real data

The heritability estimates $\hat{h}_r^2$, $\hat{h}_m^2$, and $\hat{H}^2$ were calculated for the six traits. Genotypic means were calculated as the best linear unbiased estimators (BLUEs) for the genetic effects, in a linear model containing fixed effects for both genotype and additional design effects (*Appendix A*). In the calculation of $\hat{h}_m^2$ the nondiagonal covariance structure of the BLUEs was taken into account. In the case of $\hat{h}_r^2$ and $\hat{H}^2$, design effects were directly included in the mixed model.

For the flowering traits from Atwell *et al.* (2010), our estimates of broad-sense heritability were 0.858 for LDV and 0.966 for LD (Table 5). For several reasons (File S2), these values were lower than those reported in supplementary table 7 of Atwell *et al.* (2010): 0.94 for LDV and 0.99 for LD. For both traits, marker-based heritability estimates based on individual plant data ($\hat{h}_r^2$) were substantially lower, but still high (0.80 for LDV and 0.93 for LD). Marker-based heritability estimates obtained from genotypic means ($\hat{h}_m^2$) were very different: much lower in case of LDV (0.51) and 1 for LD. (Table 5 and Figure 2). In the latter case, the likelihood as a function of h^2 was monotonically increasing, just as we observed for some of the simulated traits. The estimate of residual variance (not reported) was virtually zero in this case, and the confidence interval was equal to the whole unit interval. Also in the leaf area experiments there were substantial differences between $\hat{h}_r^2$ and $\hat{h}_m^2$, especially for the Swedish RegMap ($\hat{h}_r^2 = 0.21$ *vs.* $\hat{h}_m^2 = 0.09$). Heritability estimates were larger for the HapMap ($\hat{h}_r^2 = 0.38$ *vs.* $\hat{h}_m^2 = 0.34$), which has greater genetic diversity. Again, confidence intervals associated with $\hat{h}_m^2$ were wider than those associated with $\hat{h}_r^2$. In the final experiment, heritability
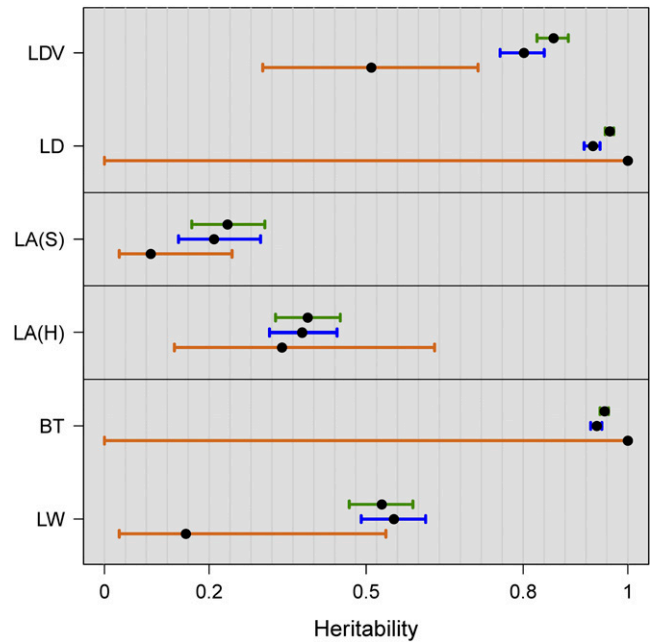
Figure 2 Heritability estimates and confidence intervals for two flowering traits from Atwell *et al.* (2010) (LDV and LD) and four traits from new experiments. Three estimators were used: the ANOVA-based estimator of broad-sense heritability ($\hat{H}^2$, green), the marker-based estimator using individual plant data ($\hat{h}_r^2$, blue) and the marker-based estimator using genotypic means ($\hat{h}_m^2$, brown). Traits from different experiments are separated by the black horizontal lines. Trait abbreviations are given in Table 1. Confidence intervals associated with the marker-based estimates are constructed using the logarithmic transformation described in *Materials and Methods*.

estimates for bolting time (BT) were very similar to those for LD: $\hat{h}_m^2 = 1$, and $\hat{h}_r^2$ and $\hat{H}^2$ were close to 1. For LW we found $\hat{h}_m^2 = 0.16$, much lower than $\hat{h}_r^2 = 0.55$. The latter was slightly larger than the broad-sense heritability estimate $\hat{H}^2 = 0.53$, but the confidence intervals largely overlap.

Genomic prediction with G-BLUP

The G-BLUP depends on the estimated genetic and residual variance and hence on the estimated heritability (Henderson 1975; Robinson 1991). Therefore genomic prediction with G-BLUP could potentially be improved when individual plant data are used, instead of genotypic means. We compared G-BLUP based on genotypic means and individual plant data using simulated traits, each including a training set of $n = 200$ genotypes and a validation set of $m = 50$ genotypes, for which only marker information was available. For both sets of genotypes, we obtained the G-BLUP ($\hat{G}$) predicting the true (simulated) genetic effects G and calculated the prediction accuracy in terms of the correlation (r) between $\hat{G}$ and G.

For the training and validation sets, accuracy decreased when the estimated heritability was too far from the simulated heritability (Figure 3, Figure 4, and Table S5). For the validation sets there were larger differences in accuracy across simulations, because of the additional randomness in the selection of the validation genotypes, creating varying
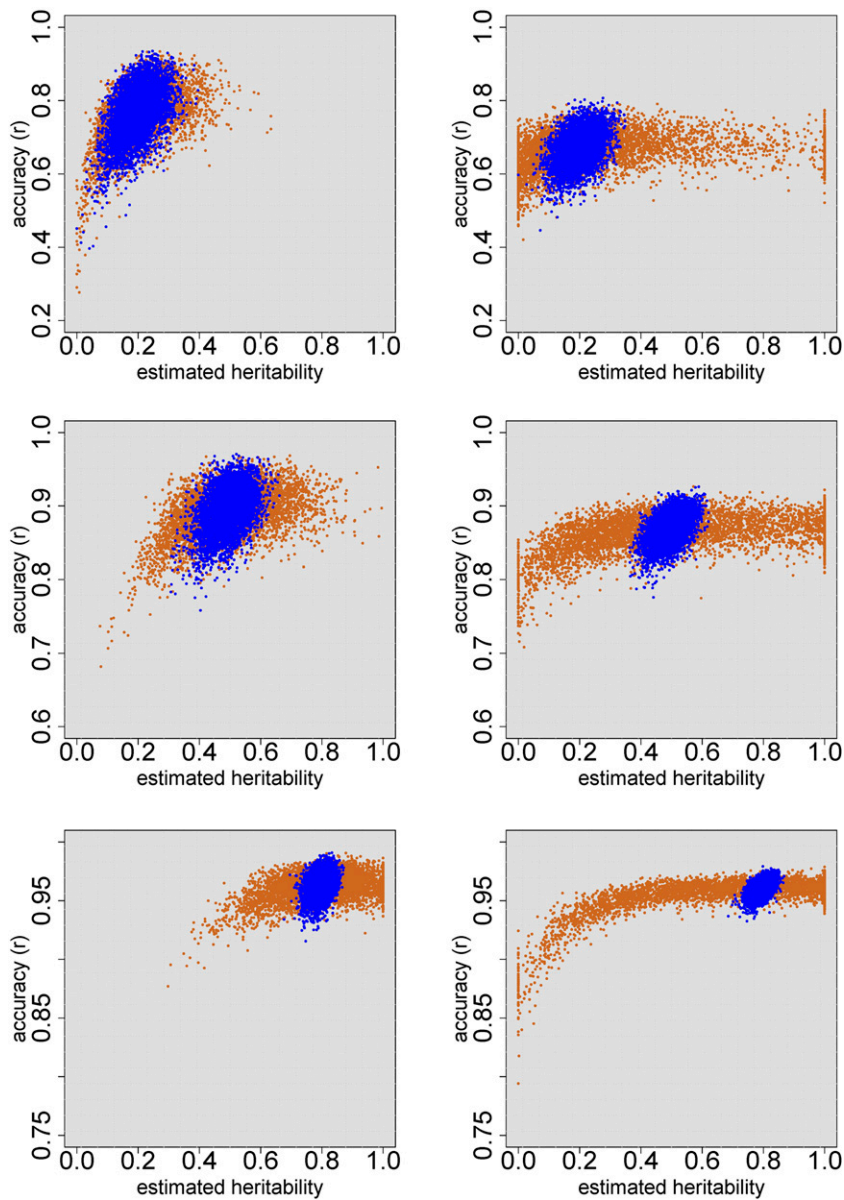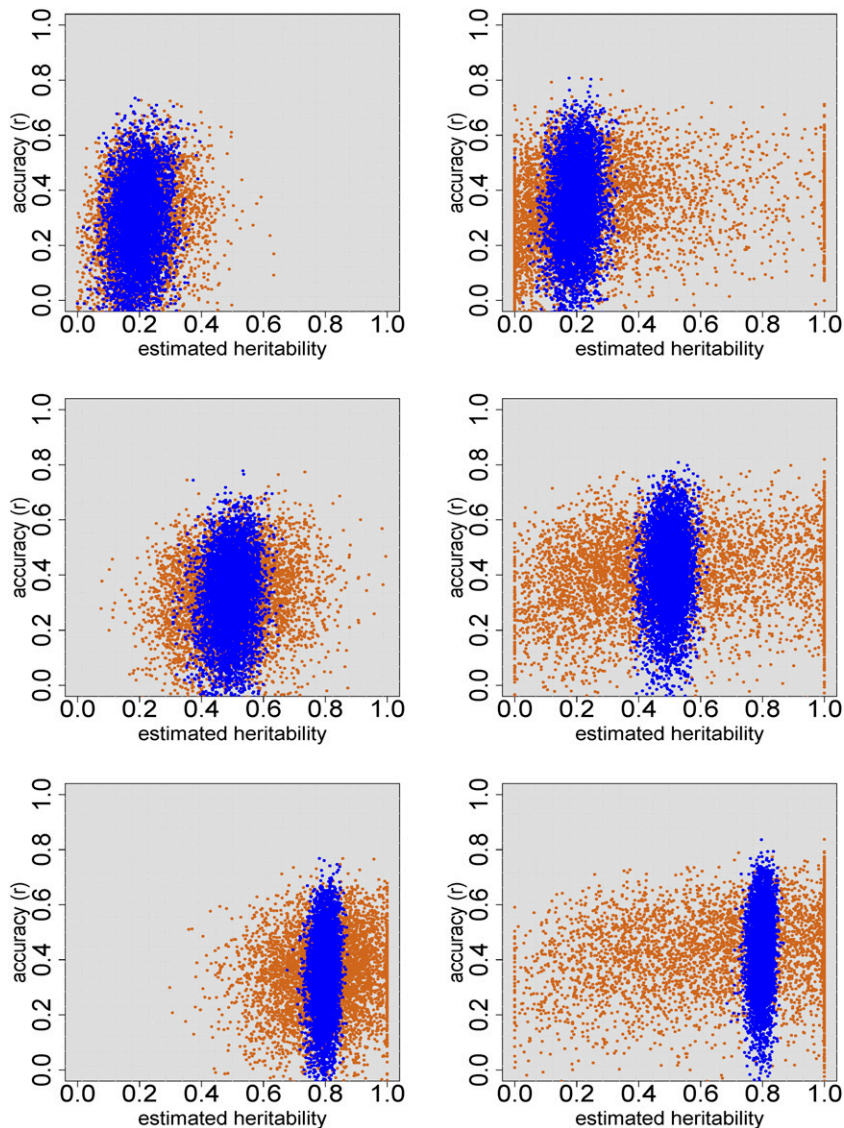
Figure 3 Prediction accuracy (*r*) of G-BLUP on training sets of 200 accessions, for 5000 simulated traits, for the structured RegMap (left) and HapMap (right). Within both populations, each trait was simulated for a randomly drawn training set of 200 accessions. Genetic effects were predicted using G-BLUP, based on a mixed model either for the individual plants (blue) or for the genotypic means (orange). Twenty QTL were simulated, which explained 50% of the genetic variance. The simulated heritability was 0.2 (top), 0.5 (middle), and 0.8 (bottom).

degrees of connectedness with the training sets. When using individual plant data, heritability estimates were never far from the simulated heritability, and prediction accuracy was close to a constant, depending on the simulated heritability. Using only genotypic means, heritability was often severely over- or underestimated, in which case accuracy decreased substantially. This decrease was largest when heritability was underestimated, which follows from the mathematical expressions for G-BLUP (*Materials and Methods* and File S5).

In the HapMap population with simulated heritability 0.8, the estimate $\hat{h}_r^2$ was between 0.7 and 0.9 for 4998 of the 5000 simulated traits, and accuracy (*r*) averaged over these simulations was 0.431 (Table 6). The heritability estimate based on means ($\hat{h}_m^2$) was, however, between 0.1 and 0.3 for 8.4% of the simulated traits, and for 2.6% of them it was even below 0.1. Averaged over the latter group of traits, accuracy was only 0.289. Consequently, for these traits an

improvement in accuracy of 49% could be realized by genomic prediction based on individual plant data instead of genotypic means. Averaged over all simulated traits, the prediction accuracy obtained from genotypic means was not much lower, because of the large "safe zone" where the approaches performed similarly: accuracy was at least 0.4 once the estimated heritability was above 0.5. It is impossible, however, to determine whether a *given* trait is within this zone, using genotypic means only. When individual plant data are available, the heritability estimates $\hat{h}_r^2$ and $\hat{h}_m^2$ could be compared, and the G-BLUP based on genotypic means could be used if $\hat{h}_r^2$ and $\hat{h}_m^2$ were similar.

Similar results were obtained for prediction accuracy observed in cross-validations on the six observed traits (Figure 5). Averaged over 500 validation sets, prediction accuracies obtained with individual plant data and means were almost the same for LDV, LD, and BT (respectively 0.77, 0.82, and

Figure 4 Prediction accuracy (*r*) of G-BLUP on validation sets of 50 accessions, for 5000 simulated traits, for the structured RegMap (left) and HapMap (right). Within both populations, each trait was simulated for a randomly drawn training set of 200 accessions. Genetic effects for a randomly drawn validation set of 50 accessions were predicted using G-BLUP, based on a mixed model either for the individual plants (blue) or for the genotypic means (orange). Twenty QTL were simulated, which explained 50% of the genetic variance. The simulated heritability was 0.2 (top), 0.5 (middle), and 0.8 (bottom).

0.67). For leaf area 13 days after sowing (Swedish RegMap) [LA(S)], leaf area 13 days after sowing (HapMap) [LA(H)], and LW, average accuracies were respectively 0.22, 0.27, and 0.04 when using genotypic means and 0.23, 0.28, and 0.11 when using individual plant data. For LDV, the heritability estimates obtained using genotypic means followed a bimodal pattern, which was largely due to accession 8233, for which only a single observation was available and genotypically is an outlier (Figure S6). Despite the fact that the standard errors and correlations among the genotypic means were taken into account, training sets including this accession produced much lower heritability estimates than in cases where it was assigned to the validation set. This, however, did not lead to lower prediction accuracy. More generally, the relation between low prediction accuracy and underestimating heritability was less clear than in Figure 4, which may be due to the additional uncertainty in the estimation of fixed effects and the fact that this concerns a subsampling of observed traits, rather than simulated traits.

Nonetheless, genomic prediction based on individual plant data again performed at least as well as the standard approach based on means and much better in the case of LW.

GWAS

Many state-of-the-art methods for GWAS (Kang *et al.* 2010; Lippert *et al.* 2011; Lipka *et al.* 2012; Zhou and Stephens 2012) use the same mixed model as in marker-based estimation of heritability, apart from the additional marker effect (*Appendix D*). When testing the significance of this marker effect, the estimated genetic and residual variances (and hence the heritability) determine the correction for population structure or genetic effects elsewhere in the genome. This suggests that poor estimation of the variance components σ_A^2 and σ_E^2 may also affect association mapping, especially when these are estimated in a model without marker effects and then kept fixed when calculating generalized least-squares (GLS) estimates of marker effects (Kang *et al.* 2010). The GLS estimate of the fixed effects (including

Table 6 Prediction accuracy (*r*) of G-BLUP for 5000 simulated traits, for the HapMap population

Interval	$\hat{h}_r^2$ (%)	$\hat{h}_m^2$ (%)	*r* (replicates) training set	*r* (means) training set	*r* (replicates) validation set	*r* (means) validation set
[0, 0.1)	0.00	2.58		0.890		0.289
[0.1, 0.3)	0.00	8.34		0.937		0.373
[0.3, 0.5)	0.00	12.34		0.954		0.409
[0.5, 0.7)	0.04	15.90	0.942	0.959	0.208	0.423
[0.7, 0.9)	99.96	15.62	0.961	0.961	0.431	0.443
[0.9, 1]	0.00	45.22		0.961		0.448
[0, 1]	100	100	0.961	0.956	0.431	0.428

Each trait was simulated for a randomly drawn training set (200 accessions) and validation set (50 accessions). Genetic effects were predicted using G-BLUP, based on a mixed model either for the individual plants (replicates) or for the genotypic means. The second and third columns contain the percentage of the 5000 traits for which the corresponding heritability estimates ($\hat{h}_r^2$ and $\hat{h}_m^2$) were contained in the intervals in the first column. The remaining columns show the correlation (*r*) between simulated and predicted genetic effects, averaged over these traits. The simulated heritability was 0.8; 20 QTL were simulated, which explained 50% of the genetic variance.

the marker effect) $\hat{\beta} = (X^t V^{-1} X)^{-1} X^t V^{-1} Y$ is unbiased since the expectation of Y is $X\beta$. Hence $E(\hat{\beta}) = \beta$, even conditionally on a poor estimate of $V = \hat{\sigma}_A^2 ZKZ^t + \hat{\sigma}_E^2 I_N$. However, P-values for the significance of marker effects will be more sensitive to poor estimates of (σ_A^2, σ_E^2) than the effect estimates themselves. We investigated this with GWAS on simulated traits for the HapMap and RegMap populations, which were performed on genotypic means (two-stage) and individual plant data (one-stage). Rank correlations between one- and two-stage $-\log_{10}$ P-values were almost 1 when $\hat{h}_m^2$ was close to the simulated heritability and decreased when $\hat{h}_m^2$ under- or overestimated heritability (Figure S7). However, even then correlations were still high, and the resulting loss in power appeared to be limited (Figure 6). For the RegMap (results not shown) differences between receiver operating characteristic curves were even smaller. In contrast to the accuracy of genomic prediction, these differences remained small when restricting the set of simulated traits to those for which $\hat{h}_m^2$ underestimated the simulated heritability (curves not shown). This may be explained by the fact that the correlation between $\hat{\beta}$ and $(\hat{\sigma}_A^2, \hat{\sigma}_E^2)$ tends to zero for large numbers of genotypes (Casella and McCulloch 2006).

Software

In most of our analyses and simulations we used the commercial R package asreml (Butler *et al.* 2009), which contains a fast implementation of the AI algorithm. We also made our own implementation of the AI algorithm, together with functions for estimating heritability. These are contained in the R package heritability, which is freely available online. In contrast to other packages such as emma and synbreed, it provides confidence intervals for heritability.

For GWAS on simulated traits we developed the command-line program scan_GLS, which is available on request. scan_GLS performs generalized least-squares calculations conditional on variance components estimated in a model without markers, as proposed by Kang *et al.* (2010), and can efficiently handle genetically identical individuals. State-of-the-art association mapping software can perform association mapping on genetically identical individuals only when these are given different identifiers, *i.e.*, as if they were different genotypes. This leads to large genotypic data files and

increases computation time. The GLS calculations in scan_GLS are more efficient since they use the fact that ZKZ^t and K have the same rank. This has been proposed in Lippert *et al.* (2011) (supplement), but to our knowledge this has not been implemented in the Fast-LMM software. Although we did not find one-stage GWAS to be more powerful in the present study, the ability to perform fast association mapping for genetically identical individuals is useful in the context of a compressed kinship matrix (Bradbury *et al.* 2007; Zhang *et al.* 2010; Lipka *et al.* 2012). scan_GLS also includes a function to perform GLS calculations with nondiagonal residual variance structure, allowing association mapping with extra (possibly nongenetic) random effects.

Discussion

We have presented new methodology for marker-based estimation of heritability for plant traits, which accounts for genetic relatedness and includes information on within-genotype variability, available from replicates. Our approach offers an alternative to mixed-model analysis based on genotypic means, which is the current practice in GWAS and genomic prediction with G-BLUP. Although mixed models can indeed estimate heritability from only kinship coefficients and genotypic means, we observed very large standard errors and sometimes unrealistic estimates of heritability, in both published data and new experiments. Using simulations and statistical arguments we showed that marker-based estimation of heritability based on genotypic means has indeed severe limitations when applied to commonly used association panels in *A. thaliana*. The main reason for this is the lack of information on within-genotype variability: heritability estimates are exclusively based on the (usually small) differences between genotypes. This is feasible in human cohorts with many thousands of individuals, but gives insufficient information in plant populations with only several hundred different genotypes, even if standard errors of genotypic means are taken into account. Much more accurate heritability estimates were obtained with mixed-model analysis at the individual plant or plot level. The resulting heritability estimates had accuracies similar to those reported for human diseases (see *e.g.*, Speed *et al.* 2012). In our simulations with 200 genotypes, the
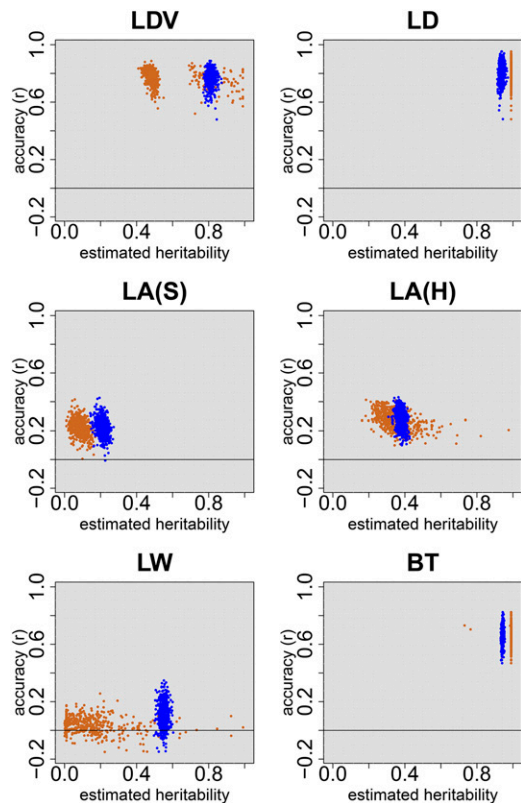
Figure 5 Prediction accuracy (*r*) of G-BLUP in 500 cross-validations, for six traits observed on *A. thaliana*. In each cross-validation, the accessions are randomly partitioned into a training set of 80% and a validation set of 20% of the accessions. Predictions for the individual plant data for the accessions in the validation set were obtained from Equation C6 in *Appendix C*. The genetic effects in (C6) were estimated either using individual plant data (blue) or using genotypic means (orange).

difference in accuracy relative to that of estimates obtained from genotypic means was larger than what was expected from the asymptotic variances. The reason for this larger difference appeared to be the monotone likelihood profile occurring in a substantial part of the simulations, giving heritability estimates of zero or one. However, even without this phenomenon, heritability estimations using individual plant data are more precise, and the asymptotic approximations appear to provide a lower bound on the gain in accuracy.

Mixed-model analysis at the individual plant level can also improve accuracy of G-BLUP, in particular when the actual heritability is underestimated. Too much shrinkage then leads to lower prediction accuracy. In GWAS, where the interest is in the estimated marker effects rather than the variance components, inclusion of the individual plant gave almost no increase in power. However, the possibility to include covariates observed at the individual plant level may be an important practical advantage. While two-stage procedures are usually considered preferable in complex multi-experiment settings (Welham *et al.* 2010; Piepho *et al.* 2012), a one-stage approach may give a more convenient and less error-prone analysis of a single experiment with a simple design.

State-of-the-art phenotyping platforms can measure plant traits with increasing accuracy and throughput. Compared to human traits, the key advantages are that phenotyping is performed under experimental conditions and can include different individuals of the same genotype. Our findings suggest that to fully exploit these advantages, statistical analysis at the individual plant level is necessary. Obviously, this requires the availability of the individual plant data, as well as covariates. For many studies in the literature this information is, however, not available, since most online resources store only genotypic means. More specifically, our results are relevant in the light of the missing heritability debate. Although the aim here was not to propose specific explanations for missing heritability, any such explanation clearly requires an accurate estimate of heritability in the first place. Standard errors for heritability are commonly reported for human traits, but usually absent in the *A. thaliana* literature. We demonstrated that these standard errors can be extremely large when using mixed models at a genotypic means level. Explaining missing heritability based on such estimates then becomes an unreasonable goal.

For two flowering traits, from both a published and a new experiment, $\hat{h}_m^2 = 1$. In these cases, also $\hat{h}_r^2$ and the broad-sense heritability estimates ($\hat{H}^2$) were very high. This can partly be explained by the discrete scale (in whole days) on which the traits were measured. Some genotypes therefore had exactly equal phenotypic values in all or many of the replicates. The heritability estimates for LD were also affected by the fact that in the Atwell *et al.* (2010) data, all nonflowering plants were given a phenotypic value of 200. However, the estimates $\hat{h}_m^2 = 1$ also occurred for some of our simulated (Gaussian) traits and have a more fundamental reason: the small number of genotypes (compared to human studies) leads to monotone likelihood profiles and very large standard errors, making it hard to make any statement about heritability.

Recently, Speed *et al.* (2012) showed that heritability estimates may become biased when linkage disequilibrium (LD) is not constant over the genome and proposed LD-adjusted kinship (LDAK) matrices to correct for this. To assess the effect on $\hat{h}_r^2$ and $\hat{h}_m^2$ we recalculated these heritability estimates, using an LD-adjusted kinship matrix (Figure S5 and Table S4). For all traits the estimates $\hat{h}_r^2$ were very close to the values obtained using the unadjusted kinship matrix. The estimates $\hat{h}_m^2$ on the other hand were quite different for several traits. Nevertheless, the same problems occurred: very large confidence intervals and estimates that appear biologically unrealistic. We conclude that $\hat{h}_m^2$ is more sensitive to the choice of kinship matrix than $\hat{h}_r^2$ and that the different behavior of $\hat{h}_r^2$ and $\hat{h}_m^2$ cannot be explained by LD being different over the genome.

The mixed-models analysis at the individual plant level proposed here can be extended in several ways, for example by partitioning the genetic variance into different chromosomal contributions, as in Yang *et al.* (2011), or by including epistatic effects. The latter has been proposed for experimental
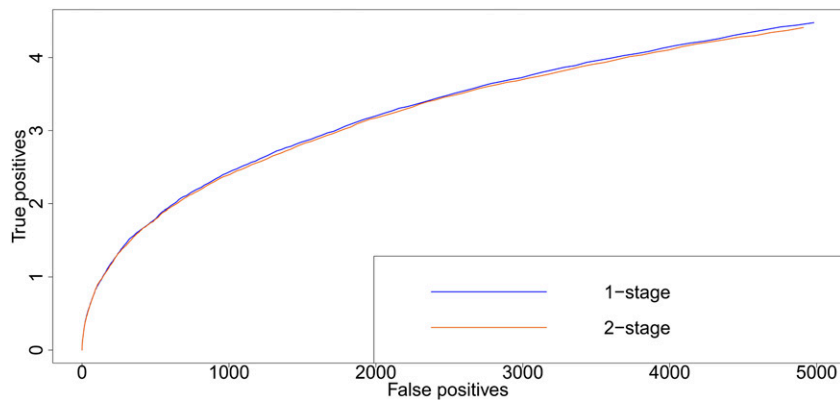
Figure 6 Receiver operating characteristics for the GWAS simulations. One thousand traits were simulated, for random samples of 200 accessions drawn from the HapMap. Ten QTL were simulated, at randomly chosen SNP positions (see *Simulations* in *Materials and Methods* for more details). The QTL explained 75% of the genetic variance. The simulated heritability was 0.5. The curves were obtained by moving from a small (conservative) significance threshold to a larger one. Simulated QTL were counted as true positive if their *P*-value was below the threshold; all other SNPs below the threshold were counted as false positive.

populations with known pedigrees (Oakey *et al.* 2007), but marker-based estimation of epistatic effects for natural populations appears possible only with more advanced statistical methodology, for example semiparametric mixed models and reproducing kernel Hilbert spaces (Gianola and Van Kaam 2008; Howard *et al.* 2014). Another direction for future research is the estimation of heritability in the presence of additional random effects, which would increase the applicability to agricultural field trials (where the raw data are usually at *plot* rather than individual plant level). Although our approach allows for missing values, it does assume that all design effects can be modeled as fixed. Field trials are often laid out in incomplete blocks, containing only a small number of the genotypes under study. Differences between such blocks are usually best modeled using random block effects. The definition of heritability in such contexts is, however, far from obvious: Oakey *et al.* (2007) proposed generalized heritability, but for natural populations this definition is not equivalent to the classical definition of heritability (Oakey *et al.* 2007, p. 813). In this case the ratio of the estimated genetic variance over the total phenotypic variance could be used as a lower bound on heritability. Nevertheless, we have demonstrated that mixed-model analysis at the individual plant or plot level offers important advantages over mixed models for genotypic means.

Acknowledgments

Literature Cited

Atwell, S., Y. S. Huang, B. J. Vilhjalmsson, G. Willems, M. Horton *et al.*, 2010 Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631.

Boer, M. P., D. Wright, L. Feng, D. W. Podlich, L. Luo *et al.*, 2007 A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177: 1801–1813.

Brachi, B., G. Morris, and J. Borevitz, 2011 Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol. 12: 1–8.

Brachi, B., R. Villoutreix, N. Faure, N. Hautekeete, Y. Piquot *et al.*, 2013 Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. Mol. Ecol. 22: 4222–4240.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss *et al.*, 2007 Tassel: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.

Browning, S. R., and B. L. Browning, 2011 Population structure can inflate SNP-based heritability estimates. Am. J. Hum. Genet. 89: 191–193.

Butler, D. G., B. R. Cullis, A. R. Gilmour, and B. J. Gogel, 2009 *ASReml-R Reference Manual.* Queensland Department of Primary Industries and Fisheries, Brisbane, Australia.

Casella, S. R. S. G., and C. E. McCulloch, 2006 *Variance Components.* John Wiley & Sons, Hoboken, NJ.

Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal *et al.*, 2010 Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. 11: 446–450.

Endelman, J. B., 2011 Ridge regression and other kernels for genomic selection with r package rrblup. Plant Genome 4: 250–255.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Prentice-Hall, Englewood Cliffs, NJ.

Fisher, R. A., 1918   The correlation between relatives on the supposition of mendelian inheritance. Trans. R. Soc. Edinb. 52: 399–433.

Fu, R., A. E. Gelfand, and K. E. Holsinger, 2003   Exact moment calculations for genetic models with migration, mutation, and drift. Theor. Popul. Biol. 63: 231–243.

Gianola, D., and J. B. C. H. M. van Kaam, 2008   Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178: 2289–2303.

Gilmour, A. R., R. Thompson, and B. R. Cullis, 1995   Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440–1450.

Goddard, M. E., N. R. Wray, K. Verbyla, and P. M. Visscher, 2010   Estimating effects and making predictions from genome-wide marker data. Stat. Sci. 24: 517–529.

Hayes, B. J., P. M. Visscher, and M. E. Goddard, 2009   Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91: 47–60.

Henderson, C. R., 1975   Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423–447.

Holland, J. B., W. E. Nyquist, and C. T. Cervantes-Martinez, 2010   *Estimating and Interpreting Heritability for Plant Breeding: An Update*, pp. 9–112. John Wiley & Sons, New York/Hoboken, NJ.

Horton, M. W., A. M. Hancock, Y. S. Huang, C. Toomajian, S. Atwell *et al.*, 2012   Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat. Genet. 44: 212–216.

Howard, R., A. L. Carriquiry, and W. D. Beavis, 2014   Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3 (Bethesda): 4: 1027–1046.

Jacquard, A., 1983   Heritability: one word, three concepts. Biometrics 39: 465–477.

Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong *et al.*, 2010   Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42: 348–354.

Kang, H. M. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman *et al.*, 2008   Efficient control of population structure in model organism association mapping. Genetics 178: 1709–1723.

Keurentjes, J. J. B., L. Bentsink, C. Alonso-Blanco, C. J. Hanhart, H. Blankestijn-De Vries *et al.*, 2007   Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175: 891–905.

Lee, S. H. H., N. R. Wray, M. E. Goddard, and P. M. Visscher, 2011   Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. 88: 294–305.

Li, Y., Y. Huang, J. Bergelson, M. Nordborg, and J. O. Borevitz, 2010   Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 107: 21199–21204.

Lipka, A. E., F. Tian, Q. Wang, J. Peiffer, M. Li *et al.*, 2012   Gapit: genome association and prediction integrated tool. Bioinformatics 28: 2397–2399.

Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson *et al.*, 2011   FaST linear mixed models for genome-wide association studies. Nat. Methods 8: 833–835.

Long, Q., F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow *et al.*, 2013   Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat. Genet. 45: 884–890.

Lynch, M., and B. Walsh, 1998   *Genetics and Analysis of Quantitative Traits*, Ed. 1. Sinauer Associates, Sunderland, MA.

Malosetti, M., J.-M. Ribaut, and F. A. van Eeuwijk, 2013   The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. 4: 44.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff *et al.*, 2009   Finding the missing heritability of complex diseases. Nature 461: 747–753.

Möhring, J., and H. P. Piepho, 2009   Comparison of weighting in two-stage analyses of series of experiments. Crop Sci. 49: 1977–1988.

Nyquist, W. E., and R. Baker, 1991   Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10: 235–322.

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.

Oakey, H., A. Verbyla, B. Cullis, X. Wei, and W. Pitchford, 2007   Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. Theor. Appl. Genet. 114: 1319–1332.

Patterson, N., A. L. Price, and D. Reich, 2006   Population structure and eigenanalysis. PLoS Genet. 2: e190.

Piepho, H., and E. Williams, 2006   A comparison of experimental designs for selection in breeding trials with nested treatment structure. Theor. Appl. Genet. 113: 1505–1513.

Piepho, H. P., A. Büchse, and B. Truberg, 2006   On the use of multiple lattice designs and alpha-designs in plant breeding trials. Plant Breed. 125: 523–528.

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

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow *et al.*, 2012   Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. 44: 217–220.

Ritland, K., 1996   A marker-based method for inferences about quantitative inheritance in natural populations. Evolution 50: 1062–1073.

Robinson, G. K., 1991   That blup is a good thing: the estimation of random effects. Stat. Sci. 6: 15–32.

Rousset, F., 2002   Inbreeding and relatedness coefficients: What do they measure? Heredity 88: 371–380.

Sillanpaa, , M. J., 2011   On statistical methods for estimating heritability in wild populations. Mol. Ecol. 20: 1324–1332.

Singh, M., S. Ceccarelli, and J. Hamblin, 1993   Estimation of heritability from varietal trials data. Theor. Appl. Genet. 86: 437–441.

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

Smith, A. B., B. R. Cullis, and R. Thompson, 2005   The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J. Agric. Sci. 143: 449–462.

Speed, D., G. Hemani, M. R. Johnson, and D. J. Balding, 2012   Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. 91: 1011–1021.

Stich, B., J. Mohring, H.-P. Piepho, M. Heckenberger, E. S. Buckler *et al.*, 2008   Comparison of mixed-model approaches for association mapping. Genetics 178: 1745–1754.

van der Vaart, A. W., 2000   *Asymptotic Statistics* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, Cambridge/London/New York.

van Eeuwijk, F. A., M. C. A. M. Bink, K. Chenu, and S. C. Chapman, 2010   Detection and use of qtl for complex traits in multiple environments. Curr. Opin. Plant Biol. 13: 193–205.

van Heerwaarden, J., M. B. Hufford, and J. Ross-Ibarra, 2012   Historical genomics of North American maize. Proc. Natl. Acad. Sci. USA 109: 12420–12425.

Vattikuti, S., J. Guo, and C. C. Chow, 2012   Heritability and genetic correlations explained by common snps for metabolic syndrome traits. PLoS Genet. 8: e1002637.

Verbyla, A., B. Cullis, and R. Thompson, 2007 The analysis of qtl by simultaneous use of the full linkage map. Theor. Appl. Genet. 116: 95–111.

Visscher, P. M., and M. E. Goddard, 2014 A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. Genetics 199: 223–232.

Visscher, P. M., S. E. Medland, M. A. R. Ferreira, K. I. Morley, G. Zhu *et al.*, 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2: e41.

Visscher, P. M., W. G. Hill, and N. R. Wray, 2008 Heritability in the genomics era–concepts and misconceptions. Nat. Rev. Genet. 9: 255–266.

Welham, S. J., B. J. Gogel, A. B. Smith, R. Thompson, and B. R. Cullis, 2010 A comparison of analysis methods for late-stage variety evaluation trials. Aust. N. Z. J. Stat. 52: 125–149.

Wijnen, C. L., and J. J. Keurentjes, 2014 Genetic resources for quantitative trait analysis: novelty and efficiency in design from an Arabidopsis perspective. Curr. Opin. Plant Biol. 18: 103–109.

Wimmer, V., T. Albrecht, H.-J. Auinger, and C.-C. Schoen, 2012 synbreed: a framework for the analysis of genomic prediction data using r. Bioinformatics 28: 2086–2087.

Wright, S., 1920 The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. Proc. Natl. Acad. Sci. USA 6: 320–332.

WTCCC, 2007 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011 Gcta: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88: 76–82.

Zhang, Z., E. Ersoz, C.-Q. Q. Lai, R. J. Todhunter, H. K. Tiwari *et al.*, 2010 Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42: 355–360.

Zhao, K., C.-W. W. Tung, G. C. Eizenga, M. H. Wright, M. L. Ali *et al.*, 2011 Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat. Commun. 2: 467.

Zhou, X., and M. Stephens, 2012 Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44: 821–824.

Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander, 2012 The mystery of missing heritability: genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA 109: 1193–1198.

Communicating editor: A. H. Paterson

Appendix A

Marker-Based Estimation of Heritability

We start with a brief summary of existing methodology and the required assumptions and then define marker-based heritability estimates given genetically identical replicates measured in an unbalanced design.

Marker-based estimation of heritability given a single observation per genotype

Phenotypic variance can be partitioned as

$$\sigma_{\text{Pheno}}^2 = \sigma_G^2 + \sigma_{\text{Env}}^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_{\text{Env}}^2,$$

where the genetic variance σ_G^2 is decomposed into additive, dominance, and interaction effects (Falconer and Mackay 1996), the dominance effect being absent for inbred lines. Heritability in the broad sense is defined as $H^2 = \sigma_G^2/\sigma_{\text{Pheno}}^2$, while the narrow-sense heritability $h^2 = \sigma_A^2/\sigma_{\text{Pheno}}^2$ takes into account only the additive genetic effects, which determine breeding values and the selection response. Defining the residual variance σ_E^2 to be the sum of the environmental and nonadditive genetic variance terms ($\sigma_D^2 + \sigma_I^2 + \sigma_{\text{Env}}^2$), we can write

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2}. \tag{A1}$$

Marker-based estimates of narrow-sense heritability can be obtained using mixed models with random genetic effects. The covariances between these effects are modeled by a genetic relatedness matrix (GRM) estimated from markers, with elements given by (1). Given a single observation per genotype, the standard infinitesimal model is

$$Y_i = \mu + x_i\beta + G_i + E_i \quad (i = 1, \ldots, n), \tag{A2}$$

where μ is the intercept, $G = (G_1, \ldots, G_n)$ has a $N(0, \sigma_A^2 K)$ distribution, and the errors E_i have independent normal distributions with variance σ_E^2. The optional term $x_i\beta = x_i^{(1)}\beta_1 + \ldots + x_i^{(k)}\beta_k$ models the effect of k additional covariates. Under model (A2), heritability can be estimated by

$$h^2 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_E^2}, \tag{A3}$$

where $\hat{\sigma}_A^2$ and $\hat{\sigma}_E^2$ are REML estimates of additive genetic and residual variance (Yang *et al.* 2011; Speed *et al.* 2012).

It has been noted (*e.g.*, Hayes *et al.* 2009) that using model (A2) with genetic relatedness matrix (1) is equivalent to assuming that the effects of the standardized marker scores are drawn independently from Gaussian distributions with variance σ_A^2/p. Consequently, the model can account only for additive genetic effects, and nonadditive effects will get into the residual variance. This is why $\hat{h}^2$ is an estimate of narrow-sense heritability. The preceding argument also highlights the fact that $\hat{h}^2$ is a marker-based estimate, which requires the (frequently made) assumptions that every causal locus is tagged by one of the markers, and that LD is constant across the genome (Speed *et al.* 2012). For the case that LD is not constant over the genome, Speed *et al.* (2012) proposed LD-adjusted kinship matrices.

The parameter σ_A^2 can be interpreted as additive genetic variance only if K is scaled appropriately. For the kinship matrix used here this is the case, but for different types of kinship matrix (*e.g.*, identity by state), it is necessary to divide each coefficient by $\text{tr}(PKP)/(n-1)$, for $P = I_n - \mathbf{11}'/n$. Under the condition that $\text{tr}(PKP) = (n-1)$,

$$E\left[\frac{1}{n-1}\sum_{i=1}^n (G_i - \overline{G})^2\right] = E\left[\frac{1}{n-1}G^t P^t P G\right] = \frac{1}{n-1}E[G^t P G]$$
$$= \frac{\sigma_A^2}{n-1}\text{tr}[PKP] = \sigma_A^2. \tag{A4}$$

Marker-based estimation of heritability given genetically identical replicates, for unbalanced designs

When for each genotype a number of genetically identical individuals are observed, (A2) generalizes to

$$Y_{i,j} = \mu + x_{i,j}\beta + G_i + E_{i,j} \quad (i = 1, \ldots, n, \quad j = 1, \ldots, r_i), \tag{A5}$$

where $Y_{i,j}$ is the phenotypic response of replicate j of genotype i, $x_{i,j}$ is a vector of fixed effects, $G = (G_1, \ldots, G_n)$ has a $N(0, \sigma_A^2 K)$ distribution, and the errors $E_{i,j}$ have independent normal distributions with variance σ_E^2. Equation A5 is similar

to Equation 2 apart from the additional covariates and the generally unbalanced design. Analogous to the case of balanced designs, we use the estimator $\hat{h}_r^2$ defined in (3) for REML estimates $\hat{\sigma}_A^2$ and $\hat{\sigma}_E^2$ obtained for model (A5).

Heritability estimates based on genotypic means can be constructed as in Equations 4 and 5, except that these genotypic means are no longer equal to the arithmetic averages $\overline{Y}_i$. We fit the linear model

$$Y = X_G\, g + X_C \beta + E, \tag{A6}$$

where $Y = (Y_{1,1}, \ldots, Y_{1,r_1}, \ldots, Y_{n,1}, \ldots, Y_{n,r_n})^t$ is the vector of phenotypic observations, r_i is the number of replicates of genotype i, X_C is the design matrix of the extra fixed effects, and $E = (E_{1,1}, \ldots, E_{n,r_n})^t$ is the vector of independent Gaussian errors with variance σ_E^2. This model is identical to (A5), apart from the factor genotype now being fixed instead of random. The intercept is included in the incidence matrix X_G, which has dimension $N \times n$, for $N = \sum_{i=1}^n r_i$.

We obtain least-squares estimates $\hat{g} = (\hat{g}_1, \ldots, \hat{g}_n)^t$ for the genotypic effects in model (A6), which form the basis for the subsequent analysis. In mixed-model terminology, this is the BLUE for g. In (A6) the only random term is E, $i.e.$, we consider an ordinary linear model, as the additional covariates are all assumed to be fixed. The least-squares estimator $(\hat{g}, \hat{\beta})$ is given by

$$\begin{pmatrix} \hat{g} \\ \hat{\beta} \end{pmatrix} = \left(X^t X\right)^{-1} X^t Y = \begin{pmatrix} X_G^t X_G & X_G^t X_C \\ X_C^t X_G & X_C^t X_C \end{pmatrix}^{-1} \begin{pmatrix} X_G^t \\ X_C^t \end{pmatrix} Y, \tag{A7}$$

where $X = [X_G\, X_C]$. This vector has covariance matrix

$$\mathrm{Var}\begin{pmatrix} \hat{g} \\ \hat{\beta} \end{pmatrix} = \left(X^t X\right)^{-1} \sigma_E^2 = \begin{pmatrix} X_G^t X_G & X_G^t X_C \\ X_C^t X_G & X_C^t X_C \end{pmatrix}^{-1} \sigma_E^2. \tag{A8}$$

Consequently, the covariance matrix of $\hat{g}$ is

$$\mathrm{Var}(\hat{g}) = \left(X_G^t X_G - X_G^t X_C \left(X_C^t X_C\right)^{-1} X_C^t X_G\right)^{-1} \sigma_E^2 = R\sigma_E^2.$$

The mixed model (4), used for the estimation of heritability, now naturally generalizes to

$$\overline{Y}_i = \mu + G_i + E_i, \quad G \sim N\left(0, \sigma_A^2 K\right), \quad E \sim N\left(0, \sigma_E^2 R\right), \tag{A9}$$

where, to avoid having twice the letter g in the same equation, we abused the notation by writing $\overline{Y}_i$ instead of $\hat{g}_i$. Hence the matrix $r^{-1} I_n$ in (4) is replaced by R. Given REML estimates $\hat{\sigma}_A^2$ and $\hat{\sigma}_E^2$ for this model, we use $\hat{h}_m^2$ in (5) as heritability estimate. As in the case of a completely randomized design, $\hat{h}_m^2$ is a so-called two-stage estimator. The fact that genotype is taken as a fixed effect in the first stage (A6) and as random in the subsequent mixed-model analysis may appear inelegant, but is common in two-stage analyses and necessary to avoid shrinking twice.

In the case of a completely randomized design with r_i replicates without further covariates, $\hat{g}_i = \sum_{j=1}^{n_i} Y_{i,j}/n_i$ and

$$R = \left(X_G^t X_G\right)^{-1} = \mathrm{diag}\left(r_1^{-1}, \ldots, r_n^{-1}\right).$$

Note that in contrast to the estimator of broad-sense heritability described below, it is not necessary to define an average number of replicates. In the case of a randomized complete block design with r replicates, X_C is the $nr \times (r-1)$ design matrix for the factor block, and $X_G^t X_C (X_C^t X_C)^{-1} X_C^t X_G$ is the $n \times n$ matrix with elements $(r-1)/n$. Hence, R is not diagonal, even though the design is balanced. The off-diagonal elements are equal in this case and small (provided n is sufficiently large). With $n = 200$ and $r = 3$ for example, the diagonal elements of R are 0.336667 and the off-diagonal elements are 0.003333. For unbalanced designs, some of the off-diagonal elements may be larger and more influential.

Appendix B

Simulations

Data for genotypes $i = 1, \ldots, n$ with replicates $j = 1, \ldots, r$ were simulated following the model

$$\begin{aligned} y_{i,j} &= \mu + \sum_{m=1}^q x_{i,m} \alpha_m + g_i + e_{i,j}, \\ g &= (g_1, \ldots, g_n)^t \sim N\left(0, \sigma_a^2 K\right), \\ e &= (e_{1,1}, \ldots, e_{n,r})^t \sim N\left(0, \sigma_e^2 I_{nr}\right), \end{aligned} \tag{B1}$$

where K was scaled such that $\operatorname{tr}(PKP) = n - 1$, for $P = I_n - \mathbf{1}\mathbf{1}'/n$ (see Equation A4).

We assumed q QTL located at marker positions, with effect sizes α_m ($m = 1, \ldots, q$) and minor allele frequencies f_m ($m = 1, \ldots, q$); i.e., genotypes AA and aa occur with frequencies f_m and $(1 - f_m)$, α_m being the effect of a single allele. Let $x_{i,m} \in \{0, 1, 2\}$ denote the marker score at QTL m for genotype i. In the case of A. thaliana inbred lines, $x_{i,m} \in \{0, 2\}$ and

$$h^2 = \frac{\sigma_a^2 + 4\sum_{m=1}^q f_m(1 - f_m)\alpha_m^2}{\sigma_a^2 + 4\sum_{m=1}^q f_m(1 - f_m)\alpha_m^2 + \sigma_e^2} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}, \tag{B2}$$

where

$$\sigma_A^2 = \sigma_a^2 + 4\sum_{m=1}^q f_m(1 - f_m)\alpha_m^2$$

is the total (additive) genetic variance. For outbreeding species under Hardy–Weinberg equilibrium, the constant 4 is replaced by a 2. In (B2) we assume linkage equilibrium between the QTL.

For each simulated trait, QTL locations were sampled randomly from all available markers whose minor allele frequency >10%. To enforce linkage equilibrium between the sampled QTL, the sampled locations were discarded and drawn again if the difference between $v_1 = 4\sum_{m=1}^q f_m(1 - f_m)\alpha_m^2$ and $v_2 = 4(\alpha_1, \ldots, \alpha_q)S(\alpha_1, \ldots, \alpha_q)'$ was too large, S being the sample covariance matrix of the marker scores at QTL positions. It was required that $\min(v_1, v_2)/\max(v_1, v_2) > 0.97$.

Let

$$\gamma = \frac{4}{\sigma_A^2} \sum_{m=1}^q f_m(1 - f_m)\alpha_m^2 \tag{B3}$$

denote the proportion of the genetic variance explained by the QTL. In our main set of simulations we choose $\sigma_e^2 = 1$, $\gamma = 0.5$, $q = 20$ QTL. For the comparison of one- and two-stage association mapping, we choose $\gamma = 0.75$, $q = 10$. To achieve the desired level of heritability (for given γ and $\sigma_e^2 = 1$), we set $\sigma_A^2 = h^2(n - 1)/((1 - h^2)n)$ and $\sigma_a^2 = (1 - \gamma)\sigma_A^2$, and $\alpha_1, \ldots, \alpha_q$ were chosen equally large, such that $4qf_m(1 - f_m)\alpha_m^2 = \gamma\sigma_A^2$ for each m. The QTL effects were given random signs. In File S4 we repeat the simulations for $\gamma = 0.1$ and $q = 1$.

For the purpose of genomic prediction, 50 additional genotypes were drawn randomly from the same subpopulation of the RegMap. These were assigned only genotypic values, defined as the sum of QTL effects $\sum_{m=1}^q x_{i,m}\alpha_m$ and polygenic effects g_i ($i = 201, \ldots, 250$). The polygenic effects were simulated such that $(g_1, \ldots, g_{250})^t \sim N(0, \sigma_a^2 K_{\text{total}})$, K_{total} being the kinship matrix of the training and validation sets combined.

Appendix C

Genomic Prediction with G-BLUP

Genomic prediction with G-BLUP relies on the same models used for marker-based estimation of heritability and can be based on individual plant data (one-stage) or genotypic means (two-stage). We now provide the expressions for one- and two-stage G-BLUP, which are based on models (A5) and (A9) in *Appendix A*.

One- and two-stage G-BLUP for the training and validation sets

First we consider the G-BLUP for the training set, for which phenotypic observations are available. Including the intercept in the design matrix, model (A5) can be rewritten as

$$Y = X\beta + ZG + E, \tag{C1}$$

where N is the total number of individuals and Z is the $N \times n$ incidence matrix assigning individuals to genotypes. The BLUP of $G = (G_1, \ldots, G_n)^t$ and the BLUE of β are given by

$$\hat{G} = \hat{\delta}KZ^t\left(\hat{\delta}ZKZ^t + I_N\right)^{-1}\left(Y - X\hat{\beta}\right), \quad \hat{\beta} = \left(X^t\left(\hat{\delta}ZKZ^t + I_N\right)^{-1}X\right)^{-1}X^t\left(\hat{\delta}ZKZ^t + I_N\right)^{-1}Y, \tag{C2}$$

where N is the total number of individuals, Z is the $N \times n$ incidence matrix assigning individuals to genotypes, and $\hat{\delta} = \hat{\sigma}_A^2/\hat{\sigma}_E^2$. Numerically $\hat{G}$ and $\hat{\beta}$ can be more conveniently obtained by solving the mixed-model equations (see, e.g.,

Henderson 1975 or Robinson 1991). Similar expressions hold for predictions based on the genotypic means following model (A9),

$$\hat{G} = \hat{\delta}K\left(\hat{\delta}K + R\right)^{-1}(\overline{Y} - 1_n\hat{\mu}), \quad \hat{\mu} = \left(1_n^t\left(\hat{\delta}K + R\right)^{-1}1_n\right)^{-1}1_n^t\left(\hat{\delta}K + R\right)^{-1}\overline{Y}, \tag{C3}$$

where, as in (A9), $\overline{Y}$ denotes the vector of genotypic means obtained from preliminary linear model analysis (not necessarily the arithmetic averages). Since we assume that all covariates have been accounted for already in this preliminary analysis, the only fixed effect in (C3) is an intercept, with design matrix $X = 1_n$.

Assuming the one-stage model (C1) and $\hat{\beta}$ given by (C2), the genetic effects $G_{\text{pred}} = (G_{n+1}, \ldots, G_{n+m})^t$ of m unobserved (but genotyped) genotypes can be predicted by the conditional means

$$\hat{G}_{\text{pred}}^{(1)} := E\left[G_{\text{pred}}|Y\right] = \hat{\delta}K_{\text{pred.obs}}Z^t\left(\hat{\delta}ZKZ^t + I_N\right)^{-1}\left(Y - X\hat{\beta}\right), \tag{C4}$$

where $K_{\text{pred.obs}}$ is the $m \times n$ matrix of kinship coefficients for the unobserved $vs.$ observed genotypes. Assuming the two-stage model (A9) and $\hat{\mu}$ given by (C3), the predictor is given by

$$\hat{G}_{\text{pred}}^{(2)} = \hat{\delta}K_{\text{pred.obs}}\left(\hat{\delta}K + R\right)^{-1}(\overline{Y} - 1_n\hat{\mu}). \tag{C5}$$

Prediction of new observations and cross-validation
For new observations Y_{pred} at the individual plant level, we have the one-stage predictor

$$\hat{Y}_{\text{pred}}^{(1)} = X_{\text{pred}}\hat{\beta} + Z_{\text{pred}}\hat{G}_{\text{pred}}^{(1)}, \tag{C6}$$

where X_{pred} and Z_{pred} are the corresponding design matrices. For predictions with genotypic means, we replace $\hat{G}_{\text{pred}}^{(1)}$ by $\hat{G}_{\text{pred}}^{(2)}$ and β by the estimate obtained within the linear model (A6) in the preliminary stage.

Appendix D

Genome-Wide Association Studies

In mixed-model-based GWAS (Kang *et al.* 2010; Lippert *et al.* 2011; Lipka *et al.* 2012; Zhou and Stephens 2012) the phenotype of genotype i is modeled as

$$Y_i = \mu + x_i\gamma + G_i + E_i, \tag{D1}$$

where x_i is the marker score, γ is the marker effect, and the genotypic effects $G = (G_1, \ldots, G_n)$ follow a $N(0, \sigma_A^2 K)$ distribution. This model assumes a single observation per genotype. If observations on genetically identical individuals are available, the Y_i's can be replaced by genotypic means $\hat{g}_i$, as in (A7)–(A9). Model (D1) then generalizes to

$$\hat{g}_i = \mu + x_i\gamma + G_i + E_i, \quad G \sim N\left(0, \sigma_A^2 K\right), \quad E \sim N\left(0, \sigma_E^2 R\right). \tag{D2}$$

This amounts to a two-stage approach: first the genotypic means are calculated using (A7), and next association mapping is performed following (D2). In practice a GWAS is often performed on the arithmetic averages $\overline{Y}_i$, which implicitly assumes a balanced completely randomized design, without any missing values or replicate effects. Here we use the more general model (D2). Apart from the additional marker effect, this is the same model we used in (A9) to construct the heritability estimate $\hat{h}_m^2$.

Alternatively, association mapping can be based on the one-stage model

$$Y_{i,j} = \mu + x_i\gamma + (X_C)_{i,j}\beta + G_i + E_{i,j}, \tag{D3}$$

where the term $(X_C)_{i,j}\beta$ models additional covariates, as in (A7). In the two-stage GWAS, this information is accounted for in the genotypic means $\hat{g}_i$. In models (D2) and (D3) we test the hypothesis $\gamma = 0$ using the F-test, conditional on estimates of the variance components obtained from a model without markers (Kang *et al.* 2010). When marker effects are small, these estimates are a good approximation of the exact estimates, obtained when genetic and residual variances are reestimated for each marker.

Appendix E

Phenotypic Data

Collection of phenotypic data
For the two flowering traits from Atwell *et al.* (2010) (LDV and LD), details can be found in the original publication. In the case of LD, all plants that had not flowered by the end of the experiment were given a phenotypic value of 200 days. The leaf area trait [LA(S)/LA(H)] was measured in two separate experiments on the Swedish RegMap and the HapMap, using the same phenotyping platform. The plants were imaged top down for projected leaf area every day. The images were taken using near-infrared light (790 nm) to not influence the photoperiod during the night measurements. In our final experiment, BT and LW were measured for the HapMap. BT was noted as the number of days after vernalization on which the plant started to bolt. LW was measured on photographs taken from the longest leaf 2 weeks after flowering.

Plant growth conditions
In the leaf area experiments, seeds from the *Arabidopsis* Swedish RegMap (298 accessions) and the *Arabidopsis* HapMap (350 accessions) were stratified at 4° for 4 days and sown on Rockwool blocks that had been covered with a black foamed PVC sheet to prevent algal growth and provide a uniform background for automated image analysis. The growth conditions were a light intensity of 200 μmol·m^{-2}·sec^{-1}, 10-hr short day, 20°/18° day/night cycle, and 70% relative humidity (RH). In the experiment measuring BT and LW on the HapMap, seeds were sown on filter paper with demi water and stratified at 4° in dark conditions for 5 days. Following stratification, seeds were transferred to a culture room (16-hr LD, 24°) to induce seed germination for 42 hr. Germinating seeds were then transplanted to wet Rockwool blocks of 4×4 cm in a climate chamber with a light intensity of 125 μmol·m^{-2}·sec^{-1} 16-hr LD, 20°/18° day/ night cycle, and 70% RH. All plants were watered every morning for 5 min at 9 AM with 1/1000 Hyponex solution (Hyponex, Osaka, Japan). Nineteen days after germination, all plants were vernalized for 6 weeks in a cold room (12 hr light, 4°). After the 6-week vernalization period plants were transferred back to the same climate chamber in the same order, but given more space to grow.

Genotypic means
All experiments were laid out as randomized complete block designs, apart from the final HapMap experiment (BT and LW), where the same randomization was used within each replicate. In all experiments we included a replicate (complete block) effect. The numbers of replicates in the different experiments were six (LD and LDV), four [LA(S)/LA(H)], and three (BT and LW). Due to nongerminating seeds and dead plants, some accessions had lower numbers of replicates. For the leaf area experiments [LA(S)/LA(H)] we additionally included a row and column effect to model the within-image position of each plant ($x = 1, 2, 3$ and $y = 1, 2, 3, 4$). These factors correct for technical artifacts, which are known to be consistent across replicates.

GENETICS

Marker-Based Estimation of Heritability in Immortal Populations

Willem Kruijer, Martin P. Boer, Marcos Malosetti, Pádraic J. Flood, Bas Engel, Rik Kooke, Joost J. B. Keurentjes, and Fred A. van Eeuwijk

File S1: confidence intervals for broad-sense heritability

Confidence intervals for the broad-sense heritability estimates obtain from the ANOVA mean sums of squares are traditionally obtained from the ratio $F = MS(G)/MS(E)$ and the quantiles of the F-distribution with the corresponding degrees of freedom. Given n genotypes with $r_1, \ldots, r_n$ replicates, the intervals are given by

$$\frac{F/F_{\text{df1},\text{df2},0.95} - 1}{F/F_{\text{df1},\text{df2},0.95} + \bar{r} - 1} < H^2 < \frac{F/F_{\text{df1},\text{df2},0.05} - 1}{F/F_{\text{df1},\text{df2},0.05} + \bar{r} - 1},$$

where $\text{df1} = n - 1$, $\text{df2} = \sum(r_i - 1)$ and $\bar{r} = (n-1)^{-1}(\sum r_i - (\sum r_i^2)/(\sum r_i))$. In case of a balanced design with $r_i = r$ replicates, this reduces to $\bar{r} = r$ and $\text{df2} = n(r - 1)$. See [1] (p.563) or [2].

File S2: analysis of flowering traits of [3]

Our broad-sense heritability estimates differ from those reported in Supplementary table 7 of [3], for the following three reasons. First, the broad-sense heritability estimates in [3] were calculated using the formula

$$\frac{MS(G)}{MS(G) + MS(E)}. \tag{1}$$

Although this quantity may be an adequate criterion to compare heritabilities of traits within the same experiment (as long as they have the same number of replicates), this is a biased estimator of broad-sense-heritability. Since the expectation of $MS(G)$ is $r\sigma_G^2 + \sigma_E^2$, $MS(G)/(MS(G) + MS(E))$ will tend to overestimate heritability. The usual estimator defined in the materials and methods section is also biased, but this bias is usually small, and (in contrast to (1)) tends to zero when the number of genotypes increases ([4], [2]).

Second, broad-sense heritability estimates in [3] were based on more accessions: 189 for LDV and 186 for LD. To allow a direct comparison with mixed model analysis we restricted our analysis to genotyped accessions, excluding 21 accessions for LDV and for 19 LD. This had little impact on heritability estimates.

Third, the analysis of variance in [3] did not include a replicate effect. In our analysis, the mean sums of squares for replicates removes some environmental variance, therefore giving higher estimates than in an analysis without a replicate effect. This however did not compensate for the use of (1); hence our heritability estimates are lower than those reported in [3].

File S3: the likelihood is constant for a kinship matrix with compound symmetry structure

Mixed model based estimation of heritability using genotypic means may become problematic when the sample size is small and the kinship matrix is close to compound symmetry, i.e. the structure where all off-diagonal elements are equal. Here we show that in the case the kinship matrix is exactly compound symmetry, the likelihood is constant in $\eta = \sigma_E^2/\sigma_A^2$. We use the notation η to avoid confusion with $\delta = \sigma_A^2/\sigma_E^2$, used in our results on genomic prediction. We write 1_n for the $n \times 1$ column vector of ones, and I_n for the n-dimensional identity matrix. Finally, let J_n be the $n \times n$ matrix of ones.

Suppose that $K = I_n + aJ_n$, for some $a > 0$. The key observation is that the covariance matrix of the data can be written as

$$\Sigma = \sigma_A^2 K + \sigma_E^2 I_n = \tilde{\sigma}_A^2 \tilde{K} + \tilde{\sigma}_E^2 I_n = \tilde{\sigma}_A^2 (\tilde{K} + \tilde{\eta} I_n),$$

where $\tilde{\sigma}_A^2 = \sigma_A^2$, $\tilde{\sigma}_E^2 = \sigma_A^2 + \sigma_E^2$, $\tilde{K} = aJ_n = a1_n 1_n^t$ and $\tilde{\eta} = (\sigma_A^2 + \sigma_E^2)/\sigma_A^2 > 0$. We can then directly apply the results in section 3 of [5], with (in their notation) $k = 1$, $d = 1$ and $X = 1_n$ (we only include an intercept, and no marker effect), and replacing σ_A^2, σ_E^2, η and K by respectively $\tilde{\sigma}_A^2$, $\tilde{\sigma}_E^2$, $\tilde{\eta}$ and $\tilde{K}$. In particular, we have the spectral decomposition

$$\tilde{K} = USU^t = [U_1, U_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} [U_1, U_2]^t$$

$$= \begin{pmatrix} n^{-\frac{1}{2}} & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ n^{-\frac{1}{2}} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} na & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix} \begin{pmatrix} n^{-\frac{1}{2}} & \dots & n^{-\frac{1}{2}} \\ 0 & \dots & 0 \\ 0 & \dots & 0 \end{pmatrix},$$

i.e. the only non-zero eigenvalue of $\tilde{K}$ is an, with eigenvector $(n^{-\frac{1}{2}}, \dots, n^{-\frac{1}{2}})$.

For this choice of X and $\tilde{K}$, the expressions for the (RE)ML estimates of β and $\tilde{\sigma}_A^2$ given in sections 3.2 and 4 of [5] greatly simplify: $\hat{\beta} = \bar{y}$ and the REML-estimate of $\tilde{\sigma}_A^2$ is $\sum_{i=1}^n (y_i - \bar{y})^2/(\tilde{\eta}(n-1))$. The extra terms

$$\frac{1}{2}\left(d\log(2\pi\tilde{\sigma}_A^2) + \log|X^t X| - \log|X^t(\tilde{K} + \tilde{\eta}I_n)^{-1}X|\right)$$

in the REML-log-likelihood (see the first equation in section 4 of [5]), now equal

$$\frac{1}{2}\left(d\log(2\pi\tilde{\sigma}_A^2) + \log n - \log\left(\frac{n}{na + \tilde{\eta}}\right)\right).$$

Combining this with their equation (3.7), it follows that the REML-log-likelihood is constant in $\tilde{\eta} = (\sigma_A^2 + \sigma_E^2)/\sigma_A^2 > 0$, and hence also constant in $\eta = \tilde{\eta} - 1 = \sigma_E^2/\sigma_A^2$.

File S4: Simulation results for a different genetic architecture.

Table 1: **Comparison of the marker-based estimators heritability estimators h_r^2 and h_m^2 for simulated data.** We simulated 5000 traits, for random samples of 200 accessions drawn from the Structured regmap and Hapmap. A single QTL was simulated, which explained 90 percent of the genetic variance. The simulated heritability was 0.2, 0.5 and 0.8. Standard errors are given relative to those of the broad sense heritability estimator (H^2).

	bias	standard error	relative standard error
Structured regmap			
$h^2 = 0.2$			
broad-sense (H^2)	-0.00127	0.04787	1.00000
replicates (h_r^2)	-0.00066	0.05102	1.06585
means (h_m^2)	0.00782	0.08626	1.80191
$h^2 = 0.5$			
broad-sense (H^2)	-0.00279	0.04500	1.00000
replicates (h_r^2)	-0.00571	0.07001	1.55569
means (h_m^2)	0.01295	0.16461	3.65791
$h^2 = 0.8$			
broad-sense (H^2)	-0.00257	0.02458	1.00000
replicates (h_r^2)	-0.01163	0.05404	2.19850
means (h_m^2)	0.00337	0.20855	8.48496
Hapmap			
$h^2 = 0.2$			
broad-sense (H^2)	-0.00110	0.04344	1.00000
replicates (h_r^2)	-0.00098	0.04320	0.99453
means (h_m^2)	0.06629	0.26168	6.02448
$h^2 = 0.5$			
broad-sense (H^2)	-0.00123	0.03437	1.00000
replicates (h_r^2)	-0.00187	0.03736	1.08695
means (h_m^2)	0.03062	0.33527	9.75477
$h^2 = 0.8$			
broad-sense (H^2)	-0.00027	0.01633	1.00000
replicates (h_r^2)	-0.00106	0.02029	1.24235
means (h_m^2)	-0.07852	0.33486	20.50621

Table 2: **Marker-based estimation of heritability: width and coverage confidence intervals obtained from the individual plant data and the genotypic means.** Results for broad sense heritability intervals are reported for comparison. We simulated 5000 traits, for random samples of 200 accessions drawn from the structured regmap (top) and Hapmap (bottom). A single QTL was simulated, which explained 90 percent of the genetic variance. The simulated heritability was 0.2, 0.5 and 0.8.

	coverage	interval width
Structured regmap		
$h^2 = 0.2$		
broad-sense	0.940	0.178
replicates (standard)	0.945	0.201
replicates (log-transformed)	0.962	0.202
means (standard)	0.911	0.315
means (log-transformed)	0.960	0.321
$h^2 = 0.5$		
broad-sense	0.926	0.160
replicates (standard)	0.837	0.194
replicates (log-transformed)	0.847	0.192
means (standard)	0.814	0.446
means (log-transformed)	0.886	0.427
$h^2 = 0.8$		
broad-sense	0.914	0.084
replicates (standard)	0.674	0.097
replicates (log-transformed)	0.666	0.097
means (standard)	0.714	0.437
means (log-transformed)	0.840	0.547
Hapmap		
$h^2 = 0.2$		
broad-sense	0.961	0.178
replicates (standard)	0.961	0.181
replicates (log-transformed)	0.972	0.182
means (standard)	0.807	0.537
means (log-transformed)	0.899	0.675
$h^2 = 0.5$		
broad-sense	0.979	0.160
replicates (standard)	0.971	0.164
replicates (log-transformed)	0.975	0.163
means (standard)	0.800	0.766
means (log-transformed)	0.967	0.819
$h^2 = 0.8$		
broad-sense	0.990	0.084
replicates (standard)	0.963	0.085
replicates (log-transformed)	0.964	0.085
means (standard)	0.820	0.840
means (log-transformed)	0.849	0.903
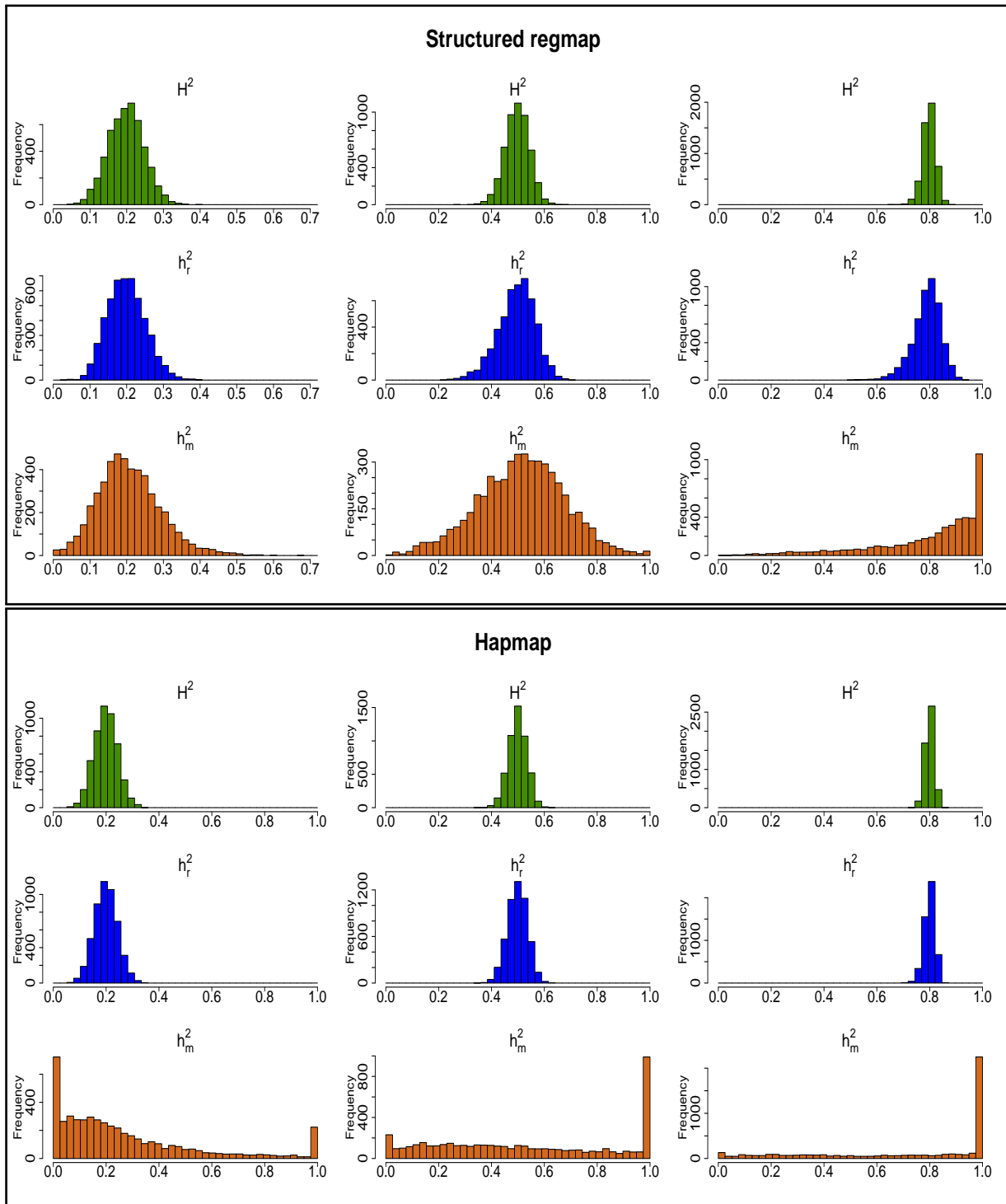
Figure 1: **Heritability estimates for 5000 simulated traits for random samples of 200 accessions drawn from the Structured regmap (top panel) and the Hapmap (bottom panel).** 1 QTL was simulated, which explained 90% of the genetic variance. The simulated heritability was 0.2 (left column), 0.5 (middle column) and 0.8 (right column). Within each panel, the first row shows the ANOVA-based estimates of broad-sense heritability, the second row the mixed model based estimates based on the individual data, and the third row the mixed model based estimates based on genotypic means.

File S5: prediction error variance in the training- and validation set.

We assume a balanced and completely random design, with n genotypes and r replicates. Given the model $y_{i,j} = \mu + G_i + E_{i,j}$, the best linear unbiased predictor (BLUP) of $G = (G_1, \ldots, G_n)^t$ and the best linear unbiased estimator (BLUE) of μ are given by

$$\hat{G} = \delta K Z^t (\delta Z K Z^t + I_N)^{-1}(y - \hat{\mu} 1_N), \quad \hat{\mu} = \frac{1_N^t (\delta Z K Z^t + I_N)^{-1} y}{1_N^t (\delta Z K Z^t + I_N)^{-1} 1_N}, \tag{2}$$

where $\delta = \sigma_A^2 / \sigma_E^2$ is the shrinkage parameter, N is the total number of individuals and Z is the $N \times n$ incidence matrix assigning individuals to genotypes. See e.g. [6] or [7], or equation (23) in the present work (Appendix C). The parameter $\delta = h^2/(1 - h^2)$ is a function of the heritability, and determines the extent to which the phenotypic data y are 'shrunk' towards zero. When the heritability is high, δ is large, and there is little shrinkage, i.e. $\hat{G}$ will be close to the observed phenotypic observations y. For low heritability, δ is small, and y will be shrunk towards the vector of zeros. When BLUPs are based on the genotypic means the same expressions hold, with $N = n$ and $Z = I_n$, and $\hat{G} = \delta_r K (\delta_r K + I_n)^{-1}((\bar{y}_1, \ldots, \bar{y}_n)^t - \hat{\mu} 1_n)$. Since the noise level is reduced from σ_E^2 to $r^{-1}\sigma_E^2$, the shrinkage parameter δ becomes $\sigma_A^2/(r^{-1}\sigma_E^2)$.

The preceding expressions assume the shrinkage parameter to be known, while it is usually estimated from the data. As a consequence, the standard error of $\hat{\mu}$ and prediction error variance of $\hat{G}$ obtained by setting $\delta = \hat{\delta} = \hat{h}^2/(1 - \hat{h}^2)$ in (2) are larger than what would be obtained when δ is known ([8], [9]). Before we give examples of too much or too little shrinkage (section), we first give expressions for the prediction error variance for the training and validation set, for the case when heritability is known ($\hat{\delta} = \delta$). These can be derived as a special case of the more general expressions in e.g. [6] or [7].

Prediction error variance when $\delta = \hat{\delta}$

First we consider the genetic effects $G = (G_1, \ldots, G_n)^t$ of the genotypes in the training sample. If we assume that $G \sim N(0, \sigma_A^2 K)$ (i.e. in equation (21) in the main text (Appendix B), γ and the QTL-effects α_m are zero), the prediction error variance is given by the diagonal elements of

$$E(\hat{G} - G)(\hat{G} - G)^t = (Z^t Z + \delta^{-1} K^{-1} - J_n)^{-1}, \tag{3}$$

where Z is the $N \times n$ incidence matrix assigning plants to genotypes, and J_n is the $n \times n$ matrix with identical elements $1/n$. In case the phenotypic data consists of genotypic means, $N = n$. For efficient computation, see [10] [11].

The genetic effects $G_{\text{pred}} = (G_{n+1}, \ldots, G_{n+m})^t$ of m unobserved (but genotyped) genotypes can be predicted with the conditional mean

$$\hat{G}_{\text{pred}} := E[G_{\text{pred}}|y] = \hat{\delta} K_{\text{pred.obs}} Z^t (\hat{\delta} Z K Z^t + I_N)^{-1}(y - \hat{\mu} 1_N), \tag{4}$$

where $K_{\text{pred.obs}}$ is the $m \times n$ matrix of kinship coefficients for the unobserved versus observed genotypes. To give expressions for the prediction error variance $E(\hat{G}_{\text{pred}} - G_{\text{pred}})_{i'}^2$ ($i' = 1, \ldots, m$) we assume again that $\gamma = 0$, all genetic signal being polygenic. Writing $K_{\text{pred.pred}}$ for the $m \times m$ kinship matrix of the unobserved genotypes, it is assumed that the kinship matrix is the $(n+m) \times (n+m)$ block matrix with K and $K_{\text{pred.pred}}$ on the diagonal and off-diagonal blocks $K_{\text{pred.obs}}$ and $K_{\text{pred.obs}}^t$. Then the conditional distribution of $G_{\text{pred}}|G$ is

$$G_{\text{pred}}|G \sim N\left(K_{\text{pred.obs}} K^{-1} G, \sigma_A^2 \left(K_{\text{pred.pred}} - K_{\text{pred.obs}} K^{-1} K_{\text{pred.obs}}^t\right)\right).$$

Since $\hat{G}_{\text{pred}} = K_{\text{pred.obs}} K^{-1} \hat{G}$ (by comparing (2) and (4)), it follows that

$$(\hat{G}_{\text{pred}} - G_{\text{pred}})|(\hat{G} - G) = K_{\text{pred.obs}} K^{-1}(\hat{G} - G) - Y,$$
$$\text{where} \quad Y \sim N\left(0, \sigma_A^2(K_{\text{pred.pred}} - K_{\text{pred.obs}} K^{-1} K_{\text{pred.obs}}^t)\right).$$

Consequently, the prediction error variances $E(\hat{G}_{\text{pred}} - G_{\text{pred}})_i^2$ are the diagonal elements of

$$\begin{aligned}
E(\hat{G}_{\text{pred}} - G_{\text{pred}})(\hat{G}_{\text{pred}} - G_{\text{pred}})^t &= E\left[E(\hat{G}_{\text{pred}} - G_{\text{pred}})(\hat{G}_{\text{pred}} - G_{\text{pred}})^t \mid (\hat{G} - G)\right] \\
&= (K_{\text{pred.obs}} K^{-1})\left[E(\hat{G} - G)(\hat{G} - G)^t\right] K^{-1} K_{\text{pred.obs}}^t \\
&\quad + \sigma_A^2(K_{\text{pred.pred}} - K_{\text{pred.obs}} K^{-1} K_{\text{pred.obs}}^t).
\end{aligned} \tag{5}$$

Hence, the prediction error variance for the validation set contains a term depending on $\delta^{-1} = \sigma_E^2/\sigma_A^2$ (see (3)), as well as a term which depends only on the genetic variance σ_A.

Prediction error variance with incorrect shrinkage ($\delta \neq \hat{\delta}$)

For the case that the amount of shrinkage is not chosen correctly ($\hat{\delta} \neq \delta = \sigma_A^2/(r^{-1}\sigma_E^2)$), we now give an expression for the prediction error variance for the training set based on genotypic means, under the additional assumption that μ is known to be zero. The BLUP for G then simplifies to

$$\hat{G} = \hat{\delta}K(\hat{\delta}K + I_n)^{-1}\bar{y}, \tag{6}$$

where we recall that we still assume a balanced and completely random design. Hence $\bar{y}_i = G_i + \bar{E}_i$, with $\bar{E}_i \sim N(0, r^{-1}\sigma_E^2)$ and $G = (G_1, \ldots, G_n)^t \sim N(0, \sigma_A^2 K)$. Since $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_n)^t \sim N(0, \sigma_A^2 K + r^{-1}\sigma_E^2 I_n) = N(0, \sigma_E^2(\delta K + r^{-1}I_n))$, the variance-covariance matrix of $\hat{G} - G$ equals

$$\mathrm{Var}(\hat{G} - G) = \sigma_A^2 K - 2\hat{\delta}K(\hat{\delta}K + I_n)^{-1}\sigma_A^2 K + \hat{\delta}K(\hat{\delta}K + I_n)^{-1}(\delta K + r^{-1}I_n)(\hat{\delta}K + I_n)^{-1}\hat{\delta}K\sigma_E^2,$$

where we used that (by the independence of G and E)

$$\mathrm{Cov}(G, \hat{G}) = \mathrm{Cov}(G, \hat{\delta}K(\hat{\delta}K + I_n)^{-1}G) = \hat{\delta}K(\hat{\delta}K + I_n)^{-1}\sigma_A^2 K$$

and that (using $\bar{y} \sim N(0, \sigma_E^2(\delta K + r^{-1}I_n))$ and the symmetry of K and I_n)

$$\hat{G} = \hat{\delta}K(\hat{\delta}K + I_n)^{-1}\bar{y} \sim N(0, \hat{\delta}K(\hat{\delta}K + I_n)^{-1}(\delta K + r^{-1}I_n)(\hat{\delta}K + I_n)^{-1}\hat{\delta}K\sigma_E^2).$$

In particular, when $\hat{\delta} = \infty$ (i.e. $\hat{h}^2 = 1$), there is no shrinkage, and $\hat{G} = \bar{y}$. The prediction error variance is then completely determined by the residual variance, since $\hat{G} - G = \bar{y} - G = \bar{E}$, and

$$E(\hat{G} - G)(\hat{G} - G)^t = r^{-1}\sigma_E^2 I_n.$$

On the other hand, when $\hat{\delta} = 0$ (i.e. $\hat{h}^2 = 0$), there is 'total' shrinkage towards zero, i.e. $\hat{G} = 0$, and

$$E(\hat{G} - G)(\hat{G} - G)^t = E(GG^t) = \sigma_A^2 K.$$

This explains the asymmetry in the observed accuracy in our simulations, in particular when $h^2 = 0.5$: when the number of replicates r is sufficiently large, overestimating the heritability will have less impact on the prediction error variance (and hence accuracy) than underestimating it.

Figure S1: histograms of the off-diagonal kinship coefficients, for 4 sub-populations of the regmap.
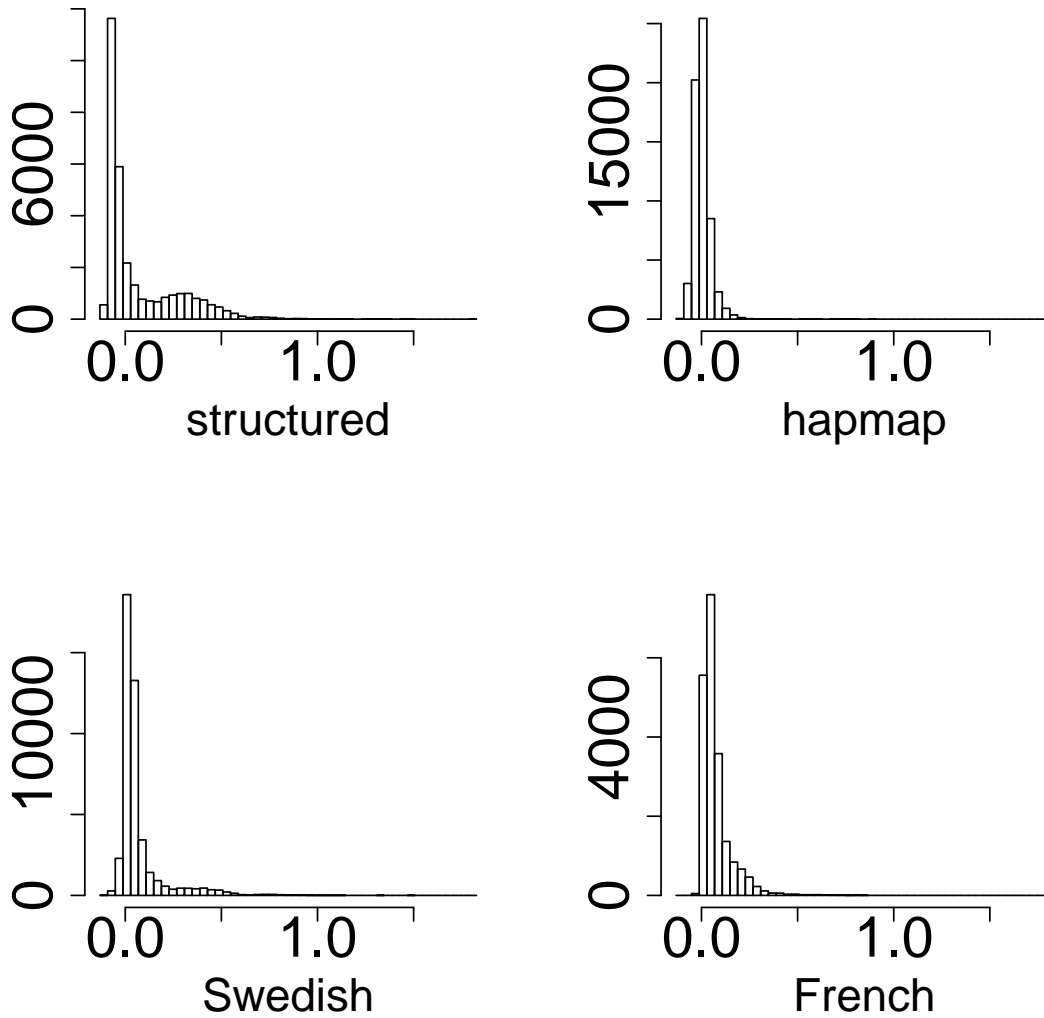
Figure S1 : **Off-diagonal coefficients of the genetic relatedness matrix (equation (1) in the main text), for 4 sub-populations of the regmap.**

Figure S2 : **Off-diagonal identity-by-state kinship coefficients, for 4 sub-populations of the regmap.**

Figure S3
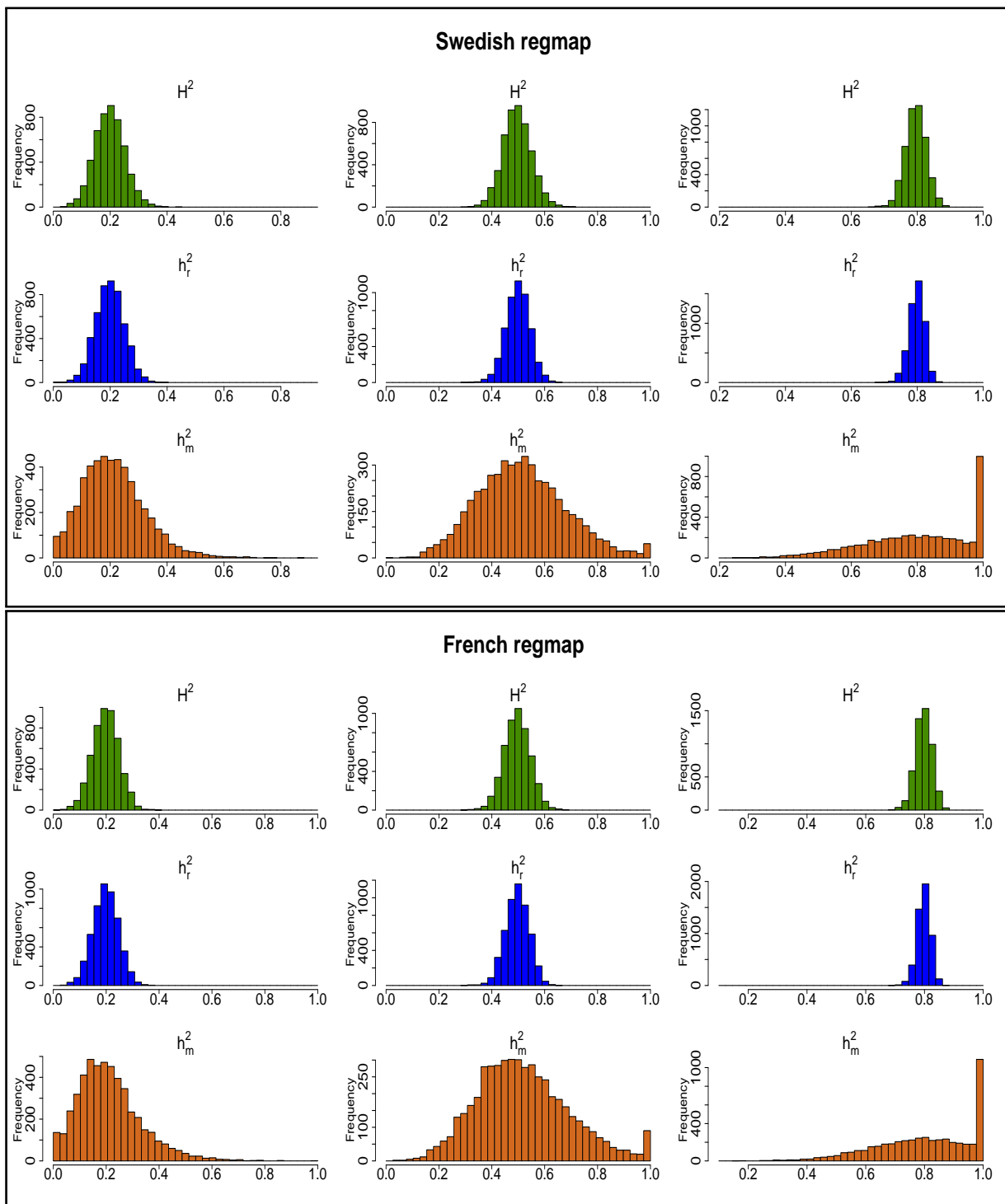
Figure S3 : **Heritability estimates for 5000 simulated traits for random samples of 200 accessions drawn from the Swedish regmap (top panel) and the French regmap (bottom panel).** 20 QTLs were simulated, which explained half of the genetic variance. The simulated heritability was 0.2 (left column), 0.5 (middle column) and 0.8 (right column). Within each panel, the first row shows the ANOVA-based estimates of broad-sense heritability, the second row the mixed model based estimates based on the individual data, and the third row the mixed model based estimates based on genotypic means.

Figure S4: Monotone likelihood
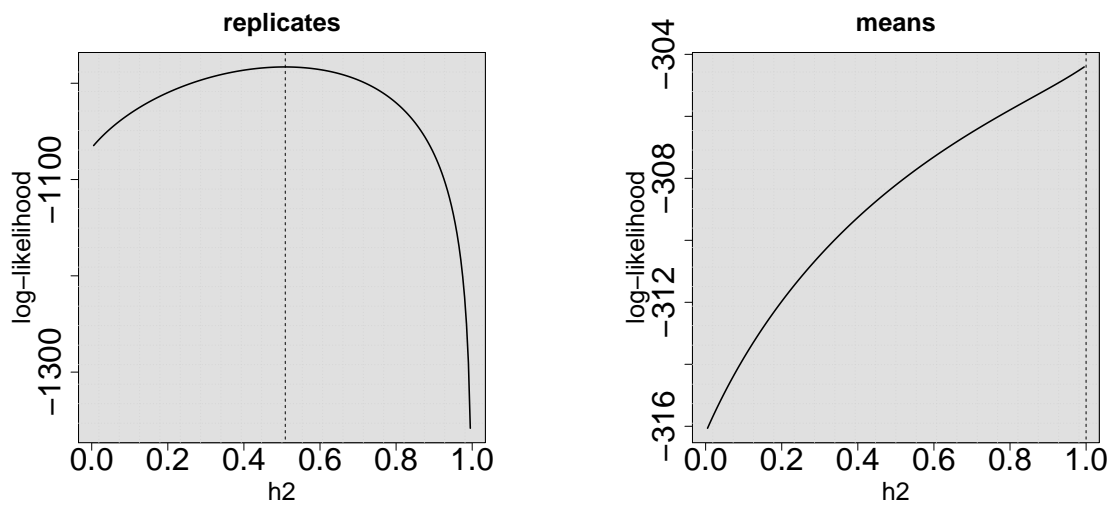
Figure S4 : **Log-likelihood as function of the heritability, for one of the 5000 simulated traits from Figure 1 (in the main text), for accessions drawn from the HapMap and a simulated heritability of 0.5** Here we choose one of the 882 traits (17.6%) for which the heritability estimate based on means ($\hat{h}_m^2$) was larger than 0.99. For these traits, the heritability estimate based on replicates ($\hat{h}_r^2$) was on average 0.502. For the trait shown here, $\hat{h}_m^2 = 0.5087$ (left) and $\hat{h}_m^2 = 0.9999$ (right).

Figure S5
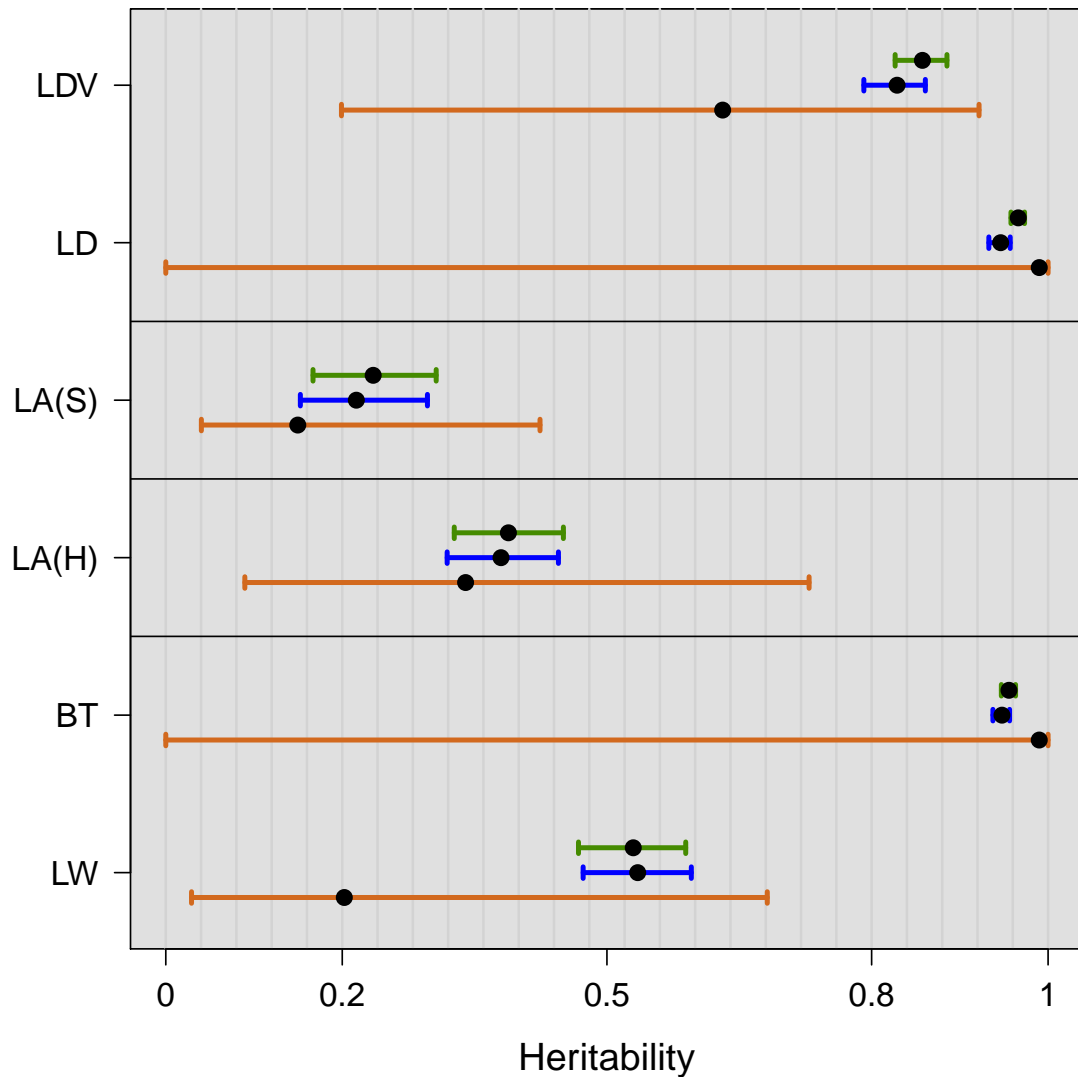
Figure S5 : **Heritability estimates and confidence intervals for two flowering traits from [3] (LDV and LD), and** 4 **traits from new experiments.** Three estimators were used: the ANOVA-based estimator of broad-sense heritability ($\hat{H}^2$, green), the marker-based estimator using individual plant data ($\hat{h}_r^2$, blue) and the marker-based estimator using genotypic means ($\hat{h}_m^2$, brown). Traits from different experiments are separated by the black horizontal lines. Trait abbreviations are given in Table 1 of the main text. The LD-adjusted kinship matrix was computed using version 2.0 of the LDAK-software [12].We used sections of 1000 SNPs, with a buffer of 200. The maximum distance considered for LD was 250kb; the 'halflife' parameter (modeling LD-decay) was set to 20kb.

Figure S6
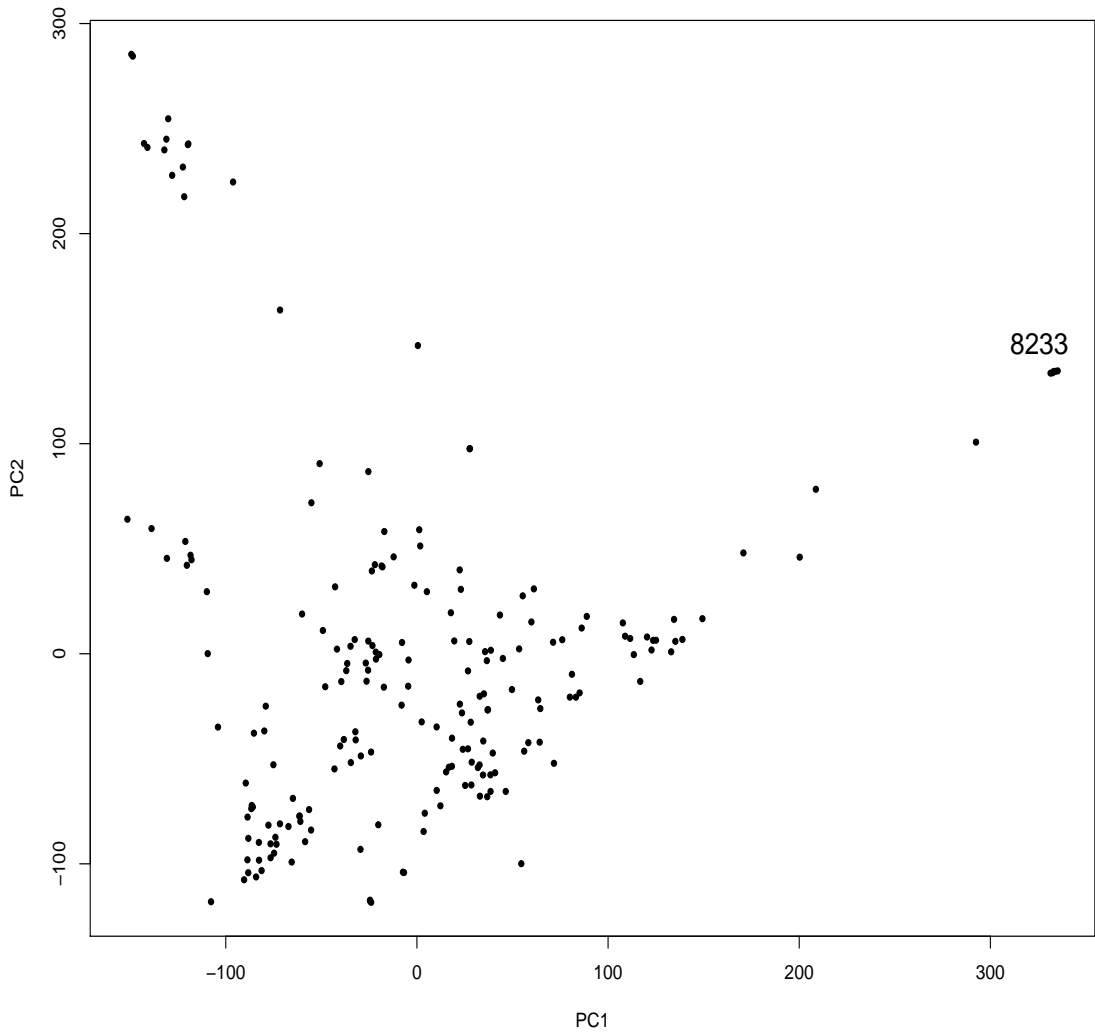
Figure S6 : **The first two principal component of the genetic markers for the panel of [3], restricted for the 168 accessions for which the trait LDV was measured.** On the very right are the accessions with ecotype ID's 8233, 7526 and 7515.

Figure S7
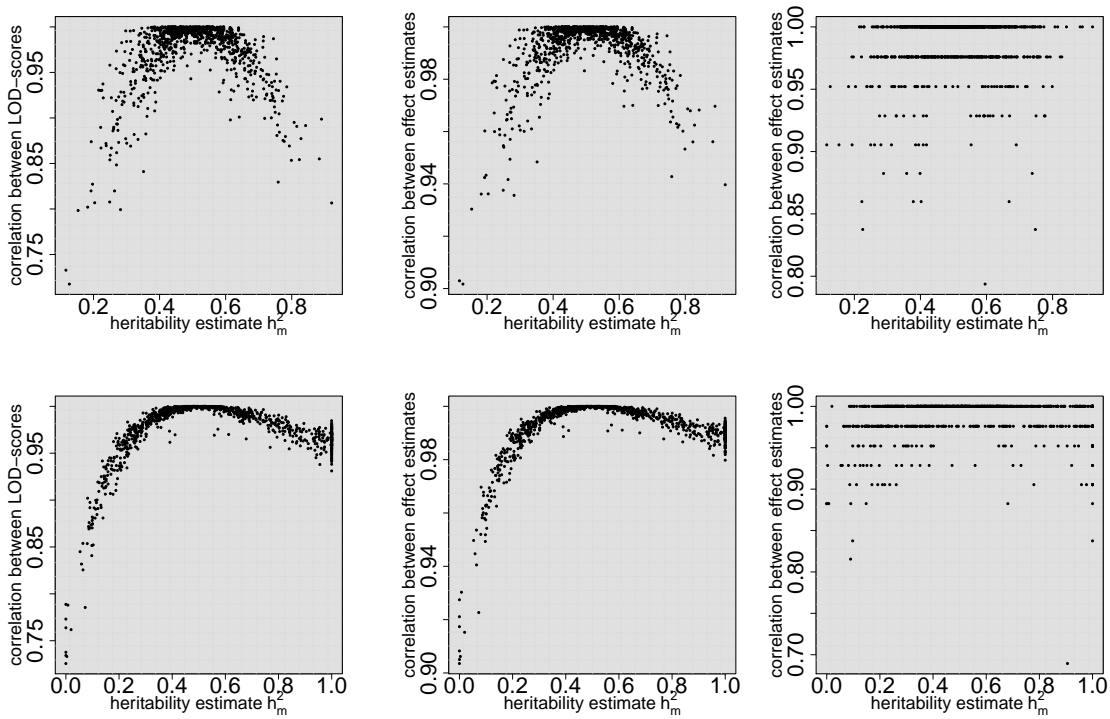
Figure S7 : **Rank correlation (Spearman ρ^2) between effect-size estimates obtained with a one- and two-stage approach, versus the heritability estimates obtained in the two-stage approach $(\hat{h}_m^2)$.** 1000 traits were simulated for the Structured RegMap (first row) and the HapMap (second row), with a simulated heritability of 0.5. 10 QTLs were simulated, which explained 75% of the genetic variance. Left column: rank correlation between LOD-scores of all SNPs. Middle column: rank correlation between effect-size estimates for all SNPs. Right column: rank correlation between effect-size estimates for the 10 simulated QTLs.

Table S1: ecotype-IDs of the structured regmap.

Accession information was taken from the Bergelson lab (`http://bergelson.uchicago.edu/Members/mhorton/resources/snps/accession_coordinates.xls`). The following table contains geographic information for 242 of the 250 accessions from our structured regmap. Geographic information was not available for eight accessions: 6909, 8428, 5712, 6143, 5708, 5730, 5829 and 8254. Accessions were selected based on the variance of the off-diagonal kinship coefficients of each row: the accessions corresponding to the rows with the 250 highest variances were chosen.

country	number of accessions
Czech Republic	7
Finland	2
France	2
Ireland	2
Sweden	83
Tajikistan	3
United Kingdom	40
United States of America	103
unknown	8

ecotype-id	country	latitude	longitude	population	native-name
1716	United States of America	42.405	-85.398	Americas	KBS-Mac-8
1718	United States of America	42.405	-85.398	Americas	KBS-Mac-15
1719	United States of America	42.405	-85.398	Americas	KBS-Mac-16
1722	United States of America	42.405	-85.398	Americas	KBS-Mac-23
1724	United States of America	42.405	-85.398	Americas	KBS-Mac-28
1726	United States of America	42.405	-85.398	Americas	KBS-Mac-33
1729	United States of America	42.405	-85.398	Americas	KBS-Mac-41
1730	United States of America	42.405	-85.398	Americas	KBS-Mac-43
1733	United States of America	42.405	-85.398	Americas	KBS-Mac-53
1736	United States of America	42.405	-85.398	Americas	KBS-Mac-58
1738	United States of America	42.405	-85.398	Americas	KBS-Mac-64
1740	United States of America	42.405	-85.398	Americas	KBS-Mac-72
1743	United States of America	42.405	-85.398	Americas	KBS-Mac-75
1744	United States of America	42.405	-85.398	Americas	KBS-Mac-76
1745	United States of America	42.405	-85.398	Americas	KBS-Mac-78
1749	United States of America	42.405	-85.398	Americas	KBS-Mac-88
1750	United States of America	42.405	-85.398	Americas	KBS-Mac-89
1751	United States of America	42.405	-85.398	Americas	KBS-Mac-91
1752	United States of America	42.405	-85.398	Americas	KBS-Mac-95
1753	United States of America	42.405	-85.398	Americas	KBS-Mac-96
1782	United States of America	42.184	-86.358	Americas	Ker-38
1829	United States of America	42.051	-86.509	Americas	Mdn-1
1850	United States of America	43.595	-86.2657	Americas	MNF-Pot-8
1853	United States of America	43.595	-86.2657	Americas	MNF-Pot-21
1858	United States of America	43.595	-86.2657	Americas	MNF-Pot-47
1864	United States of America	43.595	-86.2657	Americas	MNF-Pot-60
1871	United States of America	43.595	-86.2657	Americas	MNF-Pot-76
1872	United States of America	43.595	-86.2657	Americas	MNF-Pot-75
1873	United States of America	43.595	-86.2657	Americas	MNF-Pot-79
1874	United States of America	43.595	-86.2657	Americas	MNF-Pot-80
1938	United States of America	43.5251	-86.1843	Americas	MNF-Che-41
1941	United States of America	43.5251	-86.1843	Americas	MNF-Che-45
1948	United States of America	43.5251	-86.1843	Americas	MNF-Che-58

1960	United States of America	43.5187	-86.1739	Americas	MNF-Jac-22
1963	United States of America	43.5187	-86.1739	Americas	MNF-Jac-26
2057	United States of America	42.166	-86.412	Americas	Map-42
2148	United States of America	42.148	-86.431	Americas	Paw-1
2150	United States of America	42.148	-86.431	Americas	Paw-3
2157	United States of America	42.148	-86.431	Americas	Paw-11
2160	United States of America	42.148	-86.431	Americas	Paw-14
2180	United States of America	42.148	-86.431	Americas	Paw-40
2201	United States of America	43.7623	-86.3929	Americas	Pent-22
2204	United States of America	43.7623	-86.3929	Americas	Pent-30
2214	United States of America	43.7623	-86.3929	Americas	Pent-49
2274	United States of America	43.665	-86.496	Americas	SLSP-30
2280	United States of America	43.665	-86.496	Americas	SLSP-58
2294	United States of America	42.03	-86.514	Americas	Ste-9
2300	United States of America	42.03	-86.514	Americas	Ste-15
6927	United States of America	41.2816	-86.621	Americas	Kno-10
6983	United States of America	37.45	-119.35	Americas	Yo-0
7033	United States of America	41.3599	-122.755	Americas	Buckhorn Pass
7515	United States of America	41.5609	-86.4251	Americas	RRS-10
7523	United States of America	42.0945	-86.3253	Americas	Pna-17
7524	United States of America	42.036	-86.511	Americas	Rmx-A02
7525	United States of America	42.036	-86.511	Americas	Rmx-A180
7526	United States of America	42.0945	-86.3253	Americas	Pna-10
7566	United States of America	42.093	-86.359	Americas	627ME-13Y1
7578	United States of America	42.0945	-86.3253	Americas	627PNA-1Y1
7580	United States of America	42.0945	-86.3253	Americas	627PNA-2B3
7584	United States of America	42.0945	-86.3253	Americas	627PNA-3M4
7787	United States of America	41.273	-86.625	Americas	KNO2.77
7837	United States of America	42.093	-86.359	Americas	ME3.41
7847	United States of America	42.093	-86.359	Americas	ME3.51
7867	United States of America	42.093	-86.359	Americas	ME4.20
8122	United States of America	42.036	-86.511	Americas	RMX3.11
8233	United States of America	41.1876	-87.1923	Americas	Dem-4
8557	United States of America	42.093	-86.359	Americas	328ME032
8612	United States of America	42.093	-86.359	Americas	11ME1.34
8616	United States of America	42.093	-86.359	Americas	11ME1.41
8619	United States of America	42.093	-86.359	Americas	11ME1.44
8629	United States of America	42.093	-86.359	Americas	11ME2.10
8673	United States of America	42.0945	-86.3253	Americas	328PNA032
8724	United States of America	42.0945	-86.3253	Americas	11PNA1.15
8725	United States of America	42.0945	-86.3253	Americas	11PNA1.4
8727	United States of America	42.0945	-86.3253	Americas	11PNA1.6
8730	United States of America	42.0945	-86.3253	Americas	11PNA1.9
8760	United States of America	42.0945	-86.3253	Americas	11PNA3.19
8770	United States of America	42.0945	-86.3253	Americas	11PNA3.65
8777	United States of America	42.0945	-86.3253	Americas	11PNA3.75
8787	United States of America	42.0945	-86.3253	Americas	11PNA3.86
8796	United States of America	42.0945	-86.3253	Americas	11PNA4.101
8824	United States of America	42.0945	-86.3253	Americas	11PNA4.129
8954	United States of America	42.036	-86.511	Americas	RMX413.1
8961	United States of America	42.036	-86.511	Americas	RMX413.16
8965	United States of America	42.036	-86.511	Americas	RMX413.2
8966	United States of America	42.036	-86.511	Americas	RMX413.20
8967	United States of America	42.036	-86.511	Americas	RMX413.21
8969	United States of America	42.036	-86.511	Americas	RMX413.24
8970	United States of America	42.036	-86.511	Americas	RMX413.25
8973	United States of America	42.036	-86.511	Americas	RMX413.29

8975	United States of America	42.036	-86.511	Americas	RMX413.30
8976	United States of America	42.036	-86.511	Americas	RMX413.31
8977	United States of America	42.036	-86.511	Americas	RMX413.32
8992	United States of America	42.036	-86.511	Americas	RMX413.48
8996	United States of America	42.036	-86.511	Americas	RMX413.51
9001	United States of America	42.036	-86.511	Americas	RMX413.57
9004	United States of America	42.036	-86.511	Americas	RMX413.6
9006	United States of America	42.036	-86.511	Americas	RMX413.62
9007	United States of America	42.036	-86.511	Americas	RMX413.63
9012	United States of America	42.036	-86.511	Americas	RMX413.7
9041	United States of America	42.039	-86.5154	Americas	RMXF413.10
9045	United States of America	42.039	-86.5154	Americas	RMXF413.15
9053	United States of America	42.039	-86.5154	Americas	RMXF413.6
5991	Czech Republic	49.4112	16.2815	Austria-Hungary	DraIV 6-20
5997	Czech Republic	49.4112	16.2815	Austria-Hungary	DraIV 6-27
6427	Czech Republic	49.3853	16.2544	Austria-Hungary	ZdrI 2-1
6435	Czech Republic	49.3853	16.2544	Austria-Hungary	ZdrI 2-10
6444	Czech Republic	49.3853	16.2544	Austria-Hungary	ZdrI 2-20
6903	Czech Republic	49.4013	16.2326	Austria-Hungary	Bor-4
7461	Czech Republic	49	15	Austria-Hungary	H55
4632	United Kingdom	50.4	-4.7	British-Isles	UKSW06-025
4675	United Kingdom	50.4	-4.7	British-Isles	UKSW06-070
4862	United Kingdom	50.3	-4.9	British-Isles	UKSW06-262
5106	United Kingdom	51.3	0.5	British-Isles	UKSE06-254
5133	United Kingdom	52.2	-1.7	British-Isles	UKSE06-302
5207	United Kingdom	51.3	0.4	British-Isles	UKSE06-429
5232	United Kingdom	51.2	0.4	British-Isles	UKSE06-466
5292	United Kingdom	51.3	1.1	British-Isles	UKSE06-556
5331	United Kingdom	51.1	0.4	British-Isles	UKSE06-618
5341	United Kingdom	51.1	0.4	British-Isles	UKSE06-628
5380	United Kingdom	54.4	-3	British-Isles	UKNW06-059
5381	United Kingdom	54.4	-3	British-Isles	UKNW06-060
5385	United Kingdom	54.4	-3	British-Isles	UKNW06-078
5469	United Kingdom	54.4	-3	British-Isles	UKNW06-210
5582	United Kingdom	54.7	-3.4	British-Isles	UKNW06-410
5678	United Kingdom	54.6	-3.1	British-Isles	UKNW99-025
5709	United Kingdom	54.6	-2.6	British-Isles	UKID2
5719	Ireland	54.1335	-6.1667	British-Isles	Bur-0
5720	United Kingdom	53.3	-1.6	British-Isles	Cal-2
5723	United Kingdom	51.3	1	British-Isles	Chr-1
5724	United Kingdom	51.4	0.1	British-Isles	UKID17
5731	United Kingdom	54.9	-2.9	British-Isles	Crl-1
5732	United Kingdom	54.9	-2.9	British-Isles	UKID25
5737	United Kingdom	51.3	0.1	British-Isles	UKID32
5758	United Kingdom	50.3	-5.2	British-Isles	UKID53
5774	United Kingdom	51.1	0.6	British-Isles	Sis-1
5780	United Kingdom	53.1	-1	British-Isles	UKID75
5788	United Kingdom	50.8	-2	British-Isles	UKID83
5792	United Kingdom	50.8	-0.7	British-Isles	UKID87
5793	United Kingdom	51.3	0.6	British-Isles	UKID88
5798	United Kingdom	53.1	-3.3	British-Isles	UKID93
5804	United Kingdom	50.8	-1.1	British-Isles	UKID100
5807	United Kingdom	51.8	-0.5	British-Isles	UKID103
6905	Ireland	54.1	-6.2	British-Isles	Bur-0
6923	United Kingdom	51.4083	-0.6383	British-Isles	HR-10
6924	United Kingdom	51.4083	-0.6383	British-Isles	HR-5
6944	United Kingdom	51.4083	-0.6383	British-Isles	NFA-8

7064	United Kingdom	51.3	1.1	British-Isles	Cnt-1
7109	United Kingdom	51.3	0.5	British-Isles	Ema-1
7483	United Kingdom	51.2878	0.0565	British-Isles	PHW-14
9490	United Kingdom	55.9218	-3.17108	British-Isles	02B6
9504	United Kingdom	55.8877	-3.16377	British-Isles	12A1
6929	Tajikistan	38.48	68.49	Eastern-Range	Kondara
6962	Tajikistan	38.35	68.48	Eastern-Range	Shahdara
7168	Tajikistan	38.48	68.49	Eastern-Range	Hodja-Obi-Garm
1247	Sweden	59.4333	17.0167	Fennoscandia	Tos-31-374
1254	Sweden	59.4333	17.0167	Fennoscandia	Tos-82-387
1409	Sweden	62.8	18.2	Fennoscandia	Rd-38
1416	Sweden	62.8	18.2	Fennoscandia	Rd-45
1435	Sweden	62.8	18.2	Fennoscandia	Rd-17-319
1552	Sweden	63.0833	18.3667	Fennoscandia	Sku-30
5835	Sweden	63.324	18.484	Fennoscandia	Bil-3
5856	Sweden	63.0167	17.4914	Fennoscandia	Dr-10
5860	Sweden	62.6814	18.0165	Fennoscandia	Dra-3
6009	Sweden	62.877	18.177	Fennoscandia	Eden-1
6010	Sweden	62.877	18.177	Fennoscandia	Eden-5
6011	Sweden	62.877	18.177	Fennoscandia	Eden-6
6012	Sweden	62.877	18.177	Fennoscandia	Eden-7
6013	Sweden	62.877	18.177	Fennoscandia	Eden-9
6016	Sweden	62.9	18.4	Fennoscandia	Eds-1
6017	Sweden	62.9	18.4	Fennoscandia	Eds-9
6024	Sweden	55.7509	13.3712	Fennoscandia	Fly2-2
6025	Sweden	62.6437	17.7339	Fennoscandia	Gro-3
6030	Sweden	62.806	18.1896	Fennoscandia	Grn-5
6043	Sweden	62.801	18.079	Fennoscandia	Lv-1
6046	Sweden	62.801	18.079	Fennoscandia	Lv-5
6064	Sweden	62.9513	18.2763	Fennoscandia	Nyl-2
6069	Sweden	62.9513	18.2763	Fennoscandia	Nyl-7
6071	Sweden	62.9308	18.3448	Fennoscandia	Omn-5
6077	Sweden	55.6942	13.4504	Fennoscandia	Rev-3
6091	Sweden	55.6525	13.215	Fennoscandia	T1010
6104	Sweden	55.7	13.2	Fennoscandia	T1160
6118	Sweden	55.7	13.2	Fennoscandia	T610
6154	Sweden	62.6422	17.7406	Fennoscandia	TAA 04
6163	Sweden	62.6425	17.7356	Fennoscandia	TAA 14
6166	Sweden	62.6425	17.7372	Fennoscandia	TAA 17
6169	Sweden	62.8714	18.3447	Fennoscandia	TD 01
6170	Sweden	62.8717	18.3442	Fennoscandia	TD 02
6171	Sweden	62.8717	18.3444	Fennoscandia	TD 03
6172	Sweden	62.8717	18.3436	Fennoscandia	TD 04
6173	Sweden	62.8717	18.3419	Fennoscandia	TD 05
6174	Sweden	62.8719	18.3422	Fennoscandia	TD 06
6177	Sweden	62.6322	17.69	Fennoscandia	TL 03
6180	Sweden	62.6322	17.6906	Fennoscandia	TL 07
6184	Sweden	62.8892	18.4522	Fennoscandia	TB 01
6209	Sweden	62.8836	18.1842	Fennoscandia	TEDEN 02
6210	Sweden	62.8839	18.1836	Fennoscandia	TEDEN 03
6212	Sweden	63.0175	18.3239	Fennoscandia	TF 02
6214	Sweden	63.0175	18.3281	Fennoscandia	TF 04
6215	Sweden	63.0172	18.3281	Fennoscandia	TF 05
6216	Sweden	63.0167	18.3283	Fennoscandia	TF 06
6217	Sweden	63.0169	18.3283	Fennoscandia	TF 07
6218	Sweden	63.0172	18.3283	Fennoscandia	TF 08
6220	Sweden	62.806	18.1896	Fennoscandia	TGR 01

6221	Sweden	62.806	18.1896	Fennoscandia	TGR 02
6226	Sweden	62.7994	17.9033	Fennoscandia	TH 08
6231	Sweden	62.96	18.2844	Fennoscandia	TNY 04
6235	Sweden	62.9611	18.3589	Fennoscandia	TOM 01
6236	Sweden	62.9617	18.36	Fennoscandia	TOM 02
6237	Sweden	62.9619	18.35	Fennoscandia	TOM 03
6238	Sweden	62.9619	18.35	Fennoscandia	TOM 04
6240	Sweden	62.9622	18.35	Fennoscandia	TOM 06
6241	Sweden	62.9614	18.3608	Fennoscandia	TOM 07
6244	Sweden	62.9169	18.4728	Fennoscandia	TR 01
6900	Sweden	63.324	18.484	Fennoscandia	Bil-5
6901	Sweden	63.324	18.484	Fennoscandia	Bil-7
6913	Sweden	62.877	18.177	Fennoscandia	Eden-2
6917	Sweden	63.0165	18.3174	Fennoscandia	Fb-2
6918	Sweden	63.0165	18.3174	Fennoscandia	Fb-4
6968	Finland	60	23.5	Fennoscandia	Tamm-2
6969	Finland	60	23.5	Fennoscandia	Tamm-27
8218	Sweden	62.877	18.177	Fennoscandia	Eden-4
8227	Sweden	62.7989	17.9103	Fennoscandia	TH 03
8335	Sweden	55.71	13.2	Fennoscandia	Lund
8376	Sweden	62.69	18	Fennoscandia	Sanna-2
9321	Sweden	62.8622	18.336	Fennoscandia	dal 1
9323	Sweden	62.8622	18.336	Fennoscandia	dal 3
9332	Sweden	62.8698	18.381	Fennoscandia	Bar 1
9354	Sweden	62.8762	18.1746	Fennoscandia	Eden 15
9355	Sweden	62.8762	18.1746	Fennoscandia	Eden 16
9356	Sweden	62.8762	18.1746	Fennoscandia	Eden 17
9363	Sweden	62.9147	18.4045	Fennoscandia	EdJ 2
9371	Sweden	63.016	18.3175	Fennoscandia	FL 1
9378	Sweden	63.0165	18.3174	Fennoscandia	FU 4
9386	Sweden	62.806	18.1896	Fennoscandia	Grn 12
9388	Sweden	62.806	18.1896	Fennoscandia	Grn 14
9427	Sweden	62.8815	18.4055	Fennoscandia	Ns 2
9433	Sweden	62.9513	18.2763	Fennoscandia	Nyl 13
9434	Sweden	62.8959	18.3659	Fennoscandia	de 2
9471	Sweden	56.0648	13.9707	Fennoscandia	UllA 1
171	France	47.3833	5.31667	France	MIB-20
228	France	47.3833	5.31667	France	MIB-9

Table S2

Table S2 : **Comparison of the marker-based estimators heritability estimators** h_r^2 **and** h_m^2 **for simulated data.** We simulated 5000 traits, for random samples of 200 accessions drawn from Swedish and French regmap. 20 unlinked QTLs were simulated, which explained 50 percent of the genetic variance. The simulated heritability was 0.2, 0.5 and 0.8. Standard errors are given relative to those of the broad sense heritability estimator (H^2).

	bias	standard error	relative standard error
Swedish regmap			
$h^2 = 0.2$			
broad-sense (H^2)	-0.00162	0.05227	1.00000
replicates (h_r^2)	-0.00109	0.04991	0.95498
means (h_m^2)	0.01302	0.11018	2.10798
$h^2 = 0.5$			
broad-sense (H^2)	-0.00403	0.05373	1.00000
replicates (h_r^2)	-0.00173	0.04506	0.83860
means (h_m^2)	0.01494	0.16662	3.10123
$h^2 = 0.8$			
broad-sense (H^2)	-0.00458	0.03130	1.00000
replicates (h_r^2)	-0.00180	0.02319	0.74095
means (h_m^2)	-0.00104	0.16227	5.18435
French regmap			
$h^2 = 0.2$			
broad-sense (H^2)	-0.00183	0.04958	1.00000
replicates (h_r^2)	-0.00196	0.04780	0.96421
means (h_m^2)	0.01306	0.12049	2.43043
$h^2 = 0.5$			
broad-sense (H^2)	-0.00396	0.04930	1.00000
replicates (h_r^2)	-0.00396	0.04409	0.89431
means (h_m^2)	0.01952	0.17547	3.55941
$h^2 = 0.8$			
broad-sense (H^2)	-0.00341	0.02808	1.00000
replicates (h_r^2)	-0.00236	0.02246	0.79988
means (h_m^2)	0.00202	0.16461	5.86175

Table S3

Table S3: : **Marker-based estimation of heritability: width and coverage confidence intervals obtained from the individual plant data and the genotypic means.** Results for broad sense heritability intervals are reported for comparison. We simulated 5000 traits, for random samples of 200 accessions drawn from the Swedish regmap (top) and French (bottom). 20 unlinked QTLs were simulated, which explained 50 percent of the genetic variance. The simulated heritability was 0.2, 0.5 and 0.8.

	coverage	interval width
Swedish regmap		
$h^2 = 0.2$		
broad-sense	0.912	0.178
replicates (standard)	0.933	0.188
replicates (log-transformed)	0.961	0.189
means (standard)	0.899	0.381
means (log-transformed)	0.968	0.405
$h^2 = 0.5$		
broad-sense	0.864	0.160
replicates (standard)	0.946	0.176
replicates (log-transformed)	0.950	0.175
means (standard)	0.921	0.594
means (log-transformed)	0.970	0.560
$h^2 = 0.8$		
broad-sense	0.823	0.085
replicates (standard)	0.945	0.088
replicates (log-transformed)	0.947	0.088
means (standard)	0.960	0.635
means (log-transformed)	0.938	0.748
French regmap		
$h^2 = 0.2$		
broad-sense	0.929	0.178
replicates (standard)	0.941	0.184
replicates (log-transformed)	0.962	0.185
means (standard)	0.898	0.396
means (log-transformed)	0.960	0.431
$h^2 = 0.5$		
broad-sense	0.898	0.160
replicates (standard)	0.953	0.173
replicates (log-transformed)	0.956	0.171
means (standard)	0.927	0.619
means (log-transformed)	0.976	0.585
$h^2 = 0.8$		
broad-sense	0.866	0.084
replicates (standard)	0.947	0.088
replicates (log-transformed)	0.947	0.088
means (standard)	0.966	0.652
means (log-transformed)	0.948	0.767

Table S4

Table S4: : **Heritability estimates and confidence intervals, for two flowering traits from [3] and four traits measured in new experiments (trait abbreviations given in Table 1 of the main text).**

trait	replicates	means	broad-sense
LDV	0.829 (0.791,0.861)	0.631 (0.199,0.922)	0.858 (0.827,0.885)
LD	0.946 (0.933,0.957)	1.000 (0.000,1.000)	0.966 (0.958,0.973)
LA(S)	0.216 (0.153,0.297)	0.150 (0.040,0.424)	0.235 (0.167,0.306)
LA(H)	0.380 (0.319,0.445)	0.340 (0.090,0.729)	0.388 (0.327,0.451)
BT	0.948 (0.937,0.956)	1.000 (0.000,1.000)	0.956 (0.947,0.963)
LW	0.535 (0.473,0.596)	0.202 (0.029,0.682)	0.530 (0.468,0.589)

Three estimators were used: mixed model based on replicates ($\hat{h}_r^2$), mixed model based on genotypic means ($\hat{h}_m^2$), and the usual ANOVA-based broad-sense heritability estimator ($\hat{H}^2$). An LD-adjusted kinship matrix was used in the mixed model for $\hat{h}_r^2$ and $\hat{h}_m^2$.

The LD-adjusted kinship matrix was computed using version 2.0 of the LDAK-software [12], available at `http://dougspeed.com/ldak/`. We used sections of 1000 SNPs, with a buffer of 200. The maximum distance considered for LD was 250kb; the 'halflife' parameter (modeling LD-decay) was set to 20kb.

Table S5

In the following tables, the second and third column contain the percentage of the 5000 traits for which the corresponding heritability estimates ($\hat{h}_r^2$ and $\hat{h}_m^2$) were contained in the intervals in the first column. The remaining columns show the correlation (r) between simulated and predicted genetic effects, averaged over these traits. 20 QTLs were simulated, which explained 50 percent of the genetic variance. Each trait was simulated for a randomly drawn training (200 accessions) and validation set (50 accessions). Genetic effects were predicted using G-BLUP, based on either a mixed model for the individual plants (replicates) or for the genotypic means.

Table S5(a) : **Prediction accuracy (r) of G-BLUP for** 5000 **simulated traits, for the structured regmap population, and a simulated heritability of** 0.2.

interval	$\hat{h}_r^2$	$\hat{h}_m^2$	r (replicates) Training set	r (means) Training set	r (replicates) Validation set	r (means) Validation set
$[0, 0.1)$	3.08 %	9.88 %	0.637	0.654	0.216	0.218
$[0.1, 0.3)$	93.96 %	76.38 %	0.770	0.770	0.280	0.279
$[0.3, 0.5)$	2.96 %	13.52 %	0.816	0.803	0.325	0.313
$[0.5, 0.7)$	0 %	0.22 %		0.782		0.287
$[0.7, 0.9)$	0 %	0 %				
$[0.9, 1]$	0 %	0 %				
$[0, 1]$	100 %	100 %	0.767	0.763	0.279	0.278

Table S5(b) : **Prediction accuracy (r) of G-BLUP for** 5000 **simulated traits, for the structured regmap population and a simulated heritability of** 0.5.

interval	$\hat{h}_r^2$	$\hat{h}_m^2$	r (replicates) Training set	r (means) Training set	r (replicates) Validation set	r (means) Validation set
$[0, 0.1)$	0 %	0.04 %		0.709		0.328
$[0.1, 0.3)$	0 %	5.24 %		0.836		0.269
$[0.3, 0.5)$	51.42 %	46.54 %	0.886	0.887	0.302	0.300
$[0.5, 0.7)$	48.58 %	42.12 %	0.905	0.903	0.333	0.337
$[0.7, 0.9)$	0 %	5.84 %		0.905		0.343
$[0.9, 1]$	0 %	0.22 %		0.888		0.386
$[0, 1]$	100 %	100 %	0.895	0.892	0.317	0.317

Table S5(c) : **Prediction accuracy (r) of G-BLUP for** 5000 **simulated traits, for the structured regmap population and a simulated heritability of** 0.8.

interval	$\hat{h}_r^2$	$\hat{h}_m^2$	r (replicates) Training set	r (means) Training set	r (replicates) Validation set	r (means) Validation set
$[0, 0.1)$	0 %	0 %				
$[0.1, 0.3)$	0 %	0.02 %		0.877		0.299
$[0.3, 0.5)$	0 %	1.42 %		0.930		0.283
$[0.5, 0.7)$	0.04 %	19.26 %	0.953	0.955	0.400	0.318
$[0.7, 0.9)$	99.96 %	59.26 %	0.964	0.964	0.343	0.344
$[0.9, 1]$	0 %	20.04 %		0.965		0.365
$[0, 1]$	100 %	100 %	0.964	0.962	0.343	0.343

Table S5(d) : **Prediction accuracy (r) of G-BLUP for 5000 simulated traits, for the HapMap population and a simulated heritability of** 0.2**.**

interval	$\hat{h}_r^2$	$\hat{h}_m^2$	r (replicates) Training set	r (means) Training set	r (replicates) Validation set	r (means) Validation set
$[0, 0.1)$	1.74 %	28.64 %	0.616	0.632	0.259	0.273
$[0.1, 0.3)$	96.7 %	40.7 %	0.673	0.674	0.341	0.348
$[0.3, 0.5)$	1.56 %	17.8 %	0.711	0.684	0.364	0.382
$[0.5, 0.7)$	0 %	5.8 %		0.681		0.366
$[0.7, 0.9)$	0 %	2.84 %		0.675		0.370
$[0.9, 1]$	0 %	4.22 %		0.669		0.357
$[0, 1]$	100 %	100 %	0.672	0.664	0.340	0.335

Table S5(e) : **Prediction accuracy (r) of G-BLUP for 5000 simulated traits, for the HapMap population and a simulated heritability of** 0.5**.**

interval	$\hat{h}_r^2$	$\hat{h}_m^2$	r (replicates) Training set	r (means) Training set	r (replicates) Validation set	r (means) Validation set
$[0, 0.1)$	0 %	6 %		0.811		0.285
$[0.1, 0.3)$	0 %	21.02 %		0.851		0.366
$[0.3, 0.5)$	51.78 %	22.56 %	0.862	0.867	0.395	0.413
$[0.5, 0.7)$	48.22 %	17.5 %	0.877	0.871	0.416	0.428
$[0.7, 0.9)$	0 %	10.86 %		0.873		0.426
$[0.9, 1]$	0 %	22.06 %		0.871		0.422
$[0, 1]$	100 %	100 %	0.869	0.863	0.405	0.401

Table S5(f) (given as Table 6 in the main text) : **Prediction accuracy (r) of G-BLUP for 5000 simulated traits, for the HapMap population and a simulated heritability of** 0.8**.**

interval	$\hat{h}_r^2$	$\hat{h}_m^2$	r (replicates) Training set	r (means) Training set	r (replicates) Validation set	r (means) Validation set
$[0, 0.1)$	0 %	2.58 %		0.890		0.289
$[0.1, 0.3)$	0 %	8.34 %		0.937		0.373
$[0.3, 0.5)$	0 %	12.34 %		0.954		0.409
$[0.5, 0.7)$	0.04 %	15.9 %	0.942	0.959	0.208	0.423
$[0.7, 0.9)$	99.96 %	15.62 %	0.961	0.961	0.431	0.443
$[0.9, 1]$	0 %	45.22 %		0.961		0.448
$[0, 1]$	100 %	100 %	0.961	0.956	0.431	0.428

References

[1] Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, 1 edition. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0878934812`.

[2] Singh M, Ceccarelli S, Hamblin J (1993) Estimation of heritability from varietal trials data. Theoretical and Applied Genetics 86: 437-441.

[3] Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631.

[4] Nyquist WE, Baker R (1991) Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences 10: 235-322.

[5] Lippert C, Listgarten J, Liu Y, Kadie C, Davidson R, et al. (2011) FaST linear mixed models for genome-wide association studies: supplement. Nature Genetics 42: 833835.

[6] Henderson CR (1975) Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics 31.

[7] Robinson GK (1991) That blup is a good thing: The estimation of random effects. Statistical Science 6: pp. 15-32.

[8] Kackar RN, Harville DA (1984) Approximations for standard errors of estimators of fixed and random effect in mixed linear models. Journal of the American Statistical Association 79: pp. 853-862.

[9] Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53: pp. 983-997.

[10] Gilmour A, Cullis B, Welham S, Gogel B, Thompson R (2004) An efficient computing strategy for prediction in mixed linear models. Computational Statistics & Data Analysis 44: 571 - 586.

[11] Welham S, Cullis B, Gogel B, Gilmour A, Thompson R (2004) Prediction in linear mixed models. Australian &#38; New Zealand Journal of Statistics : 325–347.

[12] Speed D, Hemani G, Johnson MR, Balding DJ (2012) Improved Heritability Estimation from Genome-wide SNPs. The American Journal of Human Genetics 91: 1011–1021.