Theor Appl Genet (2012) 124:1315–1324
DOI 10.1007/s00122-012-1789-7

ORIGINAL PAPER

Improved LASSO priors for shrinkage quantitative trait loci mapping

Ming Fang · Dan Jiang · Dandan Li · Runqing Yang · Weixuan Fu · Lijun Pu · Huijiang Gao · Guihua Wang · Liyun Yu

Received: 15 July 2011 / Accepted: 5 January 2012 / Published online: 28 January 2012

Abstract Recently, the Bayesian least absolute shrinkage and selection operator (LASSO) has been successfully applied to multiple quantitative trait loci (QTL) mapping, which assigns the double-exponential prior and the Student's *t* prior to QTL effect that lead to the shrinkage estimate of QTL effect. However, as reported by many researchers, the Bayesian LASSO usually cannot effectively shrink the effects of zero-effect QTL very close to zero. In this study, the double-exponential prior and Student's *t* prior are modified so that the estimate of the effect for zero-effect QTL can be effectively shrunk toward zero. It is also found that the Student's *t* prior is virtually the same as the Jeffreys' prior, since both the shape and scale parameters of the scaled inverse Chi-square prior involved in the Student's *t* prior are estimated very close to zero. Besides the two modified Bayesian Markov chain Monte Carlo (MCMC) algorithms, an expectation–maximization (EM) algorithm with use of the modified double-exponential prior is also adapted. The results shows that the three new methods perform similarly on true positive rate and false positive rate for QTL detection, and all of them outperform the Bayesian LASSO.

Communicated by M. Sillanpaa.

M. Fang (✉) · D. Jiang · L. Pu · G. Wang · L. Yu
Life Science College, Heilongjiang Bayi Agricultural University, Daqing 163319, People's Republic of China
e-mail: fangming618@126.com

D. Li
Animal Quarantine Lab, Inspection and Quarantine Technology Center of Hainan Entry-Exit Inspection and Quarantine Bureau, Haikou 570311, People's Republic of China

R. Yang
School of Agriculture and Biology, Shanghai Jiaotong University, Shanghai 201101, People's Republic of China

W. Fu
Key Laboratory of Animal Genetics Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

H. Gao
Institute of Animal Science, Chinese Academy of Agricultural Science, Beijing 100193, People's Republic of China

Introduction

The least absolute shrinkage and selection operator (LASSO) is a shrinkage and selection method for linear regression, which assigns a double-exponential prior for regression coefficient. Recently, many improved LASSO methods have been developed for QTL mapping. The Bayesian LASSO was proposed by Park and Casella (2008) and applied for multiple QTL mapping by Yi and Xu (2008), who assigned two kinds of priors to QTL effect, the LASSO prior (also called double-exponential prior) and the Student's *t* prior. In their method, all the hyperparameters of the priors are treated as variables and estimated along with other parameters rather than set beforehand.

The adaptive LASSO has been proposed by Zou (2006) to practice model selection, which allows different penalization parameters for different regression coefficients, so that the estimates of regression coefficients can be shrunk differently. However, the adaptive LASSO requires consistent initial estimates for the regression coefficients, which are generally not available in oversaturated model (where the number of predictor is greater than that of observations). Sun et al. (2010) have extended the adaptive LASSO to Bayesian adaptive LASSO (BAL) and iterative

adaptive LASSO (IAL), and both methods are suitable for oversaturated model. However, the method is sensitive to the tuning of the hyperparameters δ and τ, with larger values of τ leading to bad separation between QTL and non-QTL effect (Mutshinda and Sillanpää 2010).

Mutshinda and Sillanpää (2010) developed an extended Bayesian LASSO (EBL) for multiple QTL mapping. Rather than assign a double-exponential prior $\pi(\beta_j) \propto \frac{\lambda}{2} e^{-\lambda|\beta_j|}$ with a common rate parameter λ to all regression coefficients, they assigned independent prior $\pi(\beta_j) \propto \frac{\lambda_j}{2} e^{-\lambda_j|\beta_j|}$ to each regression coefficient. Furthermore, they also factorized λ_j as $\lambda_j = \delta\eta_j$, where δ and η_j reflect the level of model sparsity common to each loci and the extent of shrinkage across loci. They claimed that the tuning in the EBL was not critical like that in Bayesian LASSO.

The LASSO prior implemented via EM algorithm (Dempster et al. 1977) was studied by Xu (2010). He treated the effects of regression coefficients as missing data and estimated other parameters by searching the maximum posterior mode of target function. His method can effectively shrink zero-effect QTL close to zero, but it usually needs some experiences or cross-validation for choosing the hyperparameter λ (Xu 2010), which may affect the performance of the method.

In this paper, we modify the double exponential and Student's t priors of Yi and Xu (2008) so that the estimates of zero-effect QTL can be very close to zero. As the two improved priors usually result in close forms of posterior distributions for model parameters, the model parameters can be sampled with efficient Gibbs sampler. In addition to the two Bayesian MCMC algorithms, an EM algorithm with use of the modified double-exponential prior is also adapted. The main feature of the EM algorithm is that the hyperparameter can be estimated with data rather than set beforehand. We both use simulated and real data to valid the proposed methods.

Method

Model

Consider n individuals derived from a backcross (BC) population and genotyped for p markers, the multiple QTL model can be expressed as

$$y_i = \mu + X_i\beta + e_i \quad (1)$$

where y_i is the phenotypes, μ is the population mean; $X_i = (x_{i1}, \ldots, x_{ip})^T$, where $x_{ij}$is the genotype of ith individual and jth marker, which equals to 1 or -1 depending on the genotypes of the marker; $\beta = (\beta, \ldots, \beta_p)^T$, where β_j is the main effect of marker j, and e_i is the residual error, which follows normal distribution, $e_i \sim N(0, \sigma^2)$. Model 1 also can be applied for other populations, but the coding of QTL genotype x_{ij} should be re-defined accordingly.

The Bayesian LASSO

Double-exponential prior

The population mean follows uniform prior, $\pi(\mu) \propto 1$, and the residual variance σ^2 follows non-informative scale-invariant prior $\pi(\sigma^2) \propto 1/\sigma^2$. The regression coefficient β_j follows double-exponential prior (Tibshirani 1996; Park and Casella 2008),

$$\pi(\beta_j) = \frac{\lambda}{2} e^{-\lambda|\beta_j|}, \quad (2)$$

where λ is the hyperparameter. The prior can be factorized into two-level priors (see Yi and Xu 2008). At the first level, β_j follows normal distribution,

$$\beta_j \left| \tau_j^2 \sim N(\beta_j \right| 0, \tau_j^2), \quad (3)$$

and at the second level, τ_j^2 follows exponential distribution,

$$\pi(\tau_j^2|\lambda) = \mathrm{Expon}\left(\tau_j^2 \middle| \frac{\lambda^2}{2}\right) = \frac{\lambda^2}{2} \mathrm{e}^{-\lambda^2\tau_j^2/2}, \quad (4)$$

where λ is treated as variable and will be estimated along with other parameters (Park and Casella 2008; and Yi and Xu 2008). A conjugate Gamma prior, Gamma(a, b), with shape parameter a and scale parameter b being small positive numbers, are assigned to $\lambda^2/2$. The method is called DE (method using double-exponential prior) here.

Student's t prior

The regression coefficient β_j follows Student's t prior,

$$\pi(\beta_j) = t_v(\beta_j|\alpha, s^2), \quad (5)$$

where v, α, and s^2 are the degrees of freedom, the location parameter and the scale parameter, respectively.

The prior (5) can be expressed as two-level priors. At the first level, β_j follows normal distribution (see prior 3); and at the second level, τ_j^2 follows scaled inverse Chi-square distribution,

$$\begin{aligned}\pi(\tau_j^2|v, s^2) &= \mathrm{Inv} - \chi^2\left(\tau_j^2|v, s^2\right) \\ &\propto (\tau_j^2)^{-(v/2+1)} \exp\left(-vs^2/\left(2\tau_j^2\right)\right), \end{aligned} \quad (6)$$

where v and s^2 are shape and the scale parameter, respectively. Both v and s^2 are treated as unknowns and estimated from data. A uniform prior U(0, A) with A being a large number is assigned to s, and thus the posterior distribution

of s^2 follows Gamma distribution, which can be sampled with Gibbs sampling. A uniform prior U(0, 1) is assigned to $1/v$. Since the posterior distribution of v cannot be derived explicitly, the Metropolis–Hasting algorithm (Metropolis et al. 1953; Hastings 1970) is used to update it. The method using the Student's t prior is called ST here.

Improved Bayesian LASSO

Improved double-exponential prior

We modified the LASSO prior by assigning an independent double-exponential prior to each marker effect, i.e., $\pi(\beta_j) \propto \frac{\lambda_j}{2} e^{-\lambda_j|\beta_j|}$, which also can be factorized into two-level priors. One is prior (3), and the other is

$$\pi(\tau_j^2|\lambda_i) = \text{Expon}\left(\tau_j^2 \middle| \frac{\lambda_i^2}{2}\right) = \frac{\lambda_i^2}{2} e^{-\lambda_i^2\tau_j^2/2}. \quad (7)$$

The method is called BIDE (Bayesian algorithm using the improved double-exponential prior). After we have finished the original experiments we found that the prior used here was also studied by Mutshinda and Sillanpää (2010). However, the prior for hyperparameter $\lambda_j^2/2$ is quite different. We assign a prior Gamma(a,b) to $\lambda_j^2/2$ with $a = 0$ and $b = 0$, whereas they factorized λ_j into $\lambda_j = \delta\eta_j$, where δ and η_j were further assigned to uniform priors.

Improved Student's t prior

The improved Student's t prior assigns a uniform prior U(0, 1) on 1/ exp (v) instead of $1/v$. The improved method is called BIST (Bayesian algorithm using improved Student's t prior). With the new prior, the domain of v is from zero to positive infinity. However, in Yi and Xu's Student's t prior, the domain of v is from 1 to positive infinity, which just neglects the values between 0 and 1. The values between 0 and 1 are crucial, since our experiments showed that the posterior estimate of v was very close to zero.

Posterior distributions

For both improved methods, the conditional posterior distributions of μ and β_j follow normal distribution and that of σ^2 follows inverted Chi-square distribution. In BIDE, the posterior distribution of τ_j^{-2} follows inverse Gaussian distribution,

$$\tau_j^{-2}|y,\mu,\beta,\sigma^2,\lambda_i^2 \sim \text{InvGauss}\left(\sqrt{\frac{\lambda_i^2}{\beta_j^2}}, \lambda_i^2\right), j = 1,\ldots,p, \quad (8)$$

and the posterior distribution of λ_j^2 follows Gamma distribution,

$$\lambda_j^2|y,\mu,\beta,\sigma^2,\tau_i^2 \sim \text{Gamma}(1 + a, \tau_i^2/2 + b/2), \quad j = 1,\ldots,p, \quad (9)$$

which is different from the posterior distribution of λ^2 in Yi and Xu (2008).

In BIST, the posterior distributions of τ_j^2 and $s^2$are the same as those in Yi and Xu (2008). However, the update of v with M–H algorithm is not provided by them, and a simple algorithm is developed and presented in Appendix 1.

Extension to EM algorithm

We adapt Xu's (2010) expectation–maximization (EM) algorithm for estimating QTL effects by using the improved double-exponential priors. The difference between the two methods is that we assign an independent double-exponential prior to each marker effect and a Gamma prior, Gamma(0.5,0), to hyperparameter $\lambda_j^2/2$, whereas he assigns a common double-exponential prior with hyperparameter λ set beforehand. It is noted that the prior Gamma(0.5,0) used here is also slightly different from the prior Gamma(0,0) in BIDE, which will be discussed later. The new EM algorithm is called EMAIL (EM algorithm for improved LASSO prior) here and the details are in Appendix 2. The feature of EMAIL is that λ_j^2 can be estimated from data rather than set beforehand.

For charity, the priors and algorithms for the three proposed methods and three related methods are shown in Table 1.

Table 1 The summary of algorithm details for each method

Method	Algorithm	Prior for τ_j^2	Prior for hyperparameter
Bayesian algorithm using the improved double-exponential prior (BIDE)	MCMC	Eq. (7)	Gamma(0,0) to $\lambda_j^2/2$
Bayesian algorithm using improved Student's t prior (BIST)	MCMC	Eq. (6)	U(0, 1) to 1/ exp (v)
EM algorithm for improved LASSO prior (EMAIL)	EM	Eq. (7)	Gamma(0.5,0) to $\lambda_j^2/2$
Double-exponential prior (DE)	MCMC	Eq. (4)	Gamma(a,b) to $\lambda^2/2$ with a and b being small positive numbers
Student's t prior (ST)	MCMC	Eq. (6)	U(0, 1) to $1/v$
Extended Bayesian LASSO (EBL)	MCMC	Eq. (7)	$\lambda_j = \delta\eta_j$ with δ and η_j assigned uniform priors

Simulation studies

We conducted a series of simulated experiments to validate the three new methods. A BC population consisted of 200 individuals was investigated. Each individual was genotyped for 501 markers. The marker interval was 5 cM and the total length of the genome was ~2,500 cM. The effects and the positions of QTL are depicted in Fig. 1. The residual variance was 1.0. As the results, the heritability explained by QTL varied from 1.3 to 22.5% and the total heritability was 85.0%.

The MCMC was run for 20,000 rounds and saved with every 10 rounds. The first 1,000 rounds were discarded as burn-in period. Six methods were used for analysis, including BIDE, BIST, EMAIL, DE, ST, and EBL. For BIST, the initial value of $1/\exp(v)$ was sampled from uniform distribution U(0,1).

General performance

The estimates of marker effects of the six methods are presented in Fig. 1. It can be seen that in BIDE, BIST, EMAIL, and EBL, the effects of most zero-effect markers were estimated very close to zero and the signals of QTL were very clear. Among these methods BIDE and BIST performed quite similar. In general, the estimated QTL effects and positions of the four methods were all very close to their true values.

The results of DE and ST showed that the positions generated clear bumps were also very near the true QTL

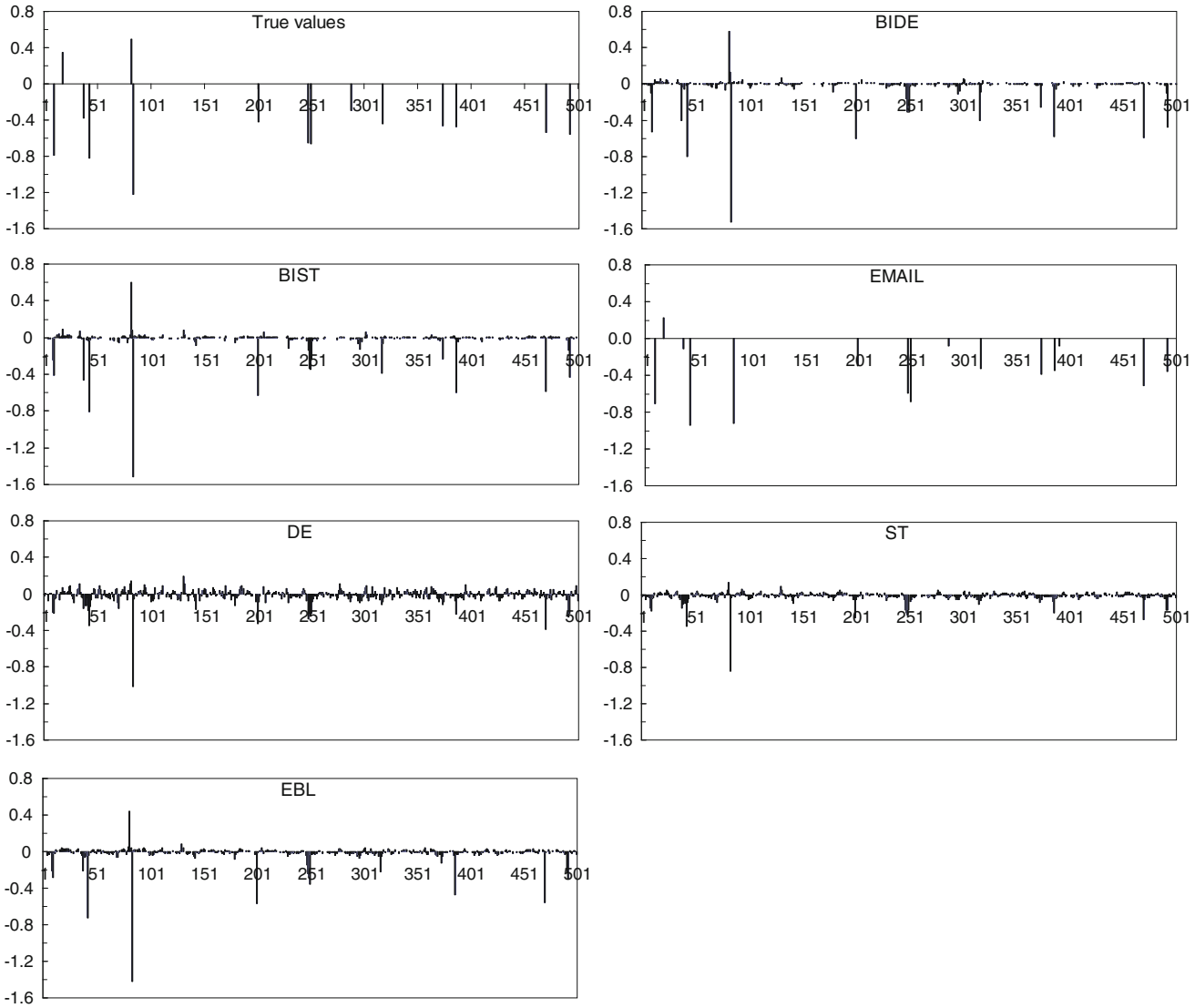

Fig. 1 The true parameters and the estimates of the marker effects with the simulated datasets. The *y*-axis presents the marker effects and the *x*-axis indicates the marker numbers

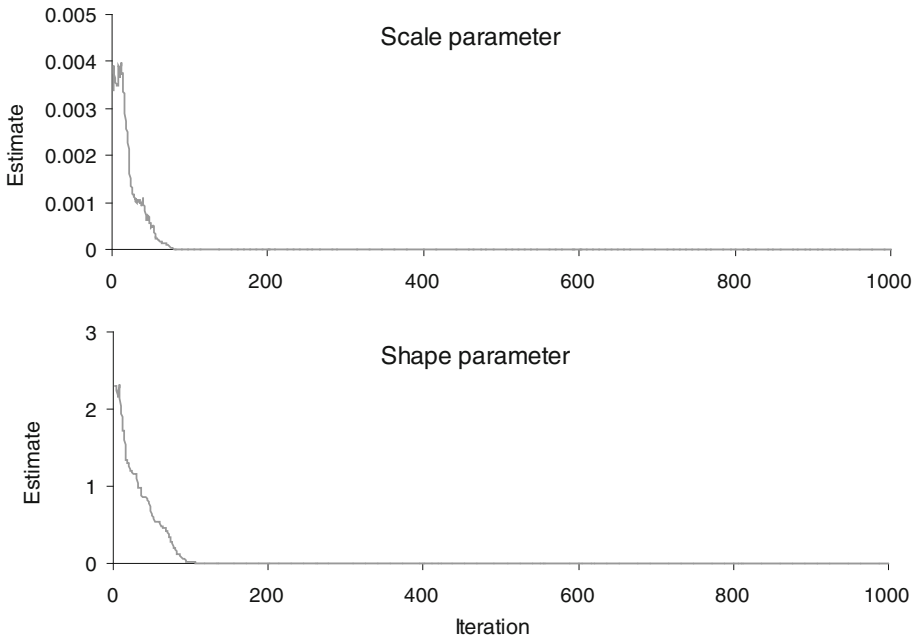

Fig. 2 The traces of the estimates of the scale parameter s^2 (*top panel*) and the shape parameter v (*bottom panel*) of the scaled inverse Chi-square prior against the first 1,000 MCMC iterations for the simulated dataset obtained from BIST

positions. However, many spurious effects were found, and the estimates of QTL effects tended to be underestimated, which might be caused by the major QTL effects absorbed by the nearby markers.

Figure 2 depicts the profiles of the shape parameter v and the scale parameter s^2 against the MCMC iterations in BIST. The values of s^2 and v closely converged to zero after hundreds of iterations. A scaled inverse Chi-square distribution $\chi^2(v, s^2)$ with v and s^2 very close to zero approximates to the Jeffreys' prior $\chi^2(0, 0)$ used by Xu (2003).

Receiver operating characteristics (ROC) profiles

The true positive rate (tp rate, or called power) and false positive rate (fp rate, or called type I error) for each method were assessed in the ROC profile (Fawcett 2006), in which tp rate and fp rate were summarized with replicated experiments under a set of successive thresholds. Following Hoti and Sillanpää (2006), we used the absolute posterior mean effect (for Bayesian methods) or the absolute effect (for EM algorithm) to declare the "significance" of a QTL. The thresholds were evenly divided into 11 successive points from 0.07 to 0.25. The markers at the true QTL positions and nearby the true QTL positions with ±1 locus were defined as QTL locus; other markers were treated as non-QTL locus. At each threshold, the tp rate was defined as the proportion of the number of the QTL detected with 100 replications to the total number of the QTL simulated (15×100). The fp rate was the proportion of the number of all the spurious markers to the total number non-QTL locus ($501–45) \times 100$, with 100 replications. Based on these definitions, tp rate and fp rate were summarized at the 11 successive thresholds for each method.

Figure 3 shows the ROC profiles of the six methods. One method with higher ROC profile is better. Therefore, approximately, the performance of these methods was ranked as BIDE ≈ BIST ≈ EMAIL > EBL > ST > DE.

Analysis of barley data

We used the barley dataset from the North American Barley Genome Mapping Project to test the performance of the proposed methods. The data were collected from a Steptoe × Morex doubled-haploid population containing 150 lines; each was grown in nine different environments for agronomic traits and 16 different environments for malting quality traits, respectively. The diastatic power averaged across nine environments was used for analysis. The dataset included 223 markers covering a genome of ~1,500 cM, which contained ~5% missing marker genotypes that were replaced by their expected values. The data can be downloaded from http://wheat.pw.usda.gov/ggpages/SxM/.

We also applied the six methods to the dataset. For the Bayesian methods, the MCMC was run for 20,000 iterations, and the first 10,000 was discarded as burn-in period.

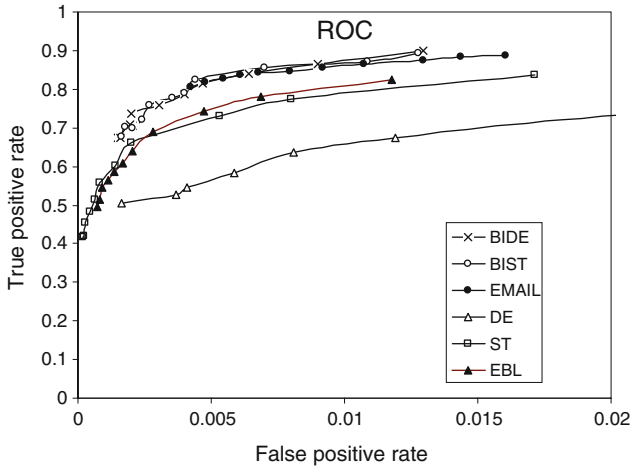

Fig. 3 The ROC profiles of the six methods. The thresholds are evenly divided into 11 successive points from 0.07 to 0.25. For each method, the true positive rate and the false positive rate are summarized at each threshold with 100 replications. It is noted that not all the points are provided for DE and ST

The thinning length was set as 10. Thus, there were 1,000 posterior samples left for posterior analysis.

To declare the "significance" of a QTL, we compared the absolute posterior mean effect (for Bayesian methods) or the absolute effect (for EM algorithm) with the threshold that was obtained from analysis of 100 reshuffled samples (permutation test). The $(1 - \alpha)100$th percentile of the distribution of the largest absolute posterior mean effects (for Bayesian methods) or the largest absolute effect (for EM algorithm) of the reshuffled sample was an approximation of the threshold, where α was taken as 0.05.

Figure 4 depicts the estimates of the marker effects of the six methods. In BIDE, BIST, EMAIL, and EBL, most of the marker effects were shrunk toward zero, so that the profiles were very clear. The profiles of BIDE and BIST were quite similar. The estimates of the shape parameter v and the scale parameter s^2 in BIST are plotted in Fig. 5, and they are very close to zero after about 200 rounds of MCMC iterations. We also tried to use some different

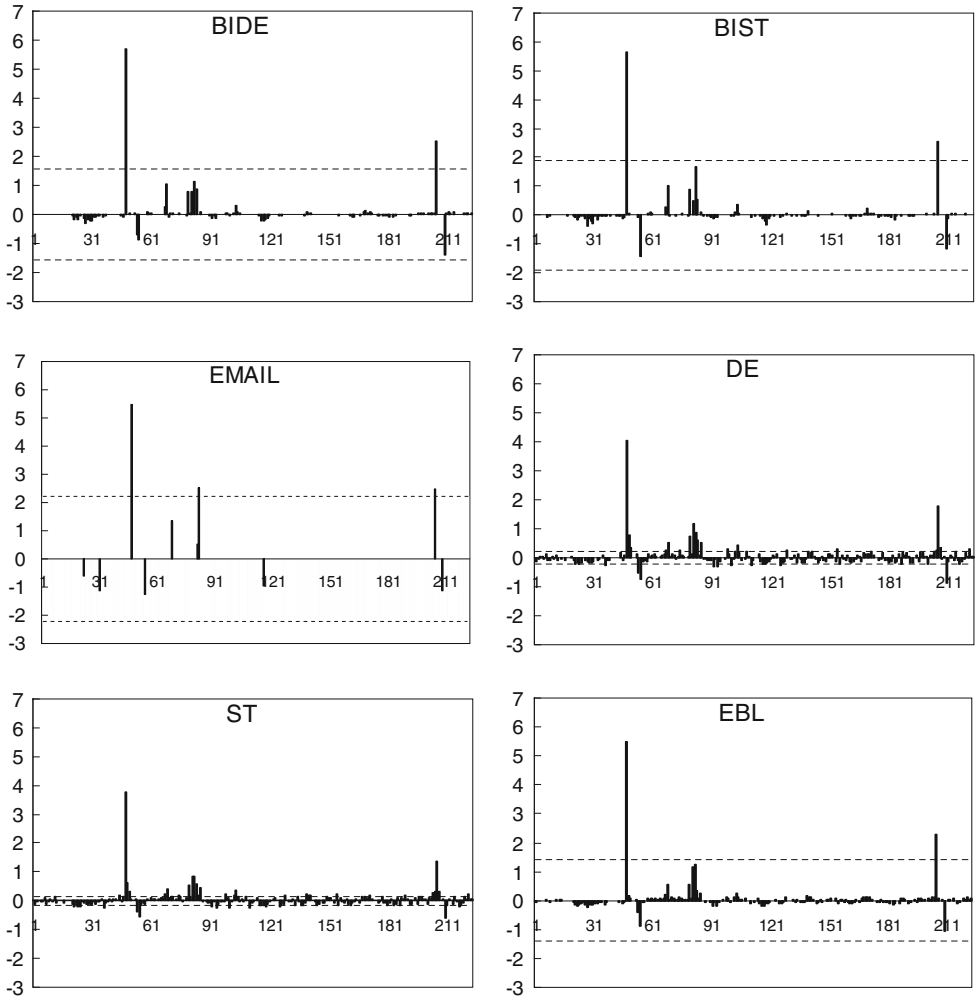

Fig. 4 The estimates of the marker effects with the barley data for each method. The *y*-axis presents the marker effects and the *x*-axis indicates the marker numbers

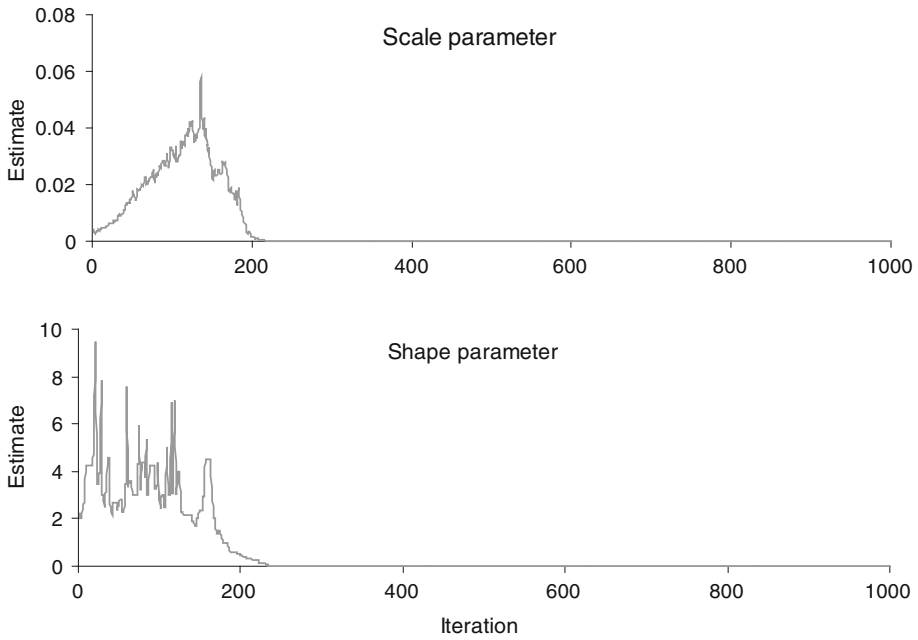

Fig. 5 The traces of the estimates of the scale parameter s^2 (*top panel*) and the shape parameter v (*bottom panel*) of the scaled inverse Chi-square prior against the first 1,000 MCMC iterations for the barley data obtained from BIST

initial values for v and s^2, but the results showed no clear difference. DE and ST generated many noisy signals (see Fig. 4), and it seemed that the major QTL effect was split into many small effects in the neighborhood of the QTL. Moreover, the estimates of the major QTL effects were lower than those with BIDE and BIST. All the conclusions are consistent with those in the simulation studies.

Two markers exceeded the threshold in BIDE, BIST, and EBL, and three in EMAIL. However, too many markers exceeded the threshold in DE and ST, and it seemed that most of them would be spurious markers with regard to the results of our simulated study. All the methods produced two most notable bumps exceeding the threshold at markers 6 and 205 (count from left to right).

Analysis of the common dataset QTL-MAS XII

To test the performance of the proposed method for handling larger number of predictors, the common dataset of QTL-MAS XII (quantitative trait loci marker assisted selection; Lund et al. 2009) was studied. The dataset aims to predict genomic breeding values (GEBV) using 6,000 genome-wide SNP (single nucleotide polymorphisms) markers. We only used the first 500 individuals to estimate the effect of 6,000 SNP markers. The phenotypic values were re-simulated with QTL genotypes, QTL effects and residual effect using model 1. In both data simulation and analysis, the genotype of each marker was coded as -1, 0, and 1, respectively. The residual variance was 1. Both true positions and effects of each QTL are depicted in Fig. 6.

We firstly applied EMAIL to the dataset. As shown in Fig. 6, most of the estimated effects of zero-effect QTL were shrunk toward zero, whereas those of non-zero QTL were close to their true values, which made the signals quite clear.

We also applied BIDE, BIST, and EBL to the dataset, but they tended to produce several false positive signals with abnormally large estimates of the effects, which made the estimate of total genetic variance larger than the total phenotypic variance (data not shown). The same strange results were also reported in Bayesian shrinkage method and stochastic search variable selection (SSVS; Yi et al. 2003) by Xu and Jia (2007) who explained that the strange results might be caused by the smaller sample size. In contrast to the above methods, the Student's t prior was the most robust to large number of predictors, and the double-exponential prior was less robust to the Student's t prior (data not shown).

Discussion

We have presented three improved LASSO priors for shrinkage QTL mapping. Two of them are implemented via Bayesian MCMC algorithm, and one is via EM algorithm. The results showed that three strategies could estimate the effect of zero-effect QTL very close to zero, while estimate

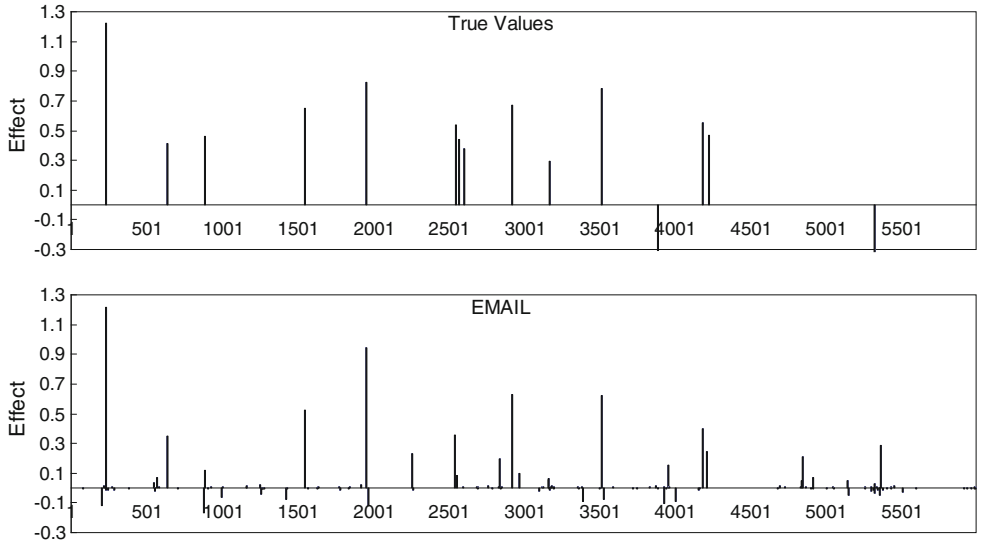

Fig. 6 The estimates of the marker effects with the re-simulated QTL-MAS dataset. The *y*-axis presents the marker effects and the *x*-axis indicates the marker numbers. The *top panel* shows the true effects and positions of the markers, and the *bottom panel* shows the estimates of marker effect with EMAIL

that of non-zero-effect QTL precisely. As a result, they displayed the advantage on true positive rate and false positive rate for QTL detection over the traditional Bayesian LASSO.

BIDE assigns an independent prior to each regression coefficient, which makes the shrinkage of each regression coefficients different. In fact, another key improvement of BIDE is that, a special prior Gamma(0, 0) is assigned to $\lambda_j^2/2$, which is an improper prior $p(\lambda_j^2) \propto 1/\lambda_j^2$. In practice, the improper prior usually leads to special results. Xu (2003) assigned an improper prior $p(\sigma_j^2) \propto 1/\sigma_j^2$ to QTL variance. He found that the improper prior could result in the estimate of the variance and the effect of zero-effect QTL very close to zero. However, in this research, we found that with the improper prior $p(\lambda_j^2) \propto 1/\lambda_j^2$, the estimate of the rate parameter λ_j^2 for zero-effect QTL was quite large. Large λ_j^2 in turn leaded to a very small variance for the prior of τ_j^2, which can be seen from $\mathrm{Var}(\tau_j^2) = 4/\lambda_j^2$; thus, β_j would be shrunk to zero, which can be seen from A3 of Yi and Xu (2008).

In some papers, the improper prior was claimed to have risk to generate improper posterior (e.g., ter Braak et al. 2005). They assigned a proper prior $p(\sigma_j^2) \propto 1/\sigma_j^{2(-1+\delta)}$ with $0 < \delta \leq 0.5$ instead of an improper prior $\chi^{-2}(0, 0)$ (also can be written as $1/\sigma_j^2$) to QTL variance σ_j^2. With two extreme values of δ, 0 and 0.5, the prior results in two priors, $p(\sigma_j^2) \propto 1/\sigma_j^2$ and $p(\sigma_j^2) \propto 1/\sigma_j$, respectively. In BIDE, the proper prior Gamma(a, 0) with $0 < a \leq 0.5$ can be assigned to $\lambda_j^2/2$. At the two extreme values of a, 0 and 0.5, the prior just leads to the two priors $p(\lambda_j^2) \propto 1/\lambda_j^2$ and $p(\lambda_j^2) \propto 1/\lambda_j$. We have attempted to use proper prior by varying a from 0 to 0.05 and conducted several experiments, but the results showed no clear difference (data not shown).

In BIST, we assign a uniform prior $\mathrm{U}(0, 1)$ on 1/exp (v). As the domain of v ranges from zero to positive infinity, the estimate of it may become close to zero. Another simple prior that assigns $\mathrm{U}(0, +\infty)$ on $1/v$ with domain of v ranging from zero to positive infinity, could also lead to the estimate of v close to zero. These results suggest that the allowance of the prior of v to cover the values close to zero is necessary.

The theory of BIDE and EBL is very similar. BIDE uses a locus-specified shrinkage factor λ_j for each marker, while EBL uses $\lambda_j = \delta\eta_j$. If δ is fixed to 1, then the two methods are very similar and the only difference lies in the prior of λ_j. BIDE uses a prior Gamma(0,0), while EBL uses a uniform prior that is equivalent to the prior Gamma(1,0). In EMAIL, a prior Gamma(0.5,0) is assigned to $\lambda_j^2/2$, which performed well in our experiments. An improper prior Gamma(0,0) is also attempted to assign to $\lambda_j^2/2$, but λ_j^2 has no solution (can be seen from Eq. 21).

We have applied the three new methods to the QTL mapping. Another topic that uses genome-wide SNP markers to predict breeding values, called genomic selection (GS), also has been studied widely. The GS firstly estimates the effects of all markers using a training dataset, and then predicts the breeding values for individuals outside the training dataset. The application of the proposed methods to GS is straightforward, but the performance is not assessed here, which is left for further investigation.

Acknowledgments The authors are grateful to the three reviewers for their insightful comments and constructive suggestions that greatly improved our manuscript; one reviewer proposed a uniform prior for BIST, which is added in "Discussion". The authors also thank Dr. Nengjun Yi for providing the BUGS code for LASSO. This work was financially supported by Scientific Research Fund of Heilongjiang Provincial Education Department Grant 11541254 and Chinese National Natural Science Foundation Grant 31001001.

Appendix 1: The M–H algorithm for updating *v*

For simplicity, we update $\omega = 1/\exp(v)$ instead of v. We firstly proposes a new value of ω, $\omega^{(*)}$, around the old one $\omega^{(0)}$ ($\omega^{(*)} = \omega^{(0)} + \delta$, where δ is tuning parameter that is sampled from uniform distribution with bound -0.05 and 0.05). The tuning is very narrow so that $\omega^{(*)}$ can be very close to 1, and thus v can be very close to 0. The acceptance probability is

$$r = \frac{\prod_{i=1}^{p} \pi(\tau_j^2 | v^{(*)}, s^2)}{\prod_{i=1}^{p} \pi(\tau_j^2 | v^{(0)}, s^2)} \times \frac{\pi(\omega^{(*)})}{\pi(\omega^{(0)})} \times \frac{q(\omega^{(0)} | \omega^{(*)})}{q(\omega^{(*)} | \omega^{(0)})} \tag{10}$$

where, $v^{(*)} = -\ln(\omega^{(*)})$and $v^{(0)} = -\ln(\omega^{(0)})$, which are derived from $\omega = 1/\exp(v)$. The first term is the probability density function of scaled inverse Chi-square distribution,

$$\pi(\tau_j^2 | v, s^2) = \frac{(vs^2/2)^{v/2}}{\Gamma(v/2)\tau_j^{2(1+v/2)}} \exp\left(\frac{-vs^2}{2\tau_j^2}\right); \tag{11}$$

the second term is the uniform prior for ω, and $\pi(\omega^{(*)}) = \pi(\omega^{(0)})$; and the third is the proposal ratio. If $\omega^{(*)}$ is not close to the end points 0 or 1, $q(\omega^{(0)} | \omega^{(*)}) = 1/(2\delta)$; if $\omega^{(*)}$ is close to the left bound 0, $q(\omega^{(0)} | \omega^{(*)}) = 1/(\delta + \omega^{(*)})$; and if $\omega^{(*)}$ is close to the right bound 1, $q(\omega^{(0)} | \omega^{(*)}) = 1/(\delta + 1 - \omega^{(*)})$. Similarly, if $\omega^{(0)}$ is not close to the end points 0 or 1, $q(\omega^{(*)} | \omega^{(0)}) = 1/(2\delta)$; if $\omega^{(0)}$ is close to the ends of the bound 0, $q(\omega^{(*)} | \omega^{(0)}) = 1/(\delta + \omega^{(0)})$; and if $\omega^{(0)}$ is close to the right bound 1, $q(\omega^{(*)} | \omega^{(0)}) = 1/(\delta + 1 - \omega^{(0)})$. If $\omega^{(*)}$ is accepted, $v^{(*)} = -\ln(\omega^{(*)})$ is also accepted.

Appendix 2: The EM algorithm using the improved double-exponential prior

Consider the linear model of Eq. 1, the total variance is

$$V = \sum_{j=1}^{p} X_j X_j' \sigma_j^2 + I\sigma^2, \tag{12}$$

where, I is a $n \times n$ identity matrix. We assign the modified double-exponential prior to β_j, which can be factorized into two-level priors, the prior (3) and the prior (7). The prior of $\lambda_j^2/2$ is assigned a Gamma distribution, Gamma(a, b). Let $\Lambda = \left\{\lambda_j^2\right\}_{j=1}^{p}$, $\Sigma = \left\{\tau_j^2\right\}_{j=1}^{p}$ and $\theta = (\sigma^2, \mu, \Lambda, \Sigma)$; then the likelihood can be expressed as

$$\begin{aligned} L(\theta|\beta, y) &= p(y|\beta, \sigma^2, \mu) p(\beta|\Sigma) p(\Sigma|\Lambda) p(\Lambda) \\ &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(y - \mu - X\beta)'(y - \mu - X\beta)\right] \\ &\quad \times \prod_{j=1}^{p} \tau_j^{-1} \exp\left[-\frac{\beta_j^2}{2\tau_j^2}\right] \\ &\quad \times \prod_{j=1}^{p} \frac{\lambda_j^2}{2} \exp\left[-\frac{\lambda_j^2}{2}\tau_j^2\right] \times \prod_{j=1}^{p} b^a \left[\frac{\lambda_j^2}{2}\right]^{a-1} \exp\left[-b\frac{\lambda_j^2}{2}\right]. \end{aligned} \tag{13}$$

The log-likelihood function is

$$\begin{aligned} \ln L(\theta|\beta, y) \propto &-\frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - \mu - X\beta)'(y - \mu - X\beta) \\ &-\frac{1}{2}\sum_{j=1}^{p}\ln(\tau_j^2) - \sum_{j=1}^{p}\frac{1}{2\tau_j^2}\beta_j^2 + \sum_{j=1}^{p}\ln\frac{\lambda_j^2}{2} \\ &-\frac{1}{2}\sum_{j=1}^{p}\lambda_j^2\tau_j^2 + (a-1)\sum_{j=1}^{p}\ln\frac{\lambda_j^2}{2} - \sum_{j=1}^{p} b\frac{\lambda_j^2}{2}. \end{aligned} \tag{14}$$

The EM algorithm developed here treats the model effect β as missing values (Xu 2010).

E-step

The E-step involves taking the expected value of the log-likelihood, conditional on y and under $\theta^{[t]}$, to get

$$\begin{aligned} Q(\theta|\theta^{[t]}) = &-\frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{j=1}^{p}\ln(\tau_j^2) - \sum_{j=1}^{p}\frac{1}{2\tau_j^2}E(\beta_j^2|y, \theta^{[t]}) \\ &-\frac{1}{2\sigma^2}E_{\beta|y,\theta^{[t]}}(y - \mu - X\beta)'(y - \mu - X\beta) \\ &+\sum_{j=1}^{p}\ln\frac{\lambda_j^2}{2} - \frac{1}{2}\sum_{j=1}^{p}\lambda_j^2\tau_j^2 + (a-1)\sum_{j=1}^{p}\ln\frac{\lambda_j^2}{2} - \sum_{j=1}^{p} b\frac{\lambda_j^2}{2}, \end{aligned} \tag{15}$$

where,

$$E_{\beta_j|y,\theta^{[t]}}(y-\mu-\sum_{j=1}^{p}X_j\beta_j)'(y-\mu-\sum_{j=1}^{p}X_j\beta_j)$$
$$=[y-\mu-\sum_{j=1}^{p}X_jE(\beta_j|y,\theta^{[t]})]'$$
$$[y-\mu-\sum_{j=1}^{p}X_jE(\beta_j|y,\theta^{[t]})] \quad (16)$$
$$+\sum_{j=1}^{p}X_j'X_j\mathrm{var}(\beta_j|y,\theta^{[t]}).$$

In Eq. 16,

$$E(\beta_j|y,\theta^{[t]})=\tau_j^{2[t]}X_j'V^{-1}(y-\mu^{[t]}1_n) \text{ and} \quad (17)$$

$$\mathrm{var}(\beta_j\Big|y,\theta^{[t]})=\tau_j^{2[t]}-\tau_j^{2[t]}X_j'V^{-1}X_j\tau_j^{2[t]}, \quad (18)$$

and in Eq. 15

$$E(\beta_j^2\Big|y,\theta^{[t]})=\left[E(\beta_j\Big|y,\theta^{[t]})\right]^2+\mathrm{var}(\beta_j|y,\theta^{[t]}). \quad (19)$$

M-step

The M-step maximizes this expression over μ, τ_j^2, λ_j^2 and σ^2 to produce the next estimate. Set $\frac{\partial}{\partial\mu}Q(\theta\Big|\theta^{[t]})=0$, $\frac{\partial}{\partial\tau_j^2}Q(\theta\Big|\theta^{[t]})=0$, $\frac{\partial}{\partial\lambda_j^2}Q(\theta\Big|\theta^{[t]})=0$ and $\frac{\partial}{\partial\sigma^2}Q(\theta\Big|\theta^{[t]})=0$ to get

$$\mu^{(t+1)}=1_n'(y-\textstyle\sum_{j=1}^{p}X_jE(\beta_j|y,\theta^{[t]})/n, \quad (20)$$

$$\tau_j^{2(t+1)}=\frac{\sqrt{1+4\lambda_j^2E(\beta_j^2|y,\theta^{[t]})}-1}{2\lambda_j^2}, \quad (21)$$

$$\lambda_j^{2(t+1)}=\frac{2a}{\tau_j^2+b}, \quad (22)$$

$$\sigma^{2(t+1)}=\frac{1}{n}\left(y-\mu-\textstyle\sum_{j=1}^{p}X_jE(\beta_j|y,\theta^{[t]})\right)'$$
$$\left(y-\mu-\textstyle\sum_{j=1}^{p}X_jE(\beta_j|y,\theta^{[t]})\right) \quad (23)$$
$$+\frac{1}{n}\textstyle\sum_{j=1}^{p}X_j'X_j\mathrm{var}(\beta_j|y,\theta^{[t]}).$$

Given the initial values for θ, then the EM algorithm proceeds with repeatedly update E-step 17–19) and M-step 20-23 until convergence is reached.

References

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B 39:1–38

Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 25:861–874

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Hoti F, Sillanpää MJ (2006) Bayesian mapping of genotype × expression interactions in quantitative and qualitative traits. Heredity 97:4–18

Lund MS, Sahana G, de Koning DJ, Su G, Carlborg O (2009) Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. BMC Proc 3(Suppl 1):S1

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1091

Mutshinda CM, Sillanpää MJ (2010) Extended Bayesian LASSO for multiple quantitative trait loci mapping and unobserved phenotype prediction. Genetics 185:645–654

Park T, Casella G (2008) The Bayesian Lasso. J Am Stat Assoc 103:681–686

Sun W, Ibrahim JG, Zou F (2010) Genome-wide multiple loci mapping in experimental crosses by the iterative adaptive penalized regression. Genetics 185:349–359

ter Braak CJF, Boer MP, Bink MAM (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 Series B 58:267–288

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

Xu S (2010) An expectation–maximization algorithm for the Lasso estimation of quantitative trait locus effects. Heredity 105: 483–494

Xu S, Jia Z (2007) Genome-wide analysis of epistatic effects for quantitative traits in barley. Genetics 175:1955–1963

Yi N, Xu S (2008) Bayesian Lasso for quantitative trait loci mapping. Genetics 179:1045–1055

Yi N, George V, Allison DB (2003) Stochastic search variable selection for identifying multiple quantitative trait loci. Genetics 164:1129–1138

Zou H (2006) The adaptive Lasso and its oracle properties. J Am Stat Assoc 101:1418–1429