IJAS Iranian Journal of Applied Animal Science

Accuracy of Genomic Prediction under Different Genetic Architectures and Estimation Methods

Research Article

A. Atefi[1], A.A. Shadparvar[1*] and N. Ghavi Hossein-Zadeh[1]

[1] Department of Animal Science, Faculty of Agricultural Science, University of Guilan, Rasht, Iran

Received on: 14 Feb 2017
Revised on: 18 Jun 2017
Accepted on: 1 Jul 2017
Online Published on: Mar 2018

*Correspondence E-mail: shad@guilan.ac.ir

Online version is available on: www.ijas.ir

ABSTRACT

The accuracy of genomic breeding value prediction was investigated in various levels of reference population size, trait heritability and the number of quantitative trait locus (QTL). Five Bayesian methods, including Bayesian Ridge regression, BayesA, BayesB, BayesC and Bayesian LASSO, were used to estimate the marker effects for each of 27 scenarios resulted from combining three levels for heritability (0.1, 0.3 and 0.5), training population size (600, 1000 and 1600) and QTL numbers (50, 100 and 150). A finite locus model was used to simulate stochastically a historical population consisting 100 animals at first 100 generations. Through next 100 generations, the population size gradually increased to 1000 individuals. Then the animals in generations 201 and 202 having both known genotypic and phenotypic records were assigned as reference population, and individuals at generations 203 and 204 were considered as validation population. The genome comprised five chromosomes of 100 cM length and 500 single nucleotide polymorphism markers for each chromosome that distributed through the genome randomly. The QTLs and markers were bi-allelic. In this study, the heritability had great significant positive effect on the accuracy ($P<0.001$). By increasing the size of the reference population, the average genomic accuracy increased from 0.64±0.03 to 0.70 ± 0.04 ($P<0.001$). The accuracy responded to increasing number of QTLs non-linearly. The highest and lowest accuracies of Bayesian methods were 0.40 ± 0.04 and 0.84 ± 0.05, respectively. The results showed having the greatest amount of information (i.e. highest heritability, highest contribution of gene action in phenotypic variation and large reference population size), the highest accuracy (0.84) was obtained, with all investigated methods of estimation.

KEY WORDS accuracy, Bayesian, genetic architecture, genomic, heritability, QTL.

INTRODUCTION

The estimation of breeding values in order to select the best animals as parents of the next generation is the main goal of animal breeding programs. Traditional methods of genetic evaluation were performed using a combination of phenotypic and pedigree information to produce estimated breeding values (EBV) (Dekkers, 2012). The rapid progress and reducing costs of genotyping of whole genome have led to a great interest in using molecular markers information to identify individuals of high genetic merit (Daetwyler *et al.* 2010). Meuwissen *et al.* (2001) proposed an approach called genomic selection (GS), which uses high density markers to estimate breeding values. Using simulations, they showed that with a dense marker panel, it is possible to accurately estimate the breeding value of animals without information about their phenotype or that of close relatives (Moser *et al.* 2009). The accuracy of GS is expected to be considerably higher than that of traditional best linear unbiased prediction (BLUP) selection (Daetwyler *et al.* 2008;

Goddard, 2009; Hayes *et al.* 2009). In addition, genome-wide selection reduces inbreeding rates due to increasing emphasis on own rather than family information, that is a better estimation of mendelian sampling term (Daetwyler *et al.* 2007; Dekkers, 2007). Genomic estimated breeding values (GEBV) can be calculated for both sexes at the early time of life. Therefore, the GS can increase the profitability through accelerated genetic gain resulted from reduced generation interval and lowering the cost of proving animals.

In whole-genome analyses, the number of marker effects to be estimated, may exceed the number of individuals (curse of dimensionality). Under this condition the models are at risk of being over parameterized. In order to deal with these problems, estimation of marker effects is often performed using penalized estimation methods such as ridge regression, the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), Bayesian methods, semi-parametric (Gianola *et al.* 2006) or non-parametric methods. Among the Bayesian methods, those using marker-specific shrinkage of effects (e.g., BayesA or BayesB of Meuwissen *et al.* 2001, or the Bayesian LASSO of Park and Casella (2008) are commonly used in animal breeding applications.

The Bayesian methods proposed by Meuwissen *et al.* (2001) differ in the way of looking at the variances of parameters. BayesA applies the same prior distribution for all of the variances of the markers. BayesB assumes that some markers contribute largely to the genetic variation, and seems more realistic for GS than Bayes A.

BayesC uses a common variance for all markers and the scale parameter of the scaled inverse chi square distribution is the user pre-specified value. Park and Casella (2008) introduced the Bayesian LASSO method for estimating the regression coefficients. They connected the LASSO method with the Bayesian analysis using Tibshirani's idea. Tibshirani (1996) noticed that the LASSO estimates of the regression coefficients can be interpreted as posterior mode estimates assuming double exponential prior distributions for the regression coefficients.

Many studies have shown that factors such as size of the reference data set (Meuwissen *et al.* 2001; VanRaden and Sullivan, 2010), trait heritability, the number of loci affecting the trait (Daetwyler *et al.* 2008), the degree of genetic relationships between training and validation samples (Habier *et al.* 2007) and distributions of allele frequencies (Clark *et al.* 2011) have great effect on accuracy of genomic prediction.

The aim of this study was to investigate the accuracy of five genomic evaluation methods under various levels of reference population size, trait heritability and the number of QTL.

MATERIALS AND METHODS

Simulation

Various scenarios were defined according to all combinations of three different levels of heritability, training population size and QTL numbers. For each scenario five Bayesian methods of estimation were compared in terms of prediction accuracy, the correlation between the predicted genomic breeding values and the true values. Parameter estimate was performed via Gibbs Sampler algorithm implemented in the BGLR package of R software (Perez and De los Campos, 2014).

A historical population of 100 effective numbers with equal sex ratio was simulated using QMSim software, assuming the heritability values of 0.1, 0.3 or 0.5. During the first 100 historical generations, mating was performed by drawing the parents of an animal randomly from the animals of the previous generation. Then, in order to arrive at a mutation-drift balance, 100 more generations were simulated while increasing the population size to 1000 individuals gradually. After the last historical generation, the recent population was constructed by random selection of 300, 500 or 800 individuals and four successive generations were generated by random mating. The animals in generations 201 and 202 with known genotypes and records for the trait constructed the training population. The animals of generations 203 and 204 formed the validation population, which assumed having no phenotypic records.
The genome is comprised of five chromosomes of 100 cM, on which 500 marker loci and QTL loci were randomly distributed.

All marker and QTL loci were bi-allelic. The number of segregating QTL affecting the trait was set at 50, 100 or 150. The Marker and QTL allele frequencies were assumed to be equal in the 200th generation. The following quality control measures were applied to the SNP data: markers with a minor allele frequencies (MAF) < 0.1 and a Hardy-Weinberg Equilibrium (HWE) P-value < 0.000001 were removed. Samples with genotype failure rate greater than 0.1 were also removed.

Linkage disequilibrium calculating

To achieve accurate genomic prediction, sufficient level of linkage disequilibrium (LD) is imperative. The extent of LD in the training populations was measured by r^2 (Hill, 1973):

$$r^2 = D^2 / (freq(A1)\times freq(A2)\times freq(B1)\times freq(B2)$$

Where:
freq (A1): frequency of A1 allele, and likewise for the other alleles in the population.

D: another statistic of linkage disequilibrium that was calculated as:

D= freq(A1-B1) × freq(A2-B2) - freq(A1-B2) × freq(A2-B1)

PLINK software and Synbreed and GGPLOT2 packages were used to calculate and display the LD properties.

Models

Following linear model was used to estimate the marker effects:

$$Y= \mu + X\beta + \varepsilon \quad [2]$$

Where:
Y: phenotypic value.
μ: population mean.
X: marker design matrix.
β: vector of marker effects
ε: error term that is assumed to be normally distributed with mean and variance equal to 0 and σ^2.

The estimator of β is:

$$(X'X+\lambda I)^{-1}X'y$$

Where:
λ: regularization parameter.

The elements of the X for each individual depended on the number of alleles present in its genotype. For example, per i^{th} individual having genotypes AA, Aa or aa at j^{th} marker locus the X_{ij} element in X was assigned equal to 2, 1 or 0, respectively.

BRR, Bayes A, B, C and Bayesian LASSO

Ridge regression best linear unbiased predictor (RR-BLUP) assumes all markers have a common variance (Meuwissen *et al.* 2001) and therefore shrinks equally for each marker effect. Bayesian Ridge regression (BRR) makes the same assumptions, but the level of shrinkage is estimated with a Bayesian hierarchical model. In a Bayesian Ridge regression, the conditional prior assigned of marker effects are independent and identically distributed (IID) and have a normal prior distribution:

$$p(\beta_j \left| \theta_{\beta_j}, \sigma^2 \right.) = N(\beta_j \left| 0, \sigma^2_\beta \right.) \quad \text{and}$$

$$p(\theta_{\beta_j} \left| \omega \right.) = \chi^{-2}(\sigma^2_\beta \left| df_\beta, s_\beta \right.)$$

Where:
β_j: marker effect.
$p(\beta_j|\theta_\beta,\sigma^2)$: prior density of the j^{th} marker effect.
$\theta_{\beta j}$: vector of parameters indexing the prior density assigned to marker effects.
$p(\theta_{\beta j}|\omega)$: prior density assigned to $\theta_{\beta j}$.
ω: parameters indexing this density.

Meuwissen *et al.* (2001) proposed Bayesian regressions including BayesA and BayesB on the genomic markers. BayesA assumes a normal prior distribution on the SNPs effects, with zero mean and variance σ_j^2 associated to each marker.
This variance is assumed to be distributed as a scaled inverted chi-squared probability distribution $X^{-2}(\nu; S^2)$ with degrees of freedom ν and scale parameter S^2 as the prior distribution.

BayesB assumes a normal prior distribution on the markers effects with zero mean and variance σ_j^2. Then, a mixture of distributions is assumed on this variance being equal to zero with probability π and distributed as in BayesA with probability 1 - π.

BayesC was proposed to compensate some of the deficiency of BayesB, as the estimation of the probability π or the distribution of mixtures, which in BayesC is applied on the SNPs effects instead of the variances. In a comparison using simulated data, Bayes BLUP, BayesA, BayesB and BayesC had the same predictive ability with correlation over 0.85 (Verbyla *et al.* 2010).

Park and Casella (2008) introduced the Bayesian LASSO method for estimating the regression coefficients. De los Campos *et al.* (2009) used the Bayesian LASSO in GS. The LASSO estimates can be viewed as the posterior mode in a Bayesian model considering a double-exponential prior for the regression coefficient estimates. The summary of investigated scenarios (Each scenario was repeated for 10 times) and statistical methods is presented in Table 1.

Prediction accuracy

The correlation coefficient between the true breeding values (BV) and the genomic predicted BV ($r_{TBV,GEBV}$) was used as a measure of the accuracy. Fitting five Bayesian methods, the GEBV values for all scenarios were predicted. An analysis of variance was performed to investigate the effect of method, heritability, number of individual in training population and number of QTL on the accuracy. The model to investigate the factors affecting the accuracy was:

$$y= \mu + method + h^2 + N_{QTL} + N_{IND} + interaction\ effects + \varepsilon \quad [3]$$

Where:

y: $r_{TBV,GEBV}$.

μ: overall mean,

method: effect of method (BRR, Bayes A, B, C and BL).

h^2: effect of heritability (0.1, 0.3 and 0.5).

N_{QTL}: effect of number of QTL (50, 100 and 150).

N_{IND}: effect of the number of individuals in each generation of training population.

interaction effects: two-way interactions between main effects.

ε: random error.

Table 1 The summary of investigated scenarios and statistical methods

N_{IND}[1]	N_{QTL}[2]	H2[3]	Model[4]
		0.1	
	50	0.3	
		0.5	
		0.1	
300	100	0.3	Five Bayesian methods including: BRR, BA, BB, BC, BL
		0.5	
		0.1	
	150	0.3	
		0.5	
		0.1	
	50	0.3	
		0.5	
		0.1	
500	100	0.3	Five Bayesian methods including: BRR, BA, BB, BC, BL
		0.5	
		0.1	
	150	0.3	
		0.5	
		0.1	
	50	0.3	
		0.5	
		0.1	
800	100	0.3	Five Bayesian methods including: BRR, BA, BB, BC, BL
		0.5	
		0.1	
	150	0.3	
		0.5	

[1] The number of individuals in each generation of training population.
[2] The number of QTLs.
[3] Heritability.
[4] Bayesian Ridge regression, BayesA, BayesB, BayesC and Bayesian LASSO.

The statistical analyze of all main and interaction effects were conducted using the GLM procedure of SAS software (SAS, 2003).

The expected accuracy of genome-wide selection has been anticipated as a function of the training population size (N), trait heritability (h^2), and the effective number of quantitative trait loci (QTL) or effective number of chromosome segments underlying the trait (Me) (Daetwyler *et al*. 2008; Daetwyler *et al*. 2010):

$$r_{g\hat{}g} = [Nh^2 / (Nh^2 + M_e)]^{1/2} \qquad [2]$$

Where:

r: expected correlation between predicted genotypic value and true genotypic value.

M_e: refers to the idealized concept of having a number of independent, bi-allelic, and additive QTL affecting the trait (Daetwyler *et al*. 2008). The M_e is a function of the breeding history of the population and of the length of the genome. The objective of this research was to investigate the accuracy of GEBV under various underlying genetic architecture using some different Bayesian methods.

RESULTS AND DISCUSSION

Marker statistics and extent of LD

The mean values of r^2 for each chromosome are shown in Table 2. An overall mean value of 0.19 was observed for r^2. The largest gap between SNPs (12.18 cM) was located on chromosome 4. The highest and lowest number of SNPs and therefore the highest and lowest mean of r^2 were located on chromosome 1 and 4 respectively (Figures 1a and 1b). The sufficient average LD over the entire genome is necessary for accurate estimations in genomic selection and whole-genome association studies. Calus *et al*. (2007) demonstrated that if the mean r^2 between adjacent SNPs was > 0.2, accurate genomic breeding values could be obtained. In Holstein-Friesian cattle, r^2 of 0.2 occurs at approximately 100 kb, suggesting that 30000 markers should be sufficient to apply genomic selection. The extent of genome-wide LD considerably depends on the past effective population size. In a simulation study, Meuwissen *et al*. (2001) demonstrated that, to get very accurate genomic estimated breeding values, $10NeL$ markers are required, where L is the length of the genome in Morgan and Ne is the effective population size. In Holstein-Friesian cattle, Ne is approximately 100, and the length of the genome is 30 Morgans, again suggesting that 30000 markers are required. In species with large effective population sizes, dense marker panels will be required. Provided the number of markers are enough (i.e. LD=0.2 that was obtained in the current study), the accuracy of GEBV will depend on the number of individuals genotyped and phenotyped in the reference population, the heritability of the trait, and the number of loci affecting the trait (Daetwyler *et al*. 2008; Goddard, 2009).

Factors affecting the prediction accuracy

Table 3 shows the result of analysis variance for accuracy and implies that the effect of all main factors, including method, heritability, number of QTL, number of individuals in each generation of training population and all interaction effects, except *Method* × h^2 and *Method* × N_{QTL} × h^2, were significant (P<0.05).

Table 2 Statistical information for genome-wide LD (measured by r^2)

Chromosome	Number of SNP pairs	Minimum r^2	Maximum r^2	Average r^2
Chr 1	2466	8.6E-8	1	0.17±0.23
Chr 2	2574	5.8E-10	1	0.18±0.22
Chr 3	2430	4.3E-10	1	0.17±0.22
Chr 4	2025	1.7E-8	1	0.22±0.25
Chr 5	1926	2.3E-7	1	0.19±0.25

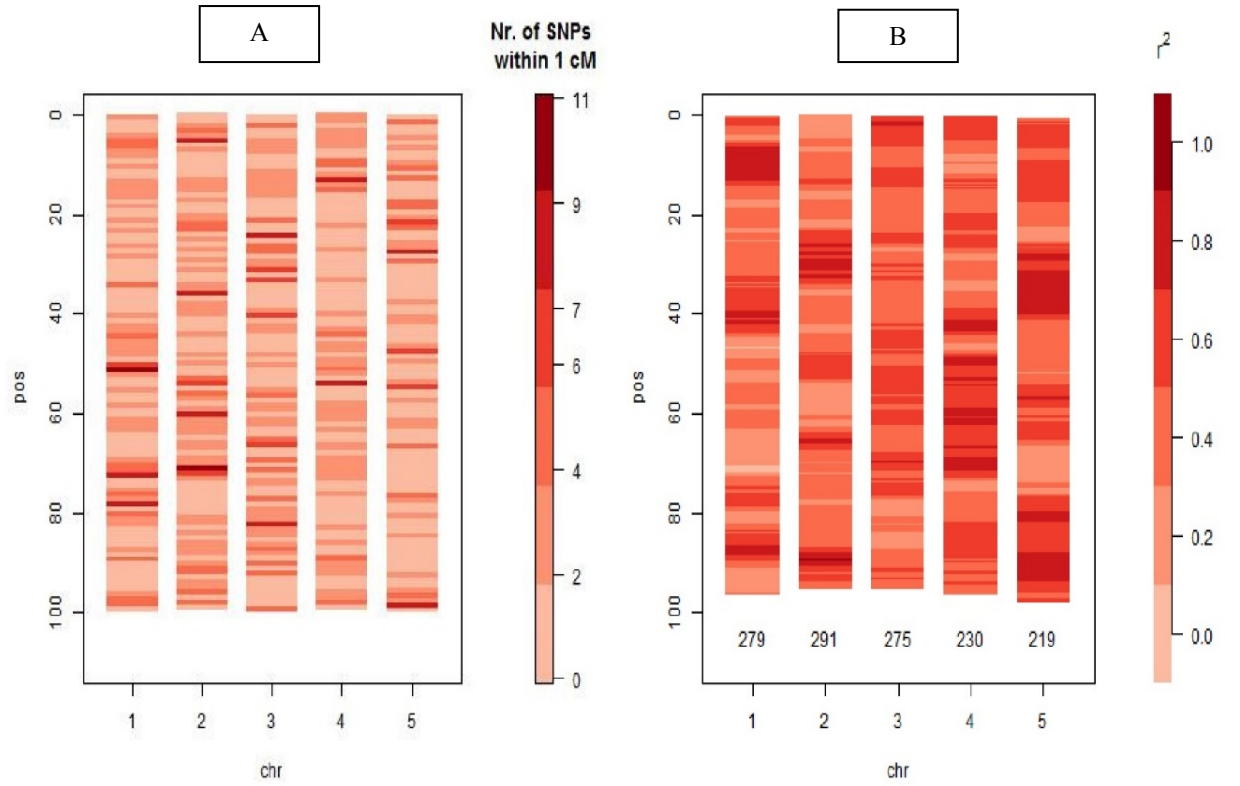

Figure 1 a) Density visualization of marker map
b) visualization of pairwise LD estimates versus marker distance

Table 3 Accuracy of Bayesian methods for different genetic architectures

Source of variation	DF	SS	MS	F-value	Pr > F
Method	4	0.00596389	0.00149097	69.05	< 0.0001
N_{IND}[1]	2	0.08591211	0.04295606	1989.43	< 0.0001
N_{QTL}[2]	2	0.01828878	0.00914439	423.5	< 0.0001
H2[3]	2	1.8334959	0.91674795	42457.4	< 0.0001
Method × N_{IND}	8	0.00044007	0.00005501	2.55	0.0212
Method × N_{QTL}	8	0.00169005	0.00021126	9.78	< 0.0001
Method × H2	8	0.0001988	0.00002485	1.15	0.348
N_{IND} × N_{QTL}	4	0.0940171	0.02350427	1088.55	< 0.0001
N_{IND} × H2	4	0.00037608	0.00009402	4.35	0.0044
N_{QTL} × H2	4	0.00955119	0.0023878	110.59	< 0.0001
Method × N_{IND} × N_{QTL}	16	0.00190706	0.00011919	5.52	< 0.0001
Method × N_{QTL} × H2	16	0.00052593	0.00003287	1.52	0.1308
N_{IND} × N_{QTL} × H2	8	0.03980656	0.00497582	230.45	< 0.0001

[1] The number of individuals in each generation of training population.
[2] The number of QTLs.
[3] Heritability.

According to the F values in Table 3, the descending order of the main factors in terms of importance was heritability, reference population size, number of QTL and the estimating method. Among the interaction effects, the effects containing the reference population size had higher importance.

Main factors affecting the genomic evaluation accuracy

Figure 2 presents the plots of correlations (R) between true breeding value and GEBV obtained for the validation population, for the different heritability (plot *a*), training population size per generation (plot *b*), number of QTLs (plot *c*) and marker effect estimating methods (plot *d*).

As shown in Figure 2a, heritability of the trait affects the accuracy of genomic breeding values severely. According to Daetwyler *et al.* (2008), a trait with a heritability of 0.8 is expected to yield the same accuracy as a trait with a heritability of 0.25 but in a reference population that includes 3.2 times more animals. For low heritability traits, such as fertility and health, for the same reference population size, lower accuracy of genomic predictions were obtained (Daetwyler *et al*. 2008; Goddard, 2009).

The training population size is the factor that is most easily controlled by the investigator. By increasing the size of the training population, from 300 to 800, there was an ascending trend in average genomic accuracy from 0.64 to 0.70 (Figure 2b). Increasing the accuracy by the size of the training population, has been anticipating using simulation studies (VanRaden and Sullivan, 2010) and also was confirmed in empirical analyses (Lorenzana and Bernardo, 2009).

Bastiaansen *et al.* (2010) and Meuwissen *et al.* (2001) showed that as the number of phenotypic records increased from 500 to 1000, correlations between true and estimated breeding values raised from 0.58 to 0.66 and 0.71 to 0.79 in BLUP and BayesB methods, respectively. Calus and Veerkamp (2007) also concluded that an increase in the number of individuals in the training population would result in higher accuracy of GEBVs of selection candidates. Muir (2007) also showed that increasing the training population size would increase the accuracy. The reasons for the effect of sample size on accuracy are: First, the accuracy of estimates of marker effects increases with sample size. This occurs because bias and variance of estimates of marker effects decrease with sample size. Additionally, in some cases an increase in sample size may also increase the extent of genetic relationships between subjects in the training and validation populations (De Los Campos *et al.* 2013).

The amount of accuracy was different for various levels of QTL numbers. The lowest value achieved for N_{QTL}= 100 but the highest value was for N_{QTL}= 150 (Figure 2c). In a study that the accuracy of genomic prediction was evaluated for five different QTL numbers, increasing the number of QTL resulted in an inverse V shape trend for accuracy. As it increased from 3 to 30, the accuracy raised, but further increasing the number of QTLs from 30 to 300 and then to 3000, resulted in decreasing trend in accuracy (Piyasatian and Dekkers, 2013). In a simulation study of investigation of factors affecting the genomic evaluation accuracy using GBLUP and BayesB methods, Daetwyler *et al.* (2010) reported that for reference population size= 1000 when the number of QTL increased from 0.03 Me to 0.05 Me and then to 0.15 Me (Me=445, 1887 and 3543 is the number of independent chromosome segments), the accuracy of GBLUP had an V shape but the accuracy of BayesB had an inverse V shape oppositely.

In a study that the effect of N_{QTL} (including 50, 100, 300, 500, 1000 and 2000) and the distribution of QTL effects (gamma and equal variance distributions) were investigated, it was shown that under gamma distribution, as the number of QTL changed from 50 to 100 and then to 150, the accuracy first increased and then decreased. However, under equal variance distribution, the trend of accuracy was completely opposite, so that the accuracy declined first and then increased. Non-linear trend observed for the effect of QTL number on accuracy in this study suggested that its effect may be related to the interaction between N_{QTL} with other components of genetic architecture, i.e. interaction between QTLs alleles and interaction between QTLs and non-genetic factors.

The comparison of five Bayesian methods showed that the highest (0.676±0.034) and lowest (0.661±0.041) mean accuracy belonged to BayesB and BayesC methods, respectively.

Clustering Bayesian methods in term of accuracy, three groups were formed: I) high accuracy group, including BayesB II) medium accuracy group, including BayesA and Bayesian LASSO and III) low accuracy group, including BRR and BayesC methods (Figure 2d). An optimal GS method should yield the highest possible accuracy, prevent over fitting on the training dataset, and be based as much as possible on marker-QTL LD rather than on kinship. Moreover, such methods must be easy to preform, consistent across a wide range of traits and datasets, and easy to compute (Habier *et al*. 2007; Heslot *et al.* 2012).

The interaction between N_{QTL} and N_{IND} was completely strong. For each level of N_{QTL} by increasing N_{IND}, an ascending trend, although non-linear, were observed for accuracies (Figure 3b). By 500 N_{IND} the accuracy from 150 QTL was the lowest and from 50 QTL was the highest one. Efficiency of increasing the number of animals was higher with 50 QTL than with 150 QTL. By 800 N_{IND} a different situation was observed and the accuracy from 100 QTL was the lowest one.

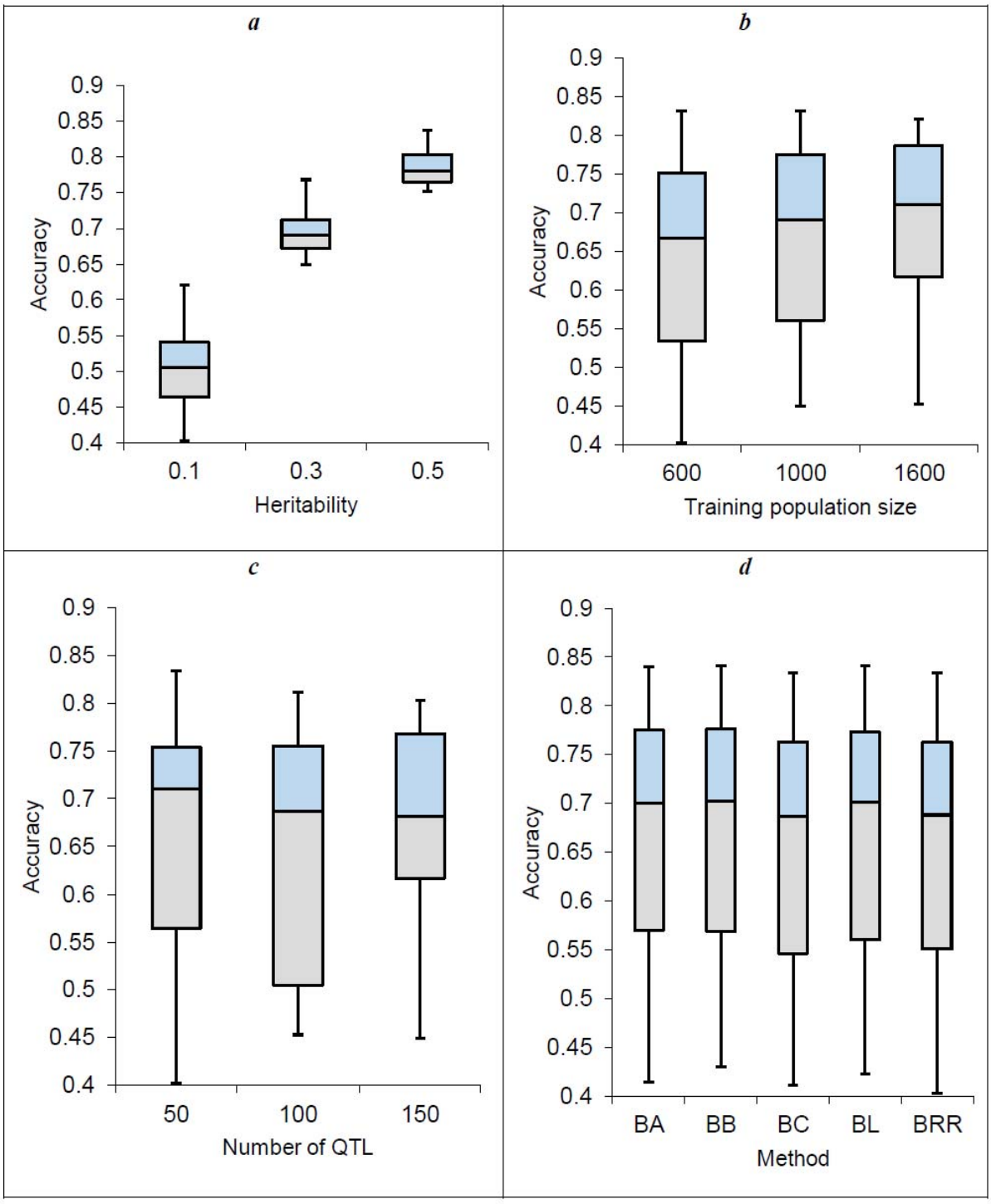

Figure 2 Accuracy of GEBV's for main factors: a) heritability; b) Training population size; c) number of QTLs and d) Bayesian methods

Investigation of the accuracy of genomic prediction of the standard marker effects using method BayesB showed that in the case of N_{QTL}= 5, doubling the N_{IND} in 1th generation after training led to decreasing in accuracy from 0.73 to 0.72 but when N_{QTL}= 50 accuracy increased from 0.63 to 0.65 (Piyasatian and Dekkers, 2013). In a simulation study of affecting factor on genomic prediction accuracy, Daetwyler *et al*. (2010) reported that for five level of N_{QTL} expressed as proportions of marker density, with increasing the N_{IND} from 500 to 2000, accuracywas increased but the highest increment occurred for the lowest level of N_{QTL} (0.03 M) while the effect of increasing of N_{IND} for the highest level of N_{QTL} (1 M) was negligible.

Using each investigated Bayesian method, increasing the size of the training population resulted an ascending trend in average genomic accuracy (Figure 3c).

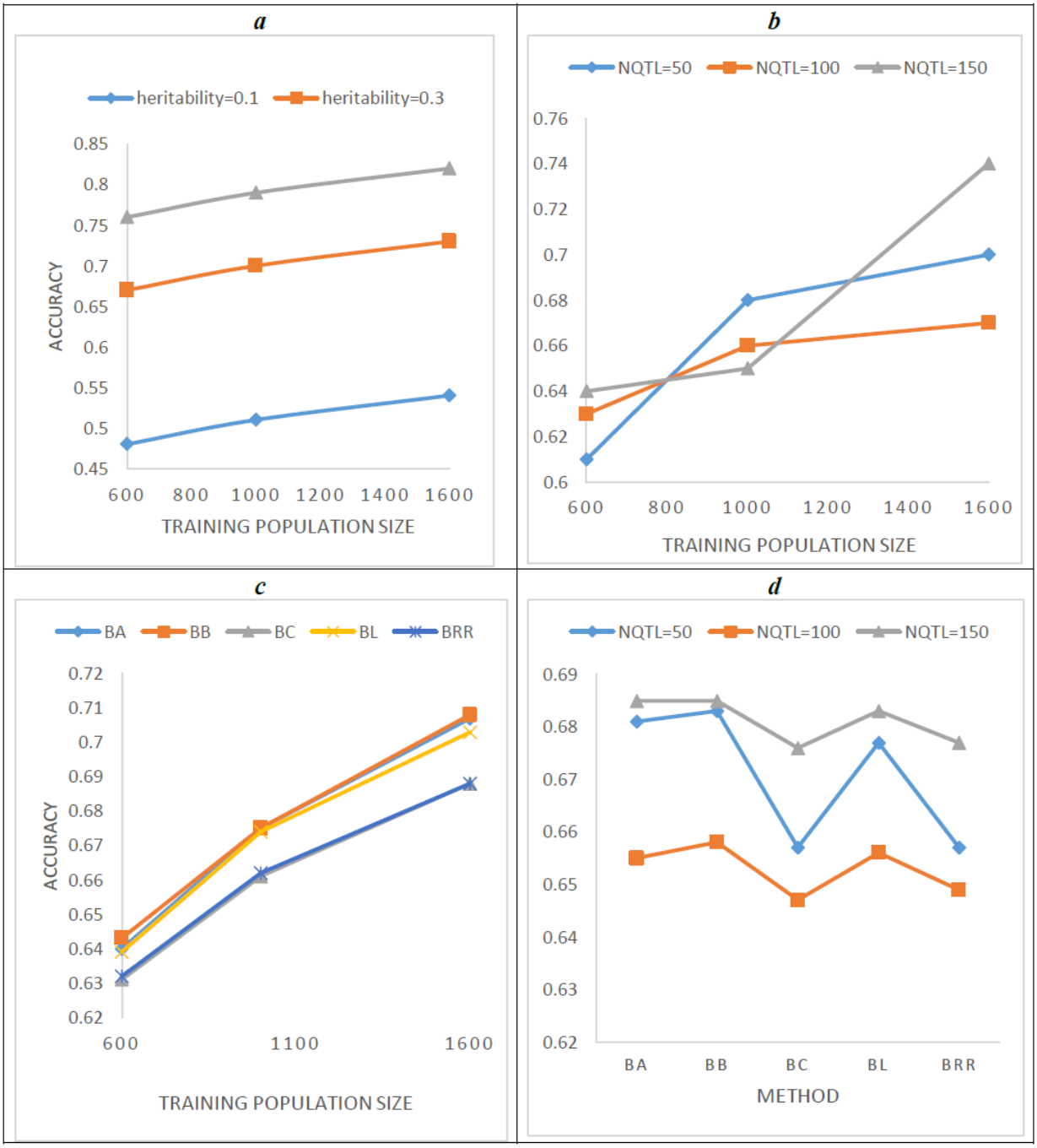

Figure 3 Accuracy of:
a) different number of individuals per generation and heritabilities
b) different number of individuals per generation and number of QTL
c) different number of individuals per generation and Bayesian methods
d) Bayesian methods and number of QTL

However, the trend was a little different among methods and this declared the existence of interaction between training population size and estimation method. In the study by Clark *et al.* (2011), it was shown that the highest accuracy was achieved by the BayesB method when genetic variation was controlled by a few QTL with relatively large effects (100 *vs*. 1000 and 10000 QTLs) but with GBLUP method differences were negligible. In this study for all methods increasing the number of QTLs, accuracies had V shape pattern so that the highest and lowest accuracies were observed for N_{QTL}= 150 and N_{QTL}= 100, respectively (Figure 3d). Coster *et al.* (2010) showed that by using Bayesian regressions and LASSO methods high accuracies achieved when the number of QTL decreased, while accuracy of par-

tial least square regression (PLS) was unaffected by the number of QTL. The same results were reported by Wientjes *et al.* (2015).

At constant heritability, RR-BLUP is insensitive to genetic architecture (i.e., the number of QTL and the distribution of their effects), while the accuracy of Bayesian methods improves as the number of QTL decreases and their effects increase (Luan *et al*. 2009; Daetwyler *et al.* 2010).

As mentioned before, three way interaction between heritability, number of QTL and marker effect estimation method was not significant and for all level of combinations of heritability and number of QTL, BayesB had the highest accuracy but when N_{QTL}= 50 the accuracy clusters were more obvious and with increasing N_{QTL} medium and high accuracy clusters merged together and displayed as a same cluster (Figure 4).

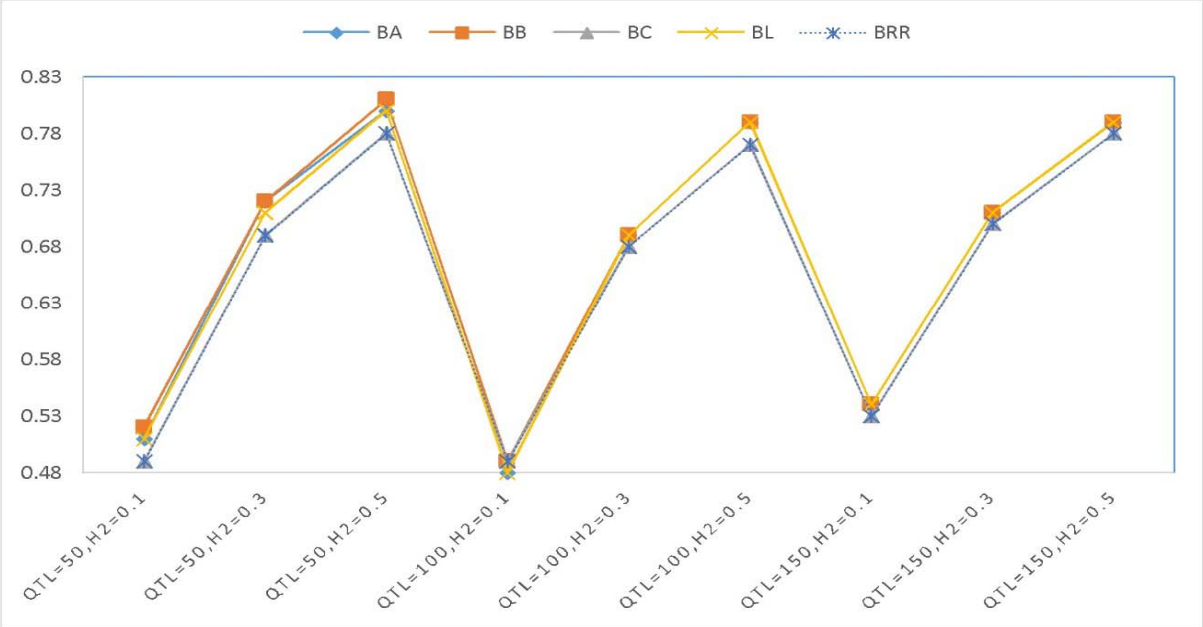

Figure 4 Accuracy of Bayesian methods for different combinations of heritability and number of QTL

CONCLUSION

Although having sufficient LD is essential for high accuracies but in the next step, other factors relating to population structure and genetic architecture of trait are important. The results of this study declared that among well-known Bayesian methods for genomic prediction, in most scenarios, well known methods introduced by Meuwissen (BayesB and BayesA) had the highest accuracies. Therefore among the Bayesian methods, we can propose these methods specially BayesB for marker effects estimation because of it's more realistic prior density assigned to marker effects. The economically important traits that involved in the breeding programs, vary in their heritability and number of QTLs. In traditional and genomic methods, the accuracy of traits with high heritability is higher than traits with low heritability due to low contribution of genes effects in phenotypic variation. Increasing the number of response variable (training population size) led to high accuracies because with more records, estimated marker effects were more accurate and using these effects in testing population give the accurate GEBVs.

REFERENCES

Bastiaansen J.W., Bink M.C., Coster A., Maliepaard C. and Calus M.P. (2010). Comparison of analyses of the QTLMAS XIII common dataset. I: genomic selection. *BMC Proc*. **4(1),** 1-11.

Calus M. and Veerkamp R. (2007). Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *J. Anim. Breed. Genet*. **124(6),** 362-368.

Clark S.A., Hickey J.M. and Van der Werf J.H. (2011). Different models of genetic variation and their effect on genomic evaluation. *Genet. Sel. Evol*. **43(18),** 12.

Coster A., Bastiaansen J.W., Calus M.P., van Arendonk J.A. and Bovenhuis H. (2010). Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance. *Genet. Sel. Evol.* **42,** 9.

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(3),** 1021-1031.

Daetwyler H.D., Villanueva B., Bijma P. and Woolliams J.A. (2007). Inbreeding in genome wide selection. *J. Anim. Breed. Genet*. **124(6),** 369-376.

Daetwyler H.D., Villanueva B. and Woolliams J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. **3(10),** e3395.

De Los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D. and Calus M.P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. **193(2),** 327-345.

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(1),** 375-385.

Dekkers J. (2007). Prediction of response to marker assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet*. **124(6),** 331-341.

Dekkers J. (2012). Application of genomics tools to animal breeding. *Curr. Genom*. **13(3),** 207-212.

Gianola D., Fernando R.L. and Stella A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*. **173(3),** 1761-1776.

Goddard M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*. **136(2),** 245-257.

Habier D., Fernando R. and Dekkers J. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. **177(4),** 2389-2397.

Hayes B., Bowman P., Chamberlain A. and Goddard M. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* **92(2),** 433-443.

Heslot N., Yang H.P., Sorrells M.E. and Jannink J.L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop. Sci.* **52(1),** 146-160.

Lorenzana R.E. and Bernardo R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* **120(1),** 151-161.

Luan T., Woolliams J.A., Lien S., Kent M., Svendsen M. and Meuwissen T.H. (2009). The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genet.* **183(3),** 1119-1126.

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(4),** 1819-1829.

Moser G., Tier B., Crump R.E., Khatkar M.S. and Raadsma H.W. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genet. Sel. Evol.* **41,** 56.

Muir W. (2007). Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* **124(6),** 342-355.

Park T. and Casella G. (2008). The bayesian lasso. *J. Am. Stat. Assoc*. **103(482),** 681-686.

Pérez P. and De Los Campos G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genet. Genet*. **114,** 164442.

Piyasatian N. and Dekkers J. (2013). Accuracy of genomic prediction when accounting for population structure and polygenic effects. *Anim. Ind. Rep*. **659(1),** 68.

Tibshirani R. (1996). Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Series B (Methodological)*. **58(1),** 267-288.

VanRaden P.M. and Sullivan P.G. (2010). International genomic evaluation methods for dairy cattle. *Genet. Sel. Evol*. **42,** 7.

Verbyla K.L., Bowman P.J., Hayes B.J. and Goddard M.E. (2010). Sensitivity of genomic selection to using different prior distributions. *BMC Proc*. **4(5),** 34-39.

Wientjes Y.C., Veerkamp R.F., Bijma P., Bovenhuis H., Schrooten C. and Calus M.P. (2015). Empirical and deterministic accuracies of across population genomic prediction. *Genet. Sel. Evol*. **47,** 5.