Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree

Gustavo de los Campos,*,[1],[2] **Hugo Naya,**†,[1] **Daniel Gianola,***,‡ **José Crossa,**§ **Andrés Legarra,****
Eduardo Manfredi,** **Kent Weigel**‡ **and José Miguel Cotes**§

**Department of Animal Sciences and* ‡*Department of Dairy Science, University of Wisconsin, Madison, Wisconsin 53706,* †*Bioinformatics Unit,
Institut Pasteur, 11400 Montevideo, Uruguay,* §*Biometrics and Statistics Unit, Crop Research Informatics Lab, International Maize
and Wheat Improvement Center (CIMMYT) , 06600 Mexico, D.F., Mexico and* ***Station d'Amélioration Génétique des Animaux,
Institut National de la Recherche Agronomique (INRA), UR 631 SAGA, F-31326 Castanet Tolosan, France*

ABSTRACT

The availability of genomewide dense markers brings opportunities and challenges to breeding programs. An important question concerns the ways in which dense markers and pedigrees, together with phenotypic records, should be used to arrive at predictions of genetic values for complex traits. If a large number of markers are included in a regression model, marker-specific shrinkage of regression coefficients may be needed. For this reason, the Bayesian least absolute shrinkage and selection operator (LASSO) (BL) appears to be an interesting approach for fitting marker effects in a regression model. This article adapts the BL to arrive at a regression model where markers, pedigrees, and covariates other than markers are considered jointly. Connections between BL and other marker-based regression models are discussed, and the sensitivity of BL with respect to the choice of prior distributions assigned to key parameters is evaluated using simulation. The proposed model was fitted to two data sets from wheat and mouse populations, and evaluated using cross-validation methods. Results indicate that inclusion of markers in the regression further improved the predictive ability of models. An R program that implements the proposed model is freely available.

GENOMEWIDE dense marker maps are now available for many species in plants and animals (*e.g.*, WANG *et al.* 2005). An important challenge is how this information should be incorporated into statistical models for prediction of genetic values in animal and plant breeding programs or prediction of diseases.

A standard quantitative genetic model assumes that genetic (u_i) and environmental (ε_i) effects act additively, to produce phenotypic outcomes (y_i) according to the rule $y_i = u_i + \varepsilon_i$. The information set now available for predicting genetic values may include, in addition to phenotypic records, a pedigree, molecular markers, or both.

Several methodologies have been proposed for incorporating dense marker data into regression models. A distinction can be made between methods that explicitly regress phenotypic records on markers via the regression function $u_i = g(\mathbf{x}_i, \boldsymbol{\beta})$, where $\mathbf{x}_i$ is a vector of marker covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients, *e.g.*, $g(\mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}$, and those that view genetic values as a function of the subject and use marker information to build a (co)variance structure between subjects. The first group of methods includes standard Bayesian regression (BR) with random coefficients, *i.e.*, a Bayesian model where regression coefficients are assigned the same Gaussian prior, and other shrinkage methods such as Bayes A or Bayes B of MEUWISSEN *et al.* (2001), and specifications described in GIANOLA *et al.* (2003). The second type of approach was suggested by GIANOLA *et al.* (2006) and GIANOLA and VAN KAAM (2008), who proposed using reproducing kernel Hilbert spaces regression (RKHS), with the information set consisting of SNP (single-nucleotide polymorphism) genotypes, possibly supplemented by genealogies. As discussed in DE LOS CAMPOS *et al.* (2009), in this approach, marker information is used to create a prior (co)variance structure between genomic values, u_i, $\mathrm{Cov}(u_i, u_j) = K(\mathbf{x}_i, \mathbf{x}_j)\sigma_\alpha^2$, where $K(., .)$ is some positive-definite function and σ_α^2 is a parameter to be estimated from the data.

The two types of approaches lead to predictions of genomic values for quantitative traits. An advantage of explicitly regressing phenotypes on marker covariates is that the model can produce information about genomic regions that may affect the trait of interest. However, a main difficulty is that the number of regression coefficients (p) is typically large, even larger than the number of records (n), with $p \gg n$. Therefore, a crucial

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Department of Animal Sciences, University of Wisconsin, 1675 Observatory Dr., Madison, WI 53706.
E-mail: gdeloscampos@wisc.edu

aspect is how this methodology can cope with the curse of dimensionality and with colinearity.

With whole-genome scans, many markers are likely to be located in regions that are not involved in the determination of traits of interest. On the other hand, some markers may be in linkage disequilibrium with some QTL or in regions harboring genes involved in the infinitesimal component of the trait. This suggests that differential shrinkage of marker effects should be a feature of the model, as noted by MEUWISSEN *et al.* (2001). TIBSHIRANI (1996) proposed a regression method (least absolute shrinkage and selection operator, LASSO) that combines the good features of subset selection (*i.e.*, variable selection) with the shrinkage produced by BR. Recently, PARK and CASELLA (2008) presented a Bayesian version of the LASSO method (Bayesian LASSO, BL) and suggested a Gibbs sampler for its implementation. Alternatives to the Gibbs sampler of Park and Casella are discussed in HANS (2008). While the BL described by Park and Casella is appealing for the reasons mentioned above, it does not accommodate pedigree information or regression on (co)variates other than the markers for which a different shrinkage approach may be desired.

Several authors have considered combining pedigree and marker data into a single model in the context of QTL analysis (*e.g.*, FERNANDO and GROSSMAN 1989; BINK *et al.* 2002, 2008). Here, in this spirit, the BL is modified and extended to accommodate pedigree information as well as covariates other than markers.

The main objectives of this article are to (1) discuss the use of BL and related methods in the context of linear regression of quantitative traits on molecular markers, (2) evaluate the sensitivity of BL with respect to the choice of the prior for the regularization parameter, (3) extend the BL so that pedigrees or regressions on covariates other than markers can also be included in the model, and (4) evaluate the methodology using data from a self-pollinated wheat population and an outcross mouse population. The article is organized as follows: the first section, BAYESIAN LASSO, introduces the BL as presented in PARK and CASELLA (2008) and discusses connections between BL and closely related methods, such as those proposed by MEUWISSEN *et al.* (2001) or variants proposed by other authors. MONTE CARLO STUDY evaluates the sensitivity of BL with respect to the choice of prior for the regularization parameter. BAYESIAN REGRESSION COUPLED WITH LASSO presents an extension of BL, treating effects of different types of regressors with different priors. In DATA ANALYSIS, the proposed methodology is applied to two data sets representing a collection of wheat lines and a population of mice. CONCLUDING REMARKS are provided in the final section of the article. An R function (R DEVELOPMENT CORE TEAM 2008) that fits the model and data sets used in this article are made available (see supporting information, File S1 and File S2).

THE BAYESIAN LASSO

TIBSHIRANI (1996) proposed using the sum of the absolute values of the regression coefficients (or L_1 norm) as a penalty in regression models, to simultaneously produce variable selection and shrinkage of coefficients; the proposed methodology was termed LASSO. In LASSO, estimates are obtained by solving the constrained optimization problem

$$\min_{\boldsymbol{\beta}}\left\{\sum(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 \text{ subject to } \sum_j |\beta_j| \leq t\right\}, \quad (1)$$

where $\mathbf{x}_i$ is a vector of covariates, $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients, and t is an arbitrary positive constant. Above, it is assumed that data are centered, *i.e.*, y_i has zero mean. Optimization problem (1) is equivalent to

$$\min_{\boldsymbol{\beta}}\left\{\sum(y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 + \lambda(t)\sum_j |\beta_j|\right\} \quad (2)$$

(*e.g.*, TIBSHIRANI 1996), for some value of the smoothing parameter $\lambda(t) \geq 0$. It is known that the solution to (2) may involve zeroing out some elements of $\boldsymbol{\beta}$, and there are many ways of illustrating why this may be so. One manner is to examine the shape of the feasible set in (1) (*e.g.*, TIBSHIRANI 1996); another way is to consider the Bayesian interpretation of the LASSO. From (2), it follows that the solution can be viewed as the posterior mode in a Bayesian model with Gaussian likelihood, $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_\varepsilon^2) = \prod_{i=1}^n N(y_i \mid \mathbf{x}_i'\boldsymbol{\beta}, \sigma_\varepsilon^2)$, and a prior on $\boldsymbol{\beta}$ that is the product of p independent, zero-mean, double-exponential (DE) densities; that is, $p(\boldsymbol{\beta} \mid \lambda) = \prod_{j=1}^p (\lambda/2)\exp(-\lambda|\beta_j|)$. In contrast, BR is obtained by assuming the same likelihood and a prior on $\boldsymbol{\beta}$ that is the product of p independent normal densities; that is, $p(\boldsymbol{\beta} \mid \sigma_\beta^2) = \prod_{j=1}^p \left(\exp\left(-(\beta_j^2/2\sigma_\beta^2)\right)/\sqrt{2\pi\sigma_\beta^2}\right)$, where σ_β^2 is a variance parameter common to all regression coefficients. The difference between these two priors is illustrated in Figure 1: the DE density places more mass at zero and has thicker tails than the Gaussian distribution. From this perspective, relative to BR, LASSO produces stronger shrinkage of regression coefficients that are close to zero and less shrinkage of those with large absolute values.

Parameter λ, sometimes referred to as a regularization parameter, plays a central role: as it approaches zero, the solution to (2) tends to ordinary least squares, while large values of λ penalize the L_1 norm of $\boldsymbol{\beta}$, $\sum_j |\beta_j|$, highly. In the Bayesian view of LASSO, λ controls the prior on $\boldsymbol{\beta}$, with large values of this parameter associated with more informative (sharper) priors.

By construction, the non-Bayesian LASSO solution admits at most $n - 1$ nonzero regression coefficients (*e.g.*, PARK and CASELLA 2008). This is not desirable in
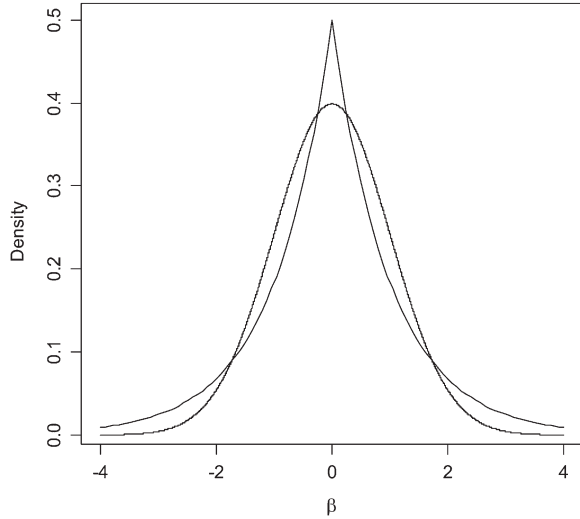
FIGURE 1.—Densities of a normal and of a double-exponential distribution (both with null mean and with unit variance).

models with dense marker-based regressions since, *a priori*, there is no reason why the number of markers with effectively nonzero effects should be smaller than the number of observations. This problem does not arise in BL, which is discussed next.

A computationally convenient hierarchical formulation of a DE distribution is obtained by exploiting the fact that the DE density can be represented as a mixture of scaled Gaussian densities (*e.g.*, ANDREWS and MALLOWS 1974; ROSA 1999), where the mixing process of the variances is an exponential distribution. Following PARK and CASELLA (2008),

$$\beta_j \sim \mathrm{DE}(\beta_j \mid \lambda) = \frac{\lambda}{2} e^{-\lambda |\beta_j|}$$
$$= \int_0^\infty \left[\frac{\exp(-(\beta_j^2/2\sigma_j^2))}{\sqrt{2\pi\sigma_j^2}} \right] \left[\frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2}{2}\sigma_j^2 \right) \right] d\sigma_j^2.$$

Above, β_j is the unknown effect of the jth marker and σ_j^2 is a variance parameter (measuring prior uncertainty) associated with β_j. Using this, PARK and CASELLA (2008) suggested the following hierarchical model (BL),

$$\text{Likelihood: } p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_\varepsilon^2) = \prod_{i=1}^n N(y_i \mid \mathbf{x}_i'\boldsymbol{\beta}, \sigma_\varepsilon^2) \qquad (3)$$

$$\text{Prior: } p(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\tau}^2, \lambda^2) = p(\boldsymbol{\beta} \mid \sigma_\varepsilon^2, \boldsymbol{\tau}^2) p(\sigma_\varepsilon^2) p(\boldsymbol{\tau}^2 \mid \lambda) p(\lambda^2)$$
$$= \left[\prod_{j=1}^p N(\beta_j \mid 0, \tau_j^2 \sigma_\varepsilon^2) \right] \chi^{-2}(\sigma_\varepsilon^2 \mid \text{d.f.}, S)$$
$$\times \left[\prod_{j=1}^p \mathrm{Exp}(\tau_j^2 \mid \lambda) \right] G(\lambda^2 \mid \alpha_1, \alpha_2). \qquad (4)$$

Above, $N(y_i \mid \mathbf{x'}_i\boldsymbol{\beta}, \sigma_\varepsilon^2)$ and $N\left(\beta_j \mid 0, \sigma_\varepsilon^2 \tau_j^2\right)$ are normal densities centered at $\mathbf{x'}_i\boldsymbol{\beta}$ and 0, with variances σ_ε^2 and

$\sigma_\varepsilon^2 \tau_j^2$, respectively; $\chi^{-2}(\sigma_\varepsilon^2 \mid \text{d.f.}, S)$ is a scaled-inverted chi-square density, with degrees of freedom d.f. and scale S, in this parameterization, $E(\sigma_\varepsilon^2 \mid \text{d.f.}, S) = S/(\text{d.f.} - 2); \mathrm{Exp}(\tau_j^2 \mid \lambda) = (\lambda^2/2)\exp(-(\lambda^2/2)\tau_j^2)$ is an exponential density indexed by a single parameter, λ, and $G(\lambda^2 \mid \alpha_1, \alpha_2)$ is a Gamma distribution, with shape parameter α_1 and rate parameter α_2.

The role of the τ_j^2's becomes more clear by changing variables in (3) and (4) from β_j to $\beta_j^* = \tau_j^{-1}\beta_j$. After this change of variables, the product of the likelihood function and of the joint prior for the regression coefficients, $N(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma_\varepsilon^2) N(\boldsymbol{\beta} \mid \mathrm{Diag}\{\tau_j^2\}\sigma_\varepsilon^2)$, becomes $N(\mathbf{y} \mid \mathbf{X}^*\boldsymbol{\beta}^*, \mathbf{I}\sigma_\varepsilon^2) N(\boldsymbol{\beta}^* \mid \mathbf{I}\sigma_\varepsilon^2)$, where $\mathbf{X}^* = \mathbf{X}\,\mathrm{Diag}\{\tau_j\}$ and $\boldsymbol{\beta}^*$ is a vector of regression coefficients with homogeneous variance. Thus, one way of viewing this class of regression models is as a standard BR model with additional unknowns, $\mathrm{Diag}\{\tau_j\}$, which assign different weights to the columns of $\mathbf{X}$, with $\tau_j \to 0$ being equivalent to removing the jth covariate from the model.

PARK and CASELLA (2008) presented a set of fully conditional distributions that allows fitting the BL model via the Gibbs sampler. Some of these distributions are discussed next, to illustrate main features of the algorithm.

Location parameters: In the Gibbs sampler of PARK and CASELLA (2008), the fully conditional distribution of the regression coefficients is multivariate normal with mean (covariance matrix) equal to the solution (inverse of the coefficient matrix) of the system of equations,

$$\left[\mathbf{X}'\mathbf{X}\sigma_\varepsilon^{-2} + \mathrm{Diag}(\tau_j^{-2}\sigma_\varepsilon^{-2}) \right]\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}\sigma_\varepsilon^{-2}. \qquad (5)$$

Recall that ordinary least-squares estimates are obtained by solving $\mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and that the counterpart of (5) in BR is $[\mathbf{X}'\mathbf{X}\sigma_\varepsilon^{-2} + \mathbf{I}\sigma_\beta^{-2}]\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}\sigma_\varepsilon^{-2}$. A key aspect of BL is that it produces a shrinkage that is marker specific, contrary to BR. Since σ_ε^2 is a scaling factor common to all regression coefficients, the differential shrinkage is due to the τ_j^2's. If τ_j^2 is large, *i.e.*, a large variance is associated with the effect of the jth marker, the quantity added to the diagonal will be small. Conversely, if a small variance is associated with the effect of the jth coefficient, τ_j^{-2} will be large. Adding a large constant to the jth diagonal element shrinks the least-squares estimates toward zero and reduces the variance of its fully conditional distribution.

Variances of the regression coefficients: An important aspect of the algorithm is how samples of the regression coefficients affect realizations of the variances of marker effects. In BL, the fully conditional posterior distributions of the τ_j^{-2}'s can be shown to be inverse Gaussian (*e.g.*, CHHIKARA and FOLKS 1989), with mean $\mu_j = (\sigma_\varepsilon\lambda/|\beta_j|)$ and scale parameter λ^2. For a given $\sigma_\varepsilon\lambda$, a small absolute value of β_j will lead to a fully conditional distribution of τ_j^{-2} with a large mean, which in turn will generate relatively small values of τ_j^2.

The λ parameter of the exponential prior: In the standard LASSO, λ controls the trade-off between goodness of fit and model complexity, and this may be crucial in defining the ability of a model to uncover signal. Small values of λ produce better fit, in the sense of the residual sum of squares (λ = 0 gives ordinary least squares); as λ increases, the penalty on model complexity increases (in optimization problem (1) the feasible set is smaller). On the other hand, in BL, λ controls the shape of the prior distribution assigned to the τ_j^2's. In general, the exponential prior assigns more density to small values of the τ_j^2's than to large ones, and this may be reasonable for most SNPs under the expectation that most of their effects are nil.

In BL λ can be treated as any other unknown. If, as in (4), a Gamma prior is assigned to λ^2, the fully conditional posterior distribution of λ^2 is also Gamma, with shape and rate parameters equal to $p + \alpha_1$ and $\frac{1}{2}\sum_{j=1}^{p}\tau_j^2 + \alpha_2$, respectively. The expectation of this Gamma distribution is $(p + \alpha_1)/\frac{1}{2}\sum_{j=1}^{p}\tau_j^2 + \alpha_2$, so a large value of $\sum_{j=1}^{p}\tau_j^2$ will lead to a relatively small λ, and the opposite will occur if the sum of the variances of the regression coefficients is small.

Relationship between LASSO and other regression models used in genomic selection: Standard BR may not be suitable for regressing phenotypes on a large number of markers because shrinkage of regression coefficients is homogeneous across markers (FERNANDO *et al.* 2007). In contrast, in BL the variance is marker specific, producing shrinkage whose extent is related to the absolute value of the estimated regression coefficient.

MEUWISSEN *et al.* (2001) recognized that marker-specific variances may be needed and suggested regression models based on marginal priors that are also mixtures of scaled-Gaussian distributions ("Bayes A") or mixtures of scaled-Gaussian distributions and of a point mass at zero ("Bayes B"). In these models, the likelihood is as in (3) and, in Bayes A, the prior is

$$p(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\sigma}_\beta^2) = p(\boldsymbol{\beta} \mid \boldsymbol{\sigma}_\beta^2)p(\sigma_\varepsilon^2)p(\boldsymbol{\sigma}_\beta^2)$$

$$= \left[\prod_{j=1}^{p} N(\beta_j \mid 0, \sigma_{\beta_j}^2)\right]\chi^{-2}(\sigma_\varepsilon^2 \mid \text{d.f.}, S)$$

$$\times \left[\prod_{j=1}^{p}\chi^{-2}(\sigma_{\beta_j}^2 \mid \text{d.f.}_\beta, S_\beta)\right].$$

$$(6)$$

The first two components of (6), $p(\boldsymbol{\beta} \mid \boldsymbol{\sigma}_\beta^2)p(\sigma_\varepsilon^2) = [\prod_{j=1}^{p} N(\beta_j|0, \sigma_{\beta_j}^2)]\chi^{-2}(\sigma_\varepsilon^2|\text{d.f.}, S)$, are the counterparts of the first two components of (4), $p(\boldsymbol{\beta}|\boldsymbol{\tau}^2)p(\sigma_\varepsilon^2) = [\prod_{j=1}^{p} N(\beta_j|0, \tau_j^2\sigma_\varepsilon^2)]\chi^{-2}(\sigma_\varepsilon^2|\text{d.f.}, S)$, with $\sigma_{\beta_j}^2 = \tau_j^2\sigma_\varepsilon^2$. The difference between BL and Bayes A (or Bayes B) is how the priors of the variances of the marker-specific regression coefficients ($\sigma_{\beta_j}^2$ in Bayes A and $\tau_j^2\sigma_\varepsilon^2$ in BL) are specified. At this level, Bayes A and BL differ in two respects:

1. In Bayes A, the prior assumption is that the marker-specific variances are independent random variables following the same scaled-inverted chi-square distribution with known prior degree of belief (d.f.$_\beta$) and scale (S_β). In BL, the assumption is that these variances are independent as well, but each following the same exponential distribution with unknown parameter λ. The conditional (given λ) marginal prior in BL $p(\boldsymbol{\beta} \mid \lambda)$ is DE, while in Bayes A $p(\beta_j \mid \text{d.f.}_\beta, S_\beta)$ is a *t*-distribution. Although a *t*-distribution may place more density at zero than the Gaussian prior of BR, the density at zero is larger in the DE. This issue was recognized by MEUWISSEN *et al.* (2001), leading to the development of Bayes B.

   XU (2003) employed an improper prior for the marker-specific variances; if d.f.$_\beta = 0$ and $S_\beta = 0$, then $p(\sigma_{\beta_j}^2) = \sigma_{\beta_j}^{-2}$. Similar to the exponential prior, this density decreases monotonically with $\sigma_{\beta_j}^2$. However, unlike the exponential distribution, where λ can be used to "tune" the shape of the distribution, this prior does not have parameters to allow any control. In addition, as noted by TER BRAAK *et al.* (2005), $p(\sigma_{\beta_j}^2) = \sigma_{\beta_j}^{-2}$ yields an improper posterior. As an alternative, these authors suggested to use $p(\sigma_{\beta_j}^2) = \sigma_{\beta_j}^{-2(\delta-1)}$ with $0 < \delta \le \frac{1}{2}$; although this prior is improper, it does not yield an improper posterior. As with the exponential prior, $p(\sigma_{\beta_j}^2) = \sigma_{\beta_j}^{-2(\delta-1)}$ is decreasing with respect to $\sigma_{\beta_j}^2$. TER BRAAK (2006) furthered discussed the role of δ, which, as λ in the BL, controls the shape of the prior density on the variance of the regression coefficients.

2. A second difference is that, in Bayes A, values of parameters d.f.$_\beta$ and S_β are specified as known *a priori*. On the other hand, in BL there is an extra level in the model: λ^2 is assigned a Gamma distribution, and information from all regression coefficients is pooled. This difference is illustrated in Figure 2: in Bayes A, d.f.$_\beta$ and S_β control, as λ does in BL, the trade-offs between goodness of fit and model complexity.

YI and XU (2008) discuss an extension of Bayes A where a prior is assigned to d.f.$_\beta$ and S_β, and these quantities are treated as nuisances, as λ is in BL. However, as argued earlier, the DE seems to be a better choice, if the assumption is that most markers have no effect on the trait of interest.

MONTE CARLO STUDY

Although in the BL λ can be treated as unknown, it is not clear how sensitive results might be with respect to the choice of hyperparameters α_1 and α_2. PARK and CASELLA (2008, p. 683) recognized that this may be an issue: "The prior density for λ^2 should approach 0 sufficiently fast as $\lambda^2 \to \infty$ (to avoid mixing problems) but should be relatively flat and place high probability near the maximum likelihood estimate." The main
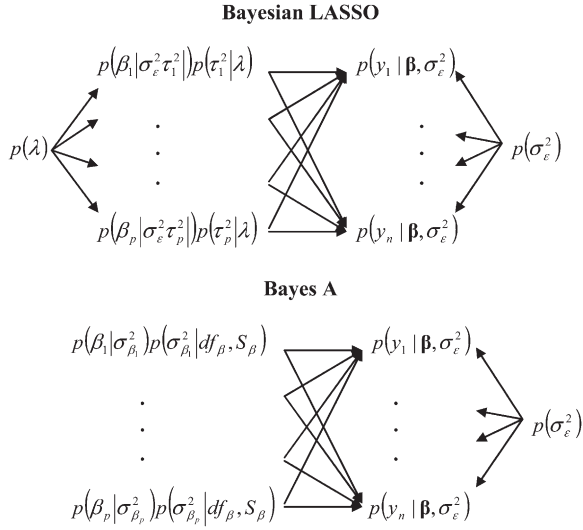
Bayesian LASSO

Bayes A

FIGURE 2.—Graphical representation of the hierarchical structure of the Bayesian LASSO (top) and Bayes A (bottom). In the Bayesian LASSO, the variances of the marker effects are $\sigma_\varepsilon^2\tau_j^2$, $j = 1, \ldots, p$, with counterparts $\sigma_{\beta_j}^2$ in Bayes A.

problem of applying this recommendation is that one does not know in advance what the maximum-likelihood estimate is.

The sensitivity of BL with respect to the choice of the prior distribution of λ^2 was investigated here by fitting the model under different priors to simulated data. In addition to the conjugate Gamma prior, we also considered (see File S1 and File S2)

$$p(\lambda \mid \alpha_3, \alpha_4, \max) \propto \text{Beta}\left(\frac{\lambda}{\max} \,\bigg|\, \alpha_3, \alpha_4\right). \quad (7)$$

The above distribution gives great flexibility for specifying a relatively flat prior over a wide range of values. The uniform prior appears as a special case when $\alpha_3 = \alpha_4 = 1$. When the Beta prior is used, the fully conditional distribution of λ does not have closed form; however, draws from the distribution can be obtained using the Metropolis–Hastings algorithm (see File S1 and File S2).

Data-generating process: Data were simulated in a simple setting, such that problems could be identified easily, while the phenotypic and genotypic structure attempted to resemble those encountered in real data sets.

Data were generated under the additive model,

$$y_i = \sum_{j=1}^{280} x_{ij}\beta_j + \varepsilon_i,$$

where y_i ($i = 1, \ldots, 250$) is the phenotype for individual i, β_j is the effect of allele substitution at marker j ($j = 1, \ldots, 280$), and x_{ij} is the code for the genotype of subject i at locus j, $x_{ij} \in \{0, 1, 2\}$. Residuals were

TABLE 1

Absolute values of the correlation between marker genotypes (average across markers and 100 Monte Carlo simulations) by scenario (X_0, low linkage disequilibrium; X_1, high linkage disequilibrium)

Scenario	Adjacency between markers			
	1	2	3	4
X_0	0.091	0.089	0.090	0.001
X_1	0.754	0.602	0.479	0.381

independently sampled from a standard normal distribution; that is, $\varepsilon_i \sim N(0, 1)$.

Two scenarios regarding the genotypic distribution were considered. In scenario X_0, markers were in low linkage disequilibrium (LD), with almost no correlation between adjacent markers (Table 1). In scenario X_1 a relatively high LD was considered (Table 1).

The effects of allele substitutions (β) were kept constant across simulations and set to zero for all markers except for 10 (Figure 3). The locations of markers with nonnull effects were chosen such that different situations regarding effects of linked markers were represented.

Choice of prior distribution of λ: For each Monte Carlo (MC) replicate, five variations of BL were fitted, and four involved a Gamma prior λ^2 with the following values of parameters: BL1, $\{\alpha_1 = 1; \alpha_2 = 1 \times 10^{-2}\}$; BL2, $\{\alpha_1 = 2; \alpha_2 = 2 \times 10^{-3}\}$; BL3, $\{\alpha_1 = 4; \alpha_2 = 1.5 \times 10^{-3}\}$; BL4, $\{\alpha_1 = 6; \alpha_2 = 1 \times 10^{-3}\}$. In BL5, the prior on λ was $p(\lambda) \propto \text{Beta}(\lambda/100) \mid \alpha_3 = 1.4, \alpha_4 = 1.4)$ (Figure 4).

Results: Table 2 shows the average (across 100 MC replicates) of posterior means of the residual variance, the regularization parameter, and the correlation between the true and the estimated quantity of several features (phenotypes, genomic values, and marker effects). $\text{Corr}(\mathbf{y}, \mathbf{X}\hat{\boldsymbol{\beta}})$ is a goodness-of-fit measure, $\text{Corr}(\mathbf{X}\boldsymbol{\beta}, \mathbf{X}\hat{\boldsymbol{\beta}})$ measures how well the model estimates genomic values, and $\text{Corr}(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}})$ evaluates how well the model estimates marker effects.

The posterior mean and standard deviation of λ were influenced by the prior (Table 2). The posterior mean was shrunk toward the prior mode, and the posterior standard deviation was larger for more dispersed priors (see Table 2 and Figure 4). These results suggest that there is not much information about λ in the type of samples evaluated. On the other hand, model goodness of fit and the ability of the model to uncover signal were not affected markedly by the choice of prior. This suggest that, while it may be difficult to learn about λ from data, inferences on quantities of interest (*e.g.*, genetic values) may be robust with respect to values of λ over a fairly wide range. For example, differences in
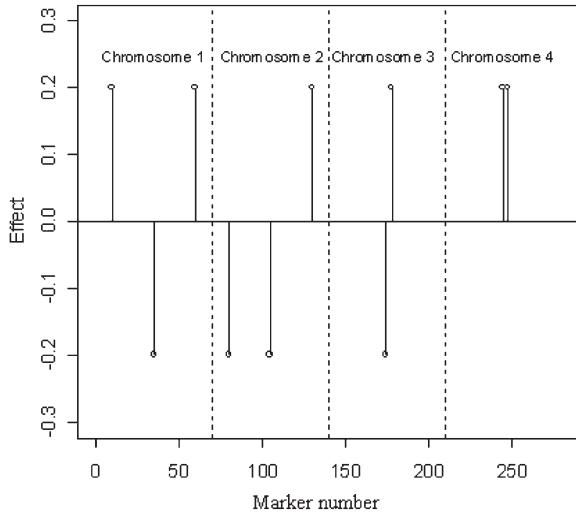
FIGURE 3.—Positions (chromosome and marker number) and effects of markers (there were 280 markers, with 270 with no effect).

Corr$(\mathbf{X\beta}, \mathbf{X\hat{\beta}})$ or in Corr$(\mathbf{\beta}, \mathbf{\hat{\beta}})$ were small when the prior was changed.

A relatively flat prior based on a Beta distribution (BL5) produced a more dispersed posterior distribution of λ, and mixing was not as good as when the sharper Gamma priors (BL1–BL4) were used. For example, the average (across MC replicates) effective sample sizes (*e.g.*, PLUMMER *et al.* 2008) for the residual variance were 1468, 1155, 1091, 1138, and 578 for BL1–BL5, respectively.

BAYESIAN REGRESSION COUPLED WITH LASSO

In practice, the information set available for prediction of genomic values may include components other than genetic markers. For example, data may cluster into known contemporary groups (*e.g.*, individuals may be measured under different experimental conditions), or a pedigree may be available in addition to genetic markers. It is natural to treat the various classes of predictors in a different way. From a penalized-likelihood point of view, this amounts to using penalty functions that are specific to each class of predictors. From a Bayesian standpoint, treating predictors differently may be achieved by assigning different priors. A straightforward extension of the BL is described next.

The data structure is denoted as $\{y_i, \mathbf{x}_{r_i}, \mathbf{x}_{l_i}, \mathrm{ID}_i\}_{i=1}^n$, where y_i is the phenotype of subject i, $\mathbf{x}_{r_i}$ is a vector of covariates that is treated as in a standard BR with a normal prior and variance common to all regressions, $\mathbf{x}_{l_i}$ is a set of covariates whose effects are assigned a double-exponential prior as in BL, and ID_i is a label that allows tracking subjects in a pedigree. The equation for the data is

$$y_i = \mu + \mathbf{x}_{r_i}'\mathbf{\beta}_r + \mathbf{x}_{l_i}'\mathbf{\beta}_l + u_i + \varepsilon_i,$$

where μ is an intercept, $\mathbf{\beta}_r$ and $\mathbf{\beta}_l$ are regressions of y_i on $\mathbf{x}_{r_i}'$ and $\mathbf{x}_{l_i}'$, respectively, u_i is an infinitesimal genetic
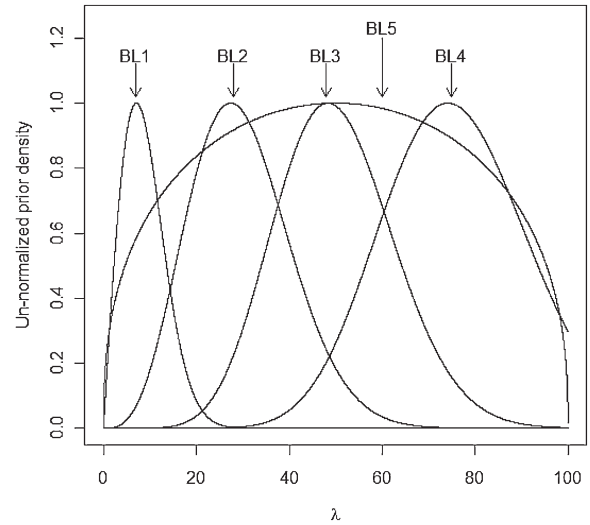

FIGURE 4.—Unnormalized density of the five priors evaluated in the MC study (BL1–BL4 use Gamma priors on λ^2, and BL5 uses a prior for λ based on a Beta distribution; the densities in this figure are the corresponding densities for λ).

effect pertaining to individual i for which the prior (co)variance structure is determined by a pedigree, and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is a model residual, assumed to be identically and independently distributed of other residuals. The likelihood function is

$$p(\mathbf{y} \mid \mu, \mathbf{\beta}_r, \mathbf{\beta}_l, u_i, \sigma_\varepsilon^2)$$
$$= \prod_{i=1}^n N(y_i \mid \mu + \mathbf{x}_{r_i}'\mathbf{\beta}_r + \mathbf{x}_{l_i}'\mathbf{\beta}_l + u_i, \sigma_\varepsilon^2). \quad (8)$$

Prior specification (4) is modified as

$$p(\mu, \mathbf{\beta}_r, \mathbf{\beta}_l, \mathbf{u}, \sigma_\varepsilon^2, \sigma_r^2, \sigma_u^2, \mathbf{\tau}^2, \lambda)$$
$$= N(\mu \mid 0, \sigma_\mu^2)N(\mathbf{\beta}_r \mid \mathbf{0}, \mathbf{I}\sigma_r^2)\prod_{j=1}^p N(\beta_{jl} \mid 0, \sigma_\varepsilon^2\tau_j^2)N(\mathbf{u} \mid \mathbf{0}, \mathbf{A}\sigma_u^2)$$
$$\times \chi^{-2}(\sigma_\varepsilon^2 \mid S_\varepsilon, \mathrm{d.f.}_\varepsilon)\chi^{-2}(\sigma_r^2 \mid S_r, \mathrm{d.f.}_r)\chi^{-2}(\sigma_u^2 \mid S_u, \mathrm{d.f.}_u)$$
$$\times \left[\prod_{j=1}^p \mathrm{Exp}(\tau_j^2 \mid \lambda)\right]p(\lambda), \quad (9)$$

where σ_μ^2, σ_r^2, and σ_u^2 are the variances of μ, β_{jr}, and u_i, respectively; d.f. and S are prior degrees of freedom and scale parameter of the corresponding distributions; $\mathbf{A}$ is a (co)variance structure computed from the genealogy (for example, a numerator-relationship matrix); and, $p(\lambda)$ is the prior on λ that may be as in (4) or (7).

In the model defined by (8) and (9) all fully conditional distributions (except that of λ if a non-conjugate prior is chosen for λ) have closed form, so a Gibbs sampler (with a Metropolis–Hastings step) can be used to draw samples from the joint posterior distribution (see File S1 and File S2). To distinguish the above model from the standard BL we refer to it as Bayesian regression coupled with LASSO (BRL).

TABLE 2

Posterior mean of residual variance, σ_ε^2, regularization parameter, λ, and correlation between the true and estimated value for several items (y, phenotypes; $X\beta$, true genomic value; β, marker effects; all quantities averaged over 100 MC replicates)

| | σ_ε^2 | | λ | | $\text{Corr}(\mathbf{y}, \mathbf{X}\hat{\boldsymbol{\beta}})$ | | $\text{Corr}(\mathbf{X}\boldsymbol{\beta}, \mathbf{X}\hat{\boldsymbol{\beta}})$ | | $\text{Corr}(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}})$ | |
	Mean[a]	SD[b]	Mean[a]	SD[b]	Mean[c]	SD[b]	Mean[c]	SD[b]	Mean[c]	SD[b]
				Low linkage disequilibrium between markers (X_0)						
BL1	0.909	0.088	24.110	1.883	0.640	0.039	0.657	0.070	0.328	0.062
BL2	0.990	0.095	35.986	4.944	0.615	0.044	0.661	0.073	0.332	0.061
BL3	1.035	0.096	45.506	6.755	0.601	0.045	0.660	0.074	0.329	0.059
BL4	1.093	0.098	63.748	9.200	0.584	0.044	0.657	0.077	0.323	0.056
BL5	1.006	0.104	41.920	11.570	0.610	0.050	0.660	0.073	0.330	0.059
				High linkage disequilibrium between markers (X_1)						
BL1	0.917	0.088	23.770	1.621	0.607	0.039	0.699	0.071	0.294	0.050
BL2	0.991	0.093	36.168	4.163	0.569	0.044	0.705	0.074	0.300	0.046
BL3	1.030	0.095	45.572	5.735	0.551	0.045	0.704	0.076	0.296	0.044
BL4	1.078	0.097	62.373	8.435	0.529	0.046	0.700	0.079	0.291	0.043
BL5	1.003	0.099	41.072	9.724	0.564	0.051	0.704	0.074	0.299	0.046

[a] Mean (across 100 MC replicates) of the posterior mean.
[b] Between-replicate standard deviation of the estimate.
[c] Mean (across MC replicates) of the correlation evaluated at the posterior mean of $\boldsymbol{\beta}$. The priors on λ^2 were as follows: BL1, $p(\lambda^2) = \text{Gamma}(\alpha_1 = 1; \alpha_2 = 1 \times 10^{-2})$; BL2, $p(\lambda^2) = \text{Gamma}(\alpha_1 = 2; \alpha_2 = 2 \times 10^{-3})$; BL3, $p(\lambda^2) = \text{Gamma}(\alpha_1 = 4; \alpha_2 = 1.5 \times 10^{-3})$; and BL4, $\text{Gamma}(\alpha_1 = 6; \alpha_2 = 1 \times 10^{-3})$. In BL5, $p(\lambda) \propto \text{Beta}((\lambda/100) \mid \alpha_3 = 1.4, \alpha_4 = 1.4)$.

DATA ANALYSIS

Two data sets were analyzed with the BRL model. The first set pertains to a collection of wheat lines (see File S1 and File S2); the second set contains information from a population of mice (publicly available at http://gscan.well.ox.ac.uk).

The wheat data set is from the Global Wheat program of the International Maize and Wheat Improvement Center (CIMMYT). This program conducted several international trials across a wide variety of environments. For this study, we took a subset of 599 wheat lines derived from 25 years of Elite Spring Wheat Yield Trials (ESWYT) conducted from 1979 through 2005. The environments represented in these trials were grouped into four macroenvironments. The phenotype consid-

ered here was average grain yield performance of the 599 wheat lines evaluated in one of the macroenvironments. An association mapping study based on a reduced number of these ESWYT trials is presented in CROSSA *et al.* (2007).

The Browse application of the International Crop Information System (ICIS), as described in http://cropwiki.irri.org/icis/index.php/TDM_GMS_Browse (MCLAREN *et al.* 2005), was used for deriving the relationship matrix **A** between the 599 lines, and it accounts for selection and inbreeding.

A total of 1447 Diversity Array Technology (DArT) markers were generated by Triticarte (Canberra, Australia; http://www.triticarte.com.au). The DArT markers may take on two values, denoted by their presence or their absence.

TABLE 3

Posterior means (standard deviations) of variance components for yield in wheat and body-mass index in mice, and of λ for each of the models, by data set

		σ_ε^2	σ_u^2	σ_r^2	λ
Wheat	P	0.561 (0.058)	0.294 (0.056)	—	—
	M	0.546 (0.046)	—	—	19.33 (2.43)
	P&M	0.410 (0.055)	0.139 (0.046)	—	18.32 (2.86)
Mice	P	0.754 (0.038)	0.092 (0.041)	0.156 (0.030)	—
	M	0.741 (0.037)	—	0.153 (0.026)	223.84 (25.78)
	P&M	0.723 (0.032)	0.021 (0.009)	0.153 (0.026)	217.90 (26.26)

P, infinitesimal models using pedigree; M, model including regressions on markers, but not pedigree information; P&M, model including an infinitesimal additive effect and regressions on markers; σ_ε^2, σ_u^2, σ_r^2, and λ, residual variance, variance of the infinitesimal additive effect, variance of cage-effects, and smoothing parameter of the BL regression, respectively.
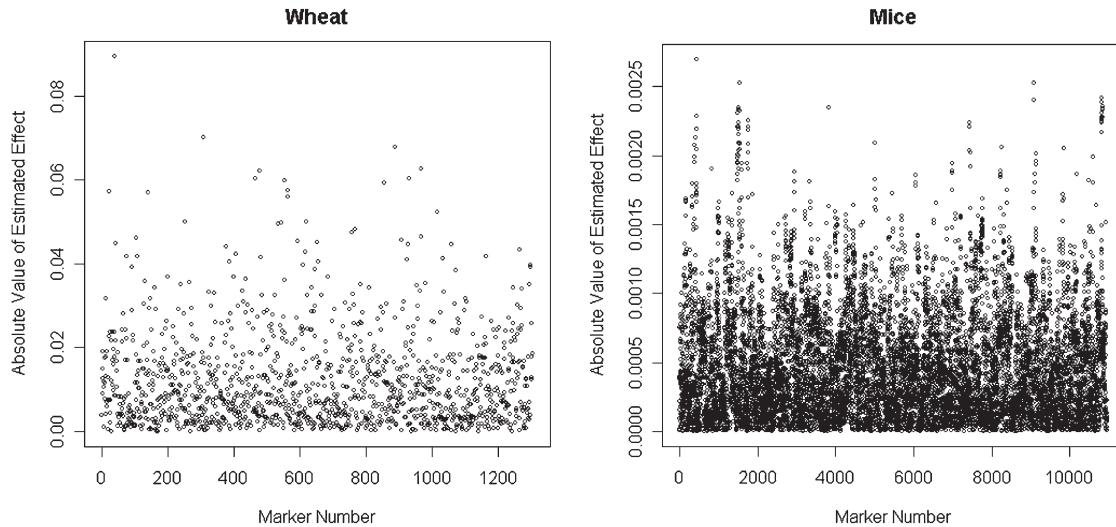
FIGURE 5.—Absolute values of the posterior means of effects of allele substitution in a model including markers and pedigree information (P&M), by data set.

The mouse data come from an experiment carried out to detect and locate QTL for complex traits in a mouse population (VALDAR *et al.* 2006a,b). These data have already been analyzed for comparing genome-assisted genetic evaluation methods (LEGARRA *et al.* 2008). The data file consists of 1884 individuals (168 full-sib families), each genotyped for 10,946 polymorphic markers. The trait analyzed here was body mass index (BMI), precorrected by body weight, season, month, and day. Mice were housed in 359 cages; on average, each litter was allocated into 2.84 cages.

Three models were fitted to each of the data sets: P (standing for pedigree) is a pedigree-based model where markers were not included; M is a model where the only genetic component is the regression on markers; P&M (standing for pedigree and markers) includes regressions on markers and an additive effect with (co)variance structure computed from the pedigree. For both data sets, phenotypes were standardized to have a sample variance equal to one, so that results are easily compared across data sets.

In the mouse data set, β_r were the effects of cages where groups of mice were reared. In the wheat data set, the component $x'_{r_i}\beta_r$ was omitted because there was no such set of regressors.

Models were first fitted to the entire data set. Subsequently, a fivefold cross-validation (CV) was carried out with assignment of individuals into folds at random. The CV yields prediction of phenotypes $\hat{y}_{i,-f}$ ($f = 1, \ldots, 5$) obtained from a model in which all observations in the fth fold were excluded. The ability of each model to predict out-of-sample data was evaluated via the correlation between phenotypes and predictions from CV. Inferences for each fit were based on 70,000 samples (after 5000 were discarded as burn-in). Convergence was checked by inspection of trace plots and with

estimates of effective sample size for (co)variance components computed using the coda package of R (PLUMMER *et al.* 2008). Parameters of the prior distributions were $S_\varepsilon = d.f._\varepsilon = S_r = d.f._r = S_u = d.f._u = 1$ and $p(\lambda) \propto \text{Beta}((\lambda/400)\,|\,\alpha_3 = 1.4,\ \alpha_4 = 1.4)$. This latter prior is flat over a wide range of values of λ.

RESULTS AND DISCUSSION

Table 3 shows summaries of the posterior distributions of the variance components and of λ by model and data set. In both populations, a moderate reduction in the posterior mean of the residual variance was observed when the P&M model was fitted, relative to P. Using model P in the wheat population gave a posterior mean of heritability of grain yield of 0.34, while in the mouse population the posterior mean of h^2 of body-mass index was 0.11. These results are in agreement with previous reports (VALDAR *et al.* 2006b and LEGARRA *et al.*

TABLE 4

Rank correlation (Spearman) between genetic values estimated from models including different sources of genetic information (pedigree, markers, and pedigree and markers), by data set (mouse data set above diagonal, wheat data set below diagonal)

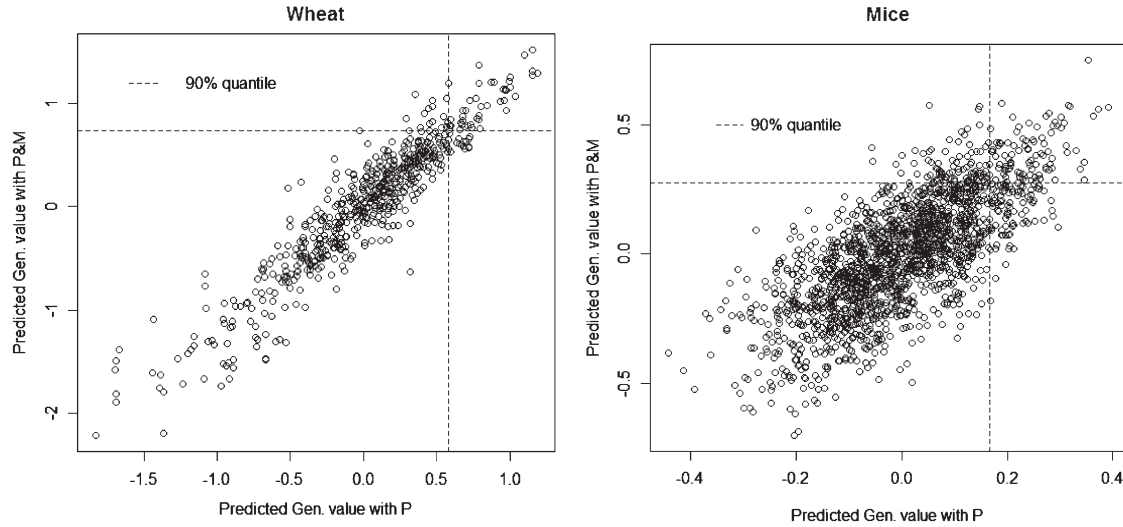| Source of genetic information	Source of genetic information		
	Pedigree	Marker	Pedigree and markers
Pedigree	—	0.715	0.729
Marker	0.802	—	0.986
Pedigree and markers	0.927	0.944	—

FIGURE 6.—Predicted genetic value using markers and pedigree (P&M) *vs.* using pedigree only (P), by data set.

2008 for the mouse data and CROSSA *et al.* 2007 for the wheat data) for these traits and populations. The inclusion of markers (P&M) reduced the estimate of the variance of the infinitesimal additive effect, relative to P. This happens because, in P&M, part of the infinitesimal additive effect is captured by the regression on markers (*e.g.*, HABIER *et al.* 2007; BINK *et al.* 2008). In the model for body-mass index in M, the variance between cages (σ_r^2) was reduced only slightly when the effects of the markers were fitted.

Figure 5 gives absolute values of the posterior means of marker effects. In the mouse data there are several regions showing groups of markers with relatively large estimated effects. This is not evident in the wheat data set where fewer markers were available.

From a breeder's perspective, a relevant question is whether or not the P, M, and P&M models lead to different ranking of individuals on the basis of the estimated genetic values. Table 4 shows the rank (Spearman) correlation of estimated genetic values. As expected, these correlations were high, but not perfect. The correlation between predicted genetic values from M and P&M was

larger than that of the estimates from P and P&M, suggesting that the inclusion of markers in the model is probably critical. This was clearer in the mouse data set, where (a) the extent of additive relationships was not as strong as in the wheat population and (b) a much larger number of markers were available.

Figure 6 shows scatter plots of predicted genomic values in P and P&M for both data sets. Although the correlation between genetic values estimated from different models was high, using P and P&M would lead to different sets of selected individuals. The difference was more marked in the mouse data set, illustrating that the impact of considering markers in breeding decisions depends on the data structure and on how informative the pedigree and markers are. Also, the dispersion of predicted genetic values was larger when markers were fitted, and this is consistent with the smaller posterior mean of the residual variance observed for P&M (Table 3). An interpretation of this result is that, in certain contexts, markers may help to uncover genetic variance that would not be captured if only pedigree-based predictions were used.

TABLE 5

Rank correlation (Spearman) between phenotypic values or corrected phenotypic records and predictions from cross-validation, by population and model (P, pedigree-based model; P&M, pedigree and marker information)

	Wheat			Mice		
	P	M	P&M	P	M	P&M
$\text{Corr}[\mathbf{y}, \hat{\mathbf{y}}']^a$	0.408	0.423	0.462	0.263	0.306	0.300
$\text{Corr}[\hat{\mathbf{y}}*, \hat{\mathbf{g}}']^b$	0.423	0.594	0.602	0.109	0.211	0.225

[a] $\hat{\mathbf{y}}' = [E(\mathbf{y}_1 \mid \hat{\boldsymbol{\theta}}_{-1})', \ldots, E(\mathbf{y}_5 \mid \hat{\boldsymbol{\theta}}_{-5})']$, where $\hat{\boldsymbol{\theta}}_{-f}$ is the posterior mean from a model with the observations in the fth fold excluded.
[b] $\hat{\mathbf{y}}*' = [(\mathbf{y}_1 - \mathbf{1}\hat{\mu}_{-1} - \mathbf{X}\hat{\boldsymbol{\beta}}_{r,-1})', \ldots, (\mathbf{y}_5 - \mathbf{1}\hat{\mu}_{-5} - \mathbf{X}\hat{\boldsymbol{\beta}}_{r,-5})']$, where $\hat{\mu}_{-f}$, $\hat{\boldsymbol{\beta}}_{r,-f}$ are the posterior mean of the intercept and of the regression coefficients of systematic effects from a model with the observations of the fth fold excluded; and $\hat{\mathbf{g}}$ is the posterior mean, from CV, of genetic effects (sum of additive value and genomic value in P&M).

The aforementioned results indicate that incorporation of markers into a genetic model can influence inferences and breeding decisions. In contrast, cross-validation allows comparing models from the standpoint of their ability to predict future outcomes. Table 5 shows the correlation between phenotypic records and predictions from cross-validation. Two CV correlations were considered:

1. $\text{Corr}[\mathbf{y}, \hat{\mathbf{y}}']$ is the correlation between phenotypic records and their prediction from CV. That is, $\hat{\mathbf{y}} = \mathbf{1}\hat{\mu} + \mathbf{X}_r\hat{\boldsymbol{\beta}}_r + \hat{\mathbf{g}}$, where $\hat{\mathbf{g}} = \hat{\mathbf{u}}$ in P, $\hat{\mathbf{g}} = \mathbf{X}_l\hat{\boldsymbol{\beta}}_l$ in M, and $\hat{\mathbf{g}} = \hat{\mathbf{u}} + \mathbf{X}_l\hat{\boldsymbol{\beta}}_l$ in P&M.
2. $\text{Corr}[\hat{\mathbf{y}}^*, \hat{\mathbf{g}}']$ is the correlation between the CV estimate of the genetic value and phenotypic records adjusted with CV estimates of nongenetic effects. That is, $\hat{\mathbf{y}}^* = \mathbf{y} - \mathbf{1}\hat{\mu}$ in the wheat population, and $\hat{\mathbf{y}}^* = \mathbf{y} - \mathbf{1}\hat{\mu} - \mathbf{X}_r\hat{\boldsymbol{\beta}}_r$ in the mouse data set.

Overall, P&M models had better predictive ability than models based on pedigrees or markers only. In the wheat data set, the increases in the correlation observed when markers were included in the model were 13% for $\text{Corr}[\mathbf{y}, \hat{\mathbf{y}}']$ and 42% for $\text{Corr}[\hat{\mathbf{y}}^*, \hat{\mathbf{g}}']$. In the mouse data set the relative increases in correlation were 14 and 100% for $\text{Corr}[\mathbf{y}, \hat{\mathbf{y}}']$ and $\text{Corr}[\hat{\mathbf{y}}^*, \hat{\mathbf{g}}']$, respectively. We conclude that there are sizable benefits from using markers for breeding decisions and that the relative impact of the contribution depends upon data structure and on how informative the pedigree and the set of markers are.

CONCLUDING REMARKS

Additive models with infinitesimal effects are ubiquitous in animal and plant breeding. For many decades, predictions of genetic values have been made using phenotypic records and pedigrees, *i.e.*, some sort of family-based evaluation. Markers capture Mendelian segregation and may enhance prediction of genomic values, independently of the mode of gene action.

With highly dense markers, marker-specific shrinkage may be needed. Priors on marker effects based on mixtures of scaled-Gaussian distributions allow this type of shrinkage and constitute a promising tool for genomic-based additive models. This family of models includes, among others, the *t* or DE distributions. Models based on marginal priors that belong to the *t* family have been proposed for marker-based regressions (*e.g.*, MEUWISSEN *et al.* 2001).

If the hypothesis that most markers do not have any effect holds, a DE prior may be a better choice than the *t*. For this reason, the Bayesian LASSO appears to be an interesting alternative for performing regressions on markers, at least under an additive model.

Our results indicate that in the type of samples that are relevant for genomic selection (*i.e.*, $p \gg n$) the choice of prior for λ matters in terms of inferences about this unknown. However, estimates of genetic values and of marker effects may be robust with respect to the choice of prior, over a wide range. To circumvent the potential influence of the prior, we proposed an alternative formulation of the BL where the prior on λ is formulated using a Beta distribution. Unlike the Gamma prior, this prior allows expressing vague prior preferences over a wide range of values of λ.

Two data analyses carried out with the proposed model showed that (a) markers may allow capturing fractions of additive variance that would be lost if pedigrees are the only source of genetic information used, (b) considering markers has a sizable impact on selection decisions, and (c) models including marker and pedigree information had better predictive ability than pedigree-based or marker-based models.

LITERATURE CITED

ANDREWS, D. F., and C. L. MALLOWS, 1974 Scale mixtures of normal distributions. J. R. Stat. Soc. Ser. B **36:** 99–102.

BINK, M. C. A. M., P. UIMARI, M. J. SILLANPÄÄ, L. L. G. JANSS and R. C. JANSEN, 2002 Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor. Appl. Genet. **104:** 751–762.

BINK, M. C. A. M., M. P. BOER, C. J. F. TER BRAAK, J. JANSEN, R. E. VOORRIPS *et al.*, 2008 Bayesian analysis of complex traits in pedigreed populations. Euphytica **161:** 85–96.

CHHIKARA, R. S, and J. L. FOLKS, 1989 *The Inverse Gaussian Distribution: Theory, Methodology and Applications.* Marcel Dekker, NY.

CROSSA, J., J. BURGUEÑO, S. DREISIGACKER, M. VARGAS, S. A. HERRERA-FOESSEL *et al.*, 2007 Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics **177:** 1889–1913.

DE LOS CAMPOS, G., D. GIANOLA and G. J. M. ROSA, 2009 Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. J. Anim. Sci. (in press).

FERNANDO, R. L., and M. GROSSMAN, 1989 Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467–477.

FERNANDO, R. L., D. HABIER, C. STRICKER, J. C. M. DEKKERS and L. R. TOTIR, 2007 Genomic selection. Acta Agric. Scand. Sect. A **57:** 192–195.

GIANOLA, D., and J. B. VAN KAAM, 2008 Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics **178:** 2289–2303.

GIANOLA, D., M. PEREZ-ENCISO and M. A. TORO, 2003 On marker-assisted prediction of genetic value: beyond the ridge. Genetics **163:** 347–365.

GIANOLA, D., R. FERNANDO and A. STELLA, 2006 Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics **173:** 1761–1776.

HABIER, D., R. L. FERNANDO and J. C. M. DEKKERS, 2007 The impact of genetic relationship information on genome-assisted breeding values. Genetics **177:** 2389–2397.

HANS, C., 2008 Bayesian LASSO regression. Technical Report No. 810. Department of Statistics, Ohio State University, Columbus, OH. (http://www.stat.osu.edu/~hans/Papers/blasso.pdf).

LEGARRA, A., C. ROBERT-GRANIÉ, E. MANFREDI and J. M. ELSEN, 2008 Performance of genomic selection in mice. Genetics **180:** 611–618.

McLaren, C. G., R. Bruskiewich, A. M. Portugal and A. B. Cosico, 2005 The international rice information system. A platform for meta-analysis of rice crop data. Plant Physiol. **139:** 637–642.

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

Park, T., and G. Casella, 2008 The Bayesian LASSO. J. Am. Stat. Assoc. **103:** 681–686.

Plummer, M., N. Best, K. Cowles and K. Vines, 2008 coda: output analysis and diagnostics for MCMC. http://cran.r-project.org/web/packages/coda/index.html

R Development Core Team, 2008 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Rosa, G. J. M., 1999 Robust mixed linear models in quantitative genetics: Bayesian analysis via Gibbs sampling. International Symposium on Animal Breeding and Genetics, September 21–24, Viçosa, Minas Gerais, Brazil, pp. 133–159.

Ter Braak, C. J. F., 2006 Bayesian sigmoid shrinkage with improper variance priors and an application to wavelet denoising. Comput. Stat. Data Anal. **51:** 1232–1242.

Ter Braak, C. J. F, M. P. Boer and M. C. A. M. Bink, 2005 Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. Genetics **170:** 1435–1438.

Tibshirani, R., 1996 Regression shrinkage and selection via the LASSO. J. R. Stat. Soc. B **58:** 267–288.

Valdar, W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman et al., 2006a Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet. **38:** 879–887.

Valdar, W., L. C. Solberg, D. Gauguier, W. O. Cookson, J. N. P. Rawlins et al., 2006b Genetic and environmental effects on complex traits in mice. Genetics **174:** 959–984.

Wang, W. Y., B. J. Barratt, D. G. Clayton and J. A. Todd, 2005 Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. **6:** 109–118.

Xu, S., 2003 Estimating polygenic effects using markers of the entire genome. Genetics **163:** 789–801.

Yi, N., and S. Xu, 2008 Bayesian LASSO for quantitative trait loci mapping. Genetics **179:** 1045–1055.

Communicating editor: E. Arjas

GENETICS

Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree

Gustavo de los Campos, Hugo Naya, Daniel Gianola, José Crossa, Andrés Legarra, Eduardo Manfredi, Kent Weigel and José Miguel Cotes

Fully Conditional Distributions for Implementing the BRL (Bayesian Regression Coupled with LASSO) Model

The fully conditional distributions of the unknowns in the Bayesian model of [8] and [9] are given. The derivation uses standard results for Bayesian linear models (e.g., SORENSEN and GIANOLA, 2002), and results for the Bayesian LASSO (BL, PARK and CASELLA, 2008).

The joint posterior distribution is

$$
\begin{aligned}
p\left(\mu,\boldsymbol{\beta}_r,\boldsymbol{\beta}_l,\mathbf{u},\sigma_\varepsilon^2,\sigma_r^2,\sigma_u^2,\boldsymbol{\tau}^2,\lambda\big|\mathbf{y}\right) \propto & \left\{\prod_{i=1}^{n} N\!\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right\} \\
& \times N\!\left(\mu|0,\sigma_\mu^2\right) N\!\left(\boldsymbol{\beta}_r|\mathbf{0},\mathbf{I}\sigma_r^2\right)\left[\prod_{j=1}^{p} N\!\left(\beta_{jl}|0,\sigma_\varepsilon^2\tau_j^2\right)\right] N\!\left(\mathbf{u}|\mathbf{0},\mathbf{A}\sigma_u^2\right) \quad [10] \\
& \times \chi^{-2}\!\left(\sigma_\varepsilon^2|S_\varepsilon,df_\varepsilon\right)\chi^{-2}\!\left(\sigma_r^2|S_r,df_r\right)\chi^{-2}\!\left(\sigma_u^2|S_u,df_u\right) \\
& \times \left[\prod_{\mathrm{J}=1}^{p} Exp\!\left(\tau_j^2|\lambda\right)\right] p(\lambda).
\end{aligned}
$$

The fully conditional distribution of any unknown is obtained by removing from the right-hand-side of [10] the components that do not involve that unknown. When conjugate priors were chosen, the remaining components are kernels of known distributions, as described next.

1. Intercept

$$
\begin{aligned}
p(\mu|else) & \propto \left[\prod_{i=1}^{n} N\!\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right] N\!\left(\mu|0,\sigma_\mu^2\right) \\
& \propto \left[\prod_{i=1}^{n} N\!\left(y_i^* \mid \mu, \sigma_\varepsilon^2\right)\right] N\!\left(\mu|0,\sigma_\mu^2\right),
\end{aligned}
$$

where, $y_i^* = y_i - \mathbf{x}'_{ri}\boldsymbol{\beta}_r - \mathbf{x}'_{li}\boldsymbol{\beta}_l - u_i$. The right-hand side of the above expression is recognized as the kernel of a normal distribution with mean $\left[n\sigma_\varepsilon^{-2} + \sigma_\mu^{-2}\right]^{-1}\sum_i y_i^*\sigma_\varepsilon^{-2}$ and variance $\left[n\sigma_\varepsilon^{-2} + \sigma_\mu^{-2}\right]^{-1}$. In practice one can set σ_μ^2 large enough so that an effectively flat prior is assigned to the intercept.

1

2. Regression coefficient with homogeneous-variance Normal prior $\left(\boldsymbol{\beta}_r\right)$.

$$p\left(\boldsymbol{\beta}_r\middle|else\right) \propto \left[\prod_{i=1}^{n} N\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right] N\left(\beta_r\middle|\mathbf{0}, \mathbf{I}\sigma_r^2\right)$$

$$\propto \left[\prod_{i=1}^{n} N\left(y_i^{**} \mid \mathbf{x}'_{ri}\boldsymbol{\beta}_r, \sigma_\varepsilon^2\right)\right] N\left(\boldsymbol{\beta}_r\middle|\mathbf{0}, \mathbf{I}\sigma_r^2\right),$$

where, $y_i^{**} = y_i - \mu - \mathbf{x}'_{li}\boldsymbol{\beta}_l - u_i$. This is recognized as a multivariate-normal distribution (MVN) with mean vector (co-variance matrix) equal to the solution (inverse of the coefficient matrix) of the system of equations,

$$\left[\mathbf{X}'_r\mathbf{X}_r\sigma_\varepsilon^{-2} + \mathbf{I}\sigma_r^{-2}\right]\hat{\boldsymbol{\beta}}_r = \mathbf{X}'_r\mathbf{y}^{**}\sigma_\varepsilon^{-2}. \qquad [11a]$$

Alternatively, if one wants to draw samples from the fully conditional distributions of each element of $\boldsymbol{\beta}_r$, one has,

$$p\left(\beta_{rj}\middle|else\right) \propto \left[\prod_{i=1}^{n} N\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right] N\left(\beta_{rj}\middle|0, \sigma_r^2\right)$$

$$\propto \left[\prod_{i=1}^{n} N\left(y_i^{\circ\circ}\middle|x_{ri}\beta_{rj}, \sigma_\varepsilon^2\right)\right] N\left(\beta_{rj}\middle|0, \sigma_r^2\right),$$

where, $y_i^{\circ\circ} = y_i^{**} - \sum_{k \neq j} x_{rki}\beta_{rk}$. The right-hand-side of the above expression is recognized as the kernel of a normal distribution with mean and variance equal to the solution (inverse of the coefficient in the left-hand-side) of,

$$\left[\sigma_\varepsilon^{-2}\sum_{i=1}^{n} x_{rji}^2 + \sigma_r^{-2}\right]\hat{\beta}_{rj} = \sigma_\varepsilon^{-2}\sum_{i=1}^{n} x_{rji}y_i^{\circ\circ}. \qquad [11b]$$

3. Regression coefficient with heterogeneous-variance Normal prior $\left(\boldsymbol{\beta}_l\right)$.

$$p\left(\boldsymbol{\beta}_l|else\right) \propto \left[\prod_{i=1}^n N\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right]\left[\prod_{j=1}^p N\left(\beta_{jl}|0, \sigma_\varepsilon^2\tau_j^2\right)\right]$$

$$\propto \left[\prod_{i=1}^n N\left(y_i^{***} \mid \mathbf{x}'_{li}\boldsymbol{\beta}_l, \sigma_\varepsilon^2\right)\right]\left[\prod_{j=1}^p N\left(\beta_{jl}|0, \sigma_\varepsilon^2\tau_j^2\right)\right],$$

where, $y_i^{***} = y_i - \mu - \mathbf{x}'_{ri}\boldsymbol{\beta}_r - u_i$. This is recognized as a MVN distribution with mean vector (co-variance matrix) equal to the solution (inverse of the coefficient matrix) of the system of equations,

$$\left[\mathbf{X}'_l\mathbf{X}_l\sigma_\varepsilon^{-2} + Diag\left\{\sigma_\varepsilon^{-2}\tau_j^{-2}\right\}\right]\hat{\boldsymbol{\beta}}_l = \mathbf{X}'_l\mathbf{y}^{***}\sigma_\varepsilon^{-2}. \qquad [12a]$$

Again, if one whishes to draw samples form the fully conditional distribution of each of the elements of $\boldsymbol{\beta}_l$, one has,

$$p\left(\beta_{lj}|else\right) \propto \left[\prod_{i=1}^n N\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right]\left[\prod_{j=1}^p N\left(\beta_{lj}|0, \sigma_\varepsilon^2\tau_j^2\right)\right]$$

$$\propto \left[\prod_{i=1}^n N\left(y_i^{\circ\circ\circ} \mid x_{lji}\beta_{lj}, \sigma_\varepsilon^2\right)\right]N\left(\beta_{lj}|0, \sigma_\varepsilon^2\tau_j^2\right),$$

where $y_i^{\circ\circ\circ} = y_i^{***} - \sum_{k \neq j} x_{lki}\beta_{lk}$. The right-hand-side of the above expression is recognized as the kernel of a normal distribution with mean and variance equal to the solution (inverse of the coefficient of the left-hand-side) of the following equation,

$$\left[\sigma_\varepsilon^{-2}\sum_{i=1}^n x_{lji}^2 + \sigma_\varepsilon^{-2}\tau_j^{-2}\right]\hat{\beta}_{lj} = \sigma_\varepsilon^{-2}\sum_{i=1}^n x_{lji}y_i^{\circ\circ\circ}. \qquad [12b]$$

4. Infinitesimal additive effects $\left(\mathbf{u}\right)$.

$$p\left(\mathbf{u}|else\right) \propto \left[\prod_{i=1}^n N\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right]N\left(\mathbf{u}|0, \mathbf{A}\sigma_u^2\right)$$

$$\propto \left[\prod_{i=1}^n N\left(y_i^{****} \mid u_i, \sigma_\varepsilon^2\right)\right]N\left(\mathbf{u}|0, \mathbf{A}\sigma_u^2\right),$$

where, $y_i^{****} = y_i - \mu - \mathbf{x}'_{ri}\boldsymbol{\beta}_r - \mathbf{x}'_{li}\boldsymbol{\beta}_l$. This is recognized as a MVN distribution with mean vector (co-variance matrix) equal to the solution (inverse of the coefficient matrix) of the system,

$$\left[\mathbf{I}\sigma_\varepsilon^{-2} + \mathbf{A}^{-1}\sigma_u^{-2}\right]\hat{\mathbf{u}} = \mathbf{y}^{****}\sigma_\varepsilon^{-2}. \qquad [13a]$$

It follows that each of the entries of $\mathbf{u}$ also has a fully conditional distribution that is normal (e.g., SORENSEN and GIANOLA, 2002), with the following mean and variance,

$$E(u_i \mid else) = c_{ii}^{-1}\left(rhs_i - \sum_{k\neq i} c_{ik}u_k\right) \; ; \; Var(u_i \mid else) = c_{ii}^{-1} \qquad [13b]$$

where, c_{ii} and rhs_i are the ith diagonal element and the ith entry of $\mathbf{C} = \left[\mathbf{I}\sigma_\varepsilon^{-2} + \mathbf{A}^{-1}\sigma_u^{-2}\right]$ and $\mathbf{rhs} = \mathbf{y}^{****}\sigma_\varepsilon^{-2}$, respectively.

5. Residual variance $\left(\sigma_\varepsilon^2\right)$.

$$p\left(\sigma_\varepsilon^2 \mid else\right) \propto \left[\prod_{i=1}^{n} N\left(y_i \mid \mu + \mathbf{x}'_{ri}\boldsymbol{\beta}_r + \mathbf{x}'_{li}\boldsymbol{\beta}_l + u_i, \sigma_\varepsilon^2\right)\right]\left[\prod_{j=1}^{p} N\left(\beta_{lj} \mid 0, \sigma_\varepsilon^2\tau_j^2\right)\right]\chi^{-2}\left(\sigma_\varepsilon^2 \mid S_\varepsilon, df_\varepsilon\right)$$

$$\propto \left[\prod_{i=1}^{n} N\left(\varepsilon_i \mid \sigma_\varepsilon^2\right)\right]\left[\prod_{j=1}^{p} N\left(\tilde{\beta}_{lj} \mid 0, \sigma_\varepsilon^2\right)\right]\chi^{-2}\left(\sigma_\varepsilon^2 \mid S_\varepsilon, df_\varepsilon\right) \qquad [14]$$

$$= \chi^{-2}\left(\sigma_\varepsilon^2 \mid S = S_\varepsilon + \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}}, df = df_\varepsilon + n + p\right)$$

where $\boldsymbol{\varepsilon} = \left(\varepsilon_1,..,\varepsilon_n\right)' = \mathbf{y} - \mathbf{1}\mu - \mathbf{X}_r\boldsymbol{\beta}_r - \mathbf{X}_l\boldsymbol{\beta}_l - \mathbf{u}$, $\tilde{\beta}_{lj} = \left(\dfrac{\beta_{lj}}{\tau_j}\right)$ and $\tilde{\boldsymbol{\beta}} = \left(\tilde{\beta}_{l1},...,\tilde{\beta}_{lp}\right)'$.

6. Additive-genetic variance $\left(\sigma_u^2\right)$.

$$p\left(\sigma_u^2 \mid else\right) \propto N\left(\mathbf{u} \mid \mathbf{0}, \mathbf{A}\sigma_u^2\right)\chi^{-2}\left(\sigma_u^2 \mid S_u, df_u\right)$$

$$= \chi^{-2}\left(\sigma_u^2 \mid S = S_u + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}, df = df_u + q\right) \qquad [15]$$

where q is the order of the square matrix $\mathbf{A}$.

7. Scaling variables associated to marker effects $\left(\boldsymbol{\tau}^2\right)$.

$$p\left(\boldsymbol{\tau}^2 \mid else\right) \propto \left[\prod_{j=1}^{p} N\left(\beta_{jl} \mid 0, \sigma_\varepsilon^2\tau_j^2\right)\right]\left[\prod_{j=1}^{p} Exp\left(\tau_j^2 \mid \lambda\right)\right],$$

so that, each of the τ_j^2's are conditionally independent,

$$p\left(\tau_j^2 \middle| else\right) \propto N\left(\beta_{jl} \middle| 0, \sigma_\varepsilon^2 \tau_j^2\right) Exp\left(\tau_j^2 \middle| \lambda\right)$$

From results presented in PARK and CASELLA, the fully conditional distributions of the reciprocal of the τ_j^2 ,s are Inverse-Gaussian.

$$p\left(\tau_j^{-2} \middle| else\right) \propto IG\left(\tau_j^{-2} \middle| \mu_j = \frac{\sigma_\varepsilon \lambda}{\left|\beta_{jl}\right|}, S = \lambda^2\right) \qquad [16]$$

8. LASSO smoothing parameter (λ).

PARK and CASELLA choose a Gamma prior for λ^2, in this case,

$$p\left(\lambda^2 \middle| else\right) \propto \left[\prod_{j=1}^p Exp\left(\tau_j^2 \middle| \lambda\right)\right] G\left(\lambda^2 \middle| \alpha, \beta\right).$$

The above is the kernel of the density of a Gamma distribution, with parameters,

$$p\left(\lambda^2 \middle| else\right) = G\left(\lambda^2 \middle| p + \alpha_1, \frac{1}{2}\sum_{j=1}^p \tau_j^2 + \alpha_2\right). \qquad [17]$$

An alternative is to use a uniform prior: $p\left(\lambda^2\right) = U[a,b]$ with $0 \le a < b$. In this case, the fully conditional distribution is,

$$p\left(\lambda^2 \middle| else\right) \propto G\left(\lambda^2 \middle| p, \frac{1}{2}\sum_{j=1}^p \tau_j^2\right) 1\left(a \le \lambda^2 \le b\right).$$

Samples from the above distribution can be obtained by using rejection sampling, which may be implemented as follows:

(i)   sample a candidate from $G\left(\lambda^2 \middle| p, \frac{1}{2}\sum_{j=1}^p \tau_j^2\right)$;

(ii)   accept the candidate if $a \le \lambda^2 \le b$, otherwise go back to (i).

A more general formulation may be obtained by using the Beta distribution. Let $\tilde{\lambda} = \frac{\lambda}{max}$, where $max > 0$ is an upper bound on λ, and $p\left(\tilde{\lambda}\right) = Beta\left(\tilde{\lambda} \middle| \alpha_3, \alpha_4\right)$. It follows that,

$$p(\lambda) = Beta\left[\tilde{\lambda}(\lambda)\,\middle|\,\alpha_3, \alpha_4\right] \left|\frac{\partial \tilde{\lambda}(\lambda)}{\partial \lambda}\right| \propto Beta\left(\frac{\lambda}{max}\,\middle|\,\alpha_3, \alpha_4\right).$$

If $\alpha_3 = \alpha_4 = 1$ this gives a uniform prior for λ (note that this is different than a uniform prior on λ^2). When a Beta distribution is used the fully conditional distribution $p(\lambda|else)$ does not have closed form, however one can use the Metropolis-Hastings algorithm to obtain samples. In our implementation, at iteration t, we use the following algorithm:

(i)   Sample a candidate, λ_c^2, from $G\left(\lambda^2\,\middle|\,p, \dfrac{1}{2}\sum_{j=1}^{p}\tau_j^2\right)$.

(ii)   Let λ_t and τ_j^2 the current sample of the corresponding parameters, evaluate:

$$r = \frac{\displaystyle\prod_{j=1}^{p} Exp\left(\tau_j^2\,\middle|\,\lambda_c\right) Beta\left(\frac{\lambda_c}{max}\,\middle|\,\alpha_3, \alpha_4\right) G\left(\lambda_t^2\,\middle|\,p, \frac{1}{2}\sum_{j=1}^{p}\tau_j^2\right)}{\displaystyle\prod_{j=1}^{p} Exp\left(\tau_j^2\,\middle|\,\lambda_t\right) Beta\left(\frac{\lambda_t}{max}\,\middle|\,\alpha_3, \alpha_4\right) G\left(\lambda_c^2\,\middle|\,p, \frac{1}{2}\sum_{j=1}^{p}\tau_j^2\right)}$$

(iii) Sample $u \sim U(u\,|\,0,1)$

(iv)   Set $\lambda^{t+1} = \lambda^c$ if $r \geq u$, otherwise set $\lambda^{t+1} = \lambda^t$.

G. de los Campos *et al.*

FILE S2

Raw data and the BRL.rda program are available in a compressed file at
http://www.genetics.org/cgi/content/full/genetics.109.101501/DC1.