Fast and general-purpose linear mixed models for genome-wide genetics

Daniel E Runcie[1,*] **and Lorin Crawford**[2]

[1]Department of Plant Sciences, University of California Davis, Davis, CA, USA
[2]Department of Biostatistics, Brown University, Providence, RI, USA
[*]deruncie@ucdavis.edu

ABSTRACT

We describe Grid-LMM, a general algorithm for fitting linear mixed effect models across a wide range of applications in quantitative genetics, including genome-wide association mapping and heritability estimation. Grid-LMM provides approximate (yet highly accurate) frequentist test statistics or Bayesian posterior summaries in a fraction of the time compared to existing general-purpose methods. Most importantly, Grid-LMM is suitable for genome-wide analyses that account for multiple sources of heterogeneity, such as additive and non-additive genetic variance, spatial heterogeneity, and genotype-environment interactions. We demonstrate that Grid-LMM can increase power in common settings using simulation studies and two real data applications. Grid-LMM is available in an R-package.

Introduction

Population stratification, genetic relatedness, ascertainment, and other sources of heterogeneity lead to spurious signals and reduced power in genetic association studies (GWAS) [e.g. 1–5]. When not properly taken into account, non-additive genetic effects and environmental variation can also bias estimates of heritability, polygenic adaptation, and genetic values in breeding programs [e.g. 5–8]. Both issues are caused by departures from a key assumption underlying linear models: that observations are independent. Non-independent samples lead to a form of pseudo-replication, effectively reducing the true sample size. Linear mixed effect models (LMMs) are widely used to account for non-independent samples in quantitative genetics [9]. The flexibility and interpretability of LMMs make them a dominant statistical tool in much of biological research [e.g. 9–18].

Random effect terms are used in LMMs to account for specific correlations among observations. Fitting an LMM requires estimating the importance of each random effect, called its variance component. General-purpose tools for this are too slow to be practical for genome-scale datasets with thousands of observations and millions of genetic markers [e.g. 19]. This lack of scalability is caused primarily by two factors: (i) closed-form solutions of maximum-likelihood (ML or REML) or posterior estimates of the variance components aren't available and numerical optimization routines require repeatedly evaluating the likelihood function many times, and (ii) each evaluation of likelihood requires inverting the covariance matrix of random effects, an operation that scales cubically with the number of observations. Repeating this whole process millions of times quickly becomes infeasible.

To this end, several specialized approaches have been developed to improve the speed of LMMs, including the use of sparse matrix operations [e.g. 20, 21] and spectral decomposition of the random effect covariance matrix [e.g. 22–26]. These algorithms are particularly useful when the same random effect structure is used many times. For example, in GWAS, each marker is tested with the same LMM. Fast and exact algorithms for fitting LMMs are limited to the case of only a single (full-rank) random effect term, besides the residuals [22–24]. Recently, approximate learning algorithms have been developed for the scalable extension to multiple random effects [e.g. 27–29], but few of these ensure guarantees in terms of estimation accuracy. One common strategy that is applicable to studies with multiple random effects is to estimate the variance components only once in a model without additional marker effects, and then perform marker tests conditional on these estimates [e.g. 30–33]. Given "known" variance components, closed-form solutions of all other parameters of a LMM can be found using a rotated version of the simple linear model. However, this approximation suffers from reduced power when marker effects are large, and doesn't permit posterior inference in a Bayesian analysis. Table 1 summarizes these different methods, and their respective computational complexities, together with relevant references.

Grid-LMM takes a different approach to fitting LMMs: rather than directly optimizing the variance components separately for each test, we define a grid spanning all valid values of the variance components and fit simple linear models at each grid location. Each evaluation involves a single Cholesky decomposition of the random effect covariance matrix, which is then reused to calculate closed-form ML solutions or Bayesian posterior summaries (under a specific conjugate prior, see

Table 1. Performance and limitations of reference models for linear mixed-model in GWAS. The time complexity of each algorithm is approximate, assuming a model with only a single marker effect and no other fixed effects. Here, l is used to index full-rank random effects; σ_l^2 are variance component parameters; n is the number of observations; p, is number of markers to test; and t is the number of iterations needed for convergence. This value is expected to vary among methods, particularly for the iterations of grid search of Grid-LMM-fast. The limiting term in common GWAS applications (where $p \gg n$) are in bold. "Method Type" describes the estimation of σ_l^2. "exact" means effectively exact, up to machine precision and subject to possible convergence of algorithms to local maxima. "null" means estimation of parameters under the null model with no marker effects. References list additional methods that are approximately equivalent to the given model classes.

Model	Max σ_l^2	Method Type	Approx. Time Complexity	References
GEMMA	1	exact	$\mathcal{O}(n^3 + \boldsymbol{pn^2} + ptn)$	[23, 24]
EMMAX	1	null	$\mathcal{O}(n^3 + tn + \boldsymbol{pn^2})$	[30, 31]
sommer	>1	exact	$\mathcal{O}(\boldsymbol{ptn^3})$	[21, 34–36]
pylmm	>1	null	$\mathcal{O}(tn^3 + \boldsymbol{pn^2})$	[33, 37]
Grid-LMM	$\sim$ 1-8	grid	$\mathcal{O}(\boldsymbol{g}(n^3 + \boldsymbol{pn^2}))$	Present Work
Grid-LMM-fast	$\sim$ 1-8	grid-sampling	$\mathcal{O}(tn^3 + \boldsymbol{pn^2} + \sum_{i=2}^{t_2} \boldsymbol{g_i}(n^3 + \boldsymbol{p_i n^2}))$	Present Work

Supplemental Methods) for all separate tests. After repeating these calculations across the whole grid, we select the highest ML (or REML) score for each marker to compute approximate likelihood ratio or Wald tests [31, 38], or analogously derive posterior distributions or Bayes factors by summing appropriate statistics across the grid. The Grid-LMM approach relies on a re-parameterization of the typical LMM from individual variance components σ_l^2 to variance component proportions $h_l^2 = \sigma_l^2/\sigma^2$, where σ^2 without the subscript denotes the total sum of all variance components (including the residual). Since the variance components must be non-negative, their proportions are restricted to the unit interval $[0,1]$ and sum to 1, forming a simplex. Therefore, a finite grid can span all valid parameter values. While the size of the grid increases rapidly with the number of random effects, for a small number of random effects ($\sim$1-8) and a moderate grid resolution, the size of the grid remains tiny relative to the number of models in a typical GWAS. As we show below, highly accurate test statistics are achieved even with a coarse grid in most reasonable situations, and further improvements to efficiency are possible by using heuristics to adaptively sample the grid or reduce the number of grid locations computed for the majority of tests. This strategy of conditioning on variance components over a grid and then combining solutions can be applied to many other tools in quantitative genetics including set tests for rare variants [39, 40], whole-genome regression models [e.g. LASSO and elastic net; 41, 42], and QTL mapping in controlled crosses [43].

Through a combination of simulation studies and analyses of well-studied datasets, we show that the approximate test-statistics and posterior quantities calculated by Grid-LMM are accurate and improve power relative to other approximation methods – while maintaining dramatically reducing computational burdens. Full derivations of the model, extensions to LASSO and QTL mapping applications, and useful heuristics are described in detail in the Methods.

Results

We used Grid-LMM to perform two types of genome-wide association studies (GWAS) that benefit from modeling multiple random effects: (1) the study of gene-environment interactions, and (2) detecting associations for phenotypes driven by non-additive genetic variation. In both cases, there are multiple sources of covariation among observations that can inflate test statistics and bias estimates of heritability if not appropriately accounted for using mixed models.

As an example of a GWAS with gene-environment interactions, we reanalyzed data on flowering times for *Arabidopsis thaliana* [44]. First, we benchmarked results from standard LMM methodologies to confirm that Grid-LMM association tests are accurate. We ran Grid-LMM either with a fine-grained grid with a step size of 0.01 h^2-units, or a larger step size of 0.1 h^2-units to test for associations between 216,130 single nucleotide polymorphisms (SNPs) and flowering times of 194 accessions measured at 10C (i.e. in a constant environment). We compared the Wald-test p-values computed by both Grid-LMM models to p-values computed using the exact method GEMMA [24], and the approximate method EMMAX [31]. Each method was applied as an LMM with only the additive relationship matrix as its single random effect. Grid-LMM p-values computed using the finer grid size (i.e. 0.01 h^2-units) were almost identical to those of GEMMA, and even p-values computed with the larger step size (i.e. 0.1 h^2-units) were more accurate than those resulting from EMMAX. There were particularly noticeable differences in performance for markers with larger scores, which were strongly underestimated by EMMAX since its approximations of h^2 are made strictly under the null model (Figure 1a). This pattern held true across all of the available 107 *Arabidopsis thaliana* phenotypes (see Figure S1). However, despite these results, we do not advocate Grid-LMM in this setting; GEMMA (and similar methods) provides exact tests and is more computationally efficient. The real advantage of Grid-LMM is

its ability to effectively model two (or more) random effects.

To demonstrate this advantage, we tested for gene-environment (G×E) interaction effects on flowering time. Here, we combined flowering time data from two conditions: constant 10C (as described above) and in the field. We limited our analysis to the 175 accessions that were grown under both conditions, yielding a total of $n = 350$ observations. When observations come from different environments, we might expect phenotypes to cluster for at least two reasons: (i) the sharing of alleles with constant effects across environments due to relatedness and population structure, commonly modeled with the additive genomic relationship matrix (**A**) as a random effect, and (ii) the sharing of alleles with effects that differ among environments (or are specific to one environment). Previous work has shown that when testing for G×E effects with GWAS, modeling the second source of covariance using a second random effect can prevent spurious signals and increase power [33], which we replicated in this setting using simulations (Figures S2 and S3).

We therefore calculated G×E *p*-values for each genetic marker using Grid-LMM (grid step size = 0.01 h^2 units) on the full dataset using both **A** and the G×E relationship matrix as two random effects. These were compared to *p*-values from an LMM that ignores the G×E covariance and only considers genetic similarity, and a model similar to `pylmm` [33, called null-LMM-G×E below] that considers both random effects but estimates variances components from the null model. For each model, we included an additional random effect to account for the repetition of accessions which could induce additional covariance among observations. In this dataset, a simpler approach to testing for GxE was also available: test for associations between markers and the *plasticity* (ie difference in flowering time between the field and 10C) of each accession, which requires only a single random effect and can be fit with GEMMA This simple approach is only possible because every genotype was measured in both environments, but is expected to produce identical tests to the full three-random-effect model and therefore serves as a validation of the Grid-LMM results. We used the plasticity GWAS as a baseline to compare the three models that use the raw data directly (Figure 1b). As expected, ignoring the G×E covariance leads to greatly inflated tests for the majority of markers. Grid-LMM replicated GEMMA's plasticity *p*-values almost exactly when run with three random effects, while a portion of the null-LMM-G×E's tests were deflated, implying lower power. The full analysis using Grid-LMM with three random effects took 2 minutes. Fitting the same model for all 216,130 markers directly using an exact method in `R` [e.g. `sommer`; 36] would take approximately 10 days.

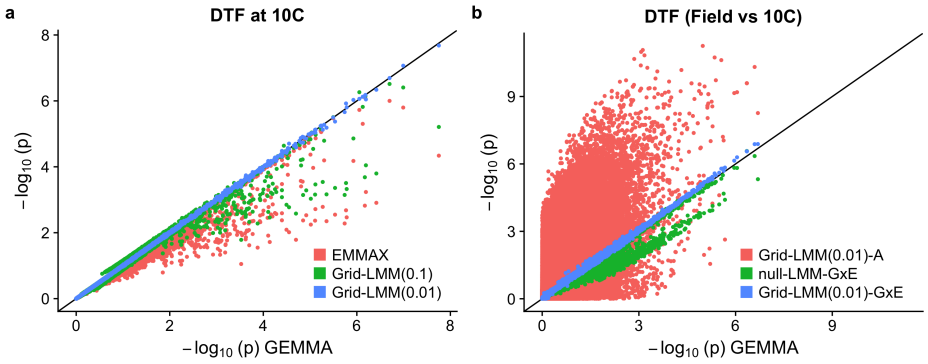

Figure 1. Comparisons of GWAS results for *Arabidopsis* flowering time. (**a**) Results for days-to-flower (DTF) at 10C. Compared are EMMAX and Grid-LMM with grid sizes 0.1 and 0.01, respectively. The exact method GEMMA is treated as a baseline. Each method was applied to the same LMM with only the additive relationship matrix as a random effect, and *p*-values were computed using the Wald test. (**b**) GWAS for the gene-environment (G×E) interactions effects on DTF between the constant 10C and field conditions. GEMMA results are for the plasticity of each accession (i.e. the difference DTF_{Field} - DTF_{10C}), fit with a single random effect of the additive relationship matrix. The other three methods considered the full data with two observations per accession. Namely, Grid-LMM-A fit a "standard" GWAS with a grid step size 0.01, but with two random effects — the additive relationship matrix and an *iid* line effect. The other two models, null-LMM-G×E and Grid-LMM-G×E, fit three random effects — the additive relationship matrix, an *iid* line effect, and a G×E-additive relationship matrix representing the background covariation in gene-environment interactions among accessions.

Even when all individuals are measured in the same environment, the typical additive relationship matrix may not account for all sources of covariation. In particular, the sharing of non-additive alleles may also induce covariance and lead to reduced power and/or inflated tests in GWAS [e.g. 45]. As an illustrative example, we tested for genetic association between 10,346 bi-allelic markers and body mass among 1756 heterogenous stock mice from the Wellcome Trust Centre for Human Genetics [46]. We computed additive, dominance, and pairwise-epistatic relationship matrices, and compared *p*-values derived from an LMM considering only the additive relationship matrix (as in a typical GWAS) to those calculated by Grid-LMM using all three

relationship matrices (Figure 2a and b). Using the three-random effect model, we identified associations on two chromosomes (numbers 4 and 11) that were not apparent when using just the additive relationship matrix as a random effect. Both loci have been identified previously as size-associated QTL [47, 48]. The three-random effect model took 11 minutes using Grid-LMM, while the full analysis on all 10,346 markers would have taken approximately 22 days using an exact approach like `sommer`. Extrapolating to a consortium sized whole-genome analysis with 1 million markers would take $\sim$ 18 hours using Grid-LMM, as opposed to 6 years using `sommer`.

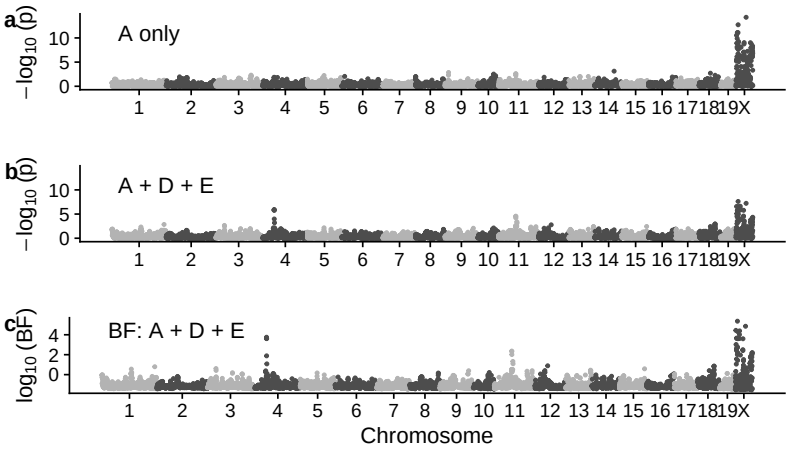

Figure 2. GWAS results for the body weight trait of the Wellcome Trust heterogenous stock mice. Grid-LMM was run using a grid size of 0.01 h^2-units for each model. (**a**) Wald test p-values organized by chromosome from a model with only the additive relationship matrix (**A**) as a random effect. (**b**) Wald test p-values from a model with three random effects: the additive, dominance (**D**) and epistatic (**E**) relationship matrices. (**c**) Approximate Bayes factors calculated from the three random effect model, assuming a standard normal $\mathrm{N}(0,1)$ prior for the marker effects and a uniform prior on the grid of variance component proportions.

To demonstrate the flexibility of Grid-LMM, we also ran an analysis on the mice body weight trait using Bayesian inference and calculated Bayes Factors by comparing the models for each marker to the null model assuming no genetic effects. Here, we used a uniform prior over the grid of variance component proportions and assumed a standard normal prior for the marker effect sizes. In this setting, the Bayes Factors were highly correlated with the frequentist p-values — they also highlight the same associations on chromosomes 4 and 11 (Figure 2c). But in general, Bayes Factors provide additional flexibility for incorporating prior information on individual markers or even combining results among multiple studies [49, 50]. The full Bayesian analysis for Grid-LMM took just 7 minutes longer than the frequentist analysis, making it practical for genome-wide studies. Finally, we confirmed that ignoring the dominance and epistatic covariance resulted in inflated heritability estimates for bodyweight in a Bayesian analysis (Figure S4).

Discussion

Grid-LMM addresses a central obstacle to the practical use of linear mixed models: the computational time needed to find optimal solutions to variance components. Our key observation is that for many common quantitative genetics analyses, optimizing variance component parameters to high precision is not necessary. From a Bayesian perspective, the posterior distribution of variance components tends to be broad even with large sample sizes, and coarse approximations can be sufficient given the uncertainty in their true values. In GWAS applications, magnitudes of variance components are of less interest than the accuracy of test statistics for the (fixed) SNP effects, and we show that these are sufficiently accurate even with approximate variance component proportions. We believe that uncertainty in the estimation of random effect covariance matrices will have a larger effect on both types of statistics than the imprecision in the estimation of the variance components themselves.

The advantage to relaxing the need for perfect variance component solutions is a vast reduction in both computational time and algorithmic complexity. This reduces the time required for a typical GWAS sized dataset with two-or-more random effects from years to hours, and provides a framework for applying LMMs to even more powerful statistical tools [e.g. 51–53]. We optimize variance component estimation with a grid search, the simplest type of optimization algorithms. At each grid vertex, after conditioning on (relative) variance component values, the LMM simplifies to a simple linear model; therefore general purpose solutions to normal linear models are available. This means that the simple LMMs we explored in this paper can easily be generalized to more complex methods used in other GWAS-type applications that currently cannot easily be extended to

experiments with heterogeneous samples (e.g. set-tests and multi-marker regressions). We outline an approach to efficiently find LASSO, ridge, and elastic net solutions with random effect error structures in the Methods. We also apply previously described methods to update the random effect covariance structure on a per SNP basis, removing highly correlated markers from the estimation of the relationship matrices to improve power (again see Methods) [54]. This is particularly useful for LMM-based QTL mapping in controlled crosses where linkage disequilibrium (LD) among markers is high. The summary statistics from each grid vertex could also be combined across studies in meta-analyses to extend mixed model GWAS to much larger sample sizes.

We demonstrated Grid-LMM using two examples that are broadly representative of many experimental settings in quantitative genetics. The first was a study of gene-environment interactions, while the second focused on the partitioning of genetic variance among additive and non-additive components. Recent studies have shown how neglecting gene-environment covariance when estimating heritability [8] or testing genetic associations [33] can lead to biased estimates. There is little reason to expect these results to be limited to controlled environmental settings, as we have studied here. Previous work has found that incorporating spatial covariance through a Gaussian radial basis function improved estimates of trait heritability within a sample of individuals from Uganda [8]. Spatial variability also exists in agricultural field trials, and two-step approaches are frequently used where spatial variation is removed first and then genetic associations tested on the residuals. This two-step procedure may be underpowered, similarly to the two-step procedure of EMMAX, when true effect sizes are large. Similarly, genomic prediction models that use multiple nonlinear genetic kernels often show increased precision [55]. Epistatic and other non-linear genetic variation are known to be large for many traits [56–58] and accounting for this variation may improve our ability to detect both the main effects of markers (as we demonstrated above), as well as possibly interacting loci [17].

The Grid-LMM approach does have limitations. The most important is that the size of the grid spanning the variance components increases nearly exponentially with the number of random effects in the linear model. Since each grid vertex requires a separate Cholesky decomposition of the observation-level covariance matrix, this grid search quickly becomes prohibitively expensive with more than ∼6-8 variance components. This is a general problem for mixed model algorithms, and it may be possible to efficiently adapt derivative-based algorithms to the grid space. Our fast-heuristic search method converges on the correct answer in most cases we have examined; however, likelihood surfaces of linear mixed models are not always convex and this algorithm may converge onto a local maximum in such cases. We note that most general-purpose algorithms for LMMs with multiple random effects are also sensitive to this issue. Furthermore, the Grid-LMM approach is limited to Gaussian linear mixed models; generalized linear mixed model algorithms rely on iteratively re-weighting the observations, a function that changes the covariance matrix in a way that cannot be discretized. Finally, we have not explored LMMs with correlated random effects, although these are commonly used in quantitative genetics. Since correlation parameters are restricted to the interval $(-1,1)$, discretizing correlations in parallel with the variance component proportions may be feasible and an avenue that is worth future study.

Methods

Linear Mixed Models

We consider the following parameterization of the standard linear mixed model:

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \sum_{l=1}^{L} \mathbf{Z}_l \boldsymbol{u}_l + \mathbf{e}, \\ \boldsymbol{u}_l &\sim \mathrm{N}(\mathbf{0}, \sigma^2 h_l^2 \mathbf{K}_l), \\ \mathbf{e} &\sim \mathrm{N}(\mathbf{0}, \sigma^2 h_e^2 \mathbf{I}), \end{aligned} \tag{1}$$

where n is the number of observations, L is the number of random effect terms (not including the residuals), $\mathbf{y}$ is an $n \times 1$ vector of quantitative traits, and $\mathbf{W}$ and $\mathbf{X}$ are $n \times c$ and $n \times p$ design matrices for covariates and marker effects, respectively, with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ corresponding $c \times 1$ and $p \times 1$ vectors of coefficients. Similarly, $\mathbf{Z}_l$ are $n \times r_l$ design matrices with corresponding random effects $\boldsymbol{u}_l$, which are normally distributed around zero and have covariance matrices proportional to the known positive semi-definite $r_l \times r_l$ matrices $\mathbf{K}_l$. Lastly, $\mathbf{e}$ is a $n \times 1$ vector of uncorrelated normally distributed errors, and $\mathrm{N}(\bullet,\bullet)$ denotes the multivariate normal distribution. The common variance of all random effects are denoted by σ^2, and the vector $\mathbf{h}^2 = (h_1^2, \ldots, h_L^2, h_e^2)$ represents the proportion of variance attributed to each random effect term or the residual error. Elements of $\mathbf{h}^2$ are all non-negative and sum to one, forming an L-dimensional simplex.

In GWAS applications, we assume that $\mathbf{W}$ is constant across markers and the values of $\boldsymbol{\alpha}$ are not of central interest; meanwhile, $\mathbf{X}$ varies for each test and we aim to perform statistical inference on a subset of $\boldsymbol{\beta}$. In heritability estimation applications, we focus on inferring the vector $\mathbf{h}^2$. In both cases, the vectors $\boldsymbol{u}_l$ and $\mathbf{e}$ are nuisance parameters and we can

integrate them out resulting in the following equivalent model:

$$\begin{aligned}\mathbf{y} &\sim \mathrm{N}(\mathbf{W}\boldsymbol{\alpha}+\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{V}),\\ \mathbf{V} &= \sum_{l=1}^{L} h_l^2\mathbf{Z}_l\mathbf{K}_l\mathbf{Z}_l^\intercal + h_e^2\mathbf{I}.\end{aligned} \tag{2}$$

If the matrix $\mathbf{V}$ is full-rank (which is guaranteed if $h_e^2 > 0$), we can use the inverse of the Cholesky decomposition $\mathbf{V} = \mathbf{L}\mathbf{L}^\intercal$ to transform model 2 to the following:

$$\mathbf{y}^* \sim \mathrm{N}(\mathbf{W}^*\boldsymbol{\alpha}+\mathbf{X}^*\boldsymbol{\beta},\sigma^2\mathbf{I}) \tag{3}$$

where $\mathbf{y}^* = \mathbf{L}^{-1}\mathbf{y}$, $\mathbf{W}^* = \mathbf{L}^{-1}\mathbf{W}$ and $\mathbf{X}^* = \mathbf{L}^{-1}\mathbf{X}$. Equation 3 is a simple linear model for $\mathbf{y}^*$, with the likelihood function:

$$l_F(\mathbf{y}^*;\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2\,|\,\mathbf{h}^2) = \frac{1}{2}\left[-n\log(2\pi\sigma^2) - \frac{1}{\sigma^2}(\mathbf{y}^*-\mathbf{W}^*\boldsymbol{\alpha}-\mathbf{X}^*\boldsymbol{\beta})^\intercal(\mathbf{y}^*-\mathbf{W}^*\boldsymbol{\alpha}-\mathbf{X}^*\boldsymbol{\beta}))\right],$$

where efficient methods for inference of $[\boldsymbol{\alpha},\boldsymbol{\beta}]$ and σ^2 are well known. We derive maximum-likelihood and restricted-maximum likelihood solutions for these parameters here, as well as posterior distributions under the conditional the normal-inverse-gamma prior below. The log-likelihood and restricted-likelihood functions (respectively) for Equation 2 are:

$$l_F(\mathbf{y};\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2,\mathbf{h}^2) = l_F(\mathbf{y}^*;\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2\,|\,\mathbf{h}^2) - \log|\mathbf{L}| \tag{4}$$

and

$$l_R(\mathbf{y};\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2,\mathbf{h}^2) = l_F(\mathbf{y};\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2,\mathbf{h}^2) + \frac{1}{2}\left[(c+p)\log(2\pi\sigma^2) + \log|\widetilde{\mathbf{X}}^\intercal\widetilde{\mathbf{X}}| - \log|\widetilde{\mathbf{X}}^{*\intercal}\widetilde{\mathbf{X}}^*|\right] \tag{5}$$

which are simple (and computationally inexpensive) updates to the likelihood function of Equation 3. Here, we denote $|\bullet|$ as the matrix determinant where $\widetilde{\mathbf{X}} = [\mathbf{W};\mathbf{X}]$ and $\widetilde{\mathbf{X}}^* = [\mathbf{W}^*;\mathbf{X}^*]$, respectively. For ML and REML applications, we can calculate the profile likelihoods $l_F(\mathbf{y};\mathbf{h}^2)$ and $l_R(\mathbf{y};\mathbf{h}^2)$ as functions of the profile likelihood of the rotated data, which is formed by maximizing $l_F(\mathbf{y}^*;\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2\,|\,\mathbf{h}^2)$ with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and σ^2. Namely:

$$l_F(\mathbf{y};\mathbf{h}^2) = \frac{n}{2}\left[\log\left(\frac{n}{2\pi}\right) - 1 - \log(\mathrm{RSS}_{\mathbf{y}^*})\right] - \log|\mathbf{L}| \tag{6}$$

where $\mathrm{RSS}_{\mathbf{y}^*} = \mathbf{y}^{*\intercal}[\mathbf{I}-\widetilde{\mathbf{P}}]\mathbf{y}^*$ is the residual sum of squares, and $\widetilde{\mathbf{P}} = \widetilde{\mathbf{X}}^*(\widetilde{\mathbf{X}}^{*\intercal}\widetilde{\mathbf{X}}^*)^{-1}\widetilde{\mathbf{X}}^{*\intercal}$ is a projection (hat) matrix.

Statistical inference by parallel grid search

We now outline the Grid-LMM approach for calculating approximate Wald-test statistics, and then we show extensions for computing Bayesian posterior distributions of variance components and marker specific Bayes Factors.

The Wald statistic for testing the null hypothesis $\mathbf{M}[\boldsymbol{\alpha}^\intercal,\boldsymbol{\beta}^\intercal]^\intercal = \mathbf{0}$ for an arbitrary $q\times(c+p)$ matrix $\mathbf{M}$ is:

$$F_{\mathrm{Wald}} = \frac{\widehat{\boldsymbol{\theta}}^\intercal\mathbf{M}^\intercal(\mathbf{M}\widetilde{\mathbf{X}}^\intercal\mathbf{V}^{-1}\widetilde{\mathbf{X}}\mathbf{M}^\intercal)\mathbf{M}\widehat{\boldsymbol{\theta}}}{q} \tag{7}$$

with q and $(n-c-p)$ degrees of freedom, where $\widehat{\boldsymbol{\theta}}$ is the estimate of $[\boldsymbol{\alpha}^\intercal,\boldsymbol{\beta}^\intercal]^\intercal$ using the REML estimate of $\mathbf{V}$ [59].

To calculate genome-wide Wald statistics, we must estimate $\widehat{\mathbf{h}}^2$ for each of the p markers tested. There are no closed-form ML (or REML) solutions for $\mathbf{h}^2$; and therefore, iterative algorithms are required. A naive approach involves repeatedly inverting $\mathbf{V}$, which scales cubically in the number of observations. Since the estimates for $\widehat{\mathbf{h}}^2$ differ for each test, the total computational complexity of in a GWAS setting is $\mathscr{O}(tpn^3)$ assuming an average of $\sim t$ inversions per run of the optimization algorithm (generally $\sim 3-100$, increasing with the number of random effect parameters). The fast algorithm used in GEMMA [24] and FaST-LMM [23] reduces this to $\mathscr{O}(n^3+pn^2+ptc^2n)$ by utilizing the eigenvalue decomposition of the similarity matrix $\mathbf{K}$. However, this only works for a single random effect. We are unaware of any exact algorithm with lower computational complexity for models with multiple random effects (and full-rank covariance matrices).

With Grid-LMM, rather than optimizing $\mathbf{h}^2$ separately for each marker, we instead define a grid of candidate values for $\mathbf{h}^2$ and calculate the restricted profile-likelihood $l_R(\mathbf{h}^2\,|\,\mathbf{y}^*)$ at every grid vertex for each marker. At each grid vertex, we must invert $\mathbf{V}$ once, but we can re-use this calculation for every marker. This has a computational complexity of $\mathscr{O}(g(n^3+cn^2+pbn^2))$ for a grid with g vertices, assuming covariates $\mathbf{W}$ are constant across markers. For all analyses reported in the main text, we use

a rectangular grid with a resolution of 0.1 or 0.01 h^2-units, with the restriction that all $h_l^2 \geq 0$ and $\sum_{l=1}^{L} h_l^2 < 1$. As described below, we either peg this grid to the origin (i.e. $h_l^2 = 0, \forall l$), or to the REML estimate $\widehat{\mathbf{h}}_0^2$ derived from the null model with no marker effects. This grid search generates a vector of g scores for each marker. We select the values of $\widehat{\mathbf{h}}^2$ that corresponds to the largest score for each marker and use Equation 7 to calculate the approximate Wald statistics.

To calculate approximate likelihood ratio test statistics, we use the Grid-LMM approach to calculate the full profile-likelihoods for models under both the null and alternative hypothesis. In the *Arabidopsis* G×E analysis, the null model included the main effect of each marker and thus was unique for each test. In the Wellcome Trust analysis with the interrogation of non-additive effects on body weight in mice, the null model was identical for each marker and so only needed to be solved once.

For Bayesian inference, rather than working with the profile likelihoods, we instead use a conditional normal-inverse-gamma prior $([\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}^\intercal]^\intercal, \sigma^2) \sim \mathrm{NIG}(\mathbf{0}, \boldsymbol{\Psi}, a_0, b_0)$, and then integrate over the prior to calculate the marginal likelihood of the data given $\mathbf{h}^2$. This results in the following:

$$p(\mathbf{y}\,|\,\mathbf{h}^2) = \frac{b_0^{a_0}}{(2\pi)^{n/2+(c+p)/2}|\mathbf{V}|^{1/2}|\boldsymbol{\Psi}|^{1/2}\Gamma(a)} \times \frac{(2\pi)^{(p+c)/2}|\boldsymbol{\Psi}^*|^{1/2}\Gamma(a^*)}{(b^*)^{a^*}}, \tag{8}$$

where $\Gamma(\bullet)$ is the gamma function, $\boldsymbol{\Psi}^* = (\boldsymbol{\Psi}^{-1} + \widetilde{\mathbf{X}}^\intercal \mathbf{V}^{-1}\widetilde{\mathbf{X}})^{-1}$, $a^* = a_0 + n/2$, and $b^* = b_0 + \mathrm{RSS}_{\mathbf{y}^*,\boldsymbol{\Psi}}/2$ with $\mathrm{RSS}_{\mathbf{y}^*,\boldsymbol{\Psi}}$ having the same form as $\mathrm{RSS}_{\mathbf{y}^*}$ in Equation 6 — except with $\widetilde{\mathbf{P}} = \widetilde{\mathbf{X}}^*\boldsymbol{\Psi}^*\widetilde{\mathbf{X}}^{*\intercal}\widetilde{\mathbf{X}}^*\boldsymbol{\Psi}^*\widetilde{\mathbf{X}}^{*\intercal}$. See the Supplemental Methods for more detail on the derivations. We calculate the marginal likelihood $p(\mathbf{y}\,|\,\mathbf{h}^2)$ at each vertex of the grid as described above. Assuming a discrete-valued prior $p(\mathbf{h}^2)$, we can then compute the posterior distribution of $\mathbf{h}^2$ as:

$$p(\mathbf{h}^2\,|\,\mathbf{y}) = \frac{|\mathbf{V}_{\mathbf{h}^2}|^{-1/2}|\boldsymbol{\Psi}^*_{\mathbf{h}^2}|^{1/2}(b^*_{\mathbf{h}^2})^{-a^*}p(\mathbf{h}^2)}{\sum_{\mathbf{h}^2}|\mathbf{V}_{\mathbf{h}^2}|^{-1/2}|\boldsymbol{\Psi}^*_{\mathbf{h}^2}|^{1/2}(b^*_{\mathbf{h}^2})^{-a^*}p(\mathbf{h}^2)} \tag{9}$$

where, for clarity, parameters that are a function of $\mathbf{h}^2$ are denoted with a subscript. Continuous target densities $\pi(\mathbf{h}^2)$ can be approximated as $p(\mathbf{h}^2)$ by assigning each grid vertex a probability equal to the integral of $\pi(\mathbf{h}^2)$ over the L-dimensional rectangle centered at the corresponding value $\mathbf{h}^2$. We assume a uniform prior over our grid for all analyses presented in the main text. Bayes factors are computed by comparing models under the alternative and null hypotheses as the ratios in Equation 9. All analytical calculations — including the summation in Equation 9 — can be performed on the log-scale to prevent numerical underflows. Terms common to both models drop out of the ratio; therefore, limiting improper priors on $\boldsymbol{\alpha}$ and σ^2 can be used, which results in scale-independence [49].

Accelerated grid searches

The full grid search described above is dramatically faster than the naive algorithm for mixed-model GWAS analyses, as long as the vertices of the grid is less than the number of genetic markers (i.e. $g < p$) and can easily be parallelized across multiple computers. However, g grows rapidly as the grid resolution and number of random effects increases:

$$g = \sum_{k=1}^{L} \binom{m}{k}\binom{L-1}{k-1}, \tag{10}$$

for a grid with m divisions per h_l^2, and therefore can still be slow. If we make two assumptions which are commonly true, we can develop heuristics for both the ML/REML and Bayesian algorithms that prevent the need to evaluate every grid vertex:

- The vast majority of markers have little association with the phenotype. This is a common hypothesis in GWAS settings [e.g. 60–63], and for our purposes, we merely require that the percentage of variance explained individually by most markers is smaller than difference between grid vertices $\approx 1/m$.
- The likelihood and/or posterior function is convex. This is not always true, since both the likelihood and posterior functions can have > 1 maximum. However, the conditions that cause these events are rare [64], and most exact LMM algorithms are also susceptible to converging to local maxima.

To search for the ML or REML solutions, we first find $\widehat{\mathbf{h}}_0^2$ under the null model with no marker effects. We calculate the profile (restricted)-likelihood scores for each test at $\widehat{\mathbf{h}}_0^2$, and then form a grid centered at this value by adding or subtracting $1/m$ to each h_l^2 in all combinations. For two random effects, this grid will be a ring around $\widehat{\mathbf{h}}_0^2$ with $g_1 \leq 8$ vertices (depending on if $\widehat{\mathbf{h}}_0^2$ is within $1/m$ of a boundary of the simplex). We calculate the scores for each test at each vertex of the grid and compare the maximum scores to the scores at $\widehat{\mathbf{h}}_0^2$. For every test, when no greater value is found, we choose $\widehat{\mathbf{h}}_0^2$ as the maximum and skip

the corresponding marker in all future calculations. For the remaining p_2 markers, we select the set $\{\widehat{\mathbf{h}}_0^2, \mathbf{h}_1^2, \ldots, \mathbf{h}_j^2\}$ of grid vertices that maximized the scores for 1+ tests, and form a new grid of size g_2 around all of them, dropping vertices already tested. This procedure is repeated T times until the new grid no-longer increases scores for any test and we are confident that all (approximate) maximums have been found. This accelerated search has total complexity $\mathcal{O}(n^3 + pn^2 + \sum_{i=1}^{T} g_i(n^3 + p_i n^2))$.

Analogously, to accelerate evaluations of posterior distributions, we evaluate $p(\mathbf{y}\,|\,\mathbf{h}^2)$ over an initial grid of resolution $1/m_1$ with discrete prior $p_{m_1}(\mathbf{h}^2)$ and estimate the posterior distribution as in Equation 9. Summary statistics, such as the posterior mean and variance, will be more accurate if the posterior mass is distributed across multiple vertices of the grid. Therefore, we identify the set $\mathscr{H} = \{\mathbf{h}_l^2\}$ of vertices that together explain 99% of the posterior mass. If the size of this set is below a threshold (say $|\mathscr{H}| = 10$), we then form a new grid with double the resolution: $m_2 = 2m_1$ and a new prior $p_{m_2}(\mathbf{h}^2)$. Vertices that overlap between the grids can be filled in directly. We then begin filling in the new grid by evaluating vertices within $1/m_2$ distance from any h_l^2 corresponding to the vertices in $\mathscr{H}$. After each iteration, we re-calculate the set $\mathscr{H}$, and continue evaluating the neighboring vertices as long as $\mathscr{H}$ continues to grow. If the size of $\mathscr{H}$ remains smaller than the threshold after whole grid is evaluated, or no new vertices are added to $\mathscr{H}$ at the end of an iteration, we double the resolution again: $m_{i+1} = 2m_i$ and repeat the grid-filling steps again. We note that this procedure is only appropriate if the posterior is convex, and therefore is limited to the case of uniform priors on $\mathbf{h}^2$.

To accelerate evaluations of GWAS Bayes Factors, we combine the two previously described algorithms. In particular, we define a grid centered on $\widehat{\mathbf{h}}_0^2$ with a resolution $1/m$ that is sufficiently fine such that we expect each posterior to be distributed across multiple vertices. We then calculate $p(\mathbf{y}\,|\,\mathbf{h}^2)$ for each test, starting from $\widehat{\mathbf{h}}_0^2$ and moving out in concentric rings on the grid. After each iteration, if the new ring contributes $< 0.01\%$ to the total posterior mass (among evaluated vertices) for that test, we assume the posterior is well characterized and stop evaluating $p(\mathbf{y}\,|\,\mathbf{h}^2)$ for that marker. As for the ML and REML solutions above, markers with little to no association with the phenotype will lead to posteriors of $\mathbf{h}^2$ that are concentrated close to $\widehat{\mathbf{h}}_0^2$; only markers with large effects will shift $p(\mathbf{h}^2\,|\,\mathbf{y})$ strongly to new regions of the grid.

Unless otherwise specified, we used accelerated grid searches for all Grid-LMM analyses presented here with grid step sizes of 0.01 h_l^2 units.

Extension to Lasso and Elastic Net

We show here how the Grid-LMM approach can be extended to more complicated LMMs, while leveraging efficient optimization algorithms from simpler models. The LASSO and elastic net are two methods for finding regularized solutions to multiple linear regression problems with high-dimensional predictors (i.e. $p \gg n$), for example to simultaneously fit all markers in a GWAS. The typical *lasso* solutions are estimates of regression parameters $\boldsymbol{\beta}$ of the model:

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}) \tag{11}$$

that maximize the penalized log-likelihood function:

$$l_p(\mathbf{y}; \boldsymbol{\beta}, \sigma^2\,|\,\lambda) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2n\lambda\sum_{j=1}^{p}|\beta_j|\right], \tag{12}$$

where λ is a tuning parameter, with larger values of λ favoring sparse solutions (see Supplemental Note). We adopt common notation to make λ independent of the sample size n [65]. For a given λ, l_p is maximized when $\widehat{\sigma}^2 = \left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2\lambda\sum_{j=1}^{p}|\beta_j|\right]/2$ is minimized. Hence yielding:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min}\,\frac{1}{2n}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|. \tag{13}$$

For *elastic net* regularization, the penalty in Equation 12 is modified to: $\lambda\left[(1-\alpha)/2\sum_{j=1}^{p}\beta_j^2 + \alpha\sum_{j=1}^{p}|\beta_j|\right]$ for the *elastic net* mixing parameter α [66].

The *lasso* can be extended to linear mixed effect models with L random effects by substituting $\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with $\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{V})$, where $\mathbf{V} = \sum_{l=1}^{L} h_l^2\mathbf{K}_l + h_e^2\mathbf{I}$ as above. The *LMM-lasso* solutions then maximize the following penalized log-likelihood function:

$$\tilde{l}_p(\mathbf{y}; \boldsymbol{\beta}, \mathbf{h}^2, \sigma^2\,|\,\lambda) = -\frac{1}{2}\log|\mathbf{V}| - \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2\mathrm{tr}(\mathbf{V}^{-1})\lambda\sum_{j=1}^{p}|\beta_j|\right], \tag{14}$$

where $\mathrm{tr}(\bullet)$ denotes the matrix trace. This objective function differs slightly from the originally proposed objective function [42] (see Supplemental Note), but still preserves the typical *lasso* solutions when $\mathbf{V} = \mathbf{I}$. Just as for the simple LMM, maximizing

$\tilde{l}_p$ requires numerical optimization involving repeatedly inverting $\mathbf{V}$ and is in general non-convex. Existing implementations are slow, or require approximations such as estimating $\mathbf{h}^2$ from a null model or ignoring the tuning of λ. On the other hand, the `glmnet` `R` package fits both lasso and elastic net penalties very efficiently for large-scale problems using a Coordinate Descent algorithm, and provide solutions for all values of the tuning parameter in a single pass [41]. It is limited, though, to models with independent errors. Using Grid-LMM, we can leverage the efficiency of algorithms like `glmnet` to efficiently find approximate regularized solutions to high-dimensional regression problems with multiple random effects. Note that if we condition on a trial value $\widehat{\mathbf{h}}^2$, we can calculate $\mathbf{V} = \mathbf{L}\mathbf{L}^\intercal$ and then rewrite $\tilde{l}_p$ in Equation 14 as:

$$\tilde{l}_p(\mathbf{y};\boldsymbol{\beta},\sigma^2\,|\,\widehat{\mathbf{h}}^2,\lambda) = -\log|\mathbf{L}| - \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta})^\intercal(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}) + 2\mathrm{tr}(\mathbf{V}^{-1})\lambda\sum_{j=1}^{p}|\beta_j|\right]. \tag{15}$$

where $\mathbf{y}^* = \mathbf{L}^{-1}\mathbf{y}$ and $\mathbf{X}^* = \mathbf{L}^{-1}\mathbf{X}$. Conditional on $\widehat{\mathbf{h}}^2$, $\tilde{l}_p$ is then maximized when $\widehat{\sigma}^2 = \left[(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta})^\intercal(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}) + 2\mathrm{tr}(\mathbf{V}^{-1})\lambda\sum_{j=1}^{p}|\beta_j|\right]/n$ is minimized, which is the general *lasso* solution for $\mathbf{y}^*$ and $\mathbf{X}^*$, with the penalty λ modified by $\mathrm{tr}(\mathbf{V}^{-1})/n$. Therefore we can use `glmnet` directly to find the regularization paths for $\boldsymbol{\beta}\,|\,\widehat{\mathbf{h}}^2$, and use this to calculate the profiled penalized log-likelihood $\tilde{l}_p(\mathbf{y};\lambda\,|\,\mathbf{h}^2)$ over a sequence of values of $\boldsymbol{\lambda} = \{\lambda_1,\ldots,\lambda_k\}$ spanning 2-3 orders of magnitude. Now, with Grid-LMM, we simply form a grid over all valid values of $\mathbf{h}^2$, and calculate the $\tilde{l}_p(\mathbf{y};\lambda\,|\,\mathbf{h}^2)$ sequence at each vertex of the grid. For each λ_i, we choose $\widehat{\mathbf{h}}^2$ that produces the maximum value of $\tilde{l}_p(\mathbf{y};\lambda_i\,|\,\mathbf{h}^2)$ across the grid, and report the corresponding *lasso* estimates $\widehat{\boldsymbol{\beta}}$.

Running `glmnet` on $\mathbf{y}^*$ and $\mathbf{X}^*$ requires care to ensure that the λ sequence is treated equivalently across grid cells. We first calculate $\widehat{\lambda}_{0g}$ — the smallest value of λ that will set all penalized coefficients to zero — at each grid vertex, where:

$$\lambda_{0g} = \max_{j\in\{1,\ldots,p\}} \rho_j\widetilde{\beta}_j/2\mathrm{tr}(\widetilde{\mathbf{V}}^{-1}) \tag{16}$$

with ρ_j denoting the penalty factor for each term in $\mathbf{X}$ (i.e. $0 =$ no penalty, $1 =$standard penalty where `glmnet` normalizes all ρ_j to sum to 1 over all coefficients), and $\widetilde{\beta}_j$ is the least squares coefficient for $\mathbf{x}_j^*$ regressed on $(\mathbf{y}^* - \widetilde{\mathbf{y}}^*)$ with $\widetilde{\mathbf{y}}^*$ being the least squares fit on all unpenalized terms to $\mathbf{y}^*$. Here, $\lambda_0 = \lambda_{0g}$ for the g that maximizes $\tilde{l}_p(\mathbf{y};\mathbf{h}_g^2\,|\,\lambda_{0g})$. Also, because `glmnet` internally standardizes the response to unit variance, we normalize both $\mathbf{y}^*$ and $\mathbf{X}^*$ by the standard deviation $\mathrm{sd}(\mathbf{y}^*)$ — which varies depending on $\mathbf{h}^2$ and is calculated with $1/n$ instead of $1/(n-1)$ — and then run `glmnet` with $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}/\mathrm{sd}(\mathbf{y}^*)^2$. For *elastic net* solutions, λ_{og} is divided by the parameter α.

Extension to account for proximal contamination in GWAS and QTL mapping

It has been shown in GWAS applications that including the testing marker (and markers in LD with the testing marker) in the calculation of the additive relationship matrix $\mathbf{K}$ leads to a deflation of tests statistics and loss of power [54]. The effect of this "proximal contamination" is greatest when the ratio between the *effective number of markers* q and the sample size n is small [67]. In most modern GWAS, $q > n$, and therefore the effect of proximal contamination is not large. However, in QTL mapping within artificial populations, LD is extensive and q can be small. Unfortunately, directly updating $\mathbf{K}$ (and therefore $\mathbf{V}$) for each marker is not practical because it would require repeatedly re-evaluating large matrix inverses. One solution to the problem of "proximal contamination" is leave out all markers on chromosome j from the calculation of $\mathbf{K}$ when testing markers on chromosome j [commonly referred to as the LOCO method; 68]. This is possible to implement in Grid-LMM, but requires re-calculating all Cholesky decompositions of $\mathbf{V}$ across the grid for each chromosome. Also, for organisms with few chromosomes (e.g. *Arabidopsis thaliana* has only five), LOCO can result in loss of power because a large proportion of the genome is excluded from $\mathbf{K}$. Alternatively, previous work presented a method for updating $\mathbf{K}$ and $\mathbf{V}^{-1}$ using low-rank operations [54]. This method is part of FaST-LMM and only directly applies to models with a single random effect. Here, we show how the same low-rank updates can be implemented in Grid-LMM for models with multiple random effects. Let the full covariance in a LMM be:

$$\sigma^2\mathbf{V} = \sum_{l=1}^{L} h_l^2\mathbf{K}_l + h_e^2\mathbf{I}, \tag{17}$$

and say that we aim to update $\mathbf{K}_1$ for each marker. A typical choice for $\mathbf{K}_1$ is:

$$\mathbf{K}_1 = \mathbf{X}\mathbf{X}^\intercal/c \tag{18}$$

where $\mathbf{X}$ has been appropriately standardized and c is a normalizing constant that is a function of the number of markers and possible allele frequencies at each marker. We wish to update $\mathbf{V}$ and $\mathbf{V}^{-1}$ to $\mathbf{V}_j$ and $\mathbf{V}_j^{-1}$ specifically for marker j, by effectively substituting $\mathbf{K}_{1j}$ for $\mathbf{K}_1$. The matrix $\mathbf{K}_{1j}$ is calculated without a set of markers in strong LD with marker j (and including

marker j itself). Let $\mathbf{j}$ be this set of markers, and define $\mathbf{X_j}$ and $\mathbf{X_{-j}}$ be the marker matrices for set $\mathbf{j}$ and all other remaining markers, respectively. If we let $\mathbf{K}_{1j} = \mathbf{X_{-j}}\mathbf{X_{-j}}^\intercal/c_j$, where c_j is the new normalizing constant, then $\mathbf{K}_{1j} = (c\mathbf{K}_1 - \mathbf{X_j}\mathbf{X_j}^\intercal)/c_j$ and $\mathbf{V}_j \approx \mathbf{V} - h_1^2\mathbf{X_j}\mathbf{X_j}^\intercal/c_j$. To make the calculation of $\mathbf{V}_j$ exact, we must calculate $\mathbf{K}_1$ with c_j instead of c, and assume that c_j is constant across markers. Now, assuming the $\mathbf{X_j}$ has fewer than n columns, $\mathbf{V} - h_1^2\mathbf{X_j}\mathbf{X_j}^\intercal/c_j$ is a low-rank update to $\mathbf{V}$. Therefore, assuming $\mathbf{V}_j$ is still full-rank (which will generally be true if $p \gg n$ and $\mathbf{j}$ contains only a few markers), we can perform a series of rank-one downdates to the Cholesky factor $\mathbf{L}$ or $\mathbf{V}$ to find $\mathbf{L}_j$ such that $\mathbf{L}_j\mathbf{L}_j^\intercal = \mathbf{V}_j$. These downdates are computationally inexpensive, requiring only vector and scalar arithmetic. Given $\mathbf{L}_j$ for marker j, we can then calculate $l_R(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \mathbf{h}^2)$ across our grid, to find the REML solution and calculate the necessary Wald statistics.

GWAS datasets

Genotype and phenotype data on 107 *Arabidopsis thaliana* traits and 216,130 genetic markers were downloaded from https://github.com/Gregor-Mendel-Institute/atpolydb/wiki. We follow practices suggested by the original authors of these data and log-transformed a subset of the phenotypes prior to analyses (except for traits that had values equal to zero) [44] . For the analysis of gene-environment interactions for flowering time, we (1) extracted the trait identifiers "7_FT10" (i.e. growth chamber at 10C) and "57_FT Field" (i.e. field setting), (2) selected data from the 175 accessions that were measured in both environments, and (3) concatenated the two datasets into a single vector of 350 observations. The two traits were individually standardized to have mean zero and standard deviation 1 prior to analysis. We used the `sommer` package in `R` to calculate an additive relationship matrix from all 216,130 markers [36], and then created a G×E kinship matrix as $\mathbf{DZKZ}^\intercal\mathbf{D}$ where $\mathbf{K}$ is the 175×175 additive genomic relationship matrix, $\mathbf{Z}$ is a 350×175 incidence matrix linking observations to accessions, and $\mathbf{D}$ is a 350×350 diagonal matrix with elements equal to -1 or 1 corresponding to observations be measured under "7_FT10" and "57_FT Field", respectively. Plasticities for each accession were calculated as the difference between "57_FT Field" - "7_FT10". We ran GEMMA (version 0.97) under "-lmm1" option for Wald tests with the $\mathbf{K}$ matrix described above. We emulated EMMAX and `pylmm` functionality by estimating variance components using a null model with no marker effects, and either a single $\mathbf{K}$ matrix (for single-trait analyses) or 3 random effects (for G×E analysis) using `sommer`. Wald test statistics were computed for each marker using our GridLMM `R` function with a grid consisting of only a single vertex. For G×E analyses, we fit a model with a main effect of the environment, a main effect for each marker of interest, and an interaction between the marker and the environment — and then calculated the Wald-F statistic only for the interaction effect.

The heterogeneous stock of mice data from the Wellcome Trust Centre for Human Genetics (http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml) consists of 1756 individuals from 85 families, all descended from eight inbred progenitor strains [46]. Data on 131 quantitative traits that are classified into 6 broad categories including behavior, diabetes, asthma, immunology, haematology, and biochemistry were downloaded from http://mtweb.cs.ucl.ac.uk/mus/www/GSCAN/index.shtml/index.old.shtml. A total of 10,346 autosomal SNPs were available for all mice. For individuals with missing genotypes, we imputed missing values by the mean genotype of that SNP in their family. We selected the body mass trait as the phenotype of interest because it had one of the largest dominance and epistatic variance components among all 131 traits in the dataset, based on a variance component analysis using `sommer` with additive, dominance, and epistatic relationship matrices (see supporting information of [69] in for more details). Wald test statistics and Bayes Factors were calculated using the heuristic (accelerated) grid search of Grid-LMM. Bayes Factors were calculated assuming flat priors on the intercept and residual variance term, a standard normal prior $\mathrm{N}(0,1)$ on the marker effects [similar to the previously proposed D_2 prior; 49], as well as a uniform prior over the grid of variance component proportions.

Power simulations

We compared the power of one and two-random effect models for detecting G×E markers based on the Arabidopsis flowering time data. We calculated additive genetic and G×E relationship matrices as described above using the actual genotypes. We then simulated phenotypes by summing together a G×E effect of a particular marker, (co)variance from the two random effects, and normally distributed error. For each combination of marker effect sizes ($\{0, 0.025, 0.5, 0.1, 0.15, 0.2\}$% of the total phenotypic variance), and random effect proportions ($h_l^2 \in \{0, 0.4, 0.8\}$ for each random effect), we ran 1000 simulations with different randomly selected markers from the Arabidopsis genotype data. We then fit five models to each dataset and calculated Wald-tests for marker G×E effects using each. The five methods included three two-random effect approaches including: (1) an exact two-random effect model fit with `sommer`, (2) the approximate two-random effect model fit with Grid-LMM with a grid size of 0.1 h^2-units, and (3) the approximate two-random effect model `pylmm` that conditions on variance components estimated under the null model [33]. We also consider two one-random effect models that could be fit with GEMMA: (4) a model that only included the additive genetic relationships and ignored the G×E covariance, and (5) a model that only included the G×E covariance and ignored the additve genetic relationships. We calculated statistical power for each GWAS method as the proportion of Wald test p-values exceeding the genome-wide significance level at the conventional Bonferroni corrected $P = 2 \times 10^{-7}$ (Figure S2). For the simulations with a marker effect size of 0%, we compared the distribution of

$-\log_{10}(p)$-values from each model to the expectation from a uniform distribution (Figure S3). The results show that Grid-LMM maintains nearly identical power to the exact method in each simulation while maintaining an accurate null-distribution. The approximate method based on `pylmm` has uniformly lower power, but maintains an accurate null-distribution. The one-random effect methods show opposite results: when including only the additive relationship matrix, power for detecting G×E is uniformly high, but the null-distribution of $-\log_{10}(p)$-values is greatly inflated when G×E-variation is large. On the other hand, power for detecting G×E is low when only G×E covariance is modeled and yet the additive genetic-variance is large. Furthermore, the null-distribution of $-\log_{10}(p)$-values for this one random effect method is greatly deflated. These results all confirm our expectations that modeling covariance accurately, but not necessarily precisely, is important for accurate association mapping in GWAS.

Software Availability

Software for computing the Grid-LMM is carried out in R code, which is freely available at https://github.com/deruncie/GridLMM.

Acknowledgements

D.E.R was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA), Hatch project 1010469. L.C. would like to acknowledge the support of Institutional Development Award Number P20GM109035 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds COBRE Center for Computational Biology of Human Disease. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the funders or supporters.

Author Contributions Statement

D.E.R. and L.C. conceived the experiment(s), D.E.R. conducted the experiment(s), D.E.R. and L.C. analyzed the results. Both authors wrote and revised the manuscript.

Competing Financial Interests

The authors have declared that no competing interests exist.

References

1. B Devlin and Kathryn Roeder. Genomic control for association studies. *Biometrics*, 55:997–1004, December 1999.
2. William Astle and David J Balding. Population structure and cryptic relatedness in genetic association studies. *Statistical Science*, 24(4):451–471, November 2009.
3. Alkes L Price, Noah A Zaitlen, David Reich, and Nick Patterson. New approaches to population stratification in genome-wide association studies. *Nature reviews. Genetics*, 11(7):459–463, 07 2010.
4. Bjarni J Vilhjálmsson and Magnus Nordborg. The nature of confounding in genome-wide association studies. *Nature Reviews Genetics*, 14(1):1–2, November 2012.
5. Jeremy J Berg, Arbel Harpak, Nicholas Sinnott-Armstrong, Anja Moltke Joergensen, Hakhamanesh Mostafavi, Yair Field, Evan A Boyle, Xinjun Zhang, Fernando Racimo, Jonathan K Pritchard, and Graham Coop. Reduced signal for polygenic adaptation of height in uk biobank. *bioRxiv*, 2018.
6. Réka Howard, Alicia L Carriquiry, and William D Beavis. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes|Genomes|Genetics*, 4(6):1027–1046, 06 2014.
7. Doug Speed and David J Balding. MultiBLUP: Improved SNP-based prediction for complex traits. *Genome Research*, 24(9):gr.169375.113–1557, June 2014.
8. David Heckerman, Deepti Gurdasani, Carl Kadie, Cristina Pomilla, Tommy Carstensen, Hilary Martin, Kenneth Ekoru, Rebecca N Nsubuga, Gerald Ssenyomo, Anatoli Kamali, Pontiano Kaleebu, Christian Widmer, and Manjinder S Sandhu. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proceedings of the National Academy of Sciences USA*, 113(27):7377–7382, July 2016.
9. M Lynch and B Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer, 1998.
10. M Lynch. Methods for the Analysis of Comparative Data in Evolutionary Biology. *Evolution; international journal of organic evolution*, 45(5):1065–1080, August 1991.

11. Lily Wang, Bing Zhang, Russell D. Wolfinger, and Xi Chen. An integrated approach for the analysis of biological pathways using mixed models. *PLoS Genet*, 4(7):e1000115–, 07 2008.

12. Alastair J Wilson, Denis Réale, Michelle N Clements, Michael M Morrissey, Erik Postma, Craig A Walling, Loeske E B Kruuk, and Daniel H Nussey. An ecologist's guide to the animal model. *Journal Of Animal Ecology*, 79(1):13–26, 2010.

13. Lily Wang, Peilin Jia, Russell D. Wolfinger, Xi Chen, Britney L. Grayson, Thomas M. Aune, and Zhongming Zhao. An efficient hierarchical generalized linear mixed model for pathway analysis of genome-wide association studies. *Bioinformatics*, 27(5):686–692, 03 2011.

14. Arthur Korte, Bjarni J Vilhjálmsson, Vincent Segura, Alexander Platt, Quan Long, and Magnus Nordborg. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature genetics*, 44(9):1066–1071, 09 2012.

15. Omer Weissbrod, Christoph Lippert, Dan Geiger, and David Heckerman. Accurate liability estimation improves power in ascertained case-control studies. *Nature Methods*, 12:332 EP –, 02 2015.

16. Diptavo Dutta, Laura Scott, Michael Boehnke, and Seunggeun Lee. Multi-skat: General framework to test multiple phenotype associations of rare variants. *bioRxiv*, 01 2017.

17. Lorin Crawford, Ping Zeng, Sayan Mukherjee, and Xiang Zhou. Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. *PLoS Genet*, 13(7):e1006869, Jul 2017.

18. Rachel Moore, Francesco Paolo Casale, Marc Jan Bonder, Danilo Horta, Lude Franke, Inês Barroso, Oliver Stegle, and BIOS Consortium. A linear mixed model approach to study multivariate gene-environment interactions. *bioRxiv*, page 270611, 2018.

19. Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3):e1001779, 2015.

20. Jarrod D Hadfield. Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22, 2010.

21. Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

22. Kateryna Mishchenko and Maya Neytcheva. New algorithms for evaluating the log-likelihood function derivatives in the AI-REML method. *Communications in Statistics - Simulation and Computation*, 38(6):1348–1364, May 2009.

23. Christoph Lippert, Jennifer Listgarten, Ying Liu, Carl M Kadie, Robert I Davidson, and David Heckerman. FaST linear mixed models for genome-wide association studies. *Nature methods*, 8(10):833–835, October 2011.

24. Xiang Zhou and Matthew Stephens. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7):821–824, June 2012.

25. Meyer Verena Hannah, Francesco Paolo Casale, Oliver Stegle, and Ewan Birney. Limmbo: a simple, scalable approach for linear mixed models in high-dimensional genetic association studies. *bioRxiv*, 01 2018.

26. Carl M. Kadie and David Heckerman. Ludicrous speed linear mixed models for genome-wide association studies. *bioRxiv*, 01 2018.

27. Gregory Darnell, Stoyan Georgiev, Sayan Mukherjee, and Barbara E Engelhardt. Adaptive randomized dimension reduction on massive data. *Journal of Machine Learning Research*, 18(140):1–30, 2017.

28. Xiang Zhou. A unified framework for variance component estimation with summary statistics in genome-wide association studies. *The Annals of Applied Statistics*, 11(4):2027–2051, December 2017.

29. Zilong Tan, Kimberly Roche, Xiang Zhou, and Sayan Mukherjee. Scalable algorithms for learning high-dimensional linear mixed models. arXiv:1803.04431, March 2018.

30. Zhiwu Zhang, Elhan Ersoz, Chao-Qiang Lai, Rory J Todhunter, Hemant K Tiwari, Michael A Gore, Peter J Bradbury, Jianming Yu, Donna K Arnett, Jose M Ordovas, and Edward S Buckler. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4):355–360, April 2010.

31. Hyun Min Kang, Jae Hoon Sul, Susan K Service, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara Sabatti, and Eleazar Eskin. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4):348–354, April 2010.

32. Barbara Rakitsch, Christoph Lippert, Oliver Stegle, and Karsten Borgwardt. A Lasso multi-marker mixed model for association mapping with population structure correction. *Bioinformatics*, 29(2):206–214, January 2013.

33. Jae Hoon Sul, Michael Bilow, Wen-Yun Yang, Emrah Kostem, Nick Furlotte, Dan He, and Eleazar Eskin. Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. *PLOS Genetics*, 12(3):e1005849, March 2016.

34. A. R. Gilmour and NSW Agriculture. *ASREML reference manual / A R Gilmour ... [et al.]*. NSW Agriculture [Orange],

1999.
35. Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. GCTA: A Tool for Genome-wide Complex Trait Analysis. *American journal of human genetics*, 88(1):76–82, January 2011.
36. G Covarrubias-Pazaran. Genome assisted prediction of quantitative traits using the r package sommer. *PLoS ONE*, 11:1–15, 2016.
37. Miaoyan Wang, Fabrice Roux, Claudia Bartoli, Carine Huard-Chauveau, Christopher Meyer, Hana Lee, Dominique Roby, Mary Sara McPeek, and Joy Bergelson. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proceedings of the National Academy of Sciences USA*, 8:201710980, May 2018.
38. Jianming Yu, Gael Pressoir, William H Briggs, Irie Vroh Bi, Masanori Yamasaki, John F Doebley, Michael D McMullen, Brandon S Gaut, Dahlia M Nielsen, James B Holland, Stephen Kresovich, and Edward S Buckler. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2):203–208, February 2006.
39. Elizabeth D Schifano, Michael P Epstein, Lawrence F Bielak, Min A Jhun, Sharon L R Kardia, Patricia A Peyser, and Xihong Lin. SNP set association analysis for familial data. *Genet Epidemiol*, 66(5):797–810, September 2012.
40. Jennifer Listgarten, Christoph Lippert, Eun Yong Kang, Jing Xiang, Carl M Kadie, and David Heckerman. A powerful and efficient set test for genetic markers that handles confounders. *Bioinformatics*, 29(12):1526–1533, June 2013.
41. Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.
42. Jürg Schelldorfer, Peter Bühlmann, and Sara van de Geer. Estimation for high-dimensional linear mixed-effects models using $\ell 1$-penalization. *Scandinavian Journal of Statistics*, 38(2):197–214, May 2011.
43. A R Gilmour. Mixed model regression mapping for QTL detection in experimental crosses. *Computational Statistics & Data Analysis*, 51(8):3749–3764, May 2007.
44. Susanna Atwell, Yu S Huang, Bjarni J Vilhjálmsson, Glenda Willems, Matthew Horton, Yan Li, Dazhe Meng, Alexander Platt, Aaron M Tarone, Tina T Hu, Rong Jiang, N Wayan Muliyati, Xu Zhang, Muhammad Ali Amer, Ivan Baxter, Benjamin Brachi, Joanne Chory, Caroline Dean, Marilyne Debieu, Juliette de Meaux, Joseph R Ecker, Nathalie Faure, Joel M Kniskern, Jonathan D G Jones, Todd Michael, Adnane Nemri, Fabrice Roux, David E Salt, Chunlao Tang, Marco Todesco, M Brian Traw, Detlef Weigel, Paul Marjoram, Justin O Borevitz, Joy Bergelson, and Magnus Nordborg. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, 465(7298):627–631, June 2010.
45. Zulma G Vitezica, Luis Varona, and Andres Legarra. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, 195(4):1223–1230, 12 2013.
46. William Valdar, Leah C Solberg, Dominique Gauguier, Stephanie Burnett, Paul Klenerman, William O Cookson, Martin S Taylor, J Nicholas P Rawlins, Richard Mott, and Jonathan Flint. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genetics*, 38(8):879–887, 2006.
47. Elizabeth A Norgard, Joseph P Jarvis, Charles C Roseman, Taylor J Maxwell, Jane P Kenney-Hunt, Kaitlin E Samocha, L Susan Pletscher, Bing Wang, Gloria L Fawcett, Christopher J Leatherwood, Jason B Wolf, and James M Cheverud. Replication of long-bone length QTL in the F9-F10 LG,SM advanced intercross. *Mammalian genome : official journal of the International Mammalian Genome Society*, 20(4):224–235, March 2009.
48. Jin Liu, Kai Wang, Shuangge Ma, and Jian Huang. Accounting for linkage disequilibrium in genome-wide association studies: A penalized regression method. *Statistics and its interface*, 6(1):99–115, January 2013.
49. Bertrand Servin and Matthew Stephens. Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLOS Genetics*, 3(7):e114, July 2007.
50. Jon Wakefield. Bayes factors for genome-wide association studies: Comparison with P-values. *Genet Epidemiol*, 33(1):79–86, January 2009.
51. Daniel Runcie and S Mukherjee. Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. *Genetics*, 194(3):753–767, July 2013.
52. Amanda J Lea, Jenny Tung, and Xiang Zhou. A flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. *PLOS Genetics*, 11(11):e1005650, November 2015.
53. Shiquan Sun, Michelle Hood, Laura Scott, Qinke Peng, Sayan Mukherjee, Jenny Tung, and Xiang Zhou. Differential expression analysis for rnaseq using poisson mixed models. *Nucleic Acids Research*, 45(11):e106–e106, 06 2017.
54. Jennifer Listgarten, Christoph Lippert, Carl M Kadie, Robert I Davidson, Eleazar Eskin, and David Heckerman. Improved linear mixed models for genome-wide association studies. *Nature methods*, 9(6):525–526, June 2012.
55. Gustavo de Los Campos, Daniel Gianola, Guilherme J M Rosa, Kent A Weigel, and Jose Crossa. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research*, 92(4):295–308, August 2010.
56. Patrick C Phillips. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat*

Rev Genet, 9(11):855–867, 10 2008.

57. Trudy F. C. Mackay. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet*, 15(1):22–33, 01 2014.

58. Rachel E Kerwin, Julie Feusier, Alise Muok, Catherine Lin, Brandon Larson, Daniel Copeland, Jason A Corwin, Matthew J Rubin, Marta Francisco, Baohua Li, Bindu Joseph, Cynthia Weinig, and Daniel J Kliebenstein. Epistasis x environment interactions among Arabidopsis thaliana glucosinolate genes impact complex traits and fitness in the field. *New Phytol*, 215(3):1249–1263, Aug 2017.

59. Hyun Min Kang, Noah A Zaitlen, Claire M Wade, Andrew Kirby, David Heckerman, Mark J Daly, and Eleazar Eskin. Efficient control of population structure in model organism association mapping. *Genetics*, 178(3):1709–1723, March 2008.

60. Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic modeling with bayesian sparse linear mixed models. *PLOS Genetics*, 9(2):e1003264–, 02 2013.

61. Po-Ru Loh, George Tucker, Brendan K Bulik-Sullivan, Bjarni J Vilhjálmsson, Hilary K Finucane, Rany M Salem, Daniel I Chasman, Paul M Ridker, Benjamin M Neale, Bonnie Berger, Nick Patterson, and Alkes L Price. Efficient bayesian mixed model analysis increases association power in large cohorts. *Nature genetics*, 47(3):284–290, 03 2015.

62. Gerhard Moser, Sang Hong Lee, Ben J Hayes, Michael E Goddard, Naomi R Wray, and Peter M Visscher. Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genetics*, 11(4):e1004969, 04 2015.

63. Zaixiang Tang, Yueping Shen, Xinyan Zhang, and Nengjun Yi. The spike-and-slab lasso generalized linear models for prediction and associated genes detection. *Genetics*, 205(1):77, 01 2017.

64. V Guiard. About the multimodality of the likelihood function when estimating the variance components in a one-way classification by means of the ML or REML method. In *Proceedings of the International Conference on Linear Statistical Inference LINSTAT '93*, pages 139–146. Springer, Dordrecht, Dordrecht, 1994.

65. Trevor Hastie, Robert Tibshirani, and Martin Martin J Wainwright. *Statistical learning with sparsity : the lasso and generalizations*. CRC Press, Boca Raton, FL, 2015.

66. Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 67(2):301–320, April 2005.

67. Jian Yang, Noah A Zaitlen, Michael E Goddard, Peter M Visscher, and Alkes L Price. Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, 46(2):100–106, February 2014.

68. Jian Yang, Noah A Zaitlen, Michael E Goddard, Peter M Visscher, and Alkes L Price. Advantages and pitfalls in the application of mixed model association methods. *Nature genetics*, 46(2):100–106, 02 2014.

69. Lorin Crawford, Kris C Wood, Xiang Zhou, and Sayan Mukherjee. Bayesian approximate kernel regression with variable selection. *Journal of the American Statistical Association*. Available online from https://doi.org/10.1080/01621459.2017.1361830, 2018.