Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(3) 2012 413-426

Application of Bayesian Elastic Net and Other Shrinkage Methods in Genomic Selection and QTL Mapping

Paulino Pérez[1], Gustavo de los Campos[2], Susanne Dreisigacker[3], Hector Sanchez-Villeda[3] and José Crossa[3*]

[1]*Colegio de Postgraduados, Km. 36.5 Carretera México, Texcoco, Montecillo, Estado de México, 56230, México*

[2]*Department of Biostatistics, University of Alabama at Birmingham, Ryals Public Health Bldg 443, Alabama, USA*

[3]*International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, México D.F., México*

Received 28 November 2011; Revised 11 July 2012; Accepted 08 August 2012

SUMMARY

The Elastic Net is a variable selection and shrinkage estimation method especially designed for regression settings with a large number of correlated predictors. Recently, a Bayesian formulation of the Elastic Net was proposed (BEN=Bayesian Elastic Net). In this article, we extend the BEN to model the combined effects of dense molecular markers and pedigree data and evaluate the performance of the proposed model using a barley data set and two large wheat data sets. The predictive power of the proposed model was compared with those of two well-established models: the Bayesian LASSO and the Bayesian Ridge Regression. Results show that the prediction assessment of BEN was as accurate as those of the other methods in all studied cases. The number of molecular markers with significant effects detected by BEN in four data sets was compared with those found by the Bayesian LASSO and Bayesian Ridge Regression models. An R-program that implements the proposed model is available.

Keywords: Bayesian Elastic Net, Shrinkage methods, Genomic selection.

1. INTRODUCTION

Genomic selection (GS, Meuwissen *et al.* 2001) uses dense molecular marker genotypes and phenotypes to predict phenotypic values of selection candidates. Several simulation (*e.g.*, Meuwissen *et al.* 2001; Habier *et al.* 2009) and empirical studies in plants (de los Campos *et al.* 2009, 2010; Crossa *et al.* 2010, 2011; Pérez *et al.* 2010; Wang *et al.* 2010) and animals (*e.g.*, Weigel *et al.* 2009; de los Campos *et al.* 2009, 2010) have demonstrated that GS can yield accurate predictions of phenotypic values; the GS is being implemented for commercial breeding in several agricultural crop species.

In a parametric model for GS (*e.g.*, Meuwissen *et al.* 2001, Xu 2003) phenotypes (y_i; $i = 1, ..., n$) are regressed on marker genotypes (x_{ij}; $j = 1, ..., p$) using a linear model of the form $y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i$, where $x_{ij} \in \{0, 1, 2\}$ represents the number of copies of a diallelic marker (*e.g.*, an SNP), and β_j is regression of y_i on the j^{th} marker. In matrix notation, the model is expressed as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{y} = \{y_i\}$, $\boldsymbol{\beta} = \{\beta_j\}$ and $\boldsymbol{\varepsilon} = \{\varepsilon_i\}$ are vectors of phenotypes, marker effects and model residuals, respectively, and $\mathbf{X} = \{x_{ij}\}$ is a matrix of marker genotypes of dimensions $n \times p$. With dense molecular markers, the number of markers exceeds the

[*]*Corresponding author* : José Crossa
E-mail address : j.crossa@cgiar.org

number of records in the reference population ($p>>n$), and therefore penalized regression estimation methods and their Bayesian counterpart are commonly used.

Ridge Regression (RR, Hoerl and Kennard 1970) is the oldest penalized estimation method that yields shrunken estimates of regression coefficients. Tibshirani (1996) proposed the LASSO (Least Absolute Selection and Shrinkage Operator) method as a way of performing variable selection and shrinkage estimation simultaneously. The LASSO was shown to perform well with uncorrelated predictors; however, its predictive performance has not been always good in settings where predictors are highly correlated. To overcome this limitation, Zou and Hastie (2005) proposed using the Elastic Net (EN), which is a penalized estimation method that represents a compromise between RR and LASSO and has been shown to have good predictive performance in regression settings with highly correlated predictors. This situation also arises in models for GS.

Most penalized estimates are equivalent to posterior modes of regression coefficients in a certain class of Bayesian models. The Bayesian equivalents of RR and LASSO are well known (*e.g.*, Goldstein 1976; Park and Casella 2008) and these models have been shown to be effective for prediction in GS (*e.g.*, de los Campos *et al.* 2009; Crossa *et al.* 2010; Perez *et al.* 2010). Recently, Li and Lin (2010) and Kyung *et al.* (2010) proposed Bayesian formulations of the EN (hereinafter BEN, for "Bayesian Elastic Net").

In this research, we extend the BEN to accommodate the combined effects of dense molecular markers and pedigree, and developed software that implements the proposed model and compares BEN's predictive performance with that of the Bayesian LASSO (BL) and Bayesian Ridge Regression (BRR). We also discuss how estimates from the BRR, BL and BEN can be used for QTL detection.

2. PENALIZED AND BAYESIAN SHRINKAGE REGRESSION METHODS

Penalized Regression Methods

The RR (Hoerl and Kennard 1970) is the oldest penalized estimation method. In RR, estimates of regression coefficients are obtained by minimizing the residual sum of squares, $RSS(\mathbf{y},\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2$, subject to the following constraint: $SS(\boldsymbol{\beta}) = \sum_{j=1}^{p}\beta_j^2 \le t$; equivalently:

$$\hat{\boldsymbol{\beta}}_{RR} = \arg\min_{\beta}\{\Sigma_{i=1}^{n}(y_i - \Sigma_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda_{RR}\Sigma_{j=1}^{p}\beta_j^2\} \quad (1)$$

where $\lambda_{RR} = \lambda(t) \ge 0$ is a non-negative constant controlling the trade-offs between goodness of fit, measured by the RSS, and model complexity, measured by the sum of squares of the regression coefficients, or the L_2 norm of $\boldsymbol{\beta}$. The quadratic L_2 penalty induces shrinkage of estimates of marker effects towards zero; this introduces bias but reduces the variance of estimates.

The LASSO (Tibshirani 1996) is another commonly used penalized estimation method that is obtained by replacing the L_2 norm in (1) with the sum of the absolute values of the regression coefficients, or L_1 norm, that is

$$\hat{\boldsymbol{\beta}}_L = \arg\min_{\beta}\left\{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}x_{ij}\beta_j)^2 + \lambda_L\sum_{j=1}^{p}|\beta_j|\right\}.$$

The solution to this problem may involve zeroing out some effects; thus, LASSO performs variable selection and shrinkage simultaneously. As the number of available markers increases, so does the number of markers located in regions of the genome that are not associated with the trait of interest. Because of this, the variable selection feature of LASSO is appealing (Usai *et al.* 2009). However, empirical evidence suggests that RR outperforms LASSO from a predictive standpoint when predictors are highly co-linear (Hastie *et al.* 2009); this situation is commonly observed in GS with dense molecular markers.

To improve the performance of LASSO in settings with correlated predictors, Zou and Hastie (2005) proposed a penalized method, the EN, whose penalty function uses a compromise between the L_2 and L_1 norms. In the EN, the optimization problem becomes

$$\min_{\beta}\{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \text{ subject to } (1-\alpha)\sum_{j=1}^{p}|\beta_j| + \alpha\sum_{j=1}^{p}\beta_j^2 \le t\},$$

where t is an arbitrary positive constant and $\alpha \in [0,1]$. Ridge Regression and the LASSO are obtained as special cases of the above problems, with α equal to one and zero, respectively. The solution to the above optimization problem is equivalent to

$$\hat{\beta}_{EN} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right\} \quad (2)$$

where λ_1 and λ_2 are non-negative constants that control the weight assigned to the L_1 and L_2 penalties, respectively. The penalty used in the EN stabilizes the LASSO solution (Zou and Hastie 2005) while keeping the variable selection feature of LASSO; however, it tends to shrink together the coefficients of correlated predictors, a feature observed in RR (Hastie *et al.* 2009). Zou and Hastie (2005) developed an efficient algorithm (LARS-EN) to solve the above optimization problem.

Bayesian Shrinkage Regression Methods

Estimates of regression coefficients derived from penalized optimization problems such as RR or LASSO are equivalent to posterior modes in certain classes of Bayesian models (*e.g.*, Goldstein 1976; Wahba 1978; Tibshirani 1996). In the Bayesian approach, inferences are based on the posterior distribution of the unknowns ($\boldsymbol{\Theta}$), given the data ($\mathbf{y}$), $p(\boldsymbol{\Theta}|\mathbf{y})$. Following Bayes' rule, this density is proportional to the product of the conditional distribution of the data given the unknowns, $p(\mathbf{y}|\boldsymbol{\Theta})$, or Bayesian likelihood times the prior density assigned to model unknowns $p(\boldsymbol{\Theta})$. In the models we are concerned, the conditional distribution of the data given the parameters, $[p(\mathbf{y}|\boldsymbol{\Theta})]$, is the product of independent normal densities centered at the regression function $E(y_i|\mathbf{x}_i,\boldsymbol{\beta}) = \sum_{j=1}^{p} x_{ij}\beta_j$, and with common residual variance (σ_ε^2), that is

$$p(\mathbf{y}|\mathbf{X},\boldsymbol{\beta},\sigma_\varepsilon^2) = \prod_{i=1}^{n} N\left(y_i \left| \sum_{j=1}^{p} x_{ij}\beta_j, \sigma_\varepsilon^2 \right.\right).$$

Marker effects are assigned identical and independent normal prior densities, $p(\boldsymbol{\beta}|\boldsymbol{\omega}) = \prod_{j=1}^{p} p(\beta_j|\boldsymbol{\omega})$, where $\boldsymbol{\omega}$ represents hyper-parameters indexing the prior density of marker effects. Following Bayes' rule, the posterior distribution is

$$p(\boldsymbol{\beta}|\mathbf{y},\sigma_\varepsilon^2,\boldsymbol{\omega}) \propto p(\mathbf{y}|\mathbf{X},\boldsymbol{\beta},\sigma_\varepsilon^2)\, p(\boldsymbol{\beta}|\boldsymbol{\omega}) = \prod_{i=1}^{n} N\left(y_i \left| \sum_{j=1}^{p} x_{ij}\beta_j, \sigma_\varepsilon^2 \right.\right) \prod_{j=1}^{p} p(\beta_j|\boldsymbol{\omega}). \quad (3)$$

The choice of prior density, $p(\beta_j|\boldsymbol{\omega})$, determines whether the posterior mode of (3) is equivalent to estimates obtained with BRR, BL or BEN. In BRR, the prior density of marker effects, $p(\beta_j|\boldsymbol{\omega})$, is Gaussian, centered at zero and with common variance, that is, $p(\beta_j|\sigma_\beta^2) = N(\beta_j|0,\sigma_\beta^2)$, where σ_β^2 is a prior-variance of marker effects. In BL, the prior assigned to marker effects is Double-Exponential (DE, *e.g.*, Tibshirani 1996), centered at zero and with inverse-scale parameter, $\frac{\lambda}{\sigma_\varepsilon^2}$, that is,

$$p(\beta_j|\lambda,\sigma_\varepsilon^2) = DE\left(\beta_j|0,\frac{\lambda}{\sigma_\varepsilon^2}\right).$$

The prior density of BEN represents a compromise between normal and DE densities. Fig. 1 displays these densities for a random variable with mean equal to zero and variance equal to one. The prior corresponding to

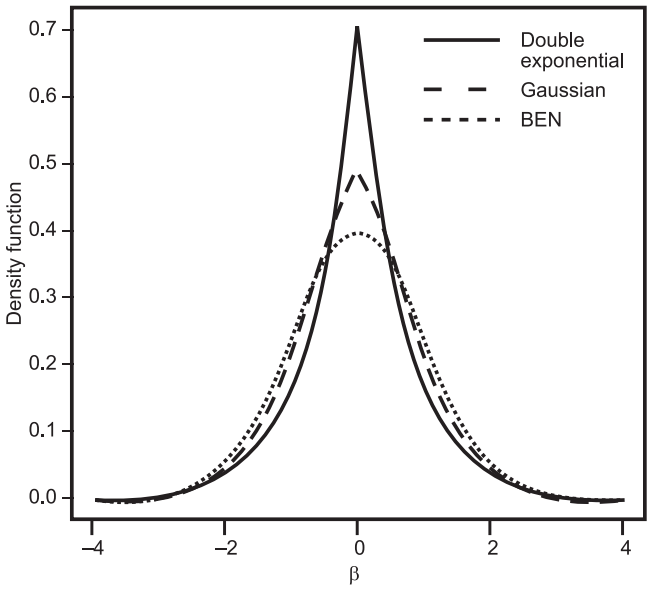

Fig. 1. Prior densities of Bayesian Ridge Regression (Gaussian, dashed-line), Bayesian LASSO (double-exponential, solid line) and Bayesian Elastic Net (BEN) derived with (dotted line).

the BEN was obtained by setting $\lambda_1 = \lambda_2 = 0.8$. Relative to the Gaussian density, the DE places higher mass at zero and have thicker tails, inducing a different type of shrinkage. In particular, relative to BRR, the prior used in the BL induces stronger shrinkage towards zero of estimates of effects of predictors that have a weak association with the response, and not as much shrinkage of estimates of effects of predictors that have a strong association with the response.

For the implementation of these Bayesian methods, the Gaussian likelihood is a conjugate of the normal prior; therefore, samples from the posterior distribution of BRR can be obtained using a Gibbs sampler. The DE prior does not conjugate with the Gaussian likelihood; however, as proposed by Park and Casella (2008), the DE density can be represented as a mixture of scaled-normal densities, which allows using a Gibbs sampler for implementing the BL. Following a strategy similar to that followed by Park and Casella (2008), Li and Lin (2010) and Kyung *et al.* (2010) proposed representing the prior density corresponding to BEN using a mixture of scaled normal densities. In the model of Kyung *et al.* (2010), joint prior density of the *j-th* marker effect and its associated scale-parameter (τ_j^2) is

$$p\left(\beta_j, \tau_j^2 | \lambda_1, \lambda_2, \sigma_\varepsilon^2\right) = N\left(\beta_j | 0, \sigma_\varepsilon^2 \left(\tau_j^{-2} + \lambda_2\right)^{-1}\right) Exp\left(\tau_j^2 | \lambda_1\right),$$

where $Exp\left(\tau_j^2 | \lambda_1\right) = \frac{\lambda_1^2}{2} \exp\left(-\frac{\lambda_1^2 \tau_j^2}{2}\right)$ is the density function of an exponential function indexed by the parameter λ_1. The marginal prior of marker effects, $p\left(\beta_j | \lambda_1, \lambda_2\right)$, can then be obtained by marginalizing with respect to the scale parameter τ_j^2,

$$p\left(\beta_j | \lambda_1, \lambda_2\right) = \int N\left(\beta_j | 0, \sigma_\varepsilon^2 \left(\tau_j^{-2} + \lambda_2\right)^{-1}\right) Exp\left(\tau_j^2 | \lambda_1\right) \partial \tau_j^2 \quad (4)$$

All the prior densities discussed above are indexed by hyper-parameters $\boldsymbol{\omega} = \{\sigma_\beta^2\}$ in BRR, $\boldsymbol{\omega} = \{\lambda, \sigma_\varepsilon^2\}$ in the BL and $\boldsymbol{\omega} = \{\lambda_1, \lambda_2, \sigma_\varepsilon^2\}$ in BEN. These can be dealt with by assigning a prior density to these unknowns, or by using empirical Bayes approaches (*e.g.*, Park and Casella 2008).

Bayesian Shrinkage Regression Methods Using Markers and Pedigree

In BRR, BL and BEN, the regression function is $E\left(y_i | \mathbf{x}_i, \boldsymbol{\beta}\right) = \sum_{j=1}^{p} x_{ij} \beta_j$. This can be extended by including in the regression function a random effect $\{u_i\}$ representing the regression of phenotypes on pedigree information (denoted as P), so that $E(y_i | \mathbf{x}_i, \boldsymbol{\beta}, P) = u_i + \sum_{j=1}^{p} x_{ij} \beta_j$. Following the standard assumption of infinitesimal models (*e.g.*, Henderson 1975), the vector of random effects is assigned a multivariate-Gaussian prior, centered at zero and with a co-variance function $Cov(u_i, u_{i'}) = a(i,i')\sigma_u^2$, where $a(i,i')$ are twice kinship coefficients, computed from the pedigree, and σ_u^2 is a variance parameter associated with the regression on the pedigree. In matrix form $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{A})$, where $\mathbf{u} = \{u_i\}$ and $A = \{a(i, i')\}$. The general form of the posterior density of a pedigree + marker (MP) Bayesian regression model is

$$p\left(\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_u^2, \mathbf{u} | \mathbf{y}, \boldsymbol{\omega}\right) = \prod_{i=1}^{n} N\left(y_i | u_i + \sum_{j=1}^{p} x_{ij} \beta_j, \sigma_\varepsilon^2\right) \prod_{j=1}^{p} p\left(\beta_j | \boldsymbol{\omega}\right) \times \chi^{-2}\left(\sigma_u^2 | S_u, df_u\right) \chi^{-2}\left(\sigma_\varepsilon^2 | S_\varepsilon, df_\varepsilon\right) \times N\left(\mathbf{u} | \mathbf{0}, \sigma_u^2 \mathbf{A}\right) \quad (5)$$

where the above variance parameters were treated as unknowns and assigned independent scaled-inverse Chi-square density with a degree of freedom and scale parameter equal to *df* and *S*, respectively. In (5), it is assumed that the parameters indexing the prior density of marker effects $(\boldsymbol{\omega})$ are known. In practice, this can be dealt with by assigning a prior to these unknowns or by using Empirical Bayes approaches. The specification of the prior density of marker effects, $p\left(\beta_j | \boldsymbol{\omega}\right)$, will define whether the above is a markers + pedigree BRR (MP-BRR; this occurs when $p(\beta_j | \sigma_\beta^2) = N(\beta_j | 0, \sigma_\beta^2)$, BL (MP-BL; this occurs

when $p(\beta_j|\lambda,\sigma^2_\varepsilon) = DE\left(\beta_j \middle| 0, \frac{\lambda}{\sigma^2_\varepsilon}\right)$, or BEN (MP-BEN), which occurs in (4). Marker-based regressions can be simply obtained by setting $u_i = 0$ in MP-BRR, MP-BL or MP-BEN, which yield M-BRR, M-BL and M-BEN, respectively. On the other and a Pedigree model (P) can be obtained by removing the markers effects in (5) and is described in Crossa *et al.* (2010).

Models MP-BL, M-BL, MP-BRR and M-BRR are fully described in Pérez *et al.* (2010) and implemented in the R package BLR (de los Campos and Pérez 2010). In this study, we implemented MP-BEN and M-BEN; the model and the algorithm used to implement them are fully described in Appendix. The software is available upon request to the first author.

Detection of Chromosome Regions Associated with Quantitative Traits

Genomic selection has traditionally focused on the problem of predicting phenotypic values; however, estimates of marker effects derived from models from GS can also be used to detect regions significantly associated with a quantitative trait. Wang *et al.* (2005) and Che and Xu (2010) have shown how to use the Bayesian shrinkage methods in QTL mapping. Che and Xu (2010) suggested using a permutation test (Fisher 1935; Churchill and Doerge 1994) to detect significant associations. The QTL mapping process consists of identifying molecular markers or genomic loci that influence the variation of complex traits (Yi and Xu 2008). Once the marker effects have been estimated, the problem is how to decide which markers are in linkage disequilibrium with QTLs; this is a model selection problem that can be solved using Bayesian shrinkage methods such as those previously shown.

Once the effects of the markers have been estimated using the BRR, BL or BEN models, a permutation "within Markov chains" (generated by using Gibbs sampling) can be performed to decide which markers have significant effects, each marker is considered to be associated with a QTL (Wang *et al.* 2005). The significance of the effect of each marker can be evaluated by using permutations. In this strategy, the original vector with phenotypic values $\mathbf{y} = (y_1, ..., y_n)'$ is permuted every *h*th iteration within a Markov chain, after all parameters are sampled (the residual variance is preserved in the permuted sample). If the genotypes do not match the phenotypes, the posterior means of the regression coefficients are expected to be very close to zero. In this strategy, only two chains are required, one for the original data and the other for the permuted sample. The reshuffled chain can be used to obtain the empirical threshold for QTL effects, $0.5\alpha\times 100\%$ and $(1-0.5\alpha)\times 100\%$. Che and Xu (2010) suggested permuting the sample at every iteration ($h = 1$).

3. EXPERIMENTAL DATA

The various BRR, BL, and BEN models with molecular markers and pedigree were evaluated using data sets of two different crops: a barley data set (*Hordeum vulgare* L.), two wheat data sets (*Triticum aestivum* L.) (Wheat1 and Wheat2 data sets) and a simulated dataset. All models were fitted using programs written in R (R Development Core Team 2010).

Barley Data Set

The barley data set is from the North American Barley Genome Mapping Project that contains $n = 145$ doubled-haploid lines; each one was grown in 25 different environments. The trait analyzed was average kernel weight. The 145 doubled-haploid individuals were genotyped with 127 molecular markers coded as 0 (absence) or 1 (presence). This data set was previously analyzed by Yi and Xu (2008).

Wheat1 Data Set

The first wheat data set is from CIMMYT's Global Wheat Program and comprises 622 wheat lines evaluated in rainfed regions of the world. The phenotypic trait considered here was grain yield. Wheat lines were genotyped using 1588 Diversity Array Technology (DArT) markers generated by Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com.au). The DArT markers may take on two values, denoted by their presence (1) or absence (0).

A pedigree tracing back many generations was available, and 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 numerator relationship matrix (**A**) among lines.

Wheat2 Data Set

This data set contains a set of 599 wheat lines from CIMMYT's Global Wheat Program that were genotyped for 1447 DArT and evaluated for grain yield (GY) in four mega-environments (E1-E4). As in the case of Wheat1, pedigree information on this data set was available and used for deriving the matrix **A** relationship among the 599 wheat lines. This data set was also used by de los Campos *et al.* (2009), Crossa *et al.* (2010) and Pérez *et al.* (2010) for evaluating the predictive power of various models for GS.

Simulated Data Set

This data set was generated and used by Zhang and Xu (2005) for estimating the epistatic effects of QTL. The simulated dataset contains $n = 600$ individuals from a backcross population. For this population, a single large chromosome 1800 cM long was simulated, covered by 121 evenly spaced markers (coded as –1 and 1) with 15 cM per marker interval. They simulated 9 QTLs with significant main effects at markers 1, 21, 31, 51, 71, 91, 101, 111 and 121.

4. DATA ANALYSIS

Prior Distributions

In the case of the Barley data set, only molecular marker models were used because pedigree information on the 145 barley lines was not available. The hyper-parameters for the M-BRR and M-BL were set using the guidelines given in Pérez *et al.* (2010). For the M-RR, $S_\beta = 0.076,\ df_\beta = 3,\ df_e = 4,\ S_e = 1;$ for the BL, λ^2 is assigned a Gamma prior with rate $\delta = 0.02$ and shape $r = 1.1$, independent Scaled Inverse Chi-squared priors were assigned to the variance parameters, and the scale and degree of freedom parameters were set to $S_\varepsilon = 1$ and $df_\varepsilon = 4$ respectively. In the case of the M-BEN, we used gamma priors for λ_1^2 and λ_2, the suggested prior given in Kyung *et al.* (2010), and we set $r_1 = r_2 = 1; \delta_1 = \delta_2 = 10.$ The hyper-parameters for M-BRR and MP-BRR were set as $S_\beta = 0.0085, df_\beta = 3;\ \ df_\varepsilon = df_u = 4,\ \ S_\varepsilon = S_u = 1$ for Wheat1; for MP-BL, λ^2 was assigned a Gamma prior with rate (δ) and shape (r), with $\delta = 1\times10^{-4}$ and $r = 0.6$; independent Scaled Inverse Chi-squared priors were assigned to the variance parameters, and the scale and degree of freedom parameters were set to $S_\varepsilon = S_u = 1$ and $df_\varepsilon = df_u = 4,$ respectively. In the case of the Wheat2 data set, the hyper-parameters for MP-BL and M-BL are given in Crossa *et al.* (2010), whereas for the BRR models, the priors are given in Pérez *et al.* (2010). The priors for λ_1^2 and λ_2 in both wheat data sets were chosen as they were in the Barley data set. Finally, in the case of the simulated data set, the hyper-parameters for M-BRR and M-BL were fixed using the same procedure as was used for the other data sets.

Some general guidelines for selecting the hyper parameters for BL, BRR can be found in Pérez *et al.* (2010), and also in de los Campos *et al.* (2009). In the case of the BEN the reader can review Kyung *et al.* (2010).

Full-data Analysis for Permutation Test

Models were fitted using all available lines for each data set in order to estimate marker effects and variance components. The inferences were based on 30,000 samples obtained after discarding the first 5,000 samples that were taken as burn-in. The hyper-parameters were previously described. The convergence was checked by inspecting trace plots of variance components.

A permutation test for QTL detection was applied to obtain an empirical threshold and detect markers associated with QTLs using the BL, BRR and BEN models. A total of 10,000 permutation samples were generated and analyzed using the same parameters as in the original data. The inferences for each fit were based on 5,000 samples (obtained after discarding 5,000 samples as burn-in).

Cross-validation (CV) for Prediction

In CV, data are randomly divided into disjoint groups (10 in our case). Each of these sets can then be used to measure predictability. For example, using the first set, the data can be divided so that the training set contains all the observations in $\{S_2, ..., S_{10}\}$ and the testing set, those in S_1. Subsequently, models are fitted using the training data $\{S_2, ..., S_{10}\}$ to obtain predictions for observations in S_1, that is, $\{\hat{y}_i : i \in S_1\}$. Repeating

Table 1: Estimates of the posterior means of parameters and (for M-BL and MP-BL) from the full data analysis of Barley, Wheat1 and Wheat2 data sets using various Bayesian shrinkage regression methods.

Data set	Model*	Parameter				
		σ_ε^{2**}	σ_u^2	σ_β^2	λ_1	λ_2
Barley	M-BL	0.2741	—	—	5.2708	—
	M-BEN	0.2289	—	—	1.8393	3.645
	M-BRR	0.2722	—	0.0240	—	—
W1	P	0.4079	0.8614	—	—	—
	M-BL	0.6246	—	—	28.7653	—
	M-BEN	0.3974	—	—	6.2742	42.207
	M-BRR	0.5905	—	0.0021	—	—
	MP-BL	0.3645	0.7077	—	39.0055	—
	MP-BEN	0.2108	0.7271	—	6.2455	41.706
	MP-BRR	0.3188	0.6365	0.0017	—	—
W2-E1	P	0.5620	0.2860	—	—	—
	M-BL	0.5536	—	—	20.1885	—
	M-BEN	0.4067	—	—	5.6382	34.11
	M-BRR	0.5346	—	0.0031	—	—
	MP-BL	0.4261	0.1421	—	20.0336	—
	MP-BEN	0.3023	0.1519	—	5.6261	33.882
	MP-BRR	0.4263	0.1343	0.0024	—	—
W2-E2	P	0.5810	0.2480	—	—	—
	M-BL	0.5742	—	—	21.9920	—
	M-BEN	0.4097	—	—	5.6478	34.282
	M-BRR	0.5627	—	0.0026	—	—
	MP-BL	0.4838	0.1197	—	23.3686	—
	MP-BEN	0.3292	0.1228	—	5.6383	34.099
	MP-BRR	0.4735	0.1063	0.0023	—	—
W2-E3	P	0.4920	0.3420	—	—	—
	M-BL	0.6711	—	—	27.8291	—
	M-BEN	0.4434	—	—	5.6544	34.385
	M-BRR	0.6492	—	0.0021	—	—
	MP-BL	0.4618	0.2344	—	31.8244	—
	MP-BEN	0.2843	0.2419	—	5.6383	34.104
	MP-BRR	0.4581	0.2090	0.0014	—	—
W2-E4	P	0.5170	0.3000	—	—	—
	M-BL	0.6122	—	—	24.9169	—
	M-BEN	0.4186	—	—	5.6501	34.334
	M-BRR	0.5841	—	0.0025	—	—
	MP-BL	0.4527	0.1809	—	26.1275	—
	MP-BEN	0.2972	0.1875	—	5.6391	34.131
	MP-BRR	0.4553	0.1613	0.0017	—	—
Simulated	M-BL	0.5012	—	—	14.3918	—
	M-BEN	0.4517	—	—	2.2101	5.556
	M-BRR	0.5369	—	0.1356	—	—

*The seven fitted models are: Pedigree model (P), molecular marker regression model using Bayesian LASSO (M-BL); pedigree model plus molecular marker model regression using Bayesian LASSO (MP-BL); Bayesian elastic net model with molecular markers (M-BEN); Bayesian elastic net model with molecular markers and pedigree (MP-BEN); Bayesian ridge regression with molecular markers (M-BRR) and Bayesian ridge regression with molecular markers and pedigree (MP-BRR).

**Phenotypes were standardized for each data set.

this exercise for the 2nd, 3rd, ..., 10th sets yields a whole set of CV predictions, $\{\hat{y}_i\}_{i=1}^n$, that can be compared with actual observations, $\{\hat{y}_i\}_{i=1}^n$, to assess predictive power. The 10-fold CV scheme was used in all the models to predict the genetic value of the missing genotypes.

5. RESULTS

This section presents the results of the estimates of posterior means of variance components, the regularization parameters for the full models, QTL mapping for the Barley and wheat data sets using the permutation test, and the findings on the predictability of the various models using the cross-validation scheme.

Variance Components and Regularization Parameters

Table 1 shows the estimates of the posterior means of $\sigma_\varepsilon^2, \sigma_u^2, \sigma_\beta^2$ and regularization parameters λ_1, λ_2. Since the phenotypes were standardized for each data set, σ_ε^2 gives an indication of the goodness-of-fit of each model. In the Barley data set, the M-BEN model best fits the data. In Wheat1, the P model fits the data better than M-BL and M-BRR. In Wheat2, marker-based models gave a better fit. It can be observed that the M-BEN models and the MP-BEN models had good fits. The estimated mean of the variance parameter σ_u^2 for MP-BL, MP-BEN and MP-BRR is smaller than that obtained with the P models, indicating that the inclusion of markers reduces the contribution of the pedigree regression (Crossa *et al.* 2010).

QTL mapping

Table 2 shows the significant markers detected by the different Bayesian shrinkage methods. The empirical threshold was selected by setting $\alpha = 0.05$ using the permutation test described in Che and Xu (2010). It can be seen that more markers are detected with the M-BL model than with the other two models; the BEN model gave rise to the same significant markers (4) as those found by the M-BRR model; they are located on chromosomes 1, 3, and 7. Tinker *et al.* (1996) detected significant markers using an approximated likelihood ratio test described in Haley and Knott (1992), with the significant thresholds selected using the permutation test given in Churchill

Table 2. Significant markers obtained with the permutation test using three Bayesian shrinkage methods: M-BL, M-BRR and M-BEN in Barley, Wheat1 (W1), Wheat2 (W2-E1, W2-E2, W2-E3, W2-E4), and simulated data sets. Chromosome numbers are indicated in parentheses for the Barley data set.

	Bayesian shrinkage regression models		
Data	M-BRR	M-BL	M-BEN
Barley	12(1), 43(3), 101(7), 102(7)	12(1), 13(1), 32(2), 37(2), 43(3), 101(7), 102(7)	12(1), 43(3), 101(7), 102(7)
W1	—	wPt.1387, wPt.5128, wPt.6780, wPt.6900, wPt.2623	wpt.1387, wPt.5128, wPt.6900
W2-E1	—	wpt.3462, wPt.3697, wPt.6047, wPt.4835, wPt.3922, wPt.9256, wPt.3393, wPt.9422, c.344809, c.346134, c.379821	wPt.9256
W2-E2	wPt.4706	wpt.3533, wPt.7024, wPt.6967, wPt.5506, wPt.1403, wPt.4706, c.343777, c.381717	wPt.4706
W2-E3	—	Wpt.1272, wPt.3533, wPt.2644, wPt.9930, wPt.1708, wPt.9814, c.345897, c.377964, c.378173	wpt.4706
W2-E4	wPt.2644	wpt.2644, wPt.5590, wPt.2755, wPt.9277, - wPt.7299, wPt.1826, wPt.9401, wPt.0194, c.348464, c.349495, c.373205, c.37808, c.378212, c.378288, c.379495, c.37969	wPt.2644
Simulated	1, 9, 21, 29, 31, 65, 101, 121	1, 21, 31, 71, 101, 121	1, 21, 31

and Doerge (1994). Tinker *et al.* (1996) found 6 major QTLs for kernel weight, 2 located on chromosome 1, 2 on chromosome 4, and 2 on chromosome 7 (See Fig. 2 in Tinker *et al.* 1996).

Fig. 2 shows the estimated marker effects for the Barley data set and 95% confidence intervals obtained using permutations. Those markers whose estimated effect is not included in the interval were declared significant. The confidence intervals obtained using permutations in the case of M-BRR and M-BEN are clearly wider than those obtained with M-BL.

In the case of Wheat1 and Wheat2, many of the detected markers were previously reported (see, for example, Quarrie *et al.* 2005; Crossa *et al.* 2007, 2010; Huang *et al.* 2003; Kumar *et al.* 2007). For example, marker wpt.9256 was found to be significant by the M-BL and M-BEN models in W2-E1; marker wPt.4706 was detected to be significant by the three models (M-BRR, M-BL, and M-BEN) in W2-E2 but was only found to be significant by model M-BEN in W2-E3; and marker wPt. 2644 was found to be significant by the three models in W2-E4.

For the simulated data, with $\alpha = 0.05$, eight markers were declared significant with M-BRR, but three of them were false positives (markers 9, 29, and 65); six markers were detected to be significant with M-BL (1, 21, 31, 71,101,121) and three with M-BEN (1, 31, 21) (Table 2).

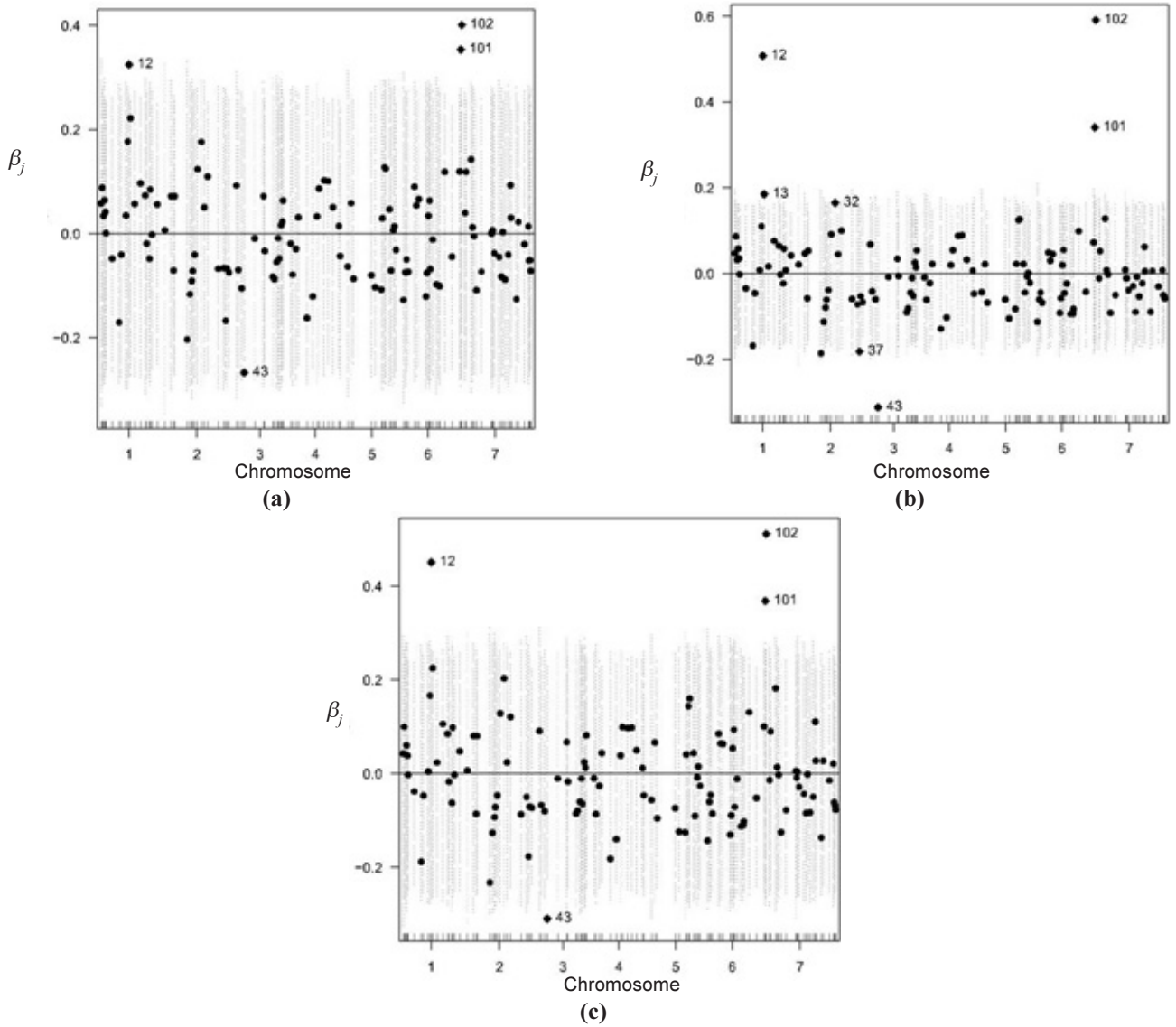

Fig. 2. Significant markers of the Barley data set using different Bayesian shrinkage models: **(a)** M-BRR; **(b)** M-BL ($\delta = 0.02$, $r = 1.1$); **(c)** M-BEN ($r_1 = r_2 = 1$; $\delta_1 = \delta_2 = 10$).

Predictability

Table 3 shows the estimated correlations between phenotypic outcomes and cross-validation (CV) predictions for grain yield data in both wheat data sets. The pedigree model (P) was included as a reference, and the details can be found in Crossa *et al.* (2010). In general for both data sets, all models including pedigree and molecular markers simultaneously (MP) gave better correlations between the predicted and the observed values than their corresponding models, which included only molecular marker data (M) or only pedigree information (P).

The three models MP-BRR, MP-BL, and MP-BEN gave a very similar level of prediction accuracy for both wheat data sets. For Wheat1 data set (W1), the P model performed very well, and only two models were slightly better predictors; the best model was MP-BEN (0.4746), closely followed by MP-BL (0.4725) with gains over the P model of 0.85% and 0.40%, respectively. For Wheat2 (W2) data set with four environments, MP-BRR gave the best correlations in E1 (0.5181), E2 (0.4799) and E4 (0.4926), whereas model MP-BL showed the best predictive accuracy in E3 (0.4347). Model MP-BEN gave a similar level of prediction accuracy as those given by the other Bayesian shrinkage models (BL and BRR).

Table 3. Predictability measured as the correlation between predicted and actual phenotypes, obtained in a 10-fold cross-validation, from data analysis of Barley, grain yield of Wheat1 (W1) and Wheat2 (W2) data sets. Seven models* were fitted to each environment (Barley, W1 and W2E1-W2E4 for the Wheat1 and Wheat2 data sets, respectively). The percent (%) change is relative to the reference pedigree model (P) model.

Data	Model						
	P	M-BL	M-BRR	M-BEN	MP-BL	MP-BRR	MP-BEN
	Correlation						
Barley	—	0.7435	0.7227	0.7264	—	—	—
W1	0.4706	0.3980	0.3942	0.3941	0.4725	0.4676	0.4746
W2-E1	0.4136	0.4943	0.4974	0.4963	0.5163	0.5181	0.5124
W2-E2	0.4049	0.4637	0.4708	0.4635	0.4720	0.4799	0.4730
W2-E3	0.4057	0.3756	0.3781	0.3555	0.4347	0.4342	0.4117
W2-E4	0.4326	0.4589	0.4648	0.4581	0.4894	0.4926	0.4901
Simulated	—	0.6441	0.5996	0.6133	—	—	—
	% change (relative to P)						
Barley	—	—	—	—	—	—	—
W1	—	–15.43	–16.24	–16.25	0.40	–0.65	0.85
W2-E1	—	19.52	20.28	20.00	24.84	25.28	23.89
W2-E2	—	14.51	16.27	14.46	16.57	18.52	16.82
W2-E3	—	–7.41	–6.79	–12.36	7.14	7.02	1.48
W2-E4	—	6.08	7.43	5.89	13.12	13.87	13.29
Simulated	—	—	—	—	—	—	—

*The seven fitted models are: Pedigree model (P), molecular marker regression model using Bayesian LASSO (M-BL); pedigree model plus molecular marker model regression using Bayesian LASSO (MP-BL); Bayesian elastic net model with molecular markers (M-BEN); Bayesian elastic net model with molecular markers and pedigree (MP-BEN); Bayesian ridge regression with molecular markers (M-BRR) and Bayesian ridge regression with molecular markers and pedigree (MP-BRR).

6. DISCUSSION

We extended the BEN model using molecular markers for GS in combination with pedigree information on the wheat lines included in two wheat trials. We applied the methodology developed by Che and Xu (2010) to use the Bayesian shrinkage models (BL, BRR and BEN) in QTL detection. Results obtained with the Barley data set and the two wheat data sets suggest that the M-BEN and M-BL models are sensitive and flexible tools for detecting significant chromosome regions. The results of this study showed that BEN had similar prediction ability than BL and BRR however it is necessary to design and study large scale simulation experiments to study the power and specificity of these models. The results also show that BEN models with molecular markers and pedigree increase the precision for estimating the breeding values of missing genotypes. Theoretically the BEN model should be preferred over the BL and BRR because it has the good properties of BL and BRR. Furthermore, the Elastic Net tends to encourage grouping effects (Zou and Hastie 2005), that is a situation commonly encountered in the case of molecular markers data. For the two wheat data sets, BEN's prediction assessment was almost equal to or better than that of BL and BRR. It has been also reported that the BEN model performs better than BL in problems of variable selection (Li and Lin 2010), which from the biological perspective is related with the QTL detection.

Two theoretical sources of GS accuracy have been reported, one originating from capturing the genetic relationships among individuals (the RR model could benefit from this source of variability) and the other due to the linkage disequilibrium (LD) between the molecular markers and the quantitative trait loci (QTL) (the Bayesian methods could benefit from this source) (Habier *et al.* 2007; Jannink *et al.* 2010). Results of this study using two wheat data sets with pedigree and MM showed that all three kinds of Bayesian shrinkage models, BL, BRR, and BEN, did benefit from using the pedigree alone as well as pedigree and MM information simultaneously; in fact, models BL and BEN effectively increase the accuracy of GS when using molecular markers alone, or pedigree and molecular markers simultaneously.

ACKNOWLEDGMENTS

The authors would like to thank the numerous cooperators in National Agricultural Research Institutes who carried out all the wheat international trials and provided the phenotypic data analyzed in this article. We also thank the International Nursery and Seed Distribution Units in CIMMYT-Mexico for preparing and distributing the seed and computerizing the data.

REFERENCES

Che, X. and Xu, S. (2010). Significance test and genome selection in Bayesian shrinkage analysis. *Intern. J. Plant Geno.* doi:10.1155/2010/893206.

Churchill, G.A. and Doerge, R.W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics,* **138**, 963-971.

Crossa, J., Burgueño, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S.A., Lillemo, M., Singh, R.P., Trethowan, R., Warburton, M., Franco, J., Reynolds, M., Crouch, J. and Ortiz, R. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics,* **177**, 1889-1913.

Crossa, J., de los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J.L., Makumbi, D., Dreisigacker, S., Yan, J., Arief, V., Banziger, M. and Braun, H-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics,* **186**, 713-724.

Crossa, J., Pérez, P., de los Campos, G., Mahuku, G., Dreisigacker, S., and Magorokosho, C. (2011). Genomic selection and prediction in plant breeding. *J. Crop Improvement* (in press).

de los Campos, G., Gianola, D., Rosa, G.J.M., Weigel, K.A. and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Res.,* **92**, 295-308.

de los Campos, G. and Pérez, P. (2010). BLR: Bayesian Linear Regression. R package version 1.1., http://www.r-project.org/.

de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K. and Cotes, J.M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigrees. *Genetics,* **182**, 375-385.

Fisher, R.A. (1935). *The Design of Experiments*. Third Edition. Oliver and Boyd Ltd., London.

Goldstein, M. (1976). Bayesian analysis of regression problems. *Biometrika,* **63**, 51-58.

Habier, D., Fernando, R.L. and Dekkers, J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics,* **177**, 2389-2397.

Haley, C.S. and Knott, S.A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity,* **69**, 315-324.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*. Second Edition. Springer-Verlag, New York, USA.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-447.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics,* **12**, 55-67.

Huang, X.Q., Cöster, H., Ganal, M.W. and Röder, M.S. (2003). Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*., **106**, 1379-1389.

Jannink, J.L., Lorenz, A.J. and Hiroyoshi, I. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics,* **9**, 166-177.

Kumar, N., Kulwal, P.L., Balyan, H.S. and Gupta, P.K. (2007). QTL mapping for yield and yield contribution traits in two mapping populations of bread wheat. *Mol. Breeding,* **19**, 163-177.

Kyung, M., Gill, J., Ghosh, M. and Casella, G. (2010). Penalized regression, standard errors, and Bayesian Lasso. *Bayesian Analy.,* **5(2)**, 369-412.

Li, Q. and Lin, N. (2010). The bayesian elastic net. *Bayesian Analy.,* **5(1)**, 151-170.

McLaren, C.G., Bruskiewich, R., Portugal, A.M. and Cosico, A.B. (2005). The International Rice Information System: A platform for meta-analysis of rice crop data. *Plant Physiology,* **139**, 637-642.

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

Park, T. and Casella, G. (2008). The Bayesian LASSO. *J. Amer. Stat. Assoc*., **103**, 681-686.

Pérez, P., de los Campos, G., Crossa, J. and Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the BLR package in R. *The Plant Genome,* **3(2)**, 106-116.

Quarrie, S.A., Steed, A., Calestani, C., Semikhodskii, A. Lebreton , C., Chinoy, C., Steele N., Pljevljakusiæ, D., Waterman, E., Weyen, J., Schondelmaier, J., Habash, D. Z., Farmer, P., Saker, L., Clarkson, D.T., Abugalieva, A., Yessimbekova, M., Turuspekov, Y., Abugalieva, S., Tuberosa, R., Sanguineti, M-C., Hollington, P.A. and Aragués, R. (2005). A high density genetic map of hexaplod wheat (*Triticum aestivum* L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theo. Appl. Genet,* **110**, 865-880.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org.

Sorensen, D. and Gianola, D. (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Springer-Verlag, New York.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *J. Roy. Statist. Soc.,* **B58**, 267-288.

Tinker, N.A. *et al*. (1996). Regions of the genome that affect agronomic performance in two-row barley. *Crop Science,* **36**, 1053-1062.

Usai, M.G., Goddard, M.E. and Hayes, B.J. (2009). LASSO with cross-validation for genomic selection. *Genetics Research,* **91**, 427-436.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc.,* **B40**, 364-372.

Wang, D., Eskridge, K.M. and Crossa, J. (2010). Identifying QTLs and epistasis in structured plant populations using adaptive mixed LASSO. *J. Agri. Bio. Environ. Stat. (JABES)* DOI: 10.1007/s13253-010-0046-2.

Wang H., Zhang Y.M., Li X., Masinde G.L., Mohan S., Baylink, D.J., and Xu, S. (2005). Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics,* **170**, 465-480.

Weigel, K.A., de los Campos, G., González-Recio, O., Naya, H., Wu, X.L., Long, N., Rosa, G.J.M. and Gianola, D. (2009). Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J. Dairy Sci.,* **92**, 5248-5257.

Xu, S. (2003). Estimating Polygenic Effects Using Markers of the Entire Genome. *Genetics,* **163**, 789-801.

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

Zhang, Y.M. and Xu, S. (2005). A penalized maximum likelihood method for estimating epistatic effects of QTL. *Heredity,* **95**, 96-104.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. Roy. Statist. Soc.,* **B67**, 301-320.

Appendix: Gibbs Sampler for the MP-BEN Model

From equations (4) and (5), and applying Bayes' theorem, the joint posterior distribution of $\{\boldsymbol{\beta}, \boldsymbol{\tau}^2, \mathbf{u}, \sigma_\varepsilon^2, \sigma_u^2, \lambda_1^2, \lambda_2\}$ is given by

$$
\begin{aligned}
&p(\boldsymbol{\beta}, \boldsymbol{\tau}^2, \mathbf{u}, \sigma_\varepsilon^2, \sigma_u^2, \lambda_1, \lambda_2 | \mathbf{y}) \\
&= \prod_{i=1}^{n} N(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + u_i, \sigma_\varepsilon^2) \\
&\quad \times \prod_{j=1}^{p} N(\beta_i | 0, \sigma_\varepsilon^2 (\tau_j^{-2} + \lambda_2)^{-1}) E\exp(\tau_j^2 | \lambda_1) \\
&\quad \times \chi^{-2}(\sigma_u^2 | S_u, df_u) \chi^{-2}(\sigma_\varepsilon^2 | S_\varepsilon, df_\varepsilon) \\
&\quad \times N(\mathbf{u} | \mathbf{0}, \sigma_u^2 \mathbf{A}) \times p(\lambda_1) p(\lambda_2) \qquad \text{(A1)}
\end{aligned}
$$

The conditional posterior distributions required to implement a Gibbs sampler are given below and are obtained using the results of Kyung *et al.* (2010).

1. $p(\boldsymbol{\beta} | else)$

$$
\begin{aligned}
p(\boldsymbol{\beta} | else) &\propto \prod_{j=1}^{n} N(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + u_i, \sigma_\varepsilon^2) \\
&\quad \times \prod_{j=1}^{p} N(\beta_j | 0, \sigma_\varepsilon^2 (\tau_j^{-2} + \lambda_2)^{-1}) \\
&\propto \exp\{ -\frac{1}{2\sigma_\varepsilon^2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u})^T \\
&\quad \times (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u}) + \boldsymbol{\beta}^T \mathbf{D}_\tau^{*-1} \boldsymbol{\beta}] \} \qquad \text{(A2)}
\end{aligned}
$$

where $\mathbf{D}_\tau^* = diag((\tau_1^{-2} + \lambda_2)^{-1}, \ ..., \ (\tau_p^{-2} + \lambda_2)^{-1}))$.

From (A2), the full conditional for $\boldsymbol{\beta}$ is multivariate normal with mean $\mathbf{Z}^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{u})$ and variance $\sigma_\varepsilon^2 \mathbf{Z}^{-1}$, where $\mathbf{Z} = \mathbf{X}^T\mathbf{X} + \mathbf{D}_\tau^{*-1}$. Note that it is not necessary to obtain $\mathbf{Z}^{-1}$ to draw samples from $\boldsymbol{\beta}$, since we can apply results found in Sorensen and Gianola (2002) and sampled from $\beta_j | else$, $j = 1, ..., p$.

2. $p(\tau^{-2} | else)$ (see Kyung *et al.* 2010, p. 404)

3. $p(\sigma_\varepsilon^2 | else)$

$$
\begin{aligned}
p(\sigma_\varepsilon^2 | else) &\propto \prod_{i=1}^{n} N(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + u_i, \sigma_\varepsilon^2) \\
&\quad \times \prod_{j=1}^{p} N(\beta_j | 0, \sigma_\varepsilon^2 (\tau_j^{-2} + \lambda_2)^{-1}) \chi^{-2}(\sigma_\varepsilon^2 | S_\varepsilon, df_\varepsilon) \\
&= \chi^{-2}(\sigma_\varepsilon^2 | S = S_\varepsilon + \mathbf{e}^T\mathbf{e} + \boldsymbol{\beta}^T \mathbf{D}_\tau^{*-1} \boldsymbol{\beta}, df = df_\varepsilon + n),
\end{aligned}
$$

where $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u}$.

4. $p(\mathbf{u} \mid else)$

$$
\begin{aligned}
p(\mathbf{u} \mid else) &\propto MN(\mathbf{u} | \mathbf{0}, \sigma_u^2 \mathbf{A}) \prod_{i=1}^{n} N(y_i | \mathbf{x}_i^T \boldsymbol{\beta} + u_i, \sigma_\varepsilon^2) \\
&\propto MN(\mathbf{u} | \mathbf{0}, \sigma_u^2 \mathbf{A}) \prod_{i=1}^{n} N(y_i^* | u_i, \sigma_\varepsilon^2) \qquad \text{(A3)}
\end{aligned}
$$

where $y_i^* = y_i - \mathbf{x}_i^T \boldsymbol{\beta}$. From Eq. A3, it follows that the posterior density of $\mathbf{u}$ is multivariate normal with covariance $\sigma_\varepsilon^2 \mathbf{C}^{-1}$ and mean $\mathbf{C}^{-1}\mathbf{y}^*$, where $\mathbf{C} = \left[\mathbf{I} + \frac{\sigma_\varepsilon^2}{\sigma_u^2} \mathbf{A} \right]$. Using results shown in Sorensen and Gianola (2002), it follows that each of the entries of $\mathbf{u}$ has a fully conditional distribution with mean $E(u_i | else) = c_{ii}^{-1}(rhs_i - \sum_{k \neq i} c_{ik} u_k)$ and variance $Var(u_i \mid \text{else}) = c_{ii}^{-1}$, where c_{ii}^{-1} and rhs_i are the i-th diagonal element and the i-th entry of $\mathbf{C}$ and $\mathbf{rhs} = \sigma_\varepsilon^2 \mathbf{y}^*$. Thus it is not necessary to invert the $\mathbf{C}$ matrix to draw samples from $\mathbf{u}$.

5. $p(\sigma_\varepsilon^2 \mid else)$

$$
p(\sigma_\varepsilon^2 | else) \propto MN((\mathbf{u} | \mathbf{0}, \sigma_u^2 \mathbf{A}) \chi^{-2}(S_u, df_u) \qquad \text{(A4)}
$$

From Eq. A4, it follows that the fully conditional distribution of σ_u^2 is χ^{-2} with scaling parameter $S_u + \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$ and degrees of freedom $df_u + r$, where r is the order of the square matrix $\mathbf{A}$.

6. $p(\lambda_1^2 \mid else)$, $p(\lambda_2 \mid else)$ (see Kyung *et al.* 2010).

The algorithm described above was implemented in R (R Development Core Team, 2010) and to some parts were coded in C language to speed up the entire process, so the resulting algorithm almost as fast as the one used in the BLR package (de los Campos and Pérez, 2010), for example for the wheat dataset with 599 individuals and 1279 markers, in an Intel Xeon 5530 2.4 GHz and 8 Gb of RAM memory, Bayesian Ridge regression took about 5.5 seconds for 1000 iterations, BL 12 seconds for 1000 iterations and BEN about the same that BL, 13 seconds for 1000 iterations.