Journal of Genetics (2021)100:24
https://doi.org/10.1007/s12041-021-01275-5

RESEARCH ARTICLE

Assessing the performance of a novel method for genomic selection: rrBLUP-method6

ZAHRA AHMADI, FARHAD GHAFOURI-KESBI* and POUYA ZAMANI

Department of Animal Science, Faculty of Agriculture, Bu-Ali Sina University, 6517838695 Hamedan, Iran
*For correspondence. E-mail: f.ghafouri@basu.ac.ir.

Received 11 August 2020; revised 26 October 2020; accepted 2 November 2020

Abstract. The aim of this study was to compare the predictive performance of ridge regression best linear unbiased prediction-method 6 (rrBLUPm6) with well-known genomic selection methods (rrBLUP, GBLUP and BayesA) in terms of accuracy of prediction, computing time and memory requirement. The impact of the genetic architecture and heritability on the accuracy of genomic evaluation was also studied. To this end, a genome was simulated which consisted of five chromosomes, one Morgan each, on which 5000 biallelic single-nucleotide polymorphisms (SNP) were distributed. Prediction of genomic breeding values was done in different scenarios of number of QTL (50 and 500 QTL), distribution of QTL effects (uniform, normal and gamma) and different heritability levels (0.1, 0.3 and 0.5). Pearson's correlation between true and predicted genomic breeding values ($\mathbf{r_{p,t}}$) was used as the measure of prediction accuracy. Computing time and memory requirement were also measured for studied methods. The accuracy of rrBLUPm6 was higher than GBLUP and rrBLUP, and was comparable with BayesA. In addition, regarding computing time and memory requirement, rrBLUPm6 outperformed other methods and ranked first. A significant increase in accuracy of prediction was observed following increase in heritability. However, the number and distribution of QTL effects did not affect the accuracy of prediction significantly. As rrBLUPm6 showed a great performance regarding accuracy of prediction, computing time and memory requirement, we recommend it for genomic selection.

Keywords. genomic selection; genetic architecture; ridge regression; single-nucleotide polymorphism; quantitative trait loci.

Introduction

The idea of using molecular genetic information for breeding purposes traces back to Neimann-Sorensen and Robertson (1961) and Smith (1967). They recommended the use of genetic information provided by a few number of DNA markers associated with the trait of interest in genetic evaluation and selection procedures. Their works resulted in marker assisted selection (MAS) approach (Fernando and Grossman 1989). However, due to its limitations (Wakchaure *et al.* 2015), the MAS was not adequately adopted by animal breeders. To overcome the drawbacks of MAS, genomic selection (GS) (Meuwissen *et al.* 2001) was introduced, in which the genomic breeding values of the candidate animals are predicted using hundreds and thousands of single-nucleotide polymorphisms (SNP) that cover the entire genome. The GS increases the genetic gain through increasing accuracy of selection and selection intensity, and decreasing generation interval. Garcia-Ruiz *et al.* (2016) analysed the results of applying GS in United State dairy cattle and reported 50–100% genetic gain for yield traits and from three-fold to four-fold genetic gain for lowly heritable traits such as fertility and longevity. Also they reported positive impact of GS on generation interval that decreased from 7.5 years to 2.5 years. The success of genomic selection depends on the accuracy of genomic breeding values (GEBVs) which are the genetic values predicted by genomic selection models for each individual. Several factors including the number of quantitative trait loci (QTL), their locations in the genome, effective sizes of QTL, linkage disequilibrium (LD) between QTL and SNPs, mode of gene action, size of reference population, degree of relationship between reference and validation population and marker density affect the accuracy of GEBVs (Davoudi *et al.* 2018).

For predicting GEBVs, several methods including parametric (such as Bayesian methods), semiparametric (such as RKHS) and nonparametric methods (such as machine learning methods) have been proposed (Howard *et al.* 2014). These methods predict GEBV with different accuracy because different methods have different assumptions about the distribution of marker effects, the selection of covariates

Published online: 28 April 2021

and/or the genetic variances and covariances matrix. Different combinations of these assumptions modify the genetic variation explained by markers, which directly reflects on the accuracy (Andrade *et al.* 2019). One of the methods which is currently used for predicting GEBVs is ridge-regression best linear unbiased prediction (rrBLUP) (Endelman 2011), which is equivalent to best linear unbiased prediction (BLUP) in the context of mixed models (Whittaker *et al.* 2000; Meuwissen *et al.* 2001). When the number of predictive variables (SNP) is larger than the number of phenotypes (p), the mixed model equations (MME) for SNP effects (u) can become prohibitively many. In dairy cattle, e.g. it is now common to use in the order of $p = 50{,}000$ SNPs while setting the n to a number around 1000. Even larger values of p are being common (e.g. 777,000 SNPs, Kramer *et al.* 2014). Solving the MME for $\boldsymbol{u}$ with such large values of p will be either impossible or prohibitively costly with most current mixed model packages. When $p \geq n$, there is also the possibility of linear dependencies among several marker profiles that would render Z singular. In animals, identical twins or clones will give rise to singular **Z** (VanRaden 2008). Piepho *et al.* (2012) reviewed several options for computing RR-BLUP and introduced RR-BLUP method 6 (rrBLUPm6) for situations in which p is much larger than n and/or **Z** is expected to be singular.

However, yet the accuracy of rrBLUPm6 has not been compared with other methods of genomic selection. Therefore, the aim of this study was to compare predictive performance of rrBLUPm6 with some common methods of genomic selection in different scenarios of genetic architecture of the trait (number of QTL and distribution of QTL effects) and heritability levels.

Material and methods

Simulation of genome and population

For simulation of genome and population, the *hypred* package (Technow 2013) was used. Parameters used for the simulation of genome are listed in table 1. A genome was simulated with five chromosomes, one Morgan each. On each chromosome, 1000 biallelic SNP markers were distributed (5000 SNPs in total). Coding for each genotype with alleles A_1 and A_2 were, respectively, 2 for A_1A_1, 0 for A_2A_2 and 1 for A_1A_2 or A_2A_1. QTL effects were modeled with gamma, uniform and normal distribution. In gamma distribution, the shape (β) and scale parameters were set to 0.4 and 1.66, respectively (Meuwissen *et al.* 2001). In uniform distribution, the QTL effects ranged between 0 and 1 with the same probability. In normal distribution scenario, the QTL effects were sampled from a standard normal distribution, $g_i \sim N(0,1)$, where g is the QTL effect. The genome simulated for a reference population that comprised of 1000 individual, i.e. the genotypic matrix consisted of 1000 individuals each genotyped for 5000 SNPs.

Table 1. Parameters used for simulation programme.

Genome size	500 cM
Number of chromosomes	5
Number of marker	5000
Distribution of additive QTL effects	Normal, uniform, gamma
Number of QTL	50,500
Effective population size (Ne)	100
Heritability	0.1, 0.3, 0.5

The r^2 statistic of Hill and Robertson (1968) was used to measure the linkage disequilibrium (LD) as follow:

$$r^2 = \mathrm{D}^2/\mathrm{freq}(\mathrm{A}_1)*\mathrm{freq}(\mathrm{A}_2)*\mathrm{freq}(\mathrm{B}_1)*\mathrm{freq}(\mathrm{B}_2),$$

freq(A1) is the frequency of A1 allele in the population likewise for other alleles in the population. D is the deviation of parental genotypes from the recombinant genotypes estimated as:

$$\mathrm{D} = \mathrm{freq}(\mathrm{A}_1\_\mathrm{B}_1) * \mathrm{freq}(\mathrm{A}_2\_\mathrm{B}_2)\mathrm{freq}(\mathrm{A}_1\_\mathrm{B}_2) * \mathrm{freq}(\mathrm{A}_2\_\mathrm{B}_1).$$

The true breeding value of each individual was calculated using the following formula:

$$TBV_i = \sum_{i=1}^{n} x_{ij} b_j,$$

where TBV_i is the true breeding value of the individual i, n is the number of QTLs underlying the trait, x_{ij} is the QTL genotype at position j, and b_j is the additive effect of the j_{th} QTL. The following equation was used to simulate the phenotype:

$$\mathbf{y}_i = TBV_i + e_i,$$

where $\boldsymbol{y_i}$ is the phenotype of individual i and e_i is the residual effect.

By mating animals of the reference population, another generation was generated which was tagged as validation population. The animals in the validation population had known genotypes but their phenotypic records were unknown and their genomic breeding values had to be predicted with following statistical methods.

Methods of genomic selection

***Ridge regression BLUP (rrBLUP) and ridge regression BLUP method 6(rrBLUPm6)*:** In rrBLUP, the predicted GEBVs are obtained by the summing of all the marker effects of an individual. Marker effects were estimated using the following mixed model:

$$y = 1_n\mu + Zg + e \quad (1)$$

where y is the vector of observed phenotypes, 1_n is a column vector of n ones and μ is a common intercept, Z is the design matrix for the random marker effects; g is the vector of

random marker effects. In RR-BLUP, the residuals and marker effects follow normal distributions with constant variance, i.e., $e \sim N(0, I\sigma_e^2)$ and $g \sim N(0, I\sigma_g^2)$, where I is an identity matrix. Under the assumed model the variance of the observed data is:

$$V = var(y) = ZZ^T\sigma_u^2 + R.$$

In which Z^T denotes the transpose of Z and R is the residual covariance matrix and equals to $I_n\sigma_e^2$. The solution for the marker effects is given by the following equation:

$$\hat{g} = (Z'Z + \lambda I)^{-1}Z'y$$

where $\lambda = \sigma_e^2/\sigma_g^2$ is the ridge penalization parameter. The R package BGLR (de los Campos and Perez Rodriguez 2018) was used to run rrBLUPm6. In rrBLUPm6, the marker effects are estimated by the following equation (Piepho *et al.* 2012):

$$\hat{g} = \sigma_u^2 Z^T V^{-1}(y - 1_n\mu).$$

The method is restricted to the case in which $R = I_n\sigma_e^2$. In cases where R does not meet this assumption, a linear transformation is applied to ensure $R = I_n\sigma_e^2$. Therefore, we would replace y with $L_R y$ and Z with $L_R Z$, in which $R^{-1} = (L_R)^2$ such that L_R is square and symmetric. L_R is easily obtained from a spectral decomposition of R^{-1}. With these replacements, analysis can proceed assuming that $R = I_n\sigma_e^2$ with $\sigma_e^2 = 1$, for more details refer to Piepho *et al.* (2012). The R package rrBlupMethod6 (Schulz-Streeck *et al.* 2015) was used to run rrBLUPm6.

***Genomic best linear unbiased prediction (GBLUP)*:**

The GBLUP was fitted as follow:

$$y = 1\mu + Zg + e,$$

where y is the vector of phenotypic observations, Z is the design matrix associating phenotypic observations to GEBVs, g is the vector of genomic breeding values and assumed that $g \sim N(0, G\,\delta_g^2)$ where δ_g^2 is the additive genetic variance, and G is the genomic relationship matrix whose elements estimated based on allelic similarity between individuals (VanRaden 2008). The GBLUP was run using package BGLR in R (de los Campos and Perez Rodriguez 2018).

***Bayesian method A (BayesA)*:**

The main assumption of Bayes**A** was that of total number of loci underlying a quantitative trait, only a small number have large effects and others have small effects. Bayes**A** can be fitted in the framework of the following model:

$$y = X\beta + u + \sum_{k=1}^{k} z_k a_k + e$$

where y is the vector of phenotypic observations, X is an incidence matrix associating observations to fixed effects in β, u is the vector of polygenic effects, k denotes the number of SNPs, z_k is an N×1 vector of genotypes at SNP k, a_k is the additive effect of that SNP, and e is a vector of residual effects. The prior for u is constant, the prior for δ_u^2 is assumed to follow normal distribution, $N(0, A\,\delta_u^2)$ where A is the numerator-relationship matrix and δ_u^2 is additive genetic variance apart from that explained by SNPs. The prior for a_k depends on the variance, δ_{ak}^2, and the prior probability, π:

$$ak|\pi, \delta_{ak}^2 = \begin{cases} 0 & \text{with probability } \pi \\ \sim N(0, \delta_{ak}^2) & \text{with probability } (1-\pi). \end{cases}$$

δ_{ak}^2 denotes that each SNP has its own variance. Each of these variances has a scaled inverse chi-square prior with degrees of freedom va and scale S_a^2, and thus with probability (1−π) the marginal prior of $ak|va, S_a^2$ is a univariate student's t-distribution, $t(0, va, S_a^2)$ where S_a^2 is derived from the expected value of a scaled inverse chi-square distributed random variable (Habier *et al.* 2011). The Bayes**A** was run using package BGLR in R (de los Campos and Perez Rodriguez 2018).

Accuracy of GEBV

The Pearson's correlation coefficient between the predicted GEBVs and true GEBVs ($r_{p,t}$) was used to measure accuracy of GEBVs prediction. Each scenario was analysed 10 times and the mean and standard errors were presented.

Memory requirement and computing time

To record memory usage by each method, package *pryr* (Wickham 2018) was used, which records the amount of memory occupied by objects created by executing a function in R. Computing time in each scenario was monitored and recorded with an R function.

Results

The average LD of different simulations was 0.192. The prediction accuracy ($r_{p,t}$) of the studied methods in different scenarios of distribution of QTL effects (normal, uniform and gamma), heritability (0.1, 0.3 and 0.5) and number of QTL (50, 500) are shown in figures 1, 2, and 3. In the scenario of 500 QTL, the rrBLUPm6 was superior to other methods. Even it outperformed BayesA which is one of the most accurate genomic selection methods. Of course, in most cases the observed differences were not statistically significant (P>0.05). In the scenario of 50 QTL, the rrBLUPm6 predicted GEBVs with the same accuracy as other methods. The effect of heritability on accuracy of

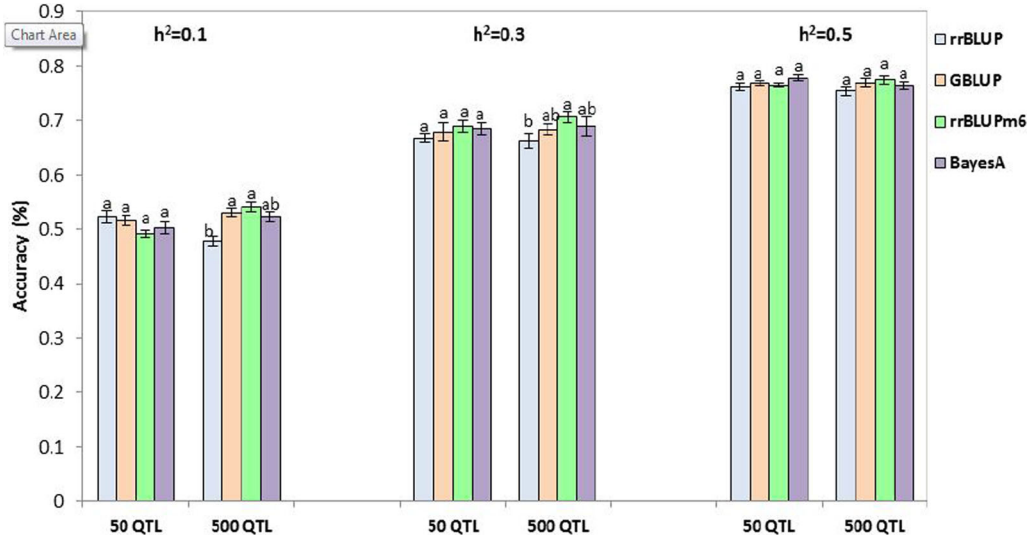

Figure 1. Predictive accuracy of studied methods in uniform distribution of QTL effects.

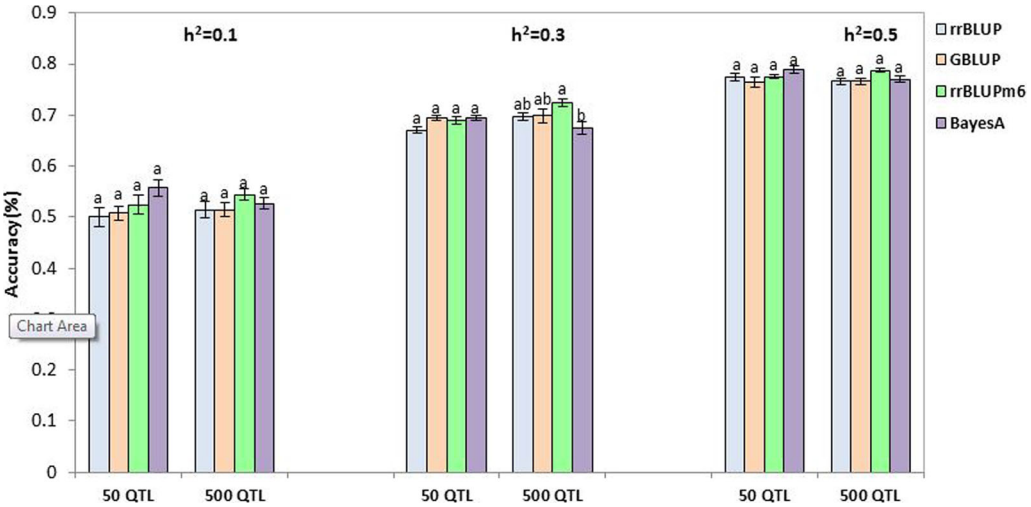

Figure 2. Predictive accuracy of studied methods in normal distribution of QTL effects.

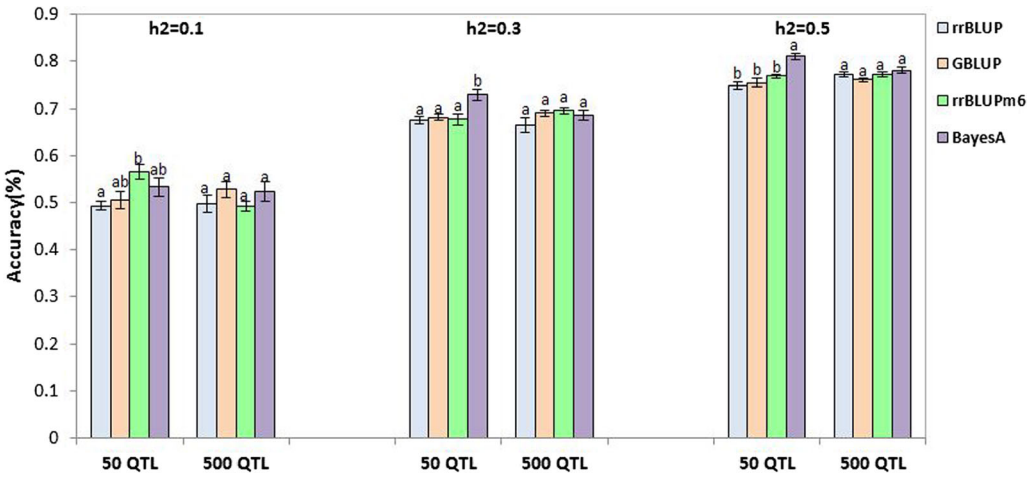

Figure 3. Predictive accuracy of studied methods in gamma distribution of QTL effects.

prediction is shown in figure 4. As shown, in all the methods studied here, heritability had a significant impact on the accuracy of prediction. By increasing heritability from 0.1 to 0.3, the $r_{p,t}$ increased by ≈25%. It was ≈0.10% when heritability increased from 0.3 to 0.5. Change in the QTL number from 50 to 500 did not affect the $r_{p,t}$ significantly (figure 5). The effect of uniform, normal and gamma distribution of QTL effects on the accuracy of genomic prediction is shown in figure 6. In most cases, accuracy of prediction was not affected by QTL effects distribution. Computational time for the method studied is shown in figure 7. In 33 s, the rrBLUPm6 was the fastest method followed by GBLUP with 38 s. rrBLUP and BayesA ranked as third and fourth with 168 s and 246 s, respectively. Memory requirement for each method is shown in figure 8. With 38 kb, rrBLUPm6 was the most efficient user of memory followed by GBLUP (38.25 kb), BayesA (584 kb) and rrBLUP (1800 kb).

Discussion

Our estimated LD was close to that reported by simulation studies including Ghafouri-Kesbi *et al.* (2016), Kasnavi *et al.* (2018) and Sahebalam *et al.* (2019). The estimated value of LD showed that each QTL is in LD with at least one marker (Kasnavi *et al.* 2018). In genomic selection, the minimum value of r^2 between markers and QTL should be 0.2 for tracking the average effect (Hayes 2007). There is no previous report regarding comparison of rrBLUPm6 predictive performance with other genomic selection methods. However, rrBLUP has been widely used for genomic selection both by empirical and simulated data. Abdollahi-Arpanahi *et al.* (2013) compared predictive performance of rrBLUP with BayeseA and GBLUP and reported a slightly lower predictive performance for rrBLUP. Similar results have been reported by Resende *et al.* (2012) and Momen *et al.* (2016). A limitation of rrBLUP, which affects its performance, is the assumption of equal contribution of all markers to the observed variation, which does not agree with the reality. This assumption may be suitable when considering an infinitesimal model (Fisher 1918), in which the characters are determined by an infinite number of unlinked and nonepistatic loci, with small effect. In the scenarios of simpler genetic architecture, with a few loci of large effect, rrBLUP and other BLUP methods will be underperformed by Bayesian methods (Resende *et al.* 2012). In addition as observed in figures 7 and 8, the computing time and memory requirement for rrBLUP was significantly higher than rrBLUPm6. It could be another limitation for rrBLUP especially where large genomic datasets are used for genomic selection. Our results showed that modifications by Piepho *et al.* (2012) which resulted in rrBLUPm6 have improved overall performance of rrBLUP significantly. Therefore, rrBLUP could be replaced with rrBLUPm6 in genomic selection studies. Superiority of rrBLUPm6 over other methods in terms of computing time and memory requirement could be another advantage for rrBLUPm6 especially in the 'big data era', where novel and cheap sequencing technologies have enabled massive genomic data generation for tens and thousands of individuals that could be exploited for genomic selection purposes. Overall, good prediction accuracy, significantly shorter running time and significantly lower memory requirement, makes rrBLUPm6 a potential alternative for current genomic selection methods.

Effect of heritability on the accuracy of prediction has been approved by several authors. For example, Hayes *et al.* (2010) showed that by increasing heritability from 0.1 to 0.9, the accuracy of genomic evaluation increased by 115% (form 0.35 to 0.75). In addition, Combs and Bernardo (2012) showed that by increasing heritability from 0.25 to 1.00, the

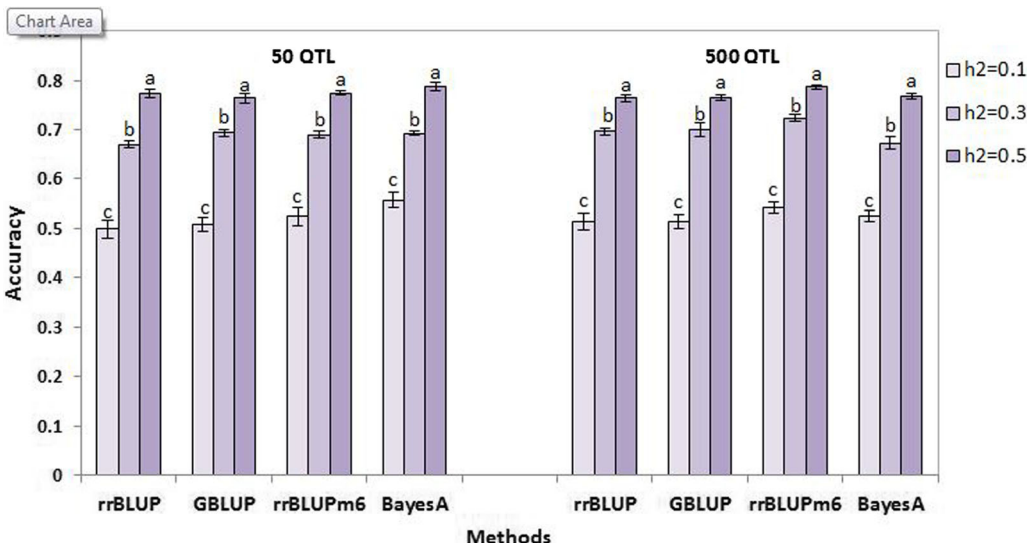

Figure 4. Effect of heritability on the accuracy of prediction in normal distribution of QTL effects.

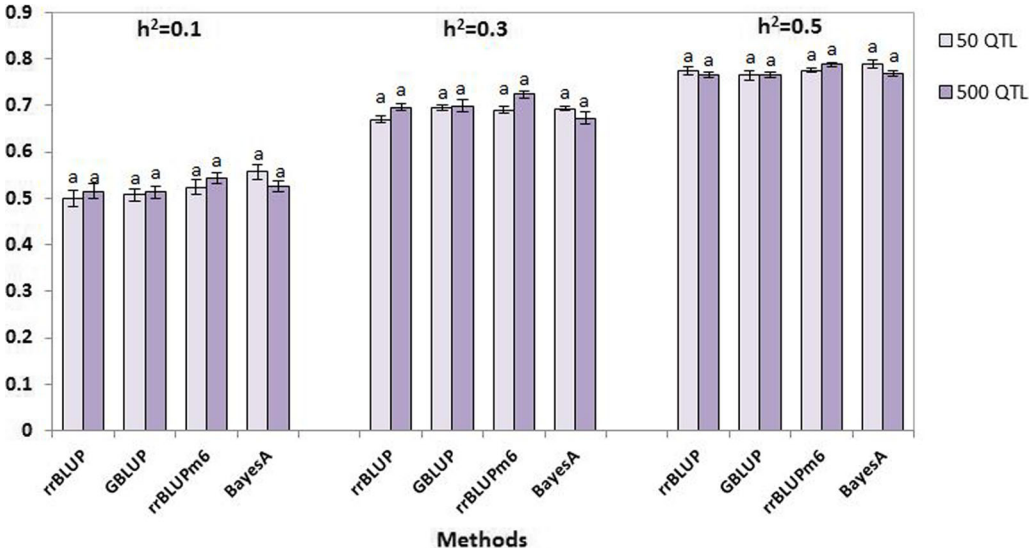

Figure 5. Effect of QTL number on the accuracy of prediction in normal distribution of QTL effects.

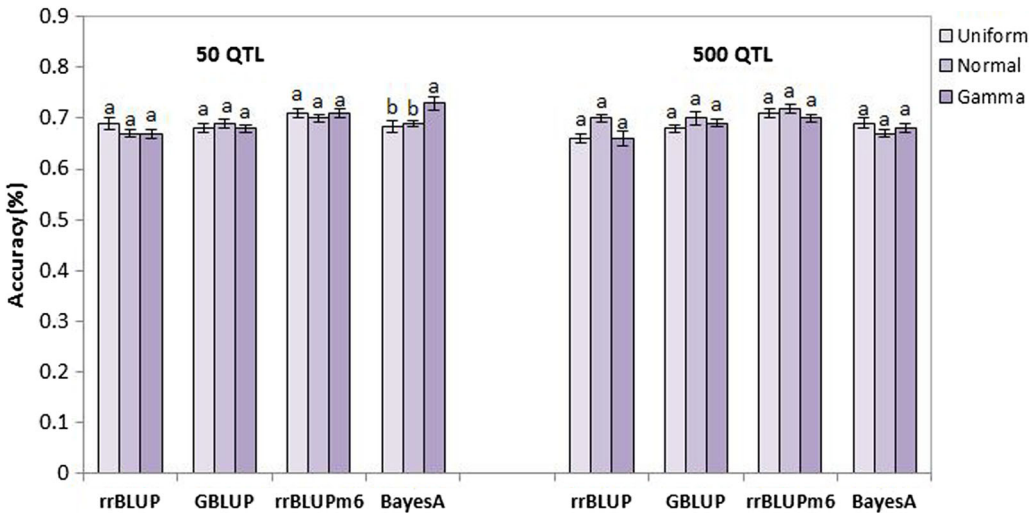

Figure 6. Effect of distribution of QTL effects on the accuracy of prediction.

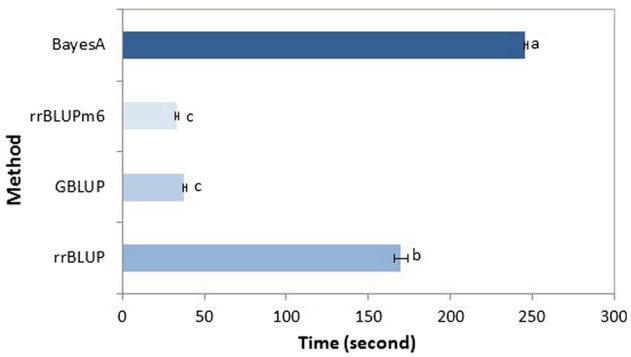

Figure 7. Computing time of the studied methods.

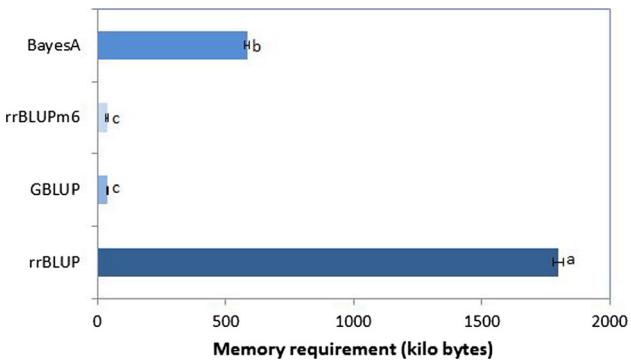

Figure 8. Memory requirement of the studied methods.

accuracy of prediction increased from 0.05 to roughly 1.00. Further, Kasnavi *et al.* (2018) made a genomic evaluation for a discrete trait with a radial-based support vector machine (SVM) and reported that at heritability of 0.3, 0.5 and 0.7, the accuracy of prediction was 0.57, 0.67 and 0.74. By increasing heritability, environmental noises decrease and

therefore most of the phenotypic variance will be additive genetic. Therefore, the additive genetic effects which is reflected by each marker increases. In such a situation, the power of model to extract such higher individual additive genetic effects increases leading to increased accuracy.

In our study, increasing the number of QTL did not affect the $r_{p,t}$. In some cases, the $r_{p,t}$ increased and in other cases it decreased. Our result was in agreement with Kasnavi *et al.* (2018) and Naderi and Mazraei (2018). However, it contradicted other reports including Coster *et al.* (2010) and Abdollahi-Arpanahi *et al.* (2013) who reported an increase in prediction accuracy following decrease in the number of QTL. They explained that when the number of QTLs is high, the additive genetic variance is divided between a greater number of QTL and consequently the efficiency of methods to estimate such small QTL effects decrease leading to decreased accuracy. There are a limited number of reports on the impact of distribution of QTL effect on the accuracy of prediction. Abdollahi-Arpanahi *et al.* (2013) studied the effect of normal, uniform and gamma distribution of QTL effects on the accuracy of prediction and reported that while normal distribution provided a slightly higher accuracy, the differences between three distributions were not significant. In addition, Kasnavi *et al.* (2018) used similar distributions for simulating QTL effects. In their study, the variation between accuracies provided by different distributions was in a small range between 0.01 and 0.03. In general, it seems that the distribution of QTL effects does not play an important role in accuracy of genomic prediction.

In conclusion, rrBLUPm6 showed a great performance regarding prediction accuracy, computing time and memory requirement. Therefore, we recommend it for genomic selection. While heritability had a significant effect on prediction accuracy, the number and distribution of QTL effects did not affect prediction accuracy significantly.

References

Abdollahi-Arpanahi R., Pakdel A., Nejati-Javaremi A. and Babak M. S. M. 2013 Comparison of different methods of genomic evaluation in traits with different genetic architecture. *J. Anim. Prod.* **15**, 65–77 (in Persian with English abstract).

Andrade L. R. B., Sousa M. B., Oliveira E. J., Resende M. D. V. and Azevedo C. F. 2019 Cassava yield traits predicted by genomic selection methods. *PLoS One* **14**, e0224920.

Combs E. and Bernardo R. 2012 Accuracy of genome wide selection for different traits with constant population size, heritability, and number of markers. *Plant. Gen.* **6**, 1–7.

Coster A., Bastiaansen J. W. M., Calus M. P. L., van Arendonk J. A. M. 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.

Davoudi P., Abdollahi-Arpanahi R. and Nejati-Javaremi A. 2018 The impact of QTL allele frequency distribution on the accuracy of genomic prediction. *Arch. Anim. Breed.* **61**, 207–213.

de los Campos G., Perez Rodriguez P. 2018 Bayesian generalized linear regression. https://cran.r-project.org/web/packages/BGLR/index.html.

Endelman J. B. 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant. Gen.* **4**, 250–255.

Fernando R. L. and Grossman M. 1989 Marker-assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* **2**, 246–477.

Fisher R. A. 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 34.

García-Ruiz A., Cole J. B., VanRaden P. M., Wiggans G. R., Ruiz-López F. J. and Van Tassell C. P. 2016 Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. USA* **113**, 3995–4004.

Ghafouri-Kesbi F., Rahimi-Mianji G., Honarvar M. and Nejati-Javaremi A. 2016 Predictive ability of random forests, boosting, support vector machines and genomic best linear unbiased predictionin different scenarios of genomic evaluation. *Anim. Prod. Sci.* **57**, 229–236.

Habier D., Fernando R. L., Kizilkaya K. and Garrick D. J. 2011 Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **12**, 186.

Hayes B. J., Daetwyler H. D., Bowman P., Moser G., Tier B., Crump R., Khatkar M., Raadsma H. W. and Goddard M. E. 2010 Accuracy of genomic selection: comparing theory and results. In *Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics*. Barossa Valley, Australia.

Hayes B. 2007 QTL mapping, MAS, and genomic selection. https://www.ans.iastate.edu/files/page/files/notes.pdf.

Hill W. and Robertson A. 1968 Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231.

Howard R., Carriquiry A. L. and Beavis W. D. 2014 Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *Genetics* **4**, 1027–1046.

Kasnavi S. A., Aminafshar M., Shariati M. M., Emam Jomeh Kashana N. and Honarvar M. 2018 The effect of kernel selection on genome wide prediction of discrete traits by support vector machine. *Gene. Rep.* **11**, 279–282.

Kramer M., Erbe M., Seefried F. R., Gredler B., Bapst B., Bieber A. and Simianer H. 2014 Accuracy of direct genomic values for functional traits in Brown Swiss cattle. *J. Dairy. Sci.* **97**, 1774–1781.

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**, 1819–1829.

Momen M., Ayatollahi Mehrgardi A., Sheikhy A., Esmailizadeh A. K. and Assadi Foozi M. 2016 Predictive ability of statistical genomic prediction methods when underlying genetic architecture of trait is purely additive. *Iran. J. Appl. Anim. Sci.* **6**, 815–822.

Naderi Y. and Mazarei M. 2018 Evaluation of genomic prediction accuracy in different genomic architectures of quantitative and threshold traits with using random forest method. The 6th Scientific Congress on the Development and Promotion of Agricultural Sciences and Natural Resources in Iran. Tehran, Iran.

Neimann-sorensen A. and Robertson A. 1961 The association between blood groups and several production characters in three Danish cattle breeds. *Acta. Agr. Scand.* **11**, 163–196.

Piepho H. P., Ogutu J. O. and Schulz-Streeck T. 2012 Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. *Crop Sci.* **52**, 1093–1104.

Resende M. F. R., Munoz P., Resende M. D. V., Garrick D. J. and Fernando R. L. 2012 Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **190**, 1503–1510.

Sahebalam H., Gholizadeh M., Hafezian H. and Farhadi A. 2019 Comparison of parametric, semiparametric and nonparametric methods in genomic evaluation. *J. Genet.* **98**, 102.

Schulz-Streeck T., Estaghvirou B. and Technow F. 2015 Re-parametrization of RR-BLUP to allow for a fixed residual variance. https://mran.microsoft.com/snapshot/2016-10-12/web/packages/rrBlupMethod6/index.html.

Smith C. 1967 Improvement of metric traits through specific genetic loci. *Anim. Prod.* **9**, 349–358.

Technow F. 2013 hypred: simulation of genomic data in applied genetics. http://cran.r-project.org/web/packages/hypred/index.html.

VanRaden P. M. 2008 Efficient methods to compute genomic predictions. *J. Dairy. Sci.* **91**, 4414–4423.

Wakchaure R., Ganguly S., Praveen P. K., Kumar A., Sharma S. and Mahajan T. 2015 Marker assisted selection (mas) in animal breeding: a review. *Drug. Metabol. Toxic.* **6**, 1000.

Wickham H. 2018 pryr: Useful tools to pry back the covers of R and understand the language at a deeper level. https://cran.r-project.org/web/packages/pryr/index.html.

Whittaker J. C., Thompson R. and Denham M. C. 2000 Marker-assisted selection using ridge regression. *Genet. Res. Camb.* **75**, 249–252.

Corresponding editor: H. A. Ranganath