

A General Bayesian Approach to Analyzing Diallel Crosses of Inbred Strains

Alan B. Lenarcic,* Karen L. Svenson,† Gary A. Churchill,† and William Valdar**¹

*Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599-7265 and †The Jackson Laboratory, Bar Harbor, Maine 04609

ABSTRACT The classic diallel takes a set of parents and produces offspring from all possible mating pairs. Phenotype values among the offspring can then be related back to their respective parentage. When the parents are diploid, sexed, and inbred, the diallel can characterize aggregate effects of genetic background on a phenotype, revealing effects of strain dosage, heterosis, parent of origin, epistasis, and sex-specific versions thereof. However, its analysis is traditionally intricate, unforgiving of unplanned missing information, and highly sensitive to imbalance, making the diallel unapproachable to many geneticists. Nonetheless, imbalanced and incomplete diallels arise frequently, albeit unintentionally, as by-products of larger-scale experiments that collect F_1 data, for example, pilot studies or multiparent breeding efforts such as the Collaborative Cross or the *Arabidopsis* MAGIC lines. We present a general Bayesian model for analyzing diallel data on dioecious diploid inbred strains that cleanly decomposes the observed patterns of variation into biologically intuitive components, simultaneously models and accommodates outliers, and provides shrinkage estimates of effects that automatically incorporate uncertainty due to imbalance, missing data, and small sample size. We further present a model selection procedure for weighing evidence for or against the inclusion of those components in a predictive model. We evaluate our method through simulation and apply it to incomplete diallel data on the founders and F_1 's of the Collaborative Cross, robustly characterizing the genetic architecture of 48 phenotypes.

THE diallel is one of the oldest designs in genetics and one whose analysis is notoriously complex. The premise is simple: given a set of J parents, generate and phenotype offspring from all $J \times J$ reciprocal crosses and from these data estimate genetic parameters that characterize how the parental genomes and sex influence phenotypic variation. Using this design one can estimate the average parental contribution to the phenotype and the effect of specific combinations with other parents. When the parents are inbred strains, one can also estimate parent-of-origin effects. Despite the potential wealth of information contained in a diallel, there has been much to discourage its use in practice. Controversies about the interpretation of estimated parameters have been inextricably confounded with controversies about the analysis methods themselves, and much of the discussant literature is steeped in terminology unfamiliar to potential users. Indeed,

to the outsider, the diallel emerges as an arcane puzzle that is perhaps best avoided in favor of simpler designs.

The diallel originated in animal and plant breeding as an extension of the idea that, from a breeding perspective, you should judge the value of an individual by the phenotypes of its offspring (Christie and Shattuck 1992 and references therein). It was originally defined by Schmidt (1919) as the set of all possible J^2 pairwise crosses and was later introduced into the mainstream genetics literature by Jinks and Hayman (1953). In the decade that followed, the diallel, whose definition quickly broadened to encompass any set of F_1 crosses between $J > 2$ parents, caught the attention of an active group of quantitative geneticists who went on to develop a series of elaborations of the design and its analysis. Among the simplest and most popular analytic decompositions is that of Griffing (1956). If η_{jk} is the mean phenotype or predicted value for the cross of parent j with parent k , then the parental effects can be modeled as

$$\eta_{jk} = \mu + g_j + g_k + s_{jk}, \quad (1)$$

where μ is the intercept, g_j is the main effect of parent j , and s_{jk} is the statistical interaction of j and k , that is, the

Copyright © 2012 by the Genetics Society of America

doi: 10.1534/genetics.111.132563

Manuscript received July 11, 2011; accepted for publication October 5, 2011

Available freely online through the author-supported open access option.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.111.132563/-/DC1>.

¹Corresponding author: 120 Mason Farm Rd., Genetic Medicine Bldg., Suite 5113, Campus Box 7264, Chapel Hill, NC 27599. E-mail: william.valdar@unc.edu

deviation from the combined main effects induced by the specific pairing of parents j and k . Following the terminology introduced by Sprague and Tatum (1942) and used throughout diallel literature, g_j is the generalized combining ability (GCA) of parent j whereas s_{jk} is the specific combining ability (SCA) of the parents j and k . GCA captures aggregate effects of additive genetics whereas SCA reflects aggregate genetic effects that lead to departures from additivity, such as dominance and epistasis.

Numerous extensions to Griffing's model have been proposed to extract more subtle effects from the diallel. These include decomposing SCA into dominance, heterosis, and epistasis components (Hayman 1957; Gardner and Eberhart 1966), into reciprocal effects (Griffing 1956), their further decomposition into maternal and paternal effects (Cockerham and Weir 1977; Zhu and Weir 1996), and sex-linked variants thereof (Carbonell *et al.* 1983). Conversely, interest in obtaining GCAs with fewer than J^2 crosses has motivated variants of the design such as the half-diallel (Griffing 1956) and the partial diallel (Kempthorne and Curnow 1961), among others (see Christie and Shattuck 1992; Lynch and Walsh 1998), which have themselves led to technical innovation (*e.g.*, Greenberg *et al.* 2010).

Disagreement about the precise meaning of parameters estimated from a diallel cross has presented a theoretical stumbling block to their interpretation. The parents could be inbred lines, independent outbreeding populations (such as open-pollinating varieties of corn), or outbred individuals (Eberhart and Gardner 1966). They could have been chosen deliberately, sampled randomly, or some compromise of these. The intention of the experiment could be to draw inferences about the parents themselves, the populations each parent represents, or the broader population from which all parents were drawn. Joint consideration of such factors weaves through much of the foundational diallel literature from 1950 to 1970 and has continued to represent a source of controversy (Baker 1978; Wright 1985).

A more practical stumbling block arises from the difficulty of estimating parameters from diallel data that are incomplete, imbalanced, or contaminated with outliers. Although some diallel crosses are created deliberately, a significant number arise as a by-product of intermediate stages in a multiparent breeding program. Such accidental diallels can contain valuable information, but their often haphazard patterns of missingness make them an imperfect match for well-studied designs. For many incomplete diallels it has been unclear how to analyze the data without discarding observations, drastically reducing the scope of inference, or making other significant compromises.

Even when traditional analysis methods accommodate the design, choice about which parameters (*e.g.*, explicit models of dominance, SCA, etc.) should be included in the model can alter the estimates and interpretation of the other parameters. The option of model selection by significance testing of individual terms, frequently proposed in older literature, provides some guidance but is unsatisfactory in

that included parameters are estimated in a way that disregards uncertainty in model choice. The notion that any *a priori* plausible effect should be excluded from modeling seems to us artificial and out of step with modern approaches to applied statistical inference (*e.g.*, Gelman and Hill 2007).

We propose a general and efficient method for the analysis of diallel crosses and applying it to a data set of 48 phenotypes collected from an incomplete eight-strain diallel that arose serendipitously from the establishment of the Collaborative Cross (Churchill *et al.* 2004; Chesler *et al.* 2008; Collaborative Cross Consortium 2012). Our methods of analysis provide an inferential framework that is robust to imbalance in the design, missing data, and outliers. We model a wide range of effects including additive, heterosis, epistatic, parent-of-origin, and sex-specific variants thereof. This structure accomplishes two important goals. The first is familiar and constant interpretation of parameters across models. The second is stable and coherent estimation and prediction that is achieved through hierarchical Bayesian shrinkage and model selection.

Statistical Models and Methods

We describe our hierarchical decomposition of the diallel in stages. Starting with the simplest submodel, which describes only additive strain dosage effects, we present successive elaborations, building up to the model depicted in Figure 1 with parameters listed in Table 1. We then state the full model in compact form, detailing its efficient estimation and methods for choosing informative submodels. Models are described by a quoted string of characters in alphabetical order in the first column of Table 1 where, following marginality restrictions (Venables and Ripley 2002), some components imply the presence of others unless otherwise stated.

The “a” model

Consider an incomplete diallel arising from crosses among J inbred strains. Index the mothers with j and the fathers with k , such that any mating pair is $(j, k) \in \{1, \dots, J\}^2$. Let y_i be the phenotype value of individual i and denote the mother, the father, and the mother–father pair relevant to individual i as $j[i]$, $k[i]$, and $(j, k)[i]$. For a continuous phenotype with normally distributed errors, we model

$$y_i = \mu + \text{mother}_{j[i]} + \text{father}_{k[i]} + \varepsilon_i, \quad (2)$$

where μ is the intercept modeled as a fixed “overall” effect, mother_j and father_k are the contributions from mother and father, and $\varepsilon_i \sim N(0, \sigma^2)$ is a deviation due to the specific individual (*i.e.*, the residual). In this simplest model, each parental contribution amounts to a “dose” of the underlying strain; *i.e.*,

$$\text{mother}_j = a_j \quad (3)$$

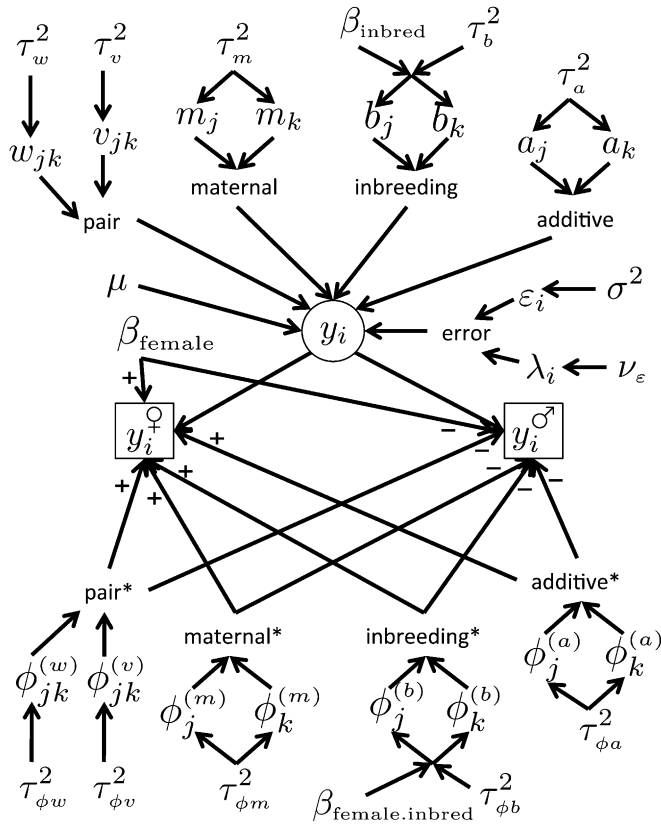


Figure 1 A directed acyclic graph depicting the hierarchy of the full model with outliers modeled at ν_ϵ t-d.f., i.e., model “B_sSa_sd_sm_sv_sw_sO_v”. Asterisks indicate components that represent sex-specific deviations from their nonasterisked counterparts, the latter representing an unsexed model.

$$\text{father}_k = a_k, \quad (4)$$

where a_j is the dose of the parental strain j (equivalently for a_k) and the effects are additive in the sense that two doses of strain j increase the expected phenotype value twice as

much as one dose. The estimated effects $\hat{\mathbf{a}}$ are used for predicting phenotypes of unobserved crosses in the incomplete design. We consider an appropriate measure of discrepancy between estimated and true values of \mathbf{a} to be the sum of squared errors (i.e., quadratic loss), $l(\hat{\mathbf{a}}, \mathbf{a}) = \sum_{j=1}^J (\hat{a}_j - a_j)^2$. In decision problems involving the estimation of three or more effects under quadratic loss, shrinkage estimators typically dominate simple means (Parmigiani and Inoue 2009). Therefore, as in Zhu and Weir (1996), we model the a_j 's hierarchically, as if drawn from a common normal distribution,

$$a_j \sim N(0, \tau_a^2), \quad (5)$$

where the variance parameter τ_a^2 is given a weakly informative prior distribution. This model leads to an estimate of each a_j that is dynamically shrunk toward zero to the extent that its supporting data are few and their dispersion τ_a^2 is estimated to be small (Gelman and Hill 2007).

In the “a” model, a_j coincides with the GCA for strain j described by Sprague and Tatum (1942). Nonetheless, we make further comparisons with GCA and SCA only parenthetically, as these constructs hinder interpretation of later models. When the “a” model is seen as an estimation problem with primary interest on \mathbf{a} , the variance of additive effects τ_a^2 does not require a biological interpretation; from a Bayesian perspective, it simply helps model how we would expect the data to appear. Nonetheless, anticipating the use of its estimate $\hat{\tau}_a^2$ as a summary statistic, we explore its relation to the concept of heritability (Lynch and Walsh 1998) in Appendix A.

Accommodating outliers

When outliers are suspected, maybe as a result of erratic measurement error, it can be desirable to model the phenotype as being sampled from a distribution with heavier tails than the normal distribution. To simultaneously accommodate and detect outliers we model the individual

Table 1 Summary of nomenclature

Model component	Description	Model variables	Implies ^a
B	Inbred penalty	β_{inbred}	
a	Strain-specific additive	$(a_1, \dots, a_J), \tau_a^2$	
b	Strain-specific inbred	$(b_1, \dots, b_J), \tau_b^2$	B
m	Strain-specific maternal	$(m_1, \dots, m_J), \tau_m^2$	
v	Strain-specific symmetric	$(v_{12}, \dots, v_{(J-1)J}), \tau_v^2$	
w	Strain-specific asymmetric	$(w_{12}, \dots, w_{(J-1)J}), \tau_w^2$	v
S	Sex effect (female advantage)	β_{female}	
B _s	Female inbred penalty	$\beta_{\text{female.inbred}}$	B, S
a _s	Strain-specific female additive	$(\phi_1^{(a)}, \dots, \phi_J^{(a)}), \tau_{\phi a}^2$	a, S
b _s	Strain-specific female dominance	$(\phi_1^{(b)}, \dots, \phi_J^{(b)}), \tau_{\phi b}^2$	b, B _s
m _s	Strain-specific female maternal	$(\phi_1^{(m)}, \dots, \phi_J^{(m)}), \tau_{\phi m}^2$	M, S
v _s	Strain-specific female symmetric	$(\phi_{12}^{(v)}, \dots, \phi_{(J-1)J}^{(v)}), \tau_{\phi v}^2$	v, S
w _s	Strain-specific female asymmetric	$(\phi_{12}^{(w)}, \dots, \phi_{(J-1)J}^{(w)}), \tau_{\phi w}^2$	w, S
O _v	Outlier model	$(\lambda_1, \dots, \lambda_n), \nu_\epsilon$	
(residual error)	Individual variation	$(\epsilon_1, \dots, \epsilon_n), \sigma^2$	

^a Components that would normally be included with this component (defined recursively).

deviations ε_i as if drawn from a scale mixture of normal densities,

$$\varepsilon_i | \lambda_i, \sigma^2 \sim N\left(0, \frac{\sigma^2}{\lambda_i}\right) \quad (6)$$

$$\lambda_i | \nu_\varepsilon \sim \frac{1}{\nu_\varepsilon} \times \chi^2(\nu_\varepsilon), \quad (7)$$

where the scale factor λ_i is modeled as $1/\nu_\varepsilon$ times a draw from a chi-square distribution with ν_ε d.f. As a result, ε_i is now t -distributed as $\varepsilon_i | \sigma^2, \nu_\varepsilon \sim t(\nu_\varepsilon, 0, \sigma^2)$. The scale λ_i for each data point has a prior mean $E(\lambda_i | \nu_\varepsilon) = 1$. *A posteriori*, λ_i acts as an indicator of the i th data point's outlier status, with $E(\lambda_i | \nu_\varepsilon, \text{Data}) \ll 1$, suggesting a highly deviant observation. Setting $\nu_\varepsilon \rightarrow \infty$ implies that residual errors closely reflect a normal distribution, while lower values of ν_ε imply increasing probability of large outliers (West 1984; Carlin and Louis 2008). For ν_ε we consider values of 15 (only slightly heavier tailed than the normal), 3 (large outliers likely; the lowest integer ν_ε for which the t has finite variance), and 6 (an intermediate value, advocated in, e.g., Greenberg *et al.* 2010).

Inbreeding and dominance effects: the “Bab” model

Hybrid vigor, or heterosis, describes the change in phenotype value due to heterozygosity when crossing two inbred lines (Lynch and Walsh 1998). It is conventional to model this effect as a dominance term that describes the deviation of hybrid individuals from the expected average of homozygous phenotypes. However, we believe the diallel is more naturally modeled in a converse manner, with inbreds, not hybrids, as the deviant type. In a full diallel, inbreds are in the minority, corresponding to $1/J$ th of the crosses. Even when the diallel is sparse, inbred crosses would seldom outnumber hybrids. We therefore define our predictive baseline as primarily modeling effects in hybrids, but accommodating inbred-specific effects through a deviation. First, Equation 2 is elaborated as

$$y_i = \mu + \text{mother}_{j[i]} + \text{father}_{k[i]} + \text{pair}_{(jk)[i]} + \varepsilon_i, \quad (8)$$

where $\text{pair}_{(jk)[i]}$ is a deviation specific to mother–father pair (jk) . In the “Bab” model, this pair effect is then modeled as a strain-specific inbred penalty b_j drawn from a common distribution centered at fixed effect β_{inbred} ; *i.e.*,

$$\text{pair}_{(jk)} = I_{\{j=k\}}(b_j + \beta_{\text{inbred}}) \quad (9)$$

$$b_j \sim N(0, \tau_b^2), \quad (10)$$

where $\tau_b^2 = 0$ implies that the inbred penalty is constant across strains. Modeling heterosis, either as an inbred or a dominance effect, inevitably changes the interpretation of a_j . In our case, a_j now estimates the dosage effect of strain j when combined with another sampled strain. Defining the additive effects this way results in more stable and precise estimates for all effects in the diallel.

Parent-of-origin effects: the “Babm” model

The full diallel includes reciprocal crosses of both mother $_j \times$ father $_k$ and father $_j \times$ mother $_k$. This allows us to estimate strain-specific effects of maternity (which could include the uterine environment for mammals, mitochondrial effects, etc.) vs. paternity. We model these parent-of-origin effects as a symmetric deviation about the “Bab” model. If m_j is the “maternal” contribution from mothers of strain j , then we revise Equations 3 and 4 to

$$\text{mother}_j = a_j + m_j$$

$$\text{father}_k = a_k - m_k$$

$$m_j \sim N(0, \tau_m^2)$$

with dispersion of maternal effects τ_m^2 .

Symmetric and asymmetric effects: “Babmvw”

Departures from the “Babm” model corresponding to statistical interactions between pairs of strains can be represented through two additional layers of effects, revising Equation 9 to

$$\text{pair}_{jk} = I_{\{j=k\}}(b_j + \beta_{\text{inbred}}) + I_{\{j \neq k\}}(v_{jk} + w_{jk}) \quad (11)$$

$$v_{jk} = v_{kj} \sim N(0, \tau_v^2) \quad (12)$$

$$w_{jk} = -w_{kj} \sim N(0, \tau_w^2). \quad (13)$$

Symmetric “v” effects model deviations that depend only on the strain labels and not how they are allocated between mother and father, whereas asymmetric “w” effects model deviations from the symmetric effect induced by differences between reciprocal crosses of the same strain pair.

Sex-specific effects

Sex-specific effects can add considerable complication to a statistical model because it can not only double the number of parameters but also change the meaning of the parameters if, for example, the effect is expressed as an offset for one sex. Here we expand our model while preserving the meaning of the existing parameters, modeling the effect of an individual being female vs. being male through a symmetric deviation about an unsexed mean. Let ψ encode femaleness,

$$\psi(\text{sex}) = \begin{cases} \frac{1}{2} & \text{if sex is female} \\ -\frac{1}{2} & \text{if sex is male,} \end{cases}$$

such that for a parameter q , adding of the term $\psi(\text{sex}) \cdot q$ to the model pushes the expected phenotype up $q/2$ for females and down $q/2$ for males. We can view females and males as modeled by Equations 14 and 15, respectively,

$$y_i^{(\text{female})} = \mu + \beta_{\text{female}} \cdot \psi(\text{female}) + \text{mother}_{j[i]}^{(\text{female})} + \text{father}_{k[i]}^{(\text{female})} + \text{pair}_{(jk)[i]}^{(\text{female})} + \varepsilon_i \quad (14)$$

$$y_i^{(\text{male})} = \mu + \beta_{\text{female}} \cdot \psi(\text{male}) + \text{mother}_{j[i]}^{(\text{male})} + \text{father}_{k[i]}^{(\text{male})} + \text{pair}_{(jk)[i]}^{(\text{male})} + \varepsilon_i \quad (15)$$

with

$$\text{mother}_j^{(\text{sex})} = a_j^{(\text{sex})} + m_j^{(\text{sex})}$$

$$\text{father}_k^{(\text{sex})} = a_k^{(\text{sex})} - m_k^{(\text{sex})}$$

$$\text{pair}_{jk}^{(\text{sex})} = I_{\{j=k\}} b_j^{(\text{sex})} + I_{\{j \neq k\}} (v_{jk}^{(\text{sex})} + w_{jk}^{(\text{sex})})$$

specified for each sex. We define for each strain-specific effect variable $q_j \in \{a_j, b_j, m_j\}$,

$$q_j^{(\text{sex})} = q_j + \psi(\text{sex}) \phi_j^{(q)},$$

and for each strain pair-specific effect $q_{jk} \in \{v_{jk}, w_{jk}\}$,

$$q_{jk}^{(\text{sex})} = q_{jk} + \psi(\text{sex}) \phi_{jk}^{(q)}.$$

Sexed effects are thus the regular unsexed effects plus a symmetric sex-specific deviation. For example, $a_j^{(\text{female})} = a_j + 0.5 \phi_j^{(a)}$. We set $f_j^{(q)} = \phi_j^{(q)}$ for $q \in \{a, m\}$ and $f_j^{(b)} = \phi_j^{(b)} + \beta_{\text{female.inbred}}$, where $\phi_j^{(q)} \sim N(0, \tau_{\phi q}^2)$ for $q \in \{a, m, b\}$; and for $q \in \{v, w\}$, $q_{jk} \sim N(0, \tau_{\phi q}^2)$ with constraints imposed as in Equations 12 and 13, where overall effects β_{female} and $\beta_{\text{female.inbred}}$ are modeled as fixed effects, and each strain-specific sex-deviation component q is modeled as random with its own variance $\tau_{\phi q}^2$.

The full model

Including fixed covariates \mathbf{x}_i and R random-effect components $u_i^{(r)} \sim N(0, \tau_r^2)$, $\forall r \in \{1, \dots, R\}$, the “full model” ($B_s S a_s b_s m_s v_s w_s$) is

$$\begin{aligned} y_i = & \underbrace{\mu + \mathbf{x}_i^T \boldsymbol{\beta}}_{\text{user fixed}} + \underbrace{\sum_{r=1}^R u_i^{(r)}}_{\text{user random}} + \underbrace{a_{j[i]} + a_{k[i]}}_{\text{additive}} + \underbrace{I_{\{j[i]=k[i]\}} (\beta_{\text{inbred}} + b_{j[i]})}_{\text{inbred penalty}} \\ & + \underbrace{m_{j[i]} - m_{k[i]}}_{\text{maternal}} + \underbrace{I_{\{j[i] \neq k[i]\}} v_{(jk)[i]}}_{\text{symmetric}} + \underbrace{I_{\{j[i] \neq k[i]\}} w_{(jk)[i]}}_{\text{asymmetric}} \\ & + \underbrace{\psi(\text{sex}_i) (\phi_{j[i]}^{(a)} + \phi_{k[i]}^{(a)})}_{\text{sex-specific additive}} + \underbrace{\psi(\text{sex}_i) I_{\{j[i]=k[i]\}} (\beta_{\text{female.inbred}} + \phi_{j[i]}^{(b)})}_{\text{sex-specific inbred penalty}} \\ & + \underbrace{\psi(\text{sex}_i) (\phi_{j[i]}^{(m)} - \phi_{k[i]}^{(m)})}_{\text{sex-specific maternal}} + \underbrace{\psi(\text{sex}_i) I_{\{j[i] \neq k[i]\}} \phi_{(jk)[i]}^{(v)}}_{\text{sex-specific symmetric}} \\ & + \underbrace{\psi(\text{sex}_i) I_{\{j[i] \neq k[i]\}} \phi_{(jk)[i]}^{(w)}}_{\text{sex-specific asymmetric}} + \varepsilon_i, \end{aligned} \quad (16)$$

with $v_\varepsilon \neq \infty$ described as the “full model with outliers” or “ $B_s S a_s b_s m_s v_s w_s O_{v_\varepsilon}$ ”.

Prior elicitation

We model prior belief about the fixed effects μ , β_{inbred} , β_{female} , and $\beta_{\text{female.inbred}}$, using diffuse normal priors $N(0, \tau_\beta^2)$. For the variance parameters, which include σ^2 and τ_q^2 parameters controlling the dispersion of strain-specific effects in the components $a, b, m, w, v, a_s, b_s, m_s, w_s, v_s$, we model prior belief using inverted chi-square distributions of the form $\tau^2 \sim \mu_\tau \times \text{Inv} - \chi^2(v_\tau)$ [equivalently described by the inverted gamma $\text{IG}(v_\tau/2, \mu_\tau/2)$], which is conjugate with our normal-likelihood and fixed-effect priors. To model vague prior information on reasonably scaled data, we apply weak priors with $\tau_\beta^2 = 10^3$ for the fixed effects and $v_\tau = 0.02$ and $\mu_\tau = 2$ for the variance parameters. The information in the variance priors is equivalent to adding 0.02 datapoints from an additional strain.

Setting shrinkage priors to beat the maximum-likelihood estimate

Within the class of unbiased regression estimators, the standard maximum-likelihood estimate $\hat{\boldsymbol{\beta}}^{\text{MLE}}$ is unbeatable (*i.e.*, admissible) when judged by mean squared error (MSE) ($E[\sum (\hat{\beta}_j - \beta_j)^2]$). However, as famously observed by Stein (1955), once the number of estimated effects exceeds 2, it is uniformly beaten (*i.e.*, dominated) by biased estimators that employ shrinkage (*e.g.*, Parmigiani and Inoue 2009). In particular, ridge regression and generalized ridge regression will dominate the maximum-likelihood estimate (MLE) under MSE loss if

$$\boldsymbol{\beta}^T [1\mathbf{Q}^{-1} + (\mathbf{X}^T \mathbf{X})^{-1}]^{-1} \boldsymbol{\beta} \leq \sigma^2, \quad (17)$$

where \mathbf{X} is a design matrix and \mathbf{Q} is any generalized ridge shrinkage matrix (using Theorem 3.19 in Gross 2003). In our case, \mathbf{X} specifies individuals’ parentage (see Appendix B), and \mathbf{Q} is a diagonal matrix with σ^2/τ_q^2 values for strain effects based upon their membership groups. A good prior for τ_q^2 is therefore one that makes \mathbf{Q}^{-1} large enough for the above condition to likely hold. Thus, for large values of β_j , one would want large τ_q^2 parameters. In this case, “flat” priors (*i.e.*, $\propto 1$, as given by $v_\tau = 2$, $\mu_\tau = 0$, after, *e.g.*, Gelman 2006, or our choice of a nonzero μ_τ) will often be preferable to the traditional Jeffreys prior [$\propto 1/\tau^2$; $\mu_\tau = 0$, $v_\tau = 0$ (Jeffreys 1946)], which includes strong prior weight around $\tau_q^2 = 0$.

Posterior estimation

We estimate posterior distributions for all parameters, using an efficient Gibbs sampling scheme. Details are provided in Appendices B and C.

Posterior inference

We obtain raw posteriors by collecting sampled values for each parameter at each MCMC iteration. However, the Gibbs

sampling scheme above gives marginal posteriors that appear overly vague for parameters that are grouped by a common variance [e.g., $\mathbf{a} = (a_1, \dots, a_j)$]. This is due to the inherent lack of identifiability in a highly parameterized mixed model. For example, a range of alternative values for \mathbf{b} , \mathbf{v} , and \mathbf{w} will produce identical predictions for y_i . Hierarchical modeling means these alternatives are distinguished by their different posterior probabilities, but less so when the posteriors of τ_b^2 , τ_v^2 , and τ_w^2 are vague. We note, however, that the purpose of the diallel experiment is most often to understand differences between strains and strain combinations, rather than estimate pure population means. Therefore, focusing on contrasts, at each iteration $t = 1, \dots, T$ we calculate recentered versions of the grouped parameters

$$\tilde{q}_j^{(t)} = q_j^{(t)} - \frac{1}{|\mathbb{J}_q|} \sum_{j' \in \mathbb{J}_q} q_{j'}^{(t)}$$

from the raw Gibbs samples, where \mathbb{J}_q denotes the group of effects assigned to the same dispersion parameter τ_q^2 , and $|\mathbb{J}_q|$ is the number of such effects. The resulting posterior intervals for \tilde{q}_j are smaller and cover their true values at the desired 95% rate. To demonstrate an alternative method for fitting mixed models that is not (explicitly) Bayesian, we also fit Equation 16 using the “hglm” procedure of Ronnegard *et al.* (2010), which gives a point estimate by finding the maximum of the hierarchical likelihood [*h*-likelihood (Lee and Nelder 1996)] suggested by our model.

Posterior prediction

Phenotype predictions for unobserved individuals that both anticipate sampling variation and incorporate all posterior uncertainty are easily obtained from MCMC output. We predict the phenotype $Y(j, k, s)$, for a future individual with mother j , father k , and sex s , by constructing the appropriate design matrix \mathbf{x}_{jks} and using the Gibbs samples to estimate the posterior predictive mean $\hat{Y}(j, k, s) = \frac{1}{T} \sum_t \mathbf{x}_{jks}^T \boldsymbol{\beta}^{(t)}$ of all draws $t \in 1, \dots, T$. Prediction intervals for $Y(j, k, s)$ are achieved from quantiles of a set of new draws $\mathbf{x}_{jks}^T \boldsymbol{\beta}^{(t)} + \sqrt{\sigma^{2(t)}} Z_t$, where $\sigma^{2(t)}$ is the t th draw of σ^2 in the Gibbs sampler, but with Z_t being a new independent draw from $N(0, 1)$ or $t(v_e)$ random noise.

Model selection by information criteria

In comparing the suitability of different submodels estimated by Gibbs sampling, we consider the popular deviance information criterion (DIC) (Spiegelhalter *et al.* 2002), defined as $2 \times \ell(\hat{\boldsymbol{\theta}}; \mathbf{y}) - 4 \times \hat{E}(\ell(\boldsymbol{\theta}; \mathbf{y}))$, where $\boldsymbol{\theta}$ collects all parameters, $\ell(\cdot | \mathbf{y})$ is the log-likelihood, and $\hat{E}(\cdot)$ is the mean over all T iterations. Lower DIC suggests a more parsimonious model containing fewer parameters that are poorly estimated.

Bayesian model selection by exclusionary Gibbs group sampling

With potentially >400 models of interest under consideration (resulting from different combinations of effects), it

is impractical to use information criteria for model selection because each model requires its own Markov chain. Furthermore, although the DIC minimizer is often a successful predictor of future $Y(j, k, s)$, it is seldom as parsimonious as the true model in simulation. In particular, our Bayesian adaptive shrinkage means that even the full model $B_s S_a s_b s_m s_v s_w$ can perform well against a model that is better informed. To deal with selection of parameter subsets $q \in 1, \dots, Q$ in a way that better identifies valuable components, we consider a zero-inflated mixture prior (George and Foster 2000; Ishwaran and Rao 2005) on τ_q^2 , namely

$$\tau_q^2 \sim \text{Bernoulli}(\pi_q) \times \mu_{\tau} \nu_{\tau} \times \text{Inv} - \chi^2(\nu_{\tau}). \quad (18)$$

This corresponds to the full model but with the elaboration that each τ_q^2 now has a prior probability $1 - \pi_q$ of being inactive, that is, equal to zero and having all corresponding effects \mathbb{J}_q equal to zero. We develop an algorithm to draw from the conditional distribution

$$\tau_q^2, \boldsymbol{\beta}_{\mathbb{J}(q)} | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)}, \mathbf{y}, \sigma^2.$$

This approach selects subsets of relevant parameters on the basis of our hierarchical model. Similarly, we model the selection of fixed effects $l \in L$ as

$$\beta_l \sim \text{Bernoulli}(\pi_l) \times N\left(0, \xi_{\beta}^2\right).$$

We calculate the posterior model inclusion probability (MIP) for each parameter subset, using the Rao-Blackwellized estimate $(1/T) \sum_t P(\tau_q^2 \neq 0 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)}, \mathbf{y})$ from the Gibbs samples (e.g., Guan and Stephens 2011). In this study, we set $\xi_{\beta}^2 = 10^2$, and $\pi_r = \pi_q = 0.5$ for all r and q . More details of the algorithm are presented in *Appendix D*.

Experimental Materials and Methods

Phenotype data from a diallel of the Collaborative Cross founders

We collected data on multiple phenotypes (Supporting Information, File S1) in a diallel of eight inbred mouse strains (abbreviated names in parentheses), A/J (AJ), C57BL/6J (B6), 129S1/SvImJ (129), NOD/LtJ (NOD), NZO/H1LtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB), which are the founder strains of the Collaborative Cross (CC) (Churchill *et al.* 2004; Chesler *et al.* 2008; Collaborative Cross Consortium 2012). It would be expected that genetic effects present in the diallel will replicate in the CC itself, which motivates our interest in this population. Table 2 lists the phenotypes collected, transformations used to normalize each phenotype before diallel analysis, and the completeness of the data for each phenotype. All CC F_1 animals had free access to standard laboratory chow containing 6% fat by weight (LabDiet 5K52) and acidified drinking water throughout the phenotyping

protocol. Mice were on a 12-hr:12-hr light:dark cycle beginning at 6:00 AM and were housed two to five animals per pen in pressurized, individually ventilated cages.

Blood composition (ADVIA)

Whole blood was obtained in the morning from the retro-orbital sinus of nonfasted animals 7 weeks of age. Collection of 200 μ l from each animal was performed using an EDTA-coated microhematocrit tube directed into a 1.5-ml microcentrifuge tube containing 2 μ l of 10% EDTA. Samples were analyzed on the Bayer Advia 120 autoanalyzer within the 4 hr following collection.

Blood pressure

Animals were acclimated to a dedicated room for blood pressure (BP) measurement on the Friday before a 5-day testing period beginning the following Monday. Systolic blood pressure and pulse were measured using the BP-2000 tail-cuff system (Visitech Systems, Apex, NC). Four unanesthetized mice were placed on a warmed platform (37°) and each was held in place using a magnetic restraining cover. The tail is placed through a cuff and held with a magnetic sensor unit that detects when blood flow stops and starts. Each day 30 measurements are obtained per mouse. Mice were trained to the apparatus for the first 3 days of the testing period and data were collected for analysis in the last 2 days, for average values based upon 60 total measurements. Animals were 10 weeks of age when blood pressure was measured.

Plasma chemistries

Whole blood was obtained from 8-week-old animals after a 4-hr period of food removal in the morning (7:00–11:00 AM). Collection of 150 μ l from each animal was performed using a heparin-coated microcapillary tube inserted into the retro-orbital sinus and directed into a 1.5-ml microcentrifuge tube containing 2 μ l of 1000 units/ml heparin. Samples were placed on ice prior to centrifugation at 10,000 rpm in a refrigerated microcentrifuge for 10 min. Plasma was collected into a clean tube for analysis on the Beckman (Fullerton, CA) CX-Delta5 Chemistry autoanalyzer.

Densitometry and body composition

Body composition was assessed when mice were 16 weeks of age by dual-energy X-ray absorptiometry (DEXA), using a Lunar PIXImus densitometer (GE Medical Systems) after mice were anesthetized intraperitoneally with tribromethanol (0.2 ml 2% solution per 10 g body weight). Because the skull is so bone dense, it is omitted from the DEXA analysis. Mice were weighed using an Ohaus Navigator scale with InCal calibration to accommodate animal movement.

Electrocardiography

Unanesthetized mice aged 12 weeks were placed on the ECGenie (Mouse Specifics, Quincy, MA) for analysis of electrocardiogram parameters. Recording is initiated when the paws of the animal contact a 3-lead electrode plate. Data are analyzed using manufacturer's software.

Simulations

We assess the performance of our methods by simulation, evaluating their ability to infer genetic parameters and to predict future phenotypes on an 8×8 diallel of these inbred strains. Two genetic architectures are considered, one simple, with additive effects only, and one more complex, with additive, inbreeding, maternal, and sex-specific maternal effects. We refer to our general Bayesian model as “BayesDiallel” and to its associated model selection procedure as “BayesSpike”. See [File S1](#).

Estimation and prediction of additive genetic effects in a simulated diallel

Gibbs sampler approaches can be difficult to compare with non-Bayesian methods or even with each other, given their indefinite approach to a point estimate. For the outlier model with low degrees of freedom, the posterior may possibly be multimodal. Furthermore, given the high dimensionality and structure of our decomposition of the diallel, not all parameters receive the same information content per complete diallel replicate, and whereas some parameters are better informed by incomplete diallels than others.

To start, we compare BayesDiallel and BayesSpike to estimates obtained from ordinary linear regression, using Griffing's model in Equation 1. We simulate an 8×8 complete diallel with five replicate individuals in each cell and assume all individuals are of the same sex. The first two columns of Table 3 (top section) list the values of the simulated parameters: eight additive strain effects (a_1, \dots, a_8), an intercept (μ), and noise variance (σ^2). Sampling from a normal distribution, these are used to generate $5 \times 64 = 320$ simulated phenotypes for the diallel. Columns 3 and 4 give MLEs and 95% confidence intervals from linear regression using Griffing's model for GCA (Equation 1 without the s_{jks} , which is directly competitive with the a model in Table 1) and for GCA + SCA (Equation 1 including s_{jks} , which is directly competitive with the “av” model, as defined in Table 1). In this setting the GCA model is at a distinct advantage: it knows *a priori* the true architecture and can thus save degrees of freedom from fitting specious parameters. The misinformed GCA + SCA model, however, risks overfitting unless the point estimates for the SCA effects happen to be small. We then consider several options for analyzing the diallel within the (unsexed) BayesDiallel framework: the full model, which we fit with and without consideration of outliers, despite the fact that outliers are not simulated; the model that minimizes the DIC among all $2^6 = 64$ unsexed models; the true model, which is a Bayesian version of the GCA; the full model fit by hglm, which maximizes the h -likelihood (see *Statistical Models and Methods*); and the component-switching BayesSpike, which attempts explicitly to learn components of value. For the BayesSpike models we report the marginal posterior median (after Meng 2008); for the hglm fit, we report the point estimate and 95% confidence interval; for all other BayesDiallel methods we report the posterior

Table 2 Phenotypes collected on a diallel of the Collaborative Cross founders

Protocol	Code	Phenotype	Units	Summary		Sample size			Filled diallel cells (of 64)		
				Mean	SD	Transform	Females	Males	Female	Male	All
ADVA	BASO	% basophil	%	0.23	0.16	x	311	315	61	61	62
ADVA	CHCM	Red cell Hgb concentration mean	g/dl	32.35	1.62	x	311	315	61	61	62
ADVA	CHGB	Calculated Hgb	g/dl	15.01	0.72	x	311	315	61	61	62
ADVA	EOS	% Eosinophil	%	2.80	2.20	x	311	315	61	61	62
ADVA	HCT	Hematocrit	%	46.47	2.60	x	311	315	61	61	62
ADVA	HDW	Hgb concentration distribution width	g/dl	1.93	0.39	x	311	315	61	61	62
ADVA	LUC	% Large unstained cell	%	0.54	0.26	x	311	315	61	61	62
ADVA	LYM	% lymphocyte	%	82.97	6.06	x	311	315	61	61	62
ADVA	MCH	Mean cell hemoglobin content	pg	15.60	0.71	x	311	315	61	61	62
ADVA	MCHC	Mean cell hemoglobin concentration	g/dl	33.67	1.22	x	311	315	61	61	62
ADVA	MCV	Mean cell volume	fl	46.37	2.19	x	311	315	61	61	62
ADVA	mHGB	Measured Hgb	g/dl	15.63	0.68	x	311	315	61	61	62
ADVA	MONO	% Monocyte	%	1.46	3.80	x	311	315	61	61	62
ADVA	MPV	Mean platelet volume	fl	5.81	1.83	x ⁻²	311	315	61	61	62
ADVA	NEUT	% neutrophil	%	12.00	3.82	log(x)	311	315	61	61	62
ADVA	PLT	Platelet count	x10E03 cells/liter	1132.25	196.07	log(x)	311	315	61	61	62
ADVA	RBC	Red blood cell count	x10E06 cells/liter	10.03	0.54	x	311	315	61	61	62
ADVA	RDW	Red cell distribution width	%	14.74	1.30	x	311	315	61	61	62
ADVA	Retic	% reticulocyte	%	2.79	1.15	log(x)	311	315	61	61	62
ADVA	WBC	White blood cell count	x10E03 cells/liter	7.01	1.91	x	311	315	61	61	62
BP	PulseMean	Pulse rate	Beats/min	577.35	99.47	x	89	99	20	24	24
BP	SystolicMean	Systolic blood pressure	mmHg	122.96	26.07	x	89	99	20	24	24
CHEM	CHOL	Total cholesterol	mg/dl	105.47	30.59	log(x)	320	311	62	61	62
CHEM	GLU	Glucose	mg/dl	185.86	33.42	x	320	311	62	61	15
CHEM	HDL	HDL Cholesterol (as HDLD)	mg/dl	8.21	7.39	x ^{1/4}	320	311	62	61	62
CHEM	TG	Triglycerides	mg/dl	137.24	59.38	log(x)	320	311	62	61	62
EKG	CV	Coefficient of variance	%	2.68	3.04	x ^{-1/2}	199	195	42	39	42
EKG	HR	Heart rate	Beats/min	768.10	65.33	x/100	199	195	42	39	42
EKG	HRV	Heart rate variability	Beats/min	19.51	20.51	x ^{-1/2}	199	195	42	39	42
EKG	Ramp1	Mean R amplitude	msec	368.33	104.68	x/100	199	195	42	39	42
EKG	SRAmpl	Mean SR amplitude	msec	416.10	165.92	x/100	199	195	42	39	42
EKG	PQ	PQ interval	msec	21.58	3.04	x	199	195	42	39	42
EKG	PR	Interval between peak of P-wave and R-wave	msec	28.56	3.49	x	199	195	42	39	42
EKG	QRS	Interval between start and end of QRS complex	msec	10.52	1.00	x	199	195	42	39	42
EKG	QT	QT interval	msec	41.61	4.55	x	199	195	42	39	42
EKG	QTD	Difference between smallest and largest QT interval	msec	20.16	8.44	x	199	195	42	39	42
EKG	QTC	Rate corrected QT interval	msec	46.85	3.47	x	199	195	42	39	42
EKG	QTCd	Rate corrected QT dispersion	msec	22.31	8.72	x	199	195	42	39	42
EKG	RR	R to R interval	msec	78.93	8.08	x	199	195	42	39	42
DEXA	B.Area	Bone area S	cm ²	5.20	1.00	x	292	302	61	61	62
DEXA	BMC	Bone mineral content	g/cm	0.27	0.06	1/x	292	302	61	61	62
DEXA	BMD	Bone mineral density	g/cm ²	0.06	0.01	x	292	302	61	61	62
DEXA	LTM	Lean tissue mass	g	9.57	3.08	x	292	302	61	61	62
DEXA	RST	R-value of soft tissue	Unitless	1.29	0.01	x	292	302	61	61	62

(continued)

Table 2. continued

Protocol	Code	Phenotype	Units	Summary		Transform	Sample size		Filled diallel cells (of 64)		
				Mean	SD		Females	Males	Female	Male	All
DEXA	TTM	Total tissue mass	g	12.82	5.22	x	292	302	61	61	62
COMP	MLNA	Mouse length,nose to anus	cm	26.71	8.89	log(x)	292	302	61	61	62
COMP	Weight	Mouseweight	g	26.71	8.89	log(x)	292	302	61	61	62
COMP	PctFat	PctFat	%	19.47	7.26	log(x)	292	302	61	61	62

ADVIA, blood composition; BP, blood pressure; CHEM, plasma chemistries; EKG, electrocardiography; DEXA, dual-energy X-ray absorptiometry; COMP, body composition.

intervals, calculated as the central 95% quantiles of the posterior distribution (Carlin and Louis 2008). For the Bayesian methods, we consider their ability to select out parameters either through severe shrinking or through formal selection. In the “top false” row in Table 3, we report for the BayesSpike the largest MIP of untrue fixed and random components, assuming a (generous) prior MIP of 0.5 per component; for the remaining Bayesian methods, we report the estimated effect for the largest untrue fixed component and the τ_q^2 estimate for the largest untrue random component q .

Diallel analysis should (at minimum) help the researcher predict phenotypes of new individuals from sampled crosses. We generate 1000 new simulated diallels on the basis of the same genetic architecture (*i.e.*, using the parameters in Table 3, column 2). Separately we use the point estimates from each method to predict the expected phenotype of the new individuals in each cell and compare these predictions with each of the 1000 subsequently observed data sets. The bottom row in Table 3 compares predicted with observed values, reporting the MSE as the average over simulations (upper row) and as its 95% central quantile (lower row).

Although μ tends to have a wide confidence/prediction interval relative to the additive effects, Table 3 shows that the BayesDiallel models can meet or improve on the MLE for GCA in terms of prediction error and point estimates, despite the fact that the full model has 82 parameters. Against the GCA + SCA MLE, however, any method using hierarchical shrinkage is twice as successful in forecasting new phenotypes. This advantage to hierarchical models would reduce with smaller σ^2 . The symmetric and asymmetric effects (v and w) tend to be the false components most likely to enter the model when using BayesSpike, although they are typically estimated with a much smaller magnitude than the additive effect.

Table 4 shows results from 100 simulation experiments, with 1000 test data sets per experiment. Values in this table measure the mean discrepancy in estimated effects and prediction MSE. Noise level, $\sigma^2 = 120$, represents a lower limit on the performance of any estimator. We see that when the model is overspecified, as in the GCA + SCA model, lack of shrinkage severely affects the consistency of the MLE.

Inferring a complex genetic architecture: a “BSabm_s” model

To investigate the performance of the Bayesian model when many effects are present, we simulated a more complex genetic architecture that included sex (“S”), maternal (“m”), sex-specific maternal (“m_s”), and inbreeding (“Bb”). In this case the nonzero effects were $\mathbf{a} = (-7.21, -5.77, -2.88, -0.72, 0.72, 2.16, 5.05, 8.65)$, $\mathbf{b} = (4.12, -3.88, -3.88, 2.12, 3.12, -2.88, 0.12, 1.12)$, $\mathbf{m} = (1, 1.63, -2.54, 7.24, 9.76, -14.18, 0.19, -3.08)$, $\phi^{(m)} = (3.75, 3.25, 4.25, -11.75, 0.75, -0.25, -4.75, 4.75)$, $\mu = 10$, $\beta_{\text{female}} = 4$,

Table 3 Performance of traditional and Bayesian models in inference and prediction for a simulated complete eight-way diallel with five replicates and additive effects

Name	True value	BayesDiallel														
		Griffing linear model					Standard normal model					Outlier model				
		GCA: "a"	GCA + SCA: "av"	Full model: "Bambvw"	Minimum DIC model: "aww"	True model: "a"	Full hglm fit: "Bambvw"	BayesSpike:	d.f. = 3: "Bambvw"	d.f. = 6: "Bambvw"	d.f. = 15: "Bambvw"	BayesSpike: d.f. = 6:				
μ	7	7.52 ^b (-4.94, 19.97)	7.52 ^b (-90.92, 105.95)	7.23 ^c (-2.72, 18.11)	7.35 ^c (-3.65, 17.5)	7.28 ^c (-3.46, 17.95)	7.44 (-2.51, 17.39)	7.32 ^d (-2.6, 17.53)	7.41 ^c (-3.91, 17.53)	7.4 ^c (-3.71, 18)	7.25 ^c (-2.84, 17.39)	7.49 ^d (-2.14, 16.91)				
a_1	-10	-9.4 (-11.61, -7.2)	-13.31 (-17.8, -8.82)	-9.12 (-11.43, -6.86)	-9.07 (-11.18, -6.77)	-9.15 (-11.5, -7.12)	-9.11 (-14.6, -3.63)	-9.05 (-11.28, -6.82)	-9.37 (-11.75, -7.13)	-9.3 (-11.68, -7.15)	-9.2 (-11.52, -6.81)	-9.69 (-11.5, -7.01)				
a_2	-8	-7.49 (-9.69, -5.29)	-7.36 (-11.85, -2.86)	-7.12 (-9.5, -4.77)	-6.97 (-9.37, -4.7)	-7.24 (-9.3, -5.14)	-6.9 (-12.38, -1.41)	-7.09 (-9.38, -4.81)	-6.64 (-8.88, -4.34)	-6.88 (-9.2, -4.62)	-6.96 (-9.1, -4.44)	-6.73 (-9.11, -4.48)				
a_3	-4	-5.26 (-7.46, -3.06)	-7.86 (-12.35, -3.37)	-5.05 (-7.21, -2.76)	-5 (-7.1, -2.72)	-5.11 (-7.26, -2.98)	-5 (-10.48, 0.49)	-5.04 (-7.4, -2.99)	-5.48 (-7.96, -3.22)	-5.3 (-7.56, -3.06)	-5.15 (-7.68, -2.9)	-5.28 (-7.57, -2.94)				
a_4	-1	-1.27 (-3.47, 0.93)	-5.9 (-10.39, -1.41)	-1.25 (-3.52, 1)	-1.33 (-3.63, 0.86)	-1.23 (-3.3, 0.98)	-1.33 (-6.81, 4.16)	-1.26 (-3.43, 0.97)	-1.25 (-3.55, 1.11)	-1.24 (-3.53, 0.99)	-1.25 (-3.61, 1.03)	-1.26 (-3.42, 0.96)				
a_5	1	2.05 (-0.15, 4.26)	-1.7 (-6.2, 2.79)	2 (-0.27, 4.28)	1.99 (-0.26, 4.28)	1.98 (-0.23, 4.14)	1.96 (-3.53, 7.44)	1.97 (-0.23, 4.21)	1.8 (-0.6, 4.14)	1.98 (-0.21, 4.31)	1.99 (-0.25, 4.41)	1.95 (-0.31, 4.25)				
a_6	3	2.24 (0.03, 4.44)	-4.62 (-9.11, -0.13)	2.07 (-0.28, 4.3)	1.94 (-0.37, 4.22)	2.17 (0.14, 4.4)	1.85 (-3.63, 7.34)	2.03 (-0.21, 4.33)	3.21 (0.58, 5.74)	2.72 (0.18, 5.18)	2.27 (-0.21, 4.68)	2.6 (0.01, 4.83)				
a_7	7	7.86 (5.66, 10.06)	5.46 (0.97, 9.95)	7.61 (5.22, 9.79)	7.65 (5.41, 9.9)	7.63 (5.61, 9.76)	7.69 (2.2, 13.17)	7.59 (5.22, 9.73)	6.95 (4.56, 9.29)	7.21 (4.8, 9.5)	7.41 (4.9, 9.6)	7.16 (4.86, 9.47)				
a_8	12	11.27 (9.07, 13.47)	7.27 (2.78, 11.76)	10.86 (8.57, 13.21)	10.79 (8.39, 12.97)	10.96 (8.73, 13)	10.83 (5.35, 16.32)	10.85 (8.52, 13.07)	10.79 (8.38, 12.92)	10.82 (8.48, 13.1)	10.89 (8.56, 13.21)	10.79 (8.39, 12.98)				
σ^2	120	107.67 (89.16, 126.17)	105.09 (86.05, 124.13)	102.51 (86.38, 118.92)	102.28 (85.76, 119.98)	108.4 (90.96, 125.84)	99.9 ^h —	103.38 (86.16, 121.14)	63.56 (50.44, 77.27)	78.89 (64.87, 95.92)	91.71 (75.84, 107.49)	79.89 (66.35, 95.48)				
Top false	Fixed ^e	—	—	B (0.381)	—	—	B (0.376)	B (0.247)	B (0.233)	B (0.296)	B (0.34)	B (0.268)				
	Random ^f	—	—	m (0.939)	—	—	v (0.125)	m (0.832)	m (1.122)	m (1.047)	m (0.976)	m (0.87)				
Predict MSE	119.66 ^g	122.5 (84.14, 166.45)	324.5 (243.59, 412.82)	124.7 (85.13, 169.98)	124.6 (85.35, 170.9)	122.5 (84.45, 166.14)	126.3 (86.96, 172.43)	124.3 (84.62, 169.96)	125.9 (85.89, 172.72)	125.2 (85.77, 171.32)	124.8 (85.42, 170.53)	124.9 (85.39, 170.81)				

^a "aww" is the model that minimized the DIC.
^b Least-squares mean with 95% confidence interval in parentheses.
^c Posterior mean with 95% posterior interval in parentheses.
^d Posterior median with 95% posterior interval in parentheses.
^e For BayesDiallel component $q \in \{\beta_{\text{inbred}}, \dots, \tau_{\text{bwd}}, \gamma(\lambda)\}$ denotes $p(q \neq 0 \mid \text{Data})$ for BayesSpike and $\min\{p(\beta > 0 \mid \text{Data}), p(\beta < 0 \mid \text{Data})\}$ otherwise.
^f Magnitude of the falsely included random-effects vector \mathbf{q} averaged over T MCMC samples, as $T^{-1}[(\mathcal{J}-1)^{-1} \sum_j (q_j^{(i)} - q^{(i)})^2]^{1/2}$.
^g Mean and 95% confidence interval of MSE from 1000 simulations (see main text).
^h Confidence intervals were not available for the residual variance.

Table 4 Performance of traditional and Bayesian models in inference and prediction for 100 simulated complete eight-way diallels with five replicates and additive effects

Simulated component	BayesDiallel											
	Griffing linear model		Standard normal model		Outlier model							
Name	Discrepancy	GCA: "a"	GCA + SCA: "av"	Full model: "Babmw"	Minimum DIC model: —	True model: "a"	Full hglm fit: "Babmw"	BayesSpike: —	Full model			BayesSpike: d.f. = 6: —
									d.f. = 3: "Babmw"	d.f. = 6: "Babmw"	d.f. = 15: "Babmw"	
Additive (a)	$\sum (\hat{a}_j - a_j)^2$	10.85 (2.5, 23.6)	144.56 (57, 266.2)	10.69 (2.7, 22.9)	10.69 (2.7, 23)	10.67 (2.8, 22.7)	10.74 (2.7, 24.7)	10.73 (2.8, 23.2)	12.02 (3.4, 28.3)	11.15 (2.8, 24.4)	10.77 (2.6, 23.2)	11.20 (2.8, 24.7)
Inbreeding (b)	$\sum (\hat{b}_j - b_j)^2$	0.00	0.00	4.60 (0.3, 29.1)	2.88 (0, 24.3)	0.00	18.47 (0, 147.5)	0.04 (0, 0.1)	8.19 (0.4, 55)	5.96 (0.3, 31.3)	4.87 (0.3, 27.9)	0.19 (0, 0.1)
Maternal (m)	$\sum (\hat{m}_j - m_j)^2$	0.00	0.00	1.37 (0.1, 5)	0.91 (0, 5.9)	0.00	0.78 (0, 7.1)	0.22 (0, 3)	1.73 (0.2, 6.8)	1.53 (0.2, 6)	1.41 (0.1, 5.8)	0.27 (0, 3.7)
Symmetric (v)	$\sum (\hat{v}_{jk} - v_{jk})^2$	0.00	2126.47 (860.2, 3899.1)	5.50 (0.9, 23.1)	3.69 (0, 24.5)	0.00	13.13 (0.1, 64.7)	0.26 (0, 0.8)	12.09 (1.3, 70.1)	7.76 (0.9, 37.1)	5.98 (0.9, 25.3)	0.91 (0, 10.3)
Prediction MSE ^a		123.09 (120.5, 125.8)	311.82 (306.7, 316.7)	124.32 (121.7, 127)	123.79 (121.2, 126.5)	123.09 (120.6, 125.8)	124.78 (122.2, 127.4)	123.43 (120.9, 126.2)	125.24 (122.6, 128)	124.65 (122, 127.4)	124.39 (121.8, 127.1)	123.69 (121.2, 126.5)

^a Mean and 95% confidence interval of MSEs from 50 independent simulation experiments.

$\beta_{\text{inbred}} = -4$, and $\sigma^2 = 120$. These were chosen so that ~20% of variation was explained by the “a” component, 20% from “m”, 12% from “m_s”, and 0.5% from “b”. We generated 295 simulated data sets of 8×8 diallels with five replicates of each sex (640 individuals per data set). To each diallel we applied all of the methods tested above, except model selection using DIC (which was intractable for multiple replications when the number of models was 400+ in sex-effect models). Table 5 provides a summary of prediction and estimation error analogous to Table 4. In this case, we see that all of the BayesDiallel models are able to adapt to the active components, producing acceptable prediction error. Predictions are weaker for strain-specific “b” and “m_s” effects, reflecting the smaller amount of data that inform them. Griffing’s models are shown for comparison, but as expected, perform poorly in this realm.

Figure 2 plots the results from fitting the Bayesian model to 50 of the simulated diallels described in Table 5. The black line in Figure 2, A and B, describes the set of parameters used for the simulation, plotting for each component $q \in \{Q, L\}$ (i.e., among the Q random components or R fixed-effect components) the parameter value β_q if it is a fixed effect or the SD of its vector \mathbf{q} if it is a random effect. In Figure 2A, each colored line summarizes estimates from the full model applied to one simulated data set, plotting $E(\beta_q | \text{Data})$ and $\text{SD}(E(\mathbf{q} | \text{Data}))$, as appropriate. Figure 2B does the same for the BayesSpike, using $\text{median}(\beta_q | \text{Data})$ and $\text{SD}(\text{median}(\mathbf{q} | \text{Data}))$. Figure 2, A and B, shows that whereas the full model shrinks spurious components, BayesSpike forces them to zero. Figure 2C plots the posterior MIP, $\hat{\pi}_q$, estimated by BayesSpike starting with a component-wise prior of $\pi_{q0} = 0.5$, with the black line now indicating the median. Figure 2D shows MIPs for BayesSpike allowing for outliers at 6 d.f. and appears similar to Figure 2C. However, Figure 2 E and F, which plot the same data in a different way, reveal an important difference. They show the \log_{10} of the Bayes factor, calculated as $\log_{10}(\hat{\pi}_q \pi_{q0}) - \log_{10}[(1 - \hat{\pi}_q)(1 - \pi_{q0})]$, which tracks the displacement of the posterior from the prior and thereby weighs the evidence provided by the data for or against a component’s inclusion (Kass and Raftery 1995; Bernardo and Smith 2000). Accommodating outliers (Figure 2F) weakens the Bayes factors for inclusion of true components, illustrating the trade-off between increased robustness and reduced power.

Application to Data

We apply our Bayesian models and BayesSpike procedure to data on 48 phenotypes, collected on mice from a diallel of founders of the Collaborative Cross. We start with an analysis of mouse weight, for which we were able to collect almost a full diallel. We then describe our semiautomated analysis of 48 phenotypes, commenting on select examples that demonstrate robustness and stability in the presence of sparsely sampled data, and richly characterize genetic and parent-of-origin architecture when data are abundant.

Table 5 Performance of traditional and Bayesian models in inference and prediction for 295 simulated "Bsbm_s" models

Name	Discrepancy	Griffing linear model			Standard normal model			BayesDiallel		
		GCA + SCA			True model:			Full model		
		GCA: "a"	GCA + SCA: "av"	"B _s Sa _s b _s m _s v _s w _s "	"B _s Sa _s b _s m _s v _s w _s "	BayesSpike:	Full hglm fit:	BayesSpike:	"B _s Sa _s b _s m _s v _s w _s O ₃ "	"B _s Sa _s b _s m _s v _s w _s O ₆ "
Additive (a)	$\sum(\hat{a}_i - a_i)^2$	9.42 (3.2, 18.3)	288.41 (177.5, 406.3)	7.25 (2, 15.3)	6.73 (2, 14.7)	7.31 (2, 15.3)	6.73 (2, 14.7)	8.20 (2, 17.4)	7.60 (1.8, 16.1)	7.36 (1.8, 15.5)
Inbreeding (b)	$\sum(\hat{b}_i - b_i)^2$	255.86	— ^b	113.20 (28.7, 216.6)	91.21 (23.9, 202.1)	144.65 (31.8, 255.9)	118.66 (30.5, 221.7)	118.66 (30.5, 221.7)	116.03 (30.4, 218.5)	113.85 (27.6, 218.3)
Maternal (m)	$\sum(\hat{m}_i - m_i)^2$	92.10	— ^b	5.45 (1.2, 11.9)	5.33 (1.1, 11.8)	5.43 (1.2, 12.1)	5.33 (1.1, 12.3)	6.05 (1.5, 12.3)	5.63 (1.4, 12.3)	5.47 (1.3, 11.8)
Sex-specific maternal (m _s)	$\sum(\hat{\phi}_i^{(m)} - \phi_i^{(m)})^2$	326.16	— ^b	20.35 (5.2, 47.2)	20.14 (5, 46.6)	20.46 (5.4, 48.8)	22.69 (5, 51.8)	22.69 (5, 51.8)	21.09 (6.2, 49.5)	20.46 (6.1, 49.8)
Prediction MSE ^a		173.68 (171.4, 175.9)	326.71 (323.2, 330)	126.60 (124.8, 128.3)	126.79 (125, 128.6)	131.72 (130, 133.5)	127.71 (125, 129.5)	126.99 (125.3, 128.7)	126.68 (124.9, 128.5)	132.03 (130.3, 133.8)

^a Mean and 95% confidence interval of MSEs from 300 independent simulation experiments.

^b These parameters are not explicitly modeled by the tested procedure.

Analysis of weight data in a diallel of the Collaborative Cross founders

Body weight data were obtained for 292 female and 302 male mice (Figure 3). Shaded cells in Figure 3, A and B, represent observed phenotypes with the degree of shading indicating the average weight of mice in each group (log₁₀ scale). Crossed boxes indicate the absence of phenotyped animals. Although mostly complete, this diallel is imbalanced: cells contained between 2 and 20 mice, with 1–13 male and 1–7 female. Visual inspection of this diallel reveals some striking trends. The dark banding in column and row 5 shows that the genome of the "New Zealand Obese" (NZO) mouse exerts a strong weight-inducing effect on its progeny (Taylor *et al.* 2001). The asymmetry of the banding, however, suggests this effect is transmitted more strongly from the mother than from the father. A similar asymmetry is apparent for strain AJ (row and column 1). Moreover, Figure 3, A and B, shows means only, unmoderated by shrinkage effects that would factor in the different numbers of mice that contribute to those estimates.

We applied the full model with outliers (B_sSa_sb_sm_sv_sw_sO₆) to the body weight data. Highest posterior density (HPD) (Box and Tiao 1973) intervals of 163 effects parameters based on 8000 posterior samples from four independent Monte Carlo Markov chains are shown in Figure 4. Parameters are divided into four groups: general effects, which include the inbreeding penalty ("inbreed.overall"; B) and strain-specific effects of additive genetics ("additive"; a), inbreeding ("inbreed"; b), and parent-of-origin effects ("maternal"; m); strain pair-specific effects, which encompass effects peculiar to specific strain pairs (v and w); sex-specific effects, which include sex-specific deviations of the general effects (S, B_s, a_s, b_s, m_s); and sex/strain pair-specific effects, encompassing sex-specific deviations from the strain pair-specific effects (v_s, w_s). The general effects clearly show the high additive dosage effect of NZO ("additive:NZO"), plus some evidence for mothers transmitting this effect more strongly ("maternal:NZO"). A more striking parent-of-origin effect is evident in CAST ("maternal: CAST"), which has its 95% HPD most displaced from zero and indicates CAST mothers transmit low body weight more strongly than CAST fathers. The sex-specific effects include an expected drop in weight for females ("female.overall") but few other strong deviations from zero. The strain pair-specific and sex/strain pair-specific effects are typically more vague. They represent fewer observations and so are more strongly subject to Bayesian adaptive shrinkage, which pulls extreme but sparsely supported means toward the middle. Figure 3, C and D, shows means of the (posterior) predictive distribution: that is, the average value of mice in a new diallel of the same strains that would be expected on the basis of the diallel model and the observed data. These Bayesian predictions incorporate all uncertainty due to finite sampling of the data and prior uncertainty about the parameters.

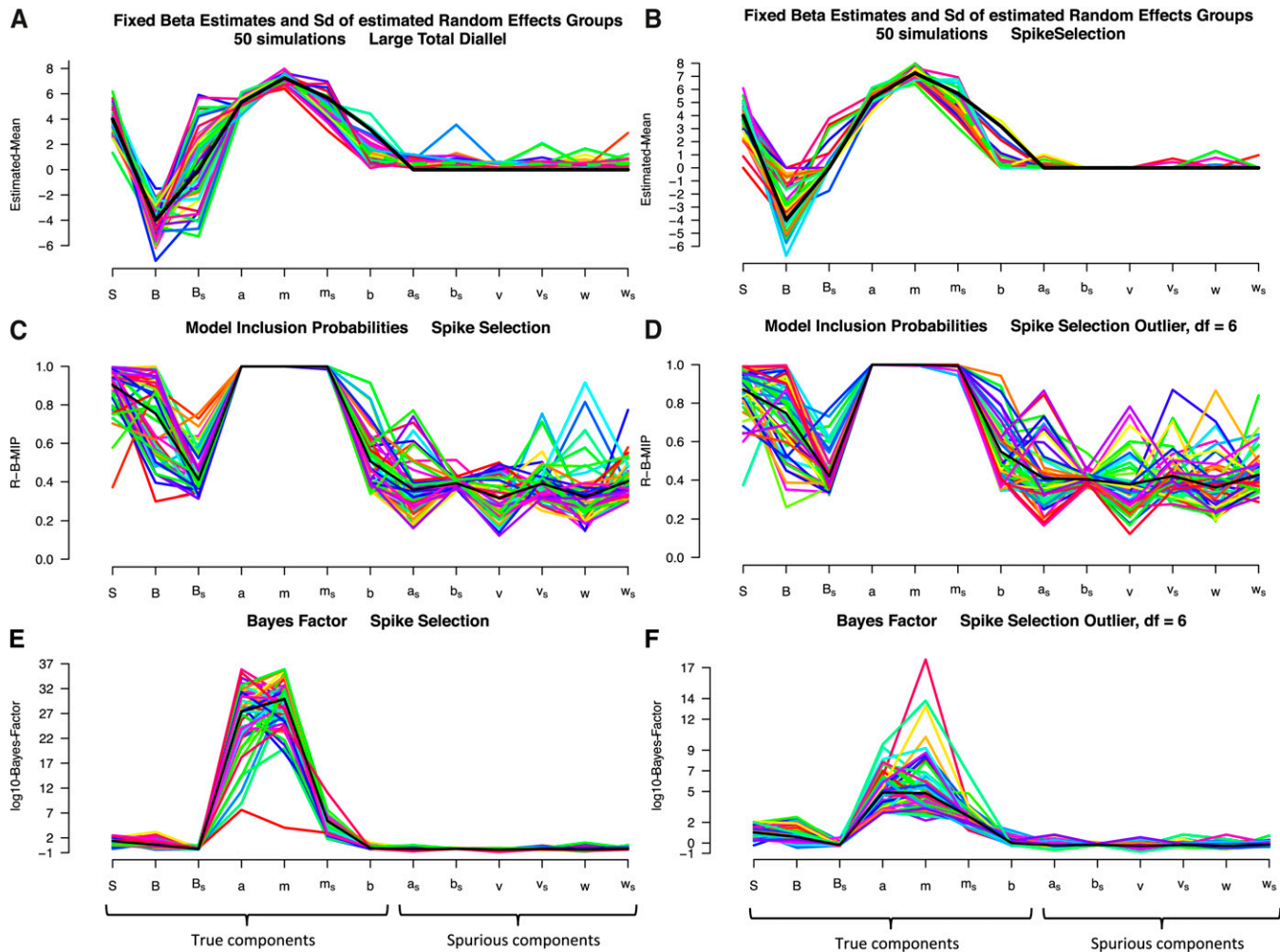


Figure 2 (A–F) Summary statistics, posterior model inclusion probabilities, and Bayes factors for Bayesian models applied to 50 simulated diallels that share a complex genetic and parent-of-origin architecture. “True components” are group effects that were simulated; “spurious components” are group effects that were absent. Each colored line depicts results from a single simulation, with A and B showing the posterior mean for fixed-effect components (S, B, B_s) and the SD of sampled random intercepts for strain- and strain pair-specific components.

Semiautomated analysis of 48 phenotypes in a diallel of the CC founders

We applied the BayesSpike procedure for automated selection of diallel components to all 48 of the phenotypes in Table 2. Figure 5 lists the posterior MIPs for each diallel component, assuming a prior MIP in each case of $\frac{1}{2}$. The “Info” column provides a measure between 0 and 1 that describes how much information the data have provided about model choice, defined by a scaled Kullback–Leibler divergence

$$\text{Info} = c_Q \sum_{q \in Q} c_q \left[\pi_q \log \left(\frac{\pi_q}{\pi_{q0}} \right) + (1 - \pi_q) \log \left(\frac{1 - \pi_q}{1 - \pi_{q0}} \right) \right],$$

where π_{q0} and π_q are the prior and posterior MIPs for component $q \in \{Q, L\}$ (where Q is the set of all effects grouped by a common variance, and L is the set of fixed effects), $c_Q = (2^{|Q|} - 1)^{-1}$, $c_q = (-\log[\min(\pi_{q0}, 1 - \pi_{q0})])^{-1}$, and $||Q, R||_0 = 13$ is the number of components considered in the selection. For each phenotype, the posterior MIPs

(rounded to 2 decimal places) are based on 10,000 samples from five independent Markov chains.

The numbers given for each effect in Figure 5 indicate how strongly, and in what direction, the data shift opinion about which components should be included. For example, the MIP values for systolic blood pressure (SystolicMean) are mostly near 0.5, indicating little posterior certainty and reflecting the fact that only 188 mice had measurements for this phenotype and these encompassed only 24 of 64 diallel combinations (Figure 6). Yet despite the sparsity of this data set, the BayesSpike returns stable, if vague, posterior opinion about inclusions, and the full model, fitted to the weight data above, provides stable, if vague, posterior distributions (Figure 7) for 175 effects and variance parameters, 188 outlier parameters, 2×64 predicted new crosses, and any further combination of parameters that is of interest. In particular, the data set contains no inbreds, which means that posterior information about inbreeding parameters (Figure 7, rightmost column) is almost as diffuse as the

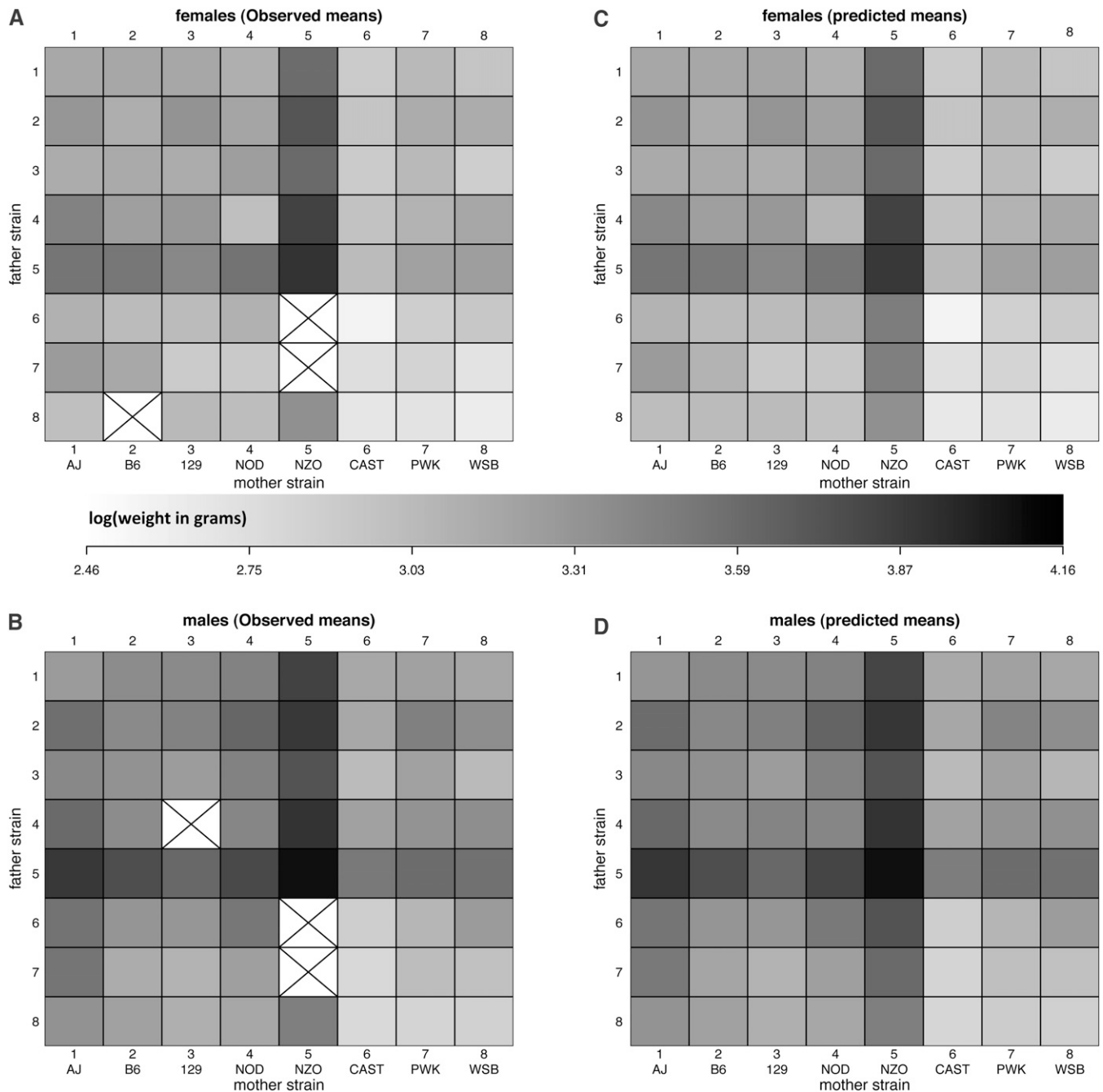


Figure 3 Weight data for 292 female and 302 male mice in an incomplete diallel of the Collaborative Cross founders. Shaded boxes indicate the average weight (on the log scale) of female (A) and male (B) mice, with crossed boxes showing missing data. C and D show the posterior predictive means after applying the full diallel model with outliers at 6 d.f.

original priors, and BayesSpike probabilities indicate the absence of evidence for or against their inclusion. In contrast, HDL cholesterol (HDL), for which there are considerably more data (631 animals with 62/64 cells covered), provides strong evidence for an overall effect of sex (“S”), strain-specific effects of additive (“a”) and dominance (as inbreeding; “b”), and symmetric effects, *i.e.*, effects that are specific to F_1 ’s between particular pairs of strains but for which the mother/father assignment does not matter. It

also shows strong evidence against any sex-specific effects (*e.g.*, sex-specific additive effects “ a_s ”) beyond that explained by an overall shift in mean.

Strong evidence for a genetic effect need not imply that the effect exerts strong influence on the phenotype. Figure 8 plots strain-specific effects for some of the phenotypes listed in Figure 5. In this “straw plot”, each colored line tracks the posterior mean for the relevant CC founder strain, with predicted means of inbreds given in the bottom two rows (these

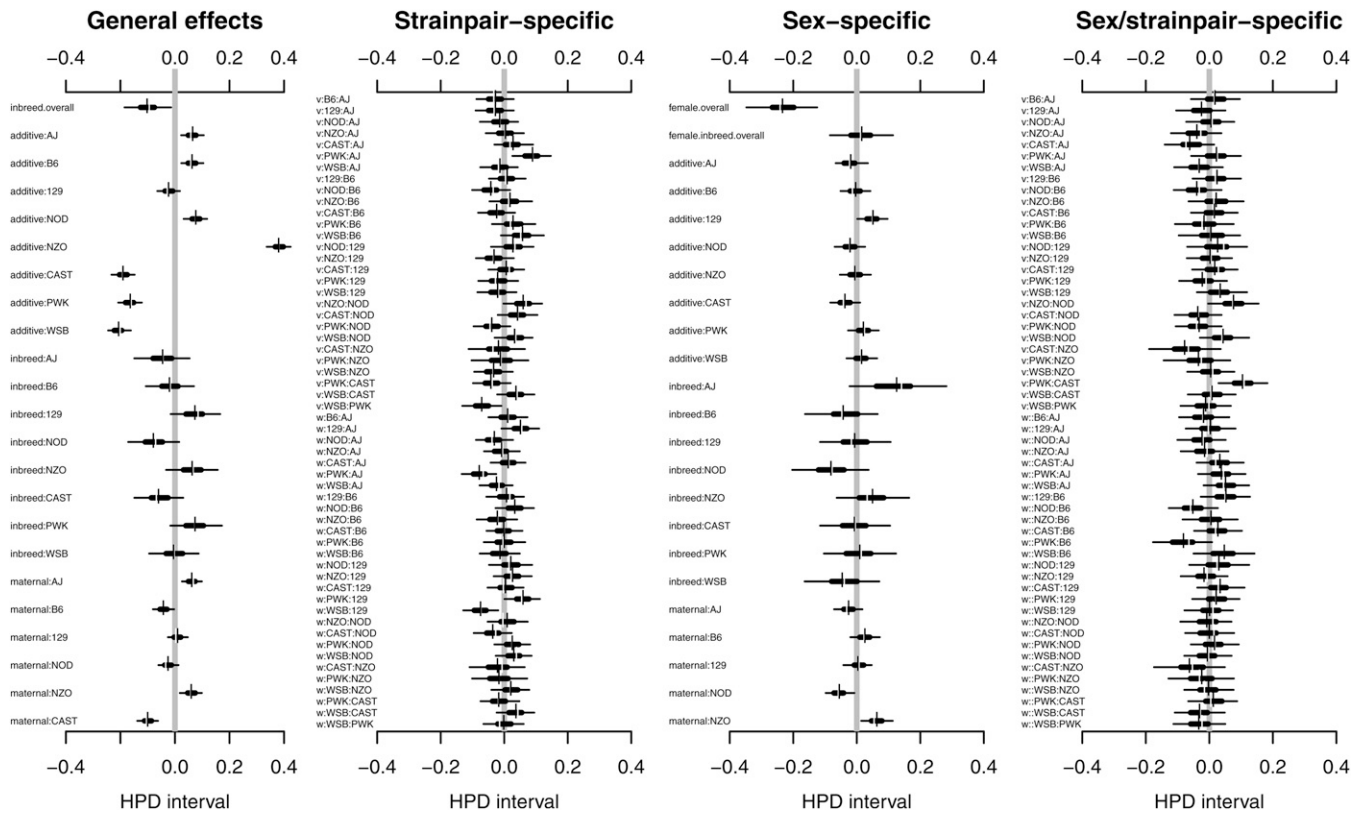


Figure 4 Highest posterior density (HPD) intervals for effects parameters fitted to weight data collected on 594 mice from a diallel of Collaborative Cross founders. Horizontal bars show for each parameter the region of highest posterior density that covers 50% (thick line) and 95% (thin line) of the posterior probability, with breaks indicating the posterior median and short vertical bars the posterior mean. The labels “additive”, “inbred”, and “maternal” in the first graph refer to the a , b , and m effects in Table 1. In the third graph they refer to sex-specific effects a_s , b_s , and m_s . The “ v ” and “ w ” labels in the second and fourth graphs refer to symmetric and asymmetric effects (as in Table 1).

contain a double dose of additive effects and so appear magnified). For ease of comparison, all values are shown on the scale of standard deviations of the transformed phenotype (transformations listed in Table 2). The percentage of eosinophils (EOS) (second straw plot) seems only moderately affected by additive genetics in this diallel. Nonetheless, as Figure 5 attests, the evidence for those effects is extremely strong, as is the evidence against substantial contributions from any other genetic, sex, or parent-of-origin effects, save a single symmetric increase attributed to the F_1 combination of PWK and CAST (95% HPD is between 0.06 and 1.7, posterior mean = 0.88, posterior probability of being $\leq 0 = 0.99$). The weight phenotype (rightmost straw plot) is the same as that described in Figures 3 and 4 and is marked by the effect of NZO escaping the ± 2 SD axis boundaries. A similar, if more moderate effect of NZO is evident for HDL cholesterol (Figure 8, fourth plot). White blood cell count (WBC) is among the few phenotypes that show strong evidence for sex-specific effects, as well as unsexed effects (Figure 5, Figure 8). In particular, examination of its posterior HPD intervals (Figure 9) reveals that the sex specificity is confined to epistatic effects (v_s , w_s) involving strains AJ, WSB, and 129, with a wider range of epistatic effects existing irrespective of sex. These posteriors accommodate, through our outlier model the potential for erratic

output from the measuring equipment. Figure 10 plots distributions of the posterior data weight (*i.e.*, the datapoint reliability) attributed to white cell counts from each individual, distinguishing one observation so incongruous as to merit down-weighting to 1/10th of a data point on average. From this robust analysis, white blood cell count emerges as a phenotype that, although apparently free of parent-of-origin effects, has an otherwise complex genetic architecture.

Discussion

We describe an efficient and general framework for modeling effects of genetics, parent-of-origin, and sex on phenotypes collected for diallels of inbred strains. By deploying a fully Bayesian approach with conjugate priors, imbalance and missing data translate to vagueness in the posterior rather than instability of the estimates. By adopting an MCMC approach to estimation, we provide a flexible environment in which the posterior distribution of arbitrary combinations of parameters, including prediction of new data, can be easily obtained. Moreover, to satisfy all inferential tastes, we describe a rapid and powerful formulation of Bayesian model selection for weighing the evidence in support of each category of effect in the diallel.

Protocol	Code	Overall effects			Strain-specific effects										Info
					Unsexed					Sex-specific					
		B	S	B _s	a	b	m	v	w	a _s	b _s	m _s	v _s	w _s	
ADVIA	BASO	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
ADVIA	CHCM	0.46	0.73	0.19	1.00	0.90	0.13	1.00	0.72	0.09	0.37	0.04	0.08	0.04	0.49
ADVIA	cHGB	0.22	0.04	0.07	1.00	0.82	0.48	1.00	0.59	0.02	0.03	0.03	0.02	0.02	0.63
ADVIA	EOS	0.06	0.36	0.06	0.93	0.13	0.00	0.99	0.01	0.01	0.05	0.00	0.08	0.00	0.73
ADVIA	HCT	0.57	0.44	0.22	0.92	0.92	0.09	0.98	0.81	0.58	0.53	0.04	0.29	0.12	0.35
ADVIA	HDW	0.43	0.01	0.01	1.00	0.97	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.89
ADVIA	LUC	0.09	0.01	0.01	0.95	0.81	0.00	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.82
ADVIA	LYM	0.74	0.14	0.33	0.99	0.95	0.17	0.97	0.85	0.43	0.40	0.10	0.79	0.87	0.39
ADVIA	MCH	0.41	0.70	0.11	1.00	0.87	0.53	1.00	0.36	0.03	0.06	0.57	0.04	0.02	0.48
ADVIA	MCHC	0.24	0.83	0.16	0.94	0.87	0.05	1.00	0.76	0.07	0.15	0.02	0.05	0.02	0.57
ADVIA	MCV	0.45	0.10	0.13	0.99	0.88	0.49	0.99	0.85	0.18	0.22	0.60	0.76	0.20	0.37
ADVIA	mHGB	0.33	0.03	0.05	0.91	0.90	0.01	1.00	0.05	0.01	0.05	0.01	0.01	0.01	0.75
ADVIA	MONO	0.46	0.93	0.09	0.58	0.50	0.00	0.55	0.00	0.02	0.08	0.00	0.00	0.00	0.59
ADVIA	MPV	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.98	0.00	0.00	0.00	0.00	1.00	0.99
ADVIA	NEUT	0.20	0.01	0.03	0.95	0.92	0.01	0.78	0.01	0.00	0.01	0.00	0.00	0.00	0.79
ADVIA	PLT	0.17	0.90	0.04	1.00	0.94	0.03	1.00	0.03	0.01	0.01	0.01	0.01	0.01	0.81
ADVIA	RBC	0.30	0.70	0.08	1.00	0.97	0.33	1.00	0.29	0.03	0.04	0.02	0.02	0.01	0.62
ADVIA	RDW	0.76	0.81	0.13	1.00	0.96	0.42	1.00	0.78	0.12	0.15	0.03	0.72	0.63	0.45
ADVIA	Retic	0.10	0.02	0.03	0.99	0.28	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.86
ADVIA	WBC	0.28	0.07	0.11	1.00	0.98	0.33	1.00	0.90	0.24	0.57	0.02	0.86	0.73	0.49
BP	PulseMean	0.46	0.63	0.46	0.77	0.40	0.41	0.61	0.80	0.43	0.40	0.38	0.43	0.40	0.06
BP	SystolicMean	0.45	0.35	0.45	0.49	0.38	0.35	0.92	0.72	0.40	0.40	0.36	0.44	0.36	0.09
CHEM	CHOL	0.70	0.62	0.17	1.00	0.95	0.17	0.96	0.06	0.12	0.09	0.02	0.08	0.02	0.56
CHEM	GLU	0.27	0.78	0.17	1.00	0.75	0.05	1.00	0.22	0.20	0.10	0.03	0.81	0.39	0.45
CHEM	HDL	0.18	0.93	0.04	1.00	1.00	0.01	1.00	0.00	0.02	0.02	0.00	0.01	0.00	0.87
CHEM	TG	0.35	0.80	0.10	0.98	0.66	0.29	0.95	0.06	0.06	0.05	0.01	0.03	0.01	0.57
EKG	CV	0.41	0.47	0.04	0.93	0.66	0.46	0.90	0.66	0.00	0.02	0.00	0.00	0.00	0.54
EKG	HR	0.36	0.24	0.11	0.92	0.52	0.60	0.79	0.65	0.05	0.08	0.03	0.05	0.04	0.41
EKG	HRV	0.45	0.06	0.01	0.59	0.21	0.01	0.28	0.06	0.00	0.00	0.00	0.00	0.00	0.66
EKG	RampI	0.59	0.15	0.08	0.66	0.45	0.14	0.52	0.33	0.04	0.11	0.02	0.71	0.09	0.34
EKG	SrampI	0.68	0.74	0.14	0.88	0.70	0.14	0.59	0.25	0.04	0.14	0.02	0.21	0.10	0.36
EKG	PQ	0.16	0.21	0.17	0.97	0.21	0.05	0.63	0.08	0.03	0.18	0.06	0.27	0.09	0.45
EKG	PR	0.17	0.22	0.23	0.97	0.20	0.05	0.76	0.09	0.06	0.20	0.04	0.46	0.09	0.43
EKG	QRS	0.08	0.03	0.07	0.93	0.11	0.04	0.15	0.06	0.02	0.02	0.01	0.01	0.02	0.72
EKG	QT	0.37	0.10	0.21	0.98	0.65	0.05	0.99	0.09	0.07	0.29	0.04	0.64	0.28	0.44
EKG	QTD	0.19	0.43	0.33	1.00	0.38	0.05	0.73	0.13	0.18	0.31	0.20	0.19	0.29	0.30
EKG	QTC	0.29	0.10	0.23	0.98	0.53	0.08	0.98	0.11	0.05	0.31	0.05	0.34	0.41	0.41
EKG	QTCD	0.20	0.26	0.34	0.99	0.36	0.05	0.82	0.15	0.14	0.32	0.22	0.20	0.33	0.31
EKG	RR	0.49	0.47	0.27	0.96	0.52	0.24	0.65	0.32	0.57	0.35	0.11	0.48	0.28	0.15
DEXA	B.Area	0.33	0.02	0.07	0.99	0.93	0.66	0.93	0.44	0.01	0.02	0.01	0.01	0.01	0.65
DEXA	BMC	0.63	0.10	0.08	1.00	0.98	0.68	1.00	0.62	0.02	0.04	0.01	0.01	0.01	0.65
DEXA	BMD	0.82	0.53	0.20	1.00	0.83	0.07	1.00	0.08	0.13	0.12	0.02	0.08	0.02	0.54
DEXA	LTM	0.32	0.84	0.07	1.00	0.97	0.71	1.00	0.65	0.03	0.03	0.00	0.01	0.00	0.66
DEXA	RST	0.06	0.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97
DEXA	TTM	0.31	0.83	0.07	1.00	0.94	0.70	1.00	0.58	0.04	0.03	0.01	0.02	0.01	0.63
COMP	MLNA	0.23	0.88	0.05	0.99	0.90	0.61	0.96	0.75	0.02	0.02	0.01	0.01	0.01	0.64
COMP	Weight	0.25	0.87	0.06	1.00	1.00	0.60	0.98	0.81	0.03	0.02	0.01	0.02	0.01	0.68
COMP	PctFat	0.34	0.74	0.05	1.00	0.87	0.46	1.00	0.33	0.01	0.02	0.01	0.02	0.01	0.61

Figure 5 Posterior model inclusion probabilities (MIPs) for genetic, sex, and parent-of-origin effects in 48 phenotypes measured in a diallel of the Collaborative Cross founders. All MIPs (colored columns except “Info”) are rounded to 2 d.p. and assume a prior of 0.5 for their components effect. Colors reflect the values and are scaled from blue (zero) to light red (one), with beige at 0.5 representing posterior belief about inclusion that is unmoved by the data. The “Info” column quantifies the gain in information about MIPs provided by the data, with values ranging from 1 (highly informative, red) to zero (uninformative, blue).

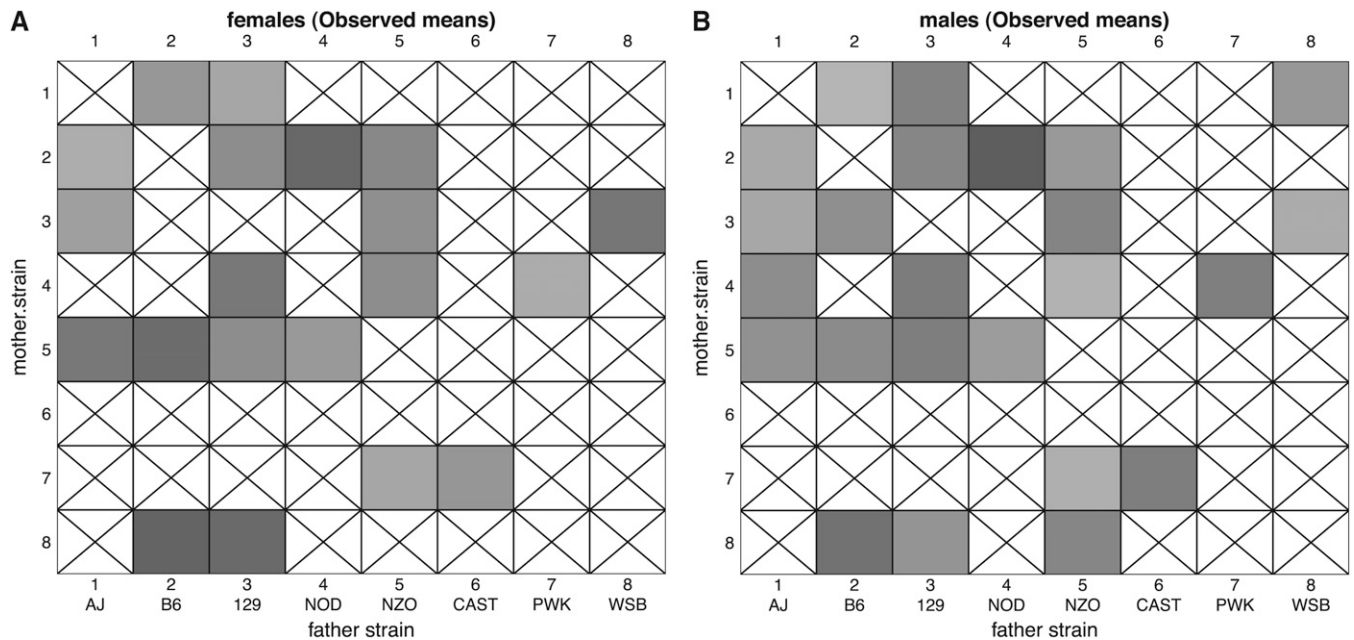


Figure 6 (A and B) Systolic blood pressure collected on 188 mice in a sparsely sampled diallel of the Collaborative Cross founders. Shaded boxes indicate the average blood pressure in a group (darker equals higher), with crossed boxes showing missing data.

Nonetheless, our approach is motivated by our bias: as geneticists focusing on model organisms, we are primarily concerned with characterizing the genetic architecture within the set of J^2 genome combinations with a view to subsequent hypothesis-driven experiments. We less often seek to infer formally parameters relating to the superpopulation of individuals or species from which those inbred strains were drawn. Interestingly, traditional literature on diallel analysis has tended to oppose the use of random effects in our context. It espouses what we call the “random parents commonplace”: that modeling parental contributions as random effects is valid only if the parents have been drawn at random from a larger population and it is the variance parameters of this larger population that the investigator seeks to estimate. It further asserts that when the above conditions are not met, for example, if we are interested in parental effects present in the cross, or if the parents were chosen deliberately, then parental contributions should be modeled as fixed effects (*e.g.*, Griffing 1956; Eberhart and Gardner 1966; Baker 1978). This commonplace persists in current literature, reiterated in, for example, Greenberg *et al.* (2010), who nonetheless develop an elegant Bayesian hierarchical model tailored to analyze a specific type of outbred diallel.

We consider this view in need of updating. Stein (1955) shocked the statistical world by showing that when simultaneously estimating three or more means as part of the same decision problem, fixed-effects modeling is dominated by biased methods that use shrinkage. Trading off errors in one dimension with those in another, the resulting estimates are drawn closer together when dimensions look similar, with useful shrinkage even when those dimensions

are unrelated (Parmigiani and Inoue 2009). In the diallel, dimensions are related, making the argument for hierarchical modeling yet stronger. Although it is sometimes hard to concoct a rationale for parental effects coming from a common distribution, it should be easy to intuit that they lie on a common scale and that knowledge about a_1, \dots, a_{j-1} provides information about how we would expect a_j to appear. Bayesian updating provides a coherent rendering of this intuition, allowing information from the data and uncertainty from the priors to propagate through the hierarchy and inform posterior estimates of parameters and predictions of new effects (Bernardo and Smith 2000; Sorensen and Gianola 2004). To make our point, we demonstrate by simulation how a fixed-effects GCA model with 10 parameters is matched or beaten by Bayesian shrinkage models with $n + 175$ parameters in its own backyard (Tables 3 and 4).

When there is prior belief that phenotypic similarity will tend to follow overall genetic similarity (*e.g.*, Kang *et al.* 2008), relatedness (*e.g.*, Lynch and Walsh 1998), spatial distribution, or some other structure that can be incorporated into an expected $J \times J$ correlation matrix \mathbf{A} , then this could aid estimation, potentially being incorporated by replacing Equation 4 with $\mathbf{a} \sim N(\mathbf{0}, \tau_a^2 \mathbf{A})$. Failing to do so, perhaps for convenience, makes inefficient use of available prior data, but discredits use of our hierarchy no more than it would for a fixed-effects model.

The use of a mixed model makes it tempting to interpret variance parameters as heritabilities. However, the very depth of genetic characterization afforded by the diallel, as well as legitimate aspects of the random parents commonplace described above, suggests a more reflective approach.

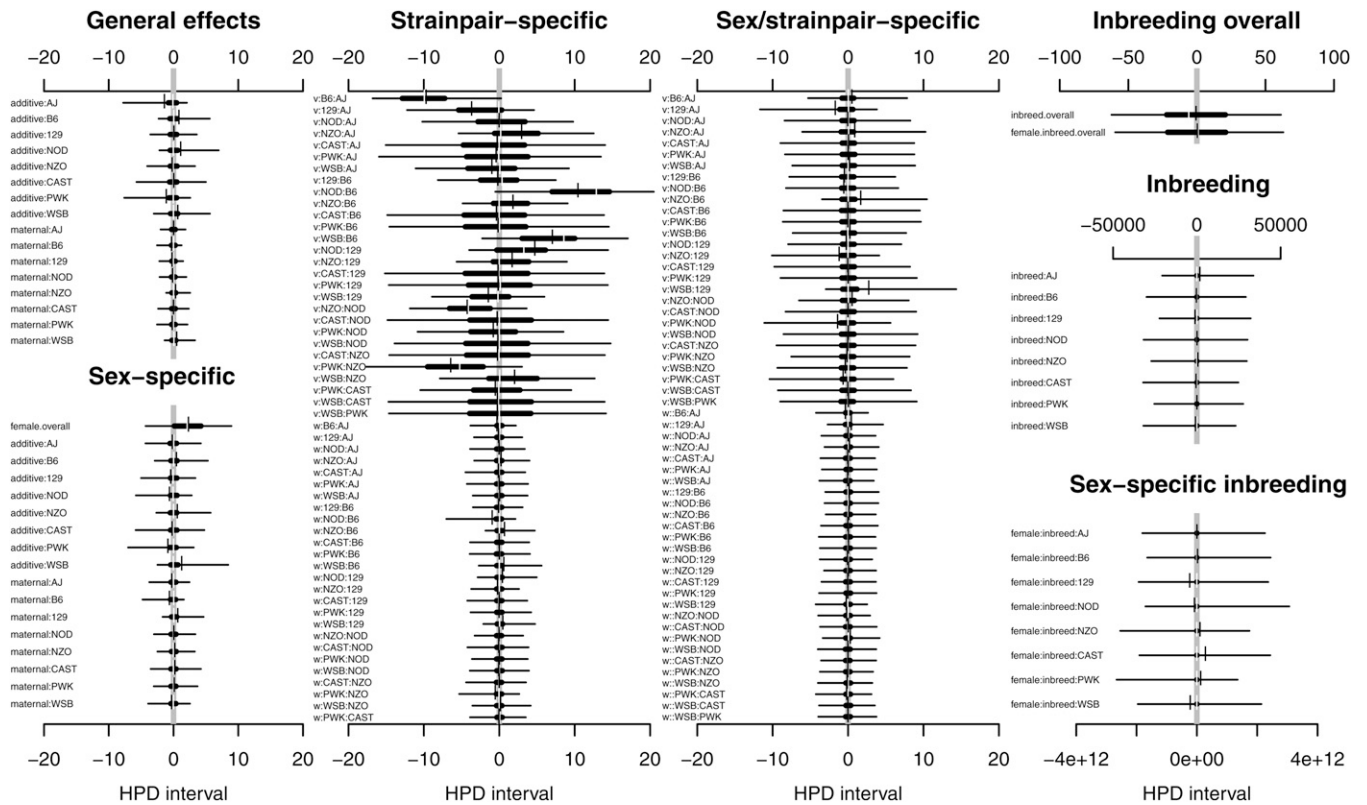


Figure 7 Highest posterior density (HPD) intervals for effects parameters fitted to systolic blood pressure data collected on 188 mice from an incomplete diallel of Collaborative Cross founders, with abbreviations and symbols as described in Figure 4. This diallel had no inbred animals. Posterior distributions for parameters relating to inbreeding (collected in the rightmost column) are therefore vague, reflecting mostly uninformative prior belief (note the x-axis scales). Nonetheless, they are stably estimated and do not disrupt estimation of the other effects.

Our decomposition of effects groups into a, b, m, v, w spans many possible reduced models for quantitative inheritance. When we further explore the heritability, or the amount of variance explained, by each group of effects, we find that, in a diallel breeding structure, the variance contributions of the groups interact. For an expedient method of calculating h_q^2 ,

the heritability attributable to effects group q , we promote a calculation using the estimated $\hat{\tau}_q$ group dispersions (Appendix A).

In our model we do not treat explicit variance heterogeneity (Rodriguez *et al.* 1993). However, our outlier approach implicitly adapts to variance heterogeneity by assigning

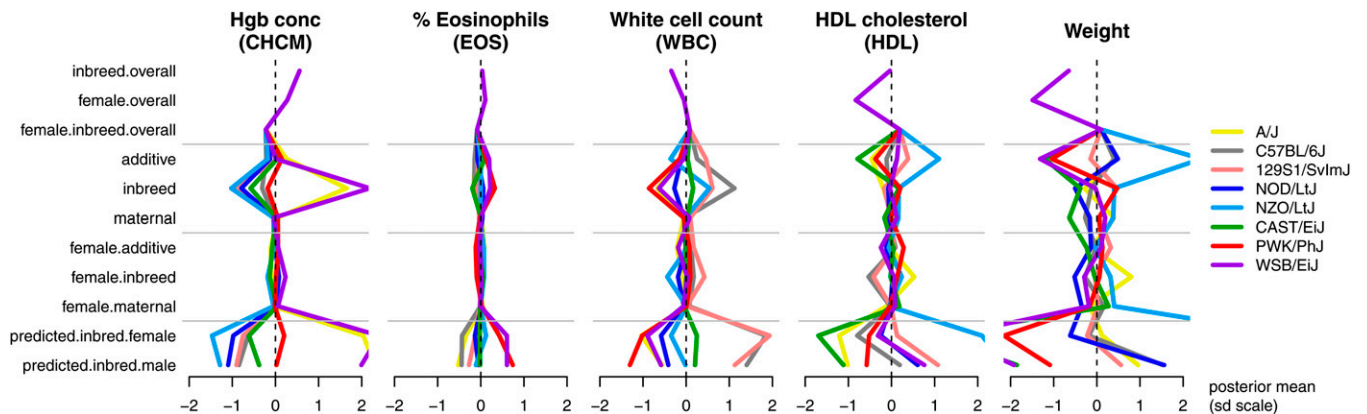


Figure 8 A “straw plot” summarizing estimated effects and predicted values in five phenotypes collected on a diallel of the Collaborative Cross founders. Gray horizontal lines group posterior means for overall (B , S , B_s), strain-specific (a , b , m), and sex/strain-specific (a_s , b_s , m_s , as $+\frac{1}{2}$ dose for females, $-\frac{1}{2}$ dose for males) effects, as well as posterior predictive means for inbreds. For ease of comparison across phenotypes, x-axes are scaled to the SD of the transformed phenotype. Values for the NZO (light blue) strain in the HDL cholesterol and weight phenotypes are extreme enough to escape the 2 SD limits of the plot.

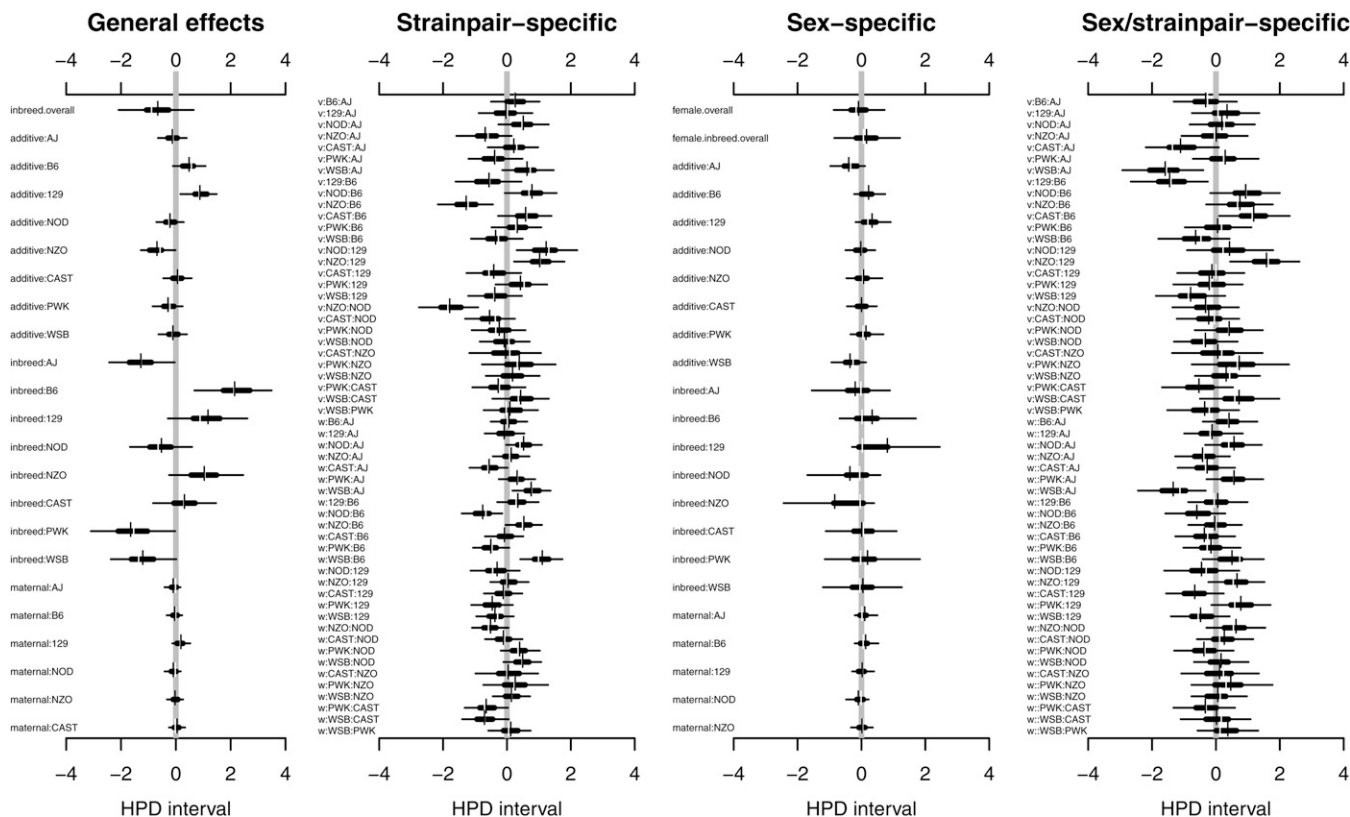


Figure 9 Highest posterior density (HPD) intervals for effects parameters fitted to white blood cell count (WBC) data collected on 626 mice from a diallel of Collaborative Cross founders, with abbreviations and symbols as described in Figure 4.

reduced weights λ_i to measurements it deems to be more variable. Postprocess exploration of $\hat{\lambda}_i | jk[i]$ may reveal strain-dependent variances. Algorithmically, an option for a parametric form of $\text{Var}(\varepsilon_i | j, k, S)$ can be achieved by modeling $1/\lambda_i \sim \exp\{x_i^T \boldsymbol{\gamma}\}$, replacing the update for λ_i with Gibbs step

$$\boldsymbol{\gamma} | Y - \mathbf{X}\boldsymbol{\beta} = \mathbf{r},$$

where drawing from a posterior using $\ell(\boldsymbol{\gamma} | \mathbf{r}) = \sum_i r_i^2 / (2 \exp\{x_i^T \boldsymbol{\gamma}\})$ can be achieved by Metropolis–Hastings, to calculate the new λ_i values. Ideally, $\boldsymbol{\gamma}$ would have a reduced form including just additive, or sex-specific additive, components to retain model parsimony. See Rönnegård and Valdar (2011) for an example of such a double generalized linear model in practice.

Bayesian models are often slower to compute and more difficult to interpret, and can require a large and ill-defined set of prior choices. Although it is often the case that estimators using reasonable hyperpriors can outperform the MLE over a large parameter space, for most loss functions neither MLE, nor ridge, nor Bayes estimators completely dominate each other. Given a mixed-model decomposition of the diallel, including the possibility of outliers, unobserved diallel combinations, and unknown sparsity among the parameters, stable MLE estimates can be difficult to achieve. Penalized estimates [e.g., from group LASSO (Hastie *et al.* 2009)] represent an opportunity, but these can often be interpreted as a Bayesian prior hypothesis. We demonstrate the use of the h -likelihood (using hglm) for obtaining mostly accurate point estimates for our diallel model and in less computational time. Although useful for this purpose, the hglm approach does not easily provide the rich flexibility of our Gibbs sampler for, among other things, modeling

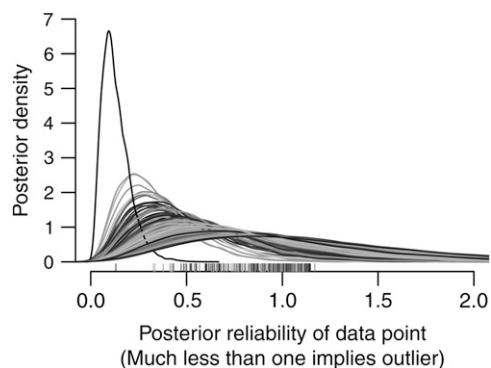


Figure 10 Simultaneous accommodation and detection of outliers in diallel data on white blood cell count (WBC), using the full model with outliers (at 6 d.f.). Each line represents the posterior distribution of the weight data (λ_i ; see *Models and Methods*) for the phenotype measurement of individual i , with corresponding posterior means given as ticks on the x-axis and line shading for visualization only. On average, individuals will have weight at 1, but posterior distributions concentrated near 0, such as the high peak at 0.1, are outliers downweighted by the model.

outliers and uncertainty about component inclusions. We supply an algorithm with reasonable default priors, but users are still permitted to choose their own, including flat Gelman or Stein priors, Jeffreys priors, or heavily informed priors from the inverse gamma family. For the BayesSpike model, one might consider *a priori* that all components have an equal 50% chance of inclusion, that they have different anticipated prior weights, or that the prior proportion of active groups is itself unknown and must be learned in the model.

Approaches that model hierarchically, borrowing strength across data sources and explicitly defining higher-order components, are becoming essential to navigate the complex high-dimensional spaces created by high-throughput data of all types. However, such hierarchies should be designed to empower researchers, not intimidate or perplex them. We emphasize interpretability of all parameters in the model in the hope they can at some level be interpreted and critiqued by nonexperts. We provide a framework for high-, medium-, and low-level analysis of imperfectly sampled diallels. In doing so, we succinctly summarize results from a vast amount of original data on the genetic architecture of phenotypes in founders of the Collaborative Cross and their F_1 hybrids.

Our diallel Gibbs sampler and BayesSpike software are provided free of charge as R packages R/BayesDiallel and R/BayesSpike as soon as practicably in [File S1](#).

Acknowledgments

We thank Vasyl Zhabotynsky and Fred Wright for helpful discussions. The authors acknowledge partial support from the National Institutes of Health (NIH) Center of Excellence in Genome Sciences grants P50 MH090338 and P50 MH006582 (to A.B.L. and W.V.), from NIH grant GM076468 (to K.L.S. and G.A.C.), from The Jackson Laboratory (G.A.C.), and from the University of North Carolina Lineberger Comprehensive Cancer Center (A.B.L. and W.V.).

Literature Cited

- Baker, R. J., 1978 Issues in diallel analysis. *Crop Sci.* 18(834): 533–536.
- Bernardo, J. M., and A. F. M. Smith, 2000 *Bayesian Theory*. John Wiley & Sons, Chichester, UK.
- Box, G. E. P., and G. C. Tiao, 1973 *Bayesian Inference in Statistical Analysis*, p. 608. Wiley-Interscience, New York.
- Carbonell, E. A., W. E. Nyquist, and A. E. Bell, 1983 Sex-linked and maternal effects in the Eberhart-Gardner general genetics model. *Biometrics* 39(3): 607–619.
- Carlin, B. P., and T. A. Louis, 2008 *Bayesian Methods for Data Analysis*, Ed. 3. Chapman & Hall/CRC, London/New York.
- Casella, G., and E. I. George, 1992 Explaining the Gibbs sampler. *Am. Stat.* 46(3): 167.
- Chesler, E. J., D. R. Miller, L. R. Branstetter, L. D. Galloway, B. L. Jackson *et al.*, 2008 The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm. Genome* 19(6): 382–389.
- Christie, B. R., and V. I. Shattuck, 1992 The diallel cross: design, analysis, and use for plant breeders. *Plant Breed. Rev.* 9: 9–36.
- Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie *et al.*, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36(11): 1133–1137.
- Cockerham, C., and B. Weir, 1977 Quadratic analyses of reciprocal crosses. *Biometrics* 33(1): 187–203.
- Collaborative Cross Consortium, 2012 The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* 190: 389–401.
- Dongarra, J., 2002 Basic linear algebra subprograms technical forum standard. *Int. J. High Performance Appl. Supercomput.* 16: 115–199.
- Eberhart, S., and C. Gardner, 1966 A general model for genetic effects. *Biometrics* 22(4): 864–881.
- Gardner, C., and S. Eberhart, 1966 Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* 22(3): 439–452.
- Gelfand, A. E., and A. F. M. Smith, 1990 Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* 85(410): 398–409.
- Gelman, A., 2006 Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Int. Soc. Bayes. Anal.* 1(3): 515–534.
- Gelman, A., and J. Hill, 2007 *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.
- Geman, S., and D. Geman, 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.* 6(6): 721–741.
- George, E., and D. Foster, 2000 Calibration and empirical Bayes variable selection. *Biometrika* 87(4): 731–747.
- Greenberg, A. J., S. R. Hackett, L. G. Harshman, and A. G. Clark, 2010 A hierarchical Bayesian model for a novel sparse partial diallel crossing design. *Genetics* 185(1): 361–373.
- Griffing, B., 1956 Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9(4): 463–493.
- Gross, J., 2003 *Linear Regression*. Springer-Verlag, Heidelberg, Germany/New York.
- Guan, Y., and M. Stephens, 2011 Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *Ann. Appl. Stat.* 5(3): 1780–1815.
- Hastie, T., R. Tibshirani, and J. Friedman, 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics, Ed. 2). Springer-Verlag, New York.
- Hayman, B. I., 1957 Interaction, heterosis and diallel crosses. *Genetics* 42: 336–355.
- Ishwaran, H., and J. S. Rao, 2005 Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Stat.* 33(2): 730–773.
- Jeffreys, H., 1946 An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A Math. Phys. Sci.* 186 (1007): 453–461.
- Jinks, J. L., and B. I. Hayman, 1953 Analysis of diallel crosses. *Maize Genet. Coop. News Lett.* 27: 48–54.
- Kang, H. 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.
- Kass, R., and A. Raftery, 1995 Bayes factors. *J. Am. Stat. Assoc.* 90 (430): 773–795.
- Kempthorne, O., and R. Curnow, 1961 The partial diallel cross. *Biometrics* 17(2): 229–250.
- Lee, Y., and J. Nelder, 1996 Hierarchical generalized linear models. *J. R. Stat. Soc. B* 58(4): 619–678.
- Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

- Meng, X.-L., 2008 Discussion: one-step sparse estimates in non-concave penalized likelihood models: Who cares if it is a white cat or a black cat? *Ann. Stat.* 36(4): 1542–1552.
- Parmigiani, G., and L. Inoue, 2009 *Decision Theory: Principles and Approaches* (Wiley Series in Probability and Statistics). John Wiley & Sons, New York.
- R Development Core Team, 2011 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Rodríguez, L. A., D. W. Fulker, and S. S. Cherny, 1993 A maximum-likelihood model-fitting approach to conducting a Hayman analysis of diallel tables with complete or missing data. *Behav. Genet.* 23(1): 69–76.
- Rönnegård, L., and W. Valdar, 2011 Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics* 188: 435–447.
- Rönnegård, L., X. Shen, and M. Alam, 2010 hglm: a package for fitting hierarchical generalized linear models. *R J.* 2(2): 20–28.
- Schmidt, J., 1919 La valeur de l'individu à titre de générateur appréciée suivant la méthode du croisement dialléle. *C. R. Trav. Lab. Carlsberg* 14(6): 1–33.
- Sorensen, D., and D. Gianola, 2004 *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health*. Springer-Verlag, New York.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde, 2002 Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64(4): 583–639.
- Sprague, G., and L. Tatum, 1942 General vs. specific combining ability in single crosses of corn. *J. Am. Soc. Agron.* 34: 923–932.
- Stein, C., 1955 Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, pp. 197–206 in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Issue 4. University of California Press, Berkeley.
- Taylor, B. A., C. Wnek, D. Schroeder, and S. J. Phillips, 2001 Multiple obesity QTLs identified in an intercross between the NZO (New Zealand obese) and the SM (small) mouse strains. *Mamm. Genome* 12(2): 95–103.
- Venables, W., and B. Ripley, 2002 *Modern Applied Statistics with S*. Springer-Verlag, New York.
- Wright, A. J., 1985 Diallel designs, analyses, and reference populations. *Heredity* 54: 307–311.
- Zhu, J., and B. Weir, 1996 Mixed model approaches for diallel analysis based on a bio-model. *Genet. Res.* 68(3): 233–240.

Edited by Lauren M. McIntyre, Dirk-Jan de Koning, and 4 dedicated Associate Editors

Appendix A: Heritability in the Diallel

Heritability is the proportion of phenotypic variance explained by genetic effects. Its exact definition depends on what effects are being considered and in which population phenotypic variance is measured. Consider any linear genetic model, $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$, where $\boldsymbol{\beta} \in \mathbb{R}^p$ is an explanatory vector of common effects, and $\varepsilon_i \sim N(0, \sigma^2)$ are iid environmental effects, for independently sampled individuals $i \in 1, \dots, N$ in the population. The vector $\mathbf{x}_i \in \mathbb{R}^{p \times 1}$ represents a draw from the genetic diversity of haplotypes/strains/alleles as distributed in the population, represented by stacked matrix $\mathbf{X} = [\mathbf{x}_1^T \dots \mathbf{x}_n^T]^T$. If both \mathbf{x}_i and $\boldsymbol{\beta}$ were randomly distributed and independent, and $\mu_Y = E[\mathbf{x}_i^T]E[\boldsymbol{\beta}]$ were the population mean, then the observed variability of $\text{Var}(y_i) = E[(y_i - \mu_Y)^2]$ has expectation that can be calculated

$$\text{Var}(Y_i) = \text{tr}\{E[\mathbf{x}_i \mathbf{x}_i^T] E[\boldsymbol{\beta} \boldsymbol{\beta}^T]\} + \sigma^2 - (E[\mathbf{x}_i^T] \boldsymbol{\beta})^2. \quad (\text{A1})$$

Equation A1 provides a method to decompose observed variability of Y to the contributions of specific effects or groups of effects. If we assume that diallels of future strains will have effects distributed $\boldsymbol{\beta}_j \sim N(0, \hat{\tau}_q^2)$, then $E(\boldsymbol{\beta} \boldsymbol{\beta}^T)$ is a diagonal matrix with $\hat{\tau}_q^2$ values for each group of factors. In a complete diallel experiment, $E[\mathbf{x}_i \mathbf{x}_i^T] \in \mathbb{R}^{p \times p}$ is a very structured design matrix. If we considered mother strain j and father strain k to be independent draws from a pool of J strains, and gender of the offspring to be an independent Bernoulli draw, then we can calculate explicit expected dosage amounts of each of the a, b, m, v, w, a_s , b_s , m_s , v_s , and w_s effects. For additive effect a_j , for strain j' , located at position $\{x_i\}_{a(j')}$ along the \mathbf{x}_i covariate vector, expected $E[\{x_i\}_{a(j')}^2] = E[I\{j = j'\} + I\{k = j'\}]^2] = 2(1+1/J)/J$. Since there are J separate a_j terms, the variance contribution of additive effects can be approximated as $\sum_{j=1}^J E[x_{ij}^2] = 2(1+1/J)\hat{\tau}_a^2$. Further expectations are calculated in Table A1. With Table A1 we can then interpret the heritability contribution of additive effects h_a^2 to be

Table A1 Expected heritability contribution of effects to $\text{Var}(Y_i)$ of the effects groups

a	b	m	v	w	a_s	b_s	m_s	v_s	w_s
$\frac{2(J+1)}{J}$	$\frac{1}{J}$	$\frac{2(J-1)}{J}$	$\frac{J-1}{J}$	$\frac{J-1}{J}$	$\frac{J+1}{2J}$	$\frac{1}{4J}$	$\frac{J-1}{2J}$	$\frac{J-1}{4J}$	$\frac{J-1}{4J}$

$$h_a^2 = \frac{2(1 + 1/J)\hat{\tau}_a^2}{2(1 + 1/J)\hat{\tau}_a^2 + 1/J\hat{\tau}_b^2 + 2(1 - 1/J)\hat{\tau}_m^2 + (1 - 1/J)\hat{\tau}_v^2 + (1 - 1/J)\hat{\tau}_w^2 + \sigma^2} \quad (\text{A2})$$

in a diallel with no sex-specific effects. We see that heritability contributions of inbreeding effects b contribute at a reduced level of $1/J$ to the variability of the phenotype. This is because in a complete diallel, inbred subjects compose only $1/J$ of the population.

Appendix B: Gibbs Sampling Scheme for the Full Model

Collecting all fixed-effect (e.g., β_{female}) and strain-specific random-effect parameters (e.g., a_j) in a single vector of p_M regression coefficients $\boldsymbol{\beta}$, we construct an $n \times p_M$ design matrix $\mathbf{X} = [\mathbf{x}_1^T \dots \mathbf{x}_n^T]^T$ and consider the regression problem

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

\mathbf{X} is both sparse and highly structured in our diallel model. For instance, in the “av” model (additive and symmetric effects), $p_M = 1 + J + (J + 1) \times J/2$, and the last $(J + 1) \times J/2$ positions in \mathbf{x}_i are mostly zeros with a single 1 at the position corresponding to mother/father pair $(j, k)[i]$. Let the parameter groups $a, a_s, m \dots$ be enumerated as groups $q \in 1, \dots, Q$ and $\mathbb{J}(q)$ denote the group of coefficients q_j assigned to the same shrinkage parameter τ_q^2 . By introducing a \mathbf{Q} matrix, as suggested from Equation 17, with diagonal terms $1/\tau_q^2$ for parameters grouped to $N(0, \tau_q^2)$, the multivariate posterior is

$$\boldsymbol{\beta} | \boldsymbol{\tau}, \sigma^2 \sim N\left((\mathbf{X}^T\mathbf{X} + \sigma^2\mathbf{Q}^{-1})^{-1}\mathbf{X}^T\mathbf{y}, \sigma^2(\mathbf{X}^T\mathbf{X} + \sigma^2\mathbf{Q}^{-1})^{-1}\right), \quad (\text{B1})$$

which can be efficiently sampled by taking the Cholesky square root of $\mathbf{X}^T\mathbf{X} + \sigma^2\mathbf{Q}^{-1}$, where the diagonal of matrix $\mathbf{X}^T\mathbf{X}$ counts the number of subjects with relevant j, k pairs for each q group. We deploy Gibbs sampling (Geman and Geman 1984; Gelfand and Smith 1990; Casella and George 1992), using C-level BLAS (Dongarra 2002) code compiled in package format for R (R Development Core Team 2011), and find that the 2011 Macintosh vecLib LAPACK (Apple Developers) is sufficient to generate as many as a 8000 draws in 26.2 sec from the posterior, given that p_M has an upper bound of 244 when the number of strains is 8. However, when p_M gets larger, doing so as the number of strains increases, it is necessary to invert conditional subset groups of $\boldsymbol{\beta}$. The advantage of a single Cholesky draw is to reduce autocorrelation of the Gibbs sampler. In this case, maximum autocorrelation at 1-lag for $p_M = 164$ variables is 0.177, and max 20-lag is 0.086, with 95% of the 1-lag being between ± 0.05 . Given $\boldsymbol{\beta}$, we draw τ_q^2 as

$$\tau_q^2 | \boldsymbol{\beta}_{\mathbb{J}(q)} \sim \left(\mu_{\tau} + \sum_{j \in \mathbb{J}(q)} \beta_j^2 \right) \times \text{Inv} - \chi^2(v_{\tau} + |\mathbb{J}(q)|),$$

where $|\mathbb{J}(q)|$ is the count of members of group q , and draw σ^2 as

$$\sigma^2 | \boldsymbol{\beta} \sim \left(v_{\sigma}\mu_{\sigma} + \sum_i (Y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 \right) \times \text{Inv} - \chi^2(v_{\sigma} + n).$$

Appendix C: Gibbs Sampling Scheme for the Full Model with Outliers

Computational complexity changes with the outlier model, where $\varepsilon_i \sim N(0, \sigma^2)/\sqrt{\chi^2(v_{\varepsilon})/v_{\varepsilon}} \sim t(0, \sigma^2, v_{\varepsilon})$. In this case, if we draw λ_i as a weight for subject i from

$$\lambda_i | \boldsymbol{\beta}, \sigma^2 \sim \left[1 + \frac{1}{\sigma^2}(y_i - \mathbf{x}_i^T\boldsymbol{\beta})^2 \right]^{-1} \times \chi^2(v_{\varepsilon} + 1),$$

then we can reweight by defining components $\widetilde{\mathbf{X}}^T\mathbf{y}_j = \sum_i \{X\}_{ij}\lambda_i y_i$ and $\{\widetilde{\mathbf{X}}^T\mathbf{X}\}_{j,k} = \sum_i \{X\}_{ij}\lambda_i \{X\}_{ik}$, in which case

$$\boldsymbol{\beta} | \mathbf{y}, \sigma^2, \boldsymbol{\lambda}, \boldsymbol{\tau}^2 \sim N\left(\left(\widetilde{\mathbf{X}}^T \mathbf{X} + \sigma^2 \mathbf{Q}\right)^{-1} \widetilde{\mathbf{X}}^T \mathbf{y}, \sigma^2 \left(\widetilde{\mathbf{X}}^T \mathbf{X} + \sigma^2 \mathbf{Q}\right)^{-1}\right)$$

becomes a new draw of $\boldsymbol{\beta}$. Since the reweighting is an $O(p_M^2 n)$ operation, it can slow down the algorithm when $n > p_M$.

Appendix D: Exclusionary Gibbs Group Sampling in the Diallel

To sample efficiently from Equation 18, we consider the residual vector $\mathbf{y} - \mathbf{X}_{\setminus \mathbb{J}(q)} \boldsymbol{\beta}_{\setminus \mathbb{J}(q)}$, which includes current information of all β_j for j not in $\mathbb{J}(q)$. Integrating $\boldsymbol{\beta}_{\setminus \mathbb{J}(q)}$ out of this draw produces an unnormalized function $f(\boldsymbol{\tau}_q^2 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)})$. This function $f(\boldsymbol{\tau}_q^2 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)})$ takes on a discrete probability at $\boldsymbol{\tau}_q^2 = 0$ that we define as f^0 and also includes a continuous density $f^+(\boldsymbol{\tau}_q^2 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)})$ with support on $\boldsymbol{\tau}_q^2 \in (0, \infty)$. If $F^+ = \int_0^\infty f^+(\boldsymbol{\tau}_q^2 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)}) d\boldsymbol{\tau}_q^2$, then a Bernoulli draw of probability

$$P\left(\boldsymbol{\tau}_q^2 \neq 0 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)}\right) = \frac{\pi_q F^+}{(1 - \pi_q) f^0 + \pi_q F^+}$$

provides a Gibbs decision to turn on or off parameters in group q . Moreover, to avoid potentially slow numerical integration, we construct a Metropolis–Hastings importance sample from $f^+(\boldsymbol{\tau}_q^2 | \boldsymbol{\beta}_{\setminus \mathbb{J}(q)})$, which approximates F^+ in expectation.

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.111.132563/-/DC1>

A General Bayesian Approach to Analyzing Diallel Crosses of Inbred Strains

Alan B. Lenarcic, Karen L. Svenson, Gary A. Churchill, and William Valdar

File S1

Compressed R Package

File S1 is available at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.111.132563/-/DC1> as a compressed R package, which contains the current version of our BayesDiallel and BayesSpike software.

This includes:

1. The code used for the simulations in the paper. On loading BayesDiallel in R, this is obtained by typing `"help(FullDiallelSimulate)"`
2. The BayesDiallel and BayesSpike R libraries, in the attached tar.gz files.
3. The data used in the paper; these are in the examples folder within the BayesDiallel.tar.gz file.

The software is also available at <http://valdar.unc.edu/>