

The Impact of Different Genetic Architectures on Accuracy of Genomic Selection Using Three Bayesian Methods

Research Article

F. Alanoshahr^{1*}, S.A. Rafat¹, R. Imany-Nabiyyi², S. Alijani¹ and C. Robert Granie³

¹ Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

² Department of Animal Science, Faculty of Mathematical Science, University of Tabriz, Tabriz, Iran

³ INRA-INPT-ENSAT-INPT-ENVT, Université de Toulouse, UMR 1388 GenPhySE, Castanet Tolosan, France

Received on: 1 Jul 2016

Revised on: 13 Sep 2016

Accepted on: 15 Sep 2016

Online Published on: Mar 2018

*Correspondence E-mail: f.Alanoshahr@gmail.com

© 2010 Copyright by Islamic Azad University, Rasht Branch, Rasht, Iran

Online version is available on: www.ijas.ir

ABSTRACT

Genome-wide evaluation uses the associations of a large number of single nucleotide polymorphism (SNP) markers across the whole genome and then combines the statistical methods with genomic data to predict the genetic values. Genomic predictions rely on linkage disequilibrium (LD) between genetic markers and quantitative trait loci (QTL) in a population. Methods that use all markers simultaneously may therefore result in greater reliabilities of predictions of the total genetic merit, indicating that a larger proportion of the genetic variance is explained. This is hypothesized that the genome-wide methods deal differently with genetic architecture of quantitative traits and genome. The genomic nonlinear Bayesian variable selection methods (BayesA, BayesB and Bayesian LASSO) are compared using the stochastic simulation across three effective population sizes (N_e). Thereby, a genome with three chromosomes, 100 cM each was simulated. For each animal, a trait was simulated with heritability of 0.50, three different marker densities (1000, 2000 and 3000 markers) and number of the QTL was assumed to be either 100, 200 or 300. The data were simulated with two different QTL distributions which were uniform and gamma ($\alpha=1.66$, $\beta=0.4$). Marker density, number of the QTL and the QTL effect distributions affected the genomic estimated breeding value accuracy with different N_e ($P<0.05$). In comparison of three methods, the greatest genomic accuracy obtained by BayesB method for traits influenced by a low number of the QTL, high marker density, gamma QTL distribution and high N_e .

KEY WORDS BayesA, BayesB, genomic accuracy, LASSO, marker density, N_e .

INTRODUCTION

The genome-wide evaluation combines traditional approaches for prediction of genetic values with using high throughput genotype data such as SNP (Meuwissen *et al.* 2001). Selection based on the genome wide distributed markers estimated breeding values (MEBVs) resulted in increased genetic progress, due to improvement in the accuracy of estimations of the MEBVs, reduction in the generation interval (Meuwissen *et al.* 2001) and reduction in in-

breeding rates, due to emphasis on the MEBVs rather than the family information (Daetwyler *et al.* 2007). The accuracy in obtaining the MEBVs determines the success rate in breeding programs. Reason of this process is that whenever marker density is high enough, the most QTL will be in high LD with some markers and estimates of marker effects will lead to accurate predictions of genetic merit for a trait. A series of the statistical methods mainly differing in the extent of regularization and variable selection has been proposed. Simulation studies revealed clear differences

between methods with respect to their predictive ability. These methods including the ridge regression (Whittaker *et al.* 2000), the Bayesian methods such as BayesA and BayesB (Meuwissen *et al.* 2001), BayesC and BayesC π (Habier *et al.* 2011), Bayesian least absolute shrinkage and selection operator (LASSO) (De los Campos *et al.* 2009), and nonparametric kernel methods (Gianola and van Kaam, 2008). Some widely used as analysis methods rely on a linear mixed model backbone (Meuwissen *et al.* 2001) in which the SNP marker effects are modeled as random effects, drawn from a normal distribution. The predictions of the marker effects are known as the best linear unbiased predictions (BLUPs), which are linear functions of the response variates. The Bayesian methods use a prior for the QTL effect distribution and another prior for the number of the QTL (Meuwissen *et al.* 2001). However, the true distribution of the QTL effects is unknown for many quantitative traits. Goddard (2009) found greater accuracies by using a gamma (1.66, 0.4) prior distribution for the QTL effects compared to a normal prior distribution. The BayesA approach assumes, all SNP to have some effect, however, assumed that some of the SNP are in LD with the QTL of moderate to large effects. But in the BayesB approach assumes, many of the SNP are in genomic regions where there are no QTL and thus have zero effects, whilst a small proportion of the SNP are in the LD with the QTL and consequently do have an effect (Meuwissen *et al.* 2001). The LASSO uses the sum of absolute values of the regression coefficients as a penalty function, which leads to a sparse solution with less than $\min(n, p)$ nonzero elements retained in the model. Several factors affecting the prediction performance of these methods such as genetic trait architecture, span of the LD, sample size, trait heritability, and marker density have been identified (Daetwyler *et al.* 2010; Habier *et al.* 2011). However, how these methods account for the respective factors is still not fully understood, causing uncertainty about the best choice of method for a given population and trait. We hypothesise in this paper that the relative utility of the genome-wide evaluation methods depends significantly on both the genomic structure of the population and the genetic trait architecture.

Thus, the main objective of this study was to compare three non-linear variable selection methods, BayesA, BayesB and Bayesian LASSO, using simulated data across a range of population and trait genetic architectures to investigate the effects of marker density, number of the QTL and N_e on the accuracy of the MEBVs.

MATERIALS AND METHODS

Simulation

The populations were simulated using the QMSim software

(Sargolzaei and Schenkel, 2009) based on forward-in-time process.

A genome consisted of three chromosomes with a length of 100cM was simulated 1000, 2000 and 3000 SNPs were equally spaced over the chromosomes. Three different numbers of QTL (100, 200 and 300) were considered and QTLs were uniformly distributed over the chromosomes. One hundred individuals, including 50 males and 50 females, were simulated for the base population (zero generation). These loci were assumed to be biallelic for both SNPs and QTL with allele frequencies equal to 0.50 (Table 1).

The simulation started with an initial population of 100 N_e individuals and followed by 0.5 N_e and 2 N_e discrete generations, denoted as historical generations. In the initial population and each historical generation, males and females were randomly selected to form N_e matings and produced N_e offspring with 0.5 N_e males and 0.5 N_e females. The parent's gametes were simulated assuming LD based on the Haldane mapping function (Haldane, 1919) to generate recombinant gametes and were randomly combined to create the individual.

The first generation structure was followed through to the 50th generation of random mating to make LD populations. Subsequent to the LD populations 10 more generations (51 to 60) were constructed. The base population consisted of 1000 unrelated animals (500 males and 500 females). In this study, generations 51 and 52 were assumed as training population and the other generations (53 to 60) as validation populations.

In simulating training and validation populations, three QTL numbers (100, 200 and 300), three marker densities (1000, 2000 and 3000) and heritability level of 0.50 were assumed to be influencing the trait of interest. Furthermore, the two different assumed distributions for the QTL effects were uniform and gamma ($\alpha=1.66$, $\beta=0.4$). Overall these assumptions for simulated traits for this study had different genetic architectures. The mutation rate of the markers and QTLs was assumed 2.5×10^{-5} per locus per generation (Solberg *et al.* 2008).

True breeding values (TBV) were generated for all individuals in the training and validation populations. For each individual, TBV was obtained by summing up the effects of all QTL (*i.e.*, $TBV = \sum_{j=1}^m z_j \alpha_j$)

Where:

α_j : effect of QTL j , which was drawn from a gamma distribution with the shape parameter $\beta=0.4$ and scale parameter $\alpha=1.66$; following Meuwissen *et al.* (2001).

m : total number of QTL.

z_j : equals -1 , 0 , or 1 for genotype 11, 12, and 22, respectively.

Table 1 Population structure and simulation parameters

Parameter	Value
Number of chromosomes	3
Genome length	300 cM
Effective population size (Ne)	50, 100 and 200
Numbers of QTL	100, 200 and 300
Number of SNP markers	1000, 2000 and 3000
QTL effects distributions	Uniform and gamma (1.66; 0.4)
Heritability	0.5
Training set	All individuals of generation 51 and 52
Validation set	All individuals of generations 53 to 60

Linkage disequilibrium

The LD measure r^2 (square of the correlation of alleles at two loci) was used for measuring LD (Hill and Robertson, 1968):

$$r^2 = \frac{D^2}{f(A) * f(a) * f(B) * f(b)}$$

Where:

$$D = f(AB) - f(A).f(B).$$

$f(AB)$, $f(A)$, $f(a)$, $f(B)$, $f(b)$: frequencies of haplotypes AB and of alleles A, a, B, b, respectively (Meuwissen *et al.* 2001).

Estimating the breeding values

Three methods, BayesA, BayesB and Bayes LASSO, were used to estimate QTL, SNPs effects and genomic breeding values. The main difference between these three applied approaches is in their assumptions regarding genetic models of the trait. The genomic estimated breeding values (GEBV) for individuals in validation generations for three BayesA, BayesB and Bayes LASSO methods were predicted using the model (Meuwissen *et al.* 2001):

$$GEBV = \sum_i^n X_i \cdot \hat{g}_i$$

Where:

n : number of SNPs across the genome.

X_i : design matrix which refers to individual genotypes for SNPs.

g_i : vector of SNPs effects in chromosome i .

BayesA method

In this model like GBLUP, all SNPs are assumed to have some effects, however, assumed that some of the SNPs are in LD with QTL of moderate to large effects. The SNP effects sampled from a normal distribution with the variance for each SNP sampled from an inverse scaled Chi-square

distribution. The BayesA method was performed using the model (Meuwissen *et al.* 2001):

$$f(\mu|a, b) \sim U(a, b)$$

$$f(e|\sigma_e^2) \sim N(0, I\sigma_e^2), f(\sigma_e^2|a, b) \sim U(a, b)$$

$$f(\sigma_{g_j}^2 | v, s^2) \sim \chi^{-2}(v, s^2)$$

$$v = 4.2, \hat{\sigma}_{SNP}^2 = \frac{\hat{\sigma}_a^2}{(1-\pi) \sum_{m=1}^M 2p_m q_m}$$

$$s^2 = \frac{\hat{\sigma}_{SNP}^2 (v - 2)}{v}$$

Where:

S: scale parameter.

v : number of degrees of freedom.

BayesB method

In BayesB assumes that many of the SNPs are in genomic regions where there are no QTL and thus have zero effects, whilst a small proportion of SNPs are in LD with QTL and consequently do have an effect. This structure means that those effects that are non-zero can be thought of as those in stronger LD with the QTL. In fact, if the number of times a SNP is included in the model (*i.e.* has a non-zero effect) is recorded, the posterior probability of that SNP being linked to a QTL can be calculated. A Gibbs sampling algorithm was implemented to obtain samples from the joint posterior distribution. Steps of the algorithm are outlined below (for details on conditional posterior distribution see Yi and Xu (2008):

1) Sample μ from $N(\mu|y, \beta, \sigma_e^2)$

2) Sample β_j from $N(\beta_j|y, \mu, \sigma_e^2, \sigma_{gi}^2)$

3) Sample σ_e^2 from $Inv - \chi^2(\sigma_e^2|y, \mu, \beta)$

4) For BayesB method: Sample σ_{gi}^2 from $Inv - \chi^2(\sigma_{gi}^2|\beta, v, s^2) -$

$f(g_j|\sigma_{gj}^2, \pi) \begin{cases} \sim N(0, \sigma_{gj}^2) & \text{with probability } (1 - \pi) \\ 0 & \text{with probability } \pi \end{cases}$

$f(\sigma_{gj}^2|v, s^2, \pi) \begin{cases} \sim \chi^{-2}(v, s^2) & \text{with probability } (1 - \pi) \\ 0 & \text{with probability } \pi \end{cases}$

5) Sample s^2 from $Gamma(s^2|\sum_j 1/\sigma_{gi}^2, v)$

6) Sample v with a Metropolis step $(v|s^2, \sum_j 1/\sigma_{gi}^2)$

7) return to step 1.

The Bayes LASSO method

The Bayes LASSO, where a large proportion of marker effects are set to zero and the Bayesian LASSO, where marker effects are modeled using a double exponential distribution, with a high peak at zero and heavy tail that accommodate SNPs with larger effects. Each marker has the same double exponential distribution and no heterogeneous variance either (Park and Casella, 2008). Better estimates are obtained where many possible QTL are estimated to have zero effect or, equivalently, excluded from the model. If all the QTL effects were from a reflected exponential distribution (i.e. without extra weight at zero), an estimator called the LASSO is the appropriate one (Tibshirani, 1996). However, in the situation where many true effects are zero, LASSO still estimates too many nonzero effects. A pragmatic alternative is to exclude from the model but the most highly significant effects.

The λ parameter in the LASSO approach was assigned a gamma(a, b) prior distribution. Values of a and b were set at 0.05 and 1.0, respectively, so that prior of λ was essentially uniform over a wide range of values. Steps of the algorithm are outlined below (Yi and Xu, 2008):

1) Sample μ from $N(\mu|y, \beta, \sigma_e^2)$

2) Sample β_j from $N(\beta_j|y, \mu, \sigma_e^2, \sigma_{gi}^2)$

3) Sample σ_e^2 from $Inv - \chi^2(\sigma_e^2|y, \mu, \beta)$

4) For LASSO method: Sample σ_{gi}^2 from $InvGauss(\sigma_{gi}^2|\beta_j, \lambda)$

5) Sample λ from $Gamma(\lambda^2|\sum_j \sigma_{gi}^2)$

6) return to step 1.

For each analysis, a markov chain monte carlo (MCMC) with 210000 cycles with WinBUGS1.4 software ran and the first 10000 cycles were discarded as burn-in period. Estimates at every 5th iteration were sorted as a sample, resulting in a total 40000 samples.

Comparison of the methods to estimate breeding values

The effects of heritability levels, marker density panels, and the number of QTLs on the accuracy of genomic predictions were evaluated using PROC GLM, and the average accuracies of GEBV were compared using the least squares means (LSM) procedure at $P < 0.05$ (SAS, 2003). The correlation between the GEBV and true genomic breeding value (TGBV) was used as measure of accuracy.

RESULTS AND DISCUSSION

This study demonstrates the accuracy of genomic selection with different numbers of the QTL, marker densities and QTL effect distribution in different N_e . Calculated average LD values between all SNPs (r^2) in the last generation of the LD population (generation 50) was 0.191 ± 0.011 . This indicates that 87% of the expected LD had been achieved in this simulation. The expected LD based on Sved (1971) formula was 0.210.

In this study, the genomic accuracy, the correlations between the TBVs and GEBVs, for different marker densities (1000, 2000 and 3000), different number of the QTLs (100, 200 and 300), the different levels of heritability (0.10, 0.30, and 0.50) with two the QTL effect distributions uniform and gamma are shown in Tables 2 and 3, respectively.

Genomic accuracy under different marker densities and N_e

The relative genomic accuracy increased with increasing of marker densities and N_e . Increasing the marker density from 1000 to 3000, increased the average genomic accuracy with two the QTL effect distributions and three N_e levels (Figure 1).

Increasing the accuracy of the genomic breeding values by increasing marker density to 3000 can be resulted to increase the LD between markers and QTL. The results of this study were agreement with the results of Solberg *et al.* (2008) and Habier *et al.* (2011).

Solberg *et al.* (2008) reported that with increasing marker density from 100 to 800 markers at each Morgan, genomic accuracy increase from 69 to 86 percent. Increasing the number of markers increase the LD between genes and markers, and thus increases the accuracy of genomic evaluations.

Also by increasing N_e , genomic accuracy of breeding values also increased (Tables 2 and 3).

Table 2 The estimated genomic accuracy for different effective population size (N_e), three marker densities and numbers of QTL (NOTL) with uniform QTL effect, $SE < 0.03$ in all scenarios

		Statistical methods								
N_e	N_{OTL}	Bayesa			BayesB			LASSO		
		Marker density								
		1000	2000	3000	1000	2000	3000	1000	2000	3000
50	100	0.441	0.462	0.474	0.563	0.579	0.592	0.558	0.577	0.587
	200	0.432	0.447	0.461	0.557	0.568	0.584	0.551	0.562	0.581
	300	0.417	0.430	0.435	0.542	0.555	0.573	0.537	0.550	0.569
100	100	0.732	0.758	0.789	0.821	0.832	0.868	0.818	0.827	0.865
	200	0.725	0.733	0.745	0.813	0.820	0.831	0.809	0.815	0.828
	300	0.706	0.711	0.718	0.733	0.745	0.751	0.726	0.741	0.746
200	100	0.757	0.774	0.791	0.831	0.847	0.879	0.827	0.844	0.871
	200	0.750	0.761	0.786	0.825	0.838	0.851	0.820	0.836	0.845
	300	0.725	0.732	0.758	0.817	0.826	0.833	0.811	0.824	0.827

Table 3 The estimated genomic accuracy for different effective population size (N_e), three marker densities and numbers of QTL (NOTL) with gamma QTL effect, $SE < 0.03$ in all scenarios

		Statistical methods								
N_e	N_{OTL}	Bayesa			BayesB			LASSO		
		Marker density								
		1000	2000	3000	1000	2000	3000	1000	2000	3000
50	100	0.458	0.473	0.481	0.574	0.585	0.599	0.570	0.578	0.592
	200	0.453	0.468	0.475	0.567	0.572	0.587	0.561	0.566	0.581
	300	0.427	0.436	0.443	0.555	0.564	0.569	0.549	0.558	0.566
100	100	0.744	0.763	0.797	0.844	0.867	0.886	0.839	0.862	0.885
	200	0.731	0.746	0.759	0.830	0.843	0.855	0.826	0.839	0.850
	300	0.712	0.719	0.725	0.818	0.827	0.839	0.815	0.826	0.833
200	100	0.793	0.811	0.825	0.849	0.875	0.893	0.846	0.871	0.887
	200	0.784	0.802	0.817	0.833	0.862	0.873	0.827	0.858	0.868
	300	0.764	0.771	0.780	0.821	0.845	0.854	0.818	0.839	0.847

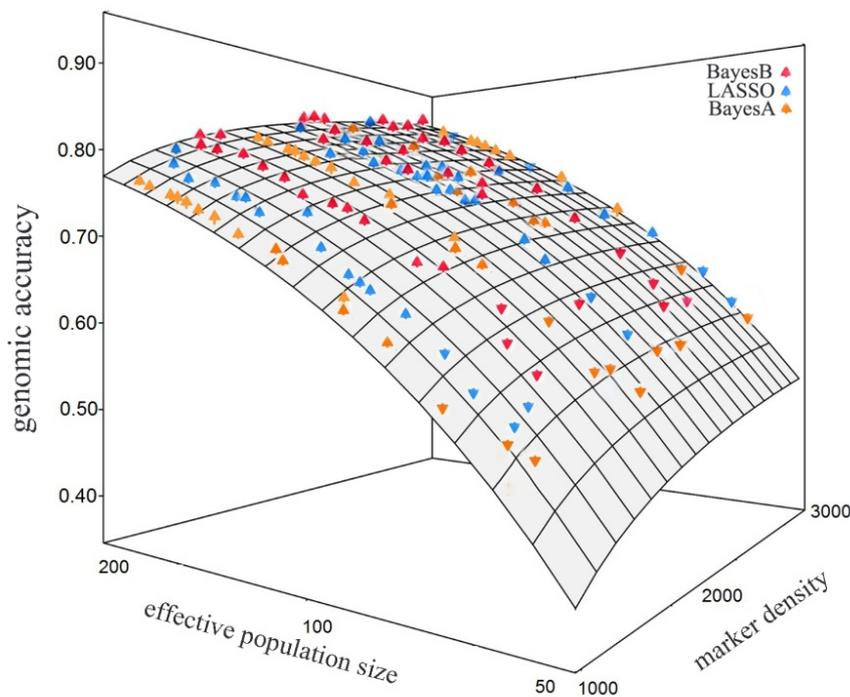


Figure 1 Genomic accuracy of three methods BayesA, BayesB and LASSO with three marker densities levels in three effective population size

The reason can be attributed to increase the number of known data (number of phenotypic records in the base population) versus the number of unknowns variables (SNP effects). When the number of observations in the base population are greater, the SNP effects more precisely estimate and eventually genomic breeding values will be greater accurate (Daetwyler *et al.* 2010). In general, in the same number of QTL and marker density in both uniform and gamma QTL effect distributions, estimated accuracies of GEBV for high N_e (200), were greater than the moderate (100) and low N_e (50), respectively.

Genomic accuracy under different numbers of the QTL and QTL effect distributions

Increasing the numbers of the QTL from 100 to 300, decreased the average genomic accuracy in all three N_e and two the QTL effect distributions (Figure 2). The results of the current study are in agreement with Daetwyler *et al.* (2010) who found a decrease in the accuracy with an increase in the numbers of QTL. By increasing the numbers of QTL for a trait, the average variance of each the QTL for the trait of interest will decrease and the estimation of the QTL effect will be less accurate. With the uniform QTL effect distribution, by increasing the numbers of the QTL, the proportional contribution of each QTL on the trait will be very low and therefore some of their effects will be missed and missing heritability will be increased.

This can be due to the fact that by increasing the numbers of the QTL, the effect of each QTL on the trait will decrease and thus estimated the QTL effects will be small and the QTL effect distribution will be more similar to a uniform distribution. In addition the gamma distributions of the QTL effects resulted in better accuracy in three methods. Shirali *et al.* (2015) also reported better accuracy using BayesC estimation for gamma distribution in the QTL effect. When the distribution of the gene effects is gamma, some genes have a major effects and a high percentage of genes are close to zero impact. So the Bayesian methods compared with non-Bayesian methods are better. These effects can be due to two possible reasons. First, the prior, the QTL effect and the QTL effect are all gamma distributions. Second, the gamma distribution captures the QTL with very high effects compared to a normal and uniform distribution, resulting in more accurate estimation of the GEBVs for traits which are influenced by a numbers of the QTL with high effects (Nadaf and Pong-Wong, 2011).

Genomic accuracy under different methods

The BayesB method was accurate in comparison with both BayesA ($P < 0.05$) and Bayesian LASSO ($P > 0.05$) methods. Among the three methods, the greatest genomic accuracy obtained in low numbers of the QTL (100), high marker density, gamma QTL effect distribution, and large numbers of N_e (200) with BayesB method.

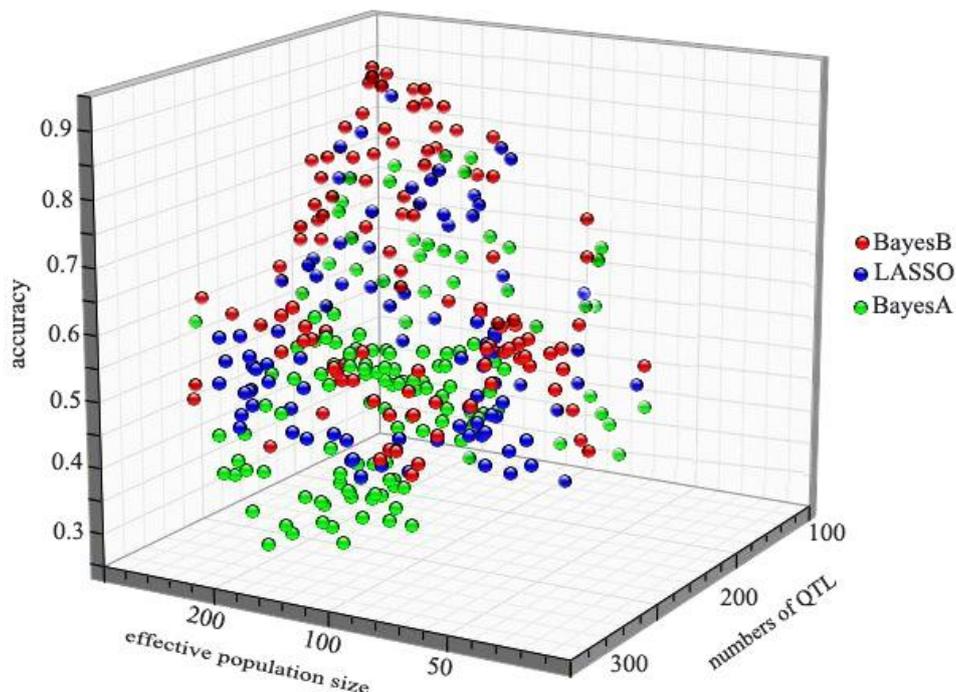


Figure 2 Genomic accuracy of three methods BayesA, BayesB and LASSO with three numbers of QTL in three effective population size

The results of this study are in agreement with Daetwyler *et al.* (2010) and Wimmer *et al.* (2013).

Habier *et al.* (2011) compared different methods of the genomic breeding values evaluation and found that the Bayesian method has high accuracy for any number of markers. Solberg *et al.* (2008) in a study used 1000 phenotypic records in the reference group for a trait with a heritability of 0.5, used Bayesian method for estimating the effects of the markers and reported that the genomic evaluation accuracy of validation set is 0.66. This advantage could be due to the statistical method used. In all scenarios, the LASSO method of analysis was as good as the BayesB method for traits influenced by a low number of the QTL, high marker density and with a gamma QTL effect distribution in high N_e .

CONCLUSION

The extent of LD have major impact on the accuracy of the MEBV. Based on the findings of this simulation study, low QTL number, as well as high dense marker panels, aiming to increase the level of LD between markers and the QTL, will likely be needed for successful implementation of the genomic selection. To implement genomic selection with the LD panels, a training population of sufficient size is necessary. Using a dense marker map covering all chromosomes, it is possible to accurately estimate the breeding value of animals that have no phenotypic record of their own. The BayesB produced estimates with greater accuracies in traits influenced by low number of the QTL and with the gamma QTL effect distribution. Also in the low N_e , the BayesB method is more efficient than others methods. Greater accuracy can be obtained with the Bayesian methods because this methods better takes into account the variable contribution of individual the QTL.

ACKNOWLEDGEMENT

The authors thank C. Robert Granie for encouraging us to write this article and for comments provided on earlier versions of this manuscript.

REFERENCES

- Daetwyler H.D., Pong-Wong R., Villanueva B. and Woolliams J.A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*. **185**, 1021-1031.
- Daetwyler H.D., Villanueva B., Bijma P. and Woolliams J.A. (2007). Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* **124**, 369-376.
- 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 pedigree. *Genetics*. **182**, 375-385.
- Gianola D. and van Kaam J. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*. **178(4)**, 2289-2303.
- Goddard M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetics*. **136**, 245-257.
- Habier D., Fernando R.L., Kizilkaya K. and Garrick D.J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **12**, 186-193.
- Haldane J.B.S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *Genetics*. **8**, 299-309.
- Hill W.G. and Robertson A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226-231.
- Meuwissen T.H.E., Hayes B.J. and Goddard M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. **157**, 321-322.
- Nadaf J. and Pong-Wong R. (2011). Applying different genomic evaluation approaches on QTLMAS2010 dataset. *BMC Proc.* **5(3)**, 9-16.
- Park T. and Casella G. (2008). The Bayesian LASSO. *J. Am. Stat. Assoc.* **103**, 681-686.
- Sargolzaei M. and Schenkel F.S. (2009). QMSim: A large-scale genome simulator for livestock. *Bioinformatics*. **25**, 680-681.
- SAS Institute. (2003). SAS[®]/STAT Software, Release 9.1. SAS Institute, Inc., Cary, NC, USA.
- Shirali M., Miraei-Ashtiani S.R., Pakdel A., Haley C. and Pong-Wong R. (2015). A comparison of the sensitivity of the BayesC and genomic best linear unbiased prediction (GBLUP) methods of estimating genomic breeding values under different quantitative trait locus (QTL) model assumptions. *Iranian J. Appl. Anim. Sci.* **5(1)**, 41-46.
- Solberg T.R., Sonesson A.K., Woolliams J.A. and Meuwissen T.H.E. (2008). Genomic selection using different marker types and densities. *J. Anim. Sci.* **86**, 2447-2454.
- Sved J.A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* **2**, 125-141.
- Tibshirani R. (1996). Regression shrinkage and selection via the LASSO. *J. Roy. Stat. Soc. B Met.* **58**, 267-288.
- Whittaker J.C., Thompson R. and Denham M.C. (2000). Marker-assisted selection using ridge regression. *Genet. Res.* **75(2)**, 249-252.
- Wimmer V., Lehermeier C., Albrecht T., Auinger H.J., Wang Y. and Schön C.C. (2013). Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics*. **195**, 573-587.
- Yi N. and Xu S. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics*. **179**, 1045-1055.