

Comparing Different Marker Densities and Various Reference Populations Using Pedigree-Marker Best Linear Unbiased Prediction (BLUP) Model

Research Article

S. Barjasteh¹, G.R. Dashab^{1*}, M. Rokouei¹, M.M. Shariati² and M. Vafaye Valleh¹

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

² Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

Received on: 6 Jun 2019

Revised on: 10 Aug 2019

Accepted on: 31 Aug 2019

Online Published on: Jun 2020

*Correspondence E-mail: dashab@uoz.ac.ir

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

Online version is available on: www.ijas.ir

ABSTRACT

In order to have successful application of genomic selection, reference population and marker density should be chosen properly. This study purpose was to investigate the accuracy of genomic estimated breeding values in terms of low (5K), intermediate (50K) and high (777K) densities in the simulated populations, when different scenarios were applied about the reference populations selecting. After simulating the historical (undergoing drift and mutation) and recent (undergoing selection) population structures, 800 individuals were remained in reference population. Three scenarios were considered for reducing the reference population number including: 1) 400 individuals which had the highest relationships with the validation set, 2) 400 individuals which had the highest inbreeding, and 3) 400 selected individuals by random. The genomic breeding values were predicted for traits with two heritability levels (0.25 and 0.5) using best linear unbiased prediction (BLUP) with different markers and pedigree information combinations of included pedigree-based BLUP (ABLUP), which was used a numerator relationships matrix (A) only, genomic best linear unbiased prediction (GBLUP) which was used a genomic relationship matrix (G) only, and BLUP|GA, which combined both A and G by using a weight parameter (λ). By increasing λ , the prediction model was changed from GBLUP ($\lambda=0$) to ABLUP ($\lambda=1$). The results indicated that without considering the panel density effects, G matrix ($\lambda=0$) and A matrix ($\lambda=1$) usages had the highest and lowest prediction accuracy, respectively. Comparative analyses of different scenarios of reference population selection revealed that all individuals' inclusion in reference population yielded the highest estimation accuracy for breeding values ($P<0.05$). On the contrary, using reduced single nucleotide polymorphism (SNP) panels considerably decreased the accuracy of breeding value prediction. Individuals selecting in the reference set with a high genetic relationship to target animals, considerably improved the reduction in genomic prediction accuracy because of small reference population size.

KEY WORDS accuracy, genomic selection, marker density, simulation.

INTRODUCTION

Using genomic selection is rapidly growing in the breeding programs of many livestock species, especially dairy cattle, with a large population size (Boison *et al.* 2017). The potential factors, which have influences on the genetic response, would also affect the genomic selection efficiency.

Selecting animals with superior quality to serve as the next generation parents can be conducted in their early life or even in the embryogenesis throughout the genomic information, which could largely reduce the generation interval in comparison with the traditional methods. In addition, many young animals can be evaluated theoretically and consequently could introduce a larger number of potential

candidates that causes to increase the selection and genetic gain intensity. In various species of livestock there is empirical evidence of increased rates of genetic gain from the use of genomic selection to target different aspects of the breeder's equation. Accurate predictions of genomic breeding value are central to this and the design of training sets is in turn central to achieving sufficient levels of accuracy.

Questions about how the genotyped population should be structured and which animals should be used in the training population are still a matter of debate in all species (Lourenco *et al.* 2015). In this study, for the first time, the effects of different population scenarios (using all individuals, randomly selected and references with high inbreeding or high relationship with validation population) in the various marker densities were studied using pedigree-based BLUP (ABLUP), GBLUP and BLUP|GA models.

Following completion of the bovine genome sequencing project, thousands SNP markers were identified and the relevant research advances were accelerated. By the DNA chips availability commercially and their cost-effective feature, it is possible to estimate the breeding values with a high accuracy (Boichard *et al.* 2016).

Detection of QTLs with small effect is feasible using the marker dense panel, and makes a real improvement in the true additive relationships calculation between relatives, which resulted in more accurate genetic merits estimations. It is also possible to categorize SNP effects into direct, indirect, and total SNP effects (Momen *et al.* 2018). Although, the genomic selection accuracy depends on linkage disequilibrium (LD) between SNPs, but the moderate to high-density arrays could often provide enough LD between marker and QTL, which influences the interested traits (Calus *et al.* 2008).

The genomic selection accuracy depends on several factors like the reference population size and structure, SNP markers map density, heritability of a trait, dependent variable quality, genomic information, genetic relationship between reference and validation populations, LD between marker and QTL, effective population size and also epigenetic effects (Habier *et al.* 2007; Calus *et al.* 2008; Solberg *et al.* 2008; Meuwissen *et al.* 2009; Amiri Roudbar *et al.* 2017; Amiri Roudbar *et al.* 2018).

Habier *et al.* (2007) established that the genomic selection uses the genomic relations between individuals and the LD between marker and QTL, in order to improve the genomic accuracy. Hayes *et al.* (2009a) reported that some factors like the heritability of a trait, genotyped animals' number with phenotypic records in reference population, validation population size, assumed statistical distribution type for the QTL effects, and the effective population size could influence the accuracy of genomic estimated breeding value (GEBV).

Daetwyler *et al.* (2008) indicated that the number of phenotypic records and the heritability of a trait effectively could influence the GEBV accuracy. They reported that for some traits, which had a low heritability, a large number of phenotypes were required in order to estimate the markers' effects.

Generating a proper reference population plays a crucial role in the genomic selection programs and it could largely influence the GEBV accuracy in the young animals and also those animals without phenotypic record in the second selection step. There are several factors in the reference population that have influence on the prediction accuracy like the number of animals in this population and their type according to their gender (Van Raden *et al.* 2009), the animals' phenotypic information reliability, the genetic relationships of the reference population and the associations between reference population with validation population (Lund *et al.* 2010). Clark *et al.* (2012) demonstrated that one of the effective factors on genomic selection accuracy is identified as the relationship between the reference and validation populations, and the increasing the relationship between two populations from 0.125 to 0.25 using the GBLUP model, could result in increase of the genomic selection accuracy from 0.41 to 0.57. The genetic relationship between the reference population and the validation population shows significant effects on the accuracy for genomic prediction. Therefore, it is very important to optimize the design of the reference population when applying genomic selection to animal breeding (Wang *et al.* 2017).

Teimourian *et al.* (2015) performed the various aspects of genomic selection and estimations accuracy in Holstein population of Iran. The production of an appropriate reference population including male and females combination along with using other populations' information would result in the estimations accuracy increasing. Estimating the marker effects in these kinds of conditions requires using a high marker density and also applying appropriate analysis models.

Vitezica *et al.* (2011) demonstrated that in the various heritability scenarios 0.05, 0.3 and 0.5 the estimated breeding values (BLUP) accuracy is lower in the traditional method in comparison with the genomic methods. According to the studies accomplished by Su *et al.* (2012), the GEBV accuracy is higher in comparison with the traditional pedigree-based method.

This research purpose was to comparing the GEBV accuracy in the different marker densities including low (5k), intermediate (50k) and high (777k) in simulated populations under various strategies, in order to select the different reference population by the use of the BLUP models with different genomic and pedigree information combination under low and high heritability.

MATERIALS AND METHODS

Population and genome simulations

The QMSim software was applied in order to generate the population structures (Sargolzaei and Schenkel, 2009). In the first simulation step, a historical population was generated included 500 animals (250 males and 250 females) as founders, in order to produce the initial LD between markers and QTL and to establish the mutation–drift equilibrium. In this population, it was assumed that only two evolutionary forces of mutation and drift have influences on the genes frequency variations; therefore, the selection and mating processes were performed randomly. This population structure continued up to 1000 generations in order to establish the essential LD between markers and QTLs. After that, 1000 generations were simulated with a population size gradual increase to 4000 animals. 50 males were considered in the last generation of the historical population. As a next step, 20 males and 200 females were randomly selected from the last generation, and recent population was generated during 10 randomly mated generations. The mating system process was based on the gametes random combination.

The genotype, phenotype and pedigree information that were associated to the generations were recorded as 8, 9 and 10, respectively. The generations 8 and 9 were selected as the reference set and the generation 10 was selected as a validation set.

A genome consisting of 29 autosomes pairs were simulated with different length and similar to the bovine chromosomes size. After that, the number of 777026 bi-allelic SNP markers and 725 bi-allelic QTL with the initial allele frequency equal to 0.5 and evenly spaced were simulated on the genome. QTL allele effects were sampled from a Gamma distribution with 0.4 shape parameter. Table 1 displays those parameters used for simulating.

In the quality control step, SNPs with less than 0.01 minor allele frequency (MAF) and monomorphic loci were deleted. In this step, 369091 SNPs were deleted, and 407935 SNPs with the known and specified loci on the autosomal chromosomes were left for further analysis. After that, 5k and 50k densities panels by the use of the C programming language were retrieved from the genotype file.

Reference population selection

In the first scenario, all of 800 individuals of the reference population were used. Different scenarios were considered in order to reduce the reference population size to 400 individuals in the various marker densities as followings: 1) animals that had the highest relationship with the validation set, 2) those animals with the highest inbreeding, and 3) the randomly selected animals.

Prediction methods

Pedigree-based BLUP (ABLUP) method: In ABLUP model, the numerator relationships matrix (**A**) was calculated due to the pedigree information using the individuals' relationship average. It is very probable that the related animals were selected and along with that, inbreeding increases.

Additionally, these estimates accuracy is partly in accordance with the pedigree's accuracy and quality (Calus, 2010). The estimated breeding values (EBVs) were derived from a linear model as followings:

$$y = 1\mu + Za + e \quad (1)$$

Where:

y: vector of interest phenotype.

1: vector of 1.

μ : average population.

a and e: breeding values and residual effects vectors, respectively.

Z: design matrix for the random effects.

Henderson's mixed-model equations (Henderson, 1984) for estimating the breeding values are as equation (2):

$$\begin{bmatrix} 1'1 & 1'Z \\ Z'1 & Z'Z + A^{-1}\alpha \end{bmatrix} \times \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'y \\ Z'y \end{bmatrix}$$

Where:

α : error variance to additive variance ratio.

Genotype-based BLUP (GBLUP) method: In GBLUP, the relationship matrix was calculated in terms of the marker genotypes after applying a series of appropriate algebraic operations like changing the scale and weighting the genotypes in the BLUP standard framework. The genomic relationship matrix (**G**) tends to measure an actual section of the common alleles between individuals, not to measure an expected section like pedigree-based relationship matrix.

In **G**, the individuals with the same genotype for a large number of markers have more genetically similarities, and also have a large value in their correspondent location of the genomic matrix. Here, the **G** was created and calculated for the different scenarios using the Van Raden's model, as followings (Van Raden, 2008):

$$G = \frac{(M-P)(M-P)'}{2 \sum_{i=1}^m p_i(1-p_i)} = \frac{QQ'}{2 \sum_{i=1}^m p_i(1-p_i)} \quad (3)$$

Where:

M: genotypes matrix (codes -1 and 1 for homozygotes and code 0 for heterozygotes).

P: minor allelic frequency (MAF) matrix and π_i shows the MAF for i^{th} marker.

Q: stands for a matrix, which was obtained from P and M subtraction.

The GEBVs were estimated from a linear model, as followings:

$$y = 1\mu + Zg + e \quad (4)$$

Where:

g: genomic effects vector, assumed $g \sim N(0, G\sigma_g^2)$. Here, σ_g^2 is the additive genomic variance.

Integrating genomic and pedigree information in BLUP (BLUP|GA): BLUP|GA was used for the markers and pedigree information integration in order to evaluate the GEBV accuracy by the use of the different marker densities and various reference population subsets under two heritabilities as 0.25 and 0.5. The genomic and pedigree information were used in the Kernel matrix (K) form. This matrix combines the pedigree information (A) and the marker information (G), as followings:

$$K = \lambda A + (1-\lambda)G \quad (5)$$

Where:

λ : limited parameter, and can be ranged between 0 and 1. In this study, we chose λ equal to 0 (GBLUP), 0.1, 0.3, 0.5, and 1 (ABLUP).

Prediction accuracy access

The predicted accuracy was calculated by evaluating the Pearson's correlation between the GEBV and the observed phenotype (y) by the use of the following equation (Hayes *et al.* 2009b):

$$\rho_{y, \text{GEBV}} = \sigma(y, \text{GEBV}) / (\sigma_y \sigma_{\text{GEBV}}) \quad (6)$$

Where:

$\sigma(y, \text{GEBV})$: covariance between y and GEBV.

σ_y and σ_{GEBV} : standard deviation of y and GEBV, respectively.

The Duncan's multiple range test (with $\alpha=0.05$) was performed for comparing the different scenarios effects on the GEBV accuracy including the reference subset selection, marker densities, heritability, and statistical methods.

RESULTS AND DISCUSSION

Using all animals in the reference population

The results of the GEBV and EBV accuracies are displayed in Table 2 in different scenarios using all of the animals in the reference population. The results indicated that by increasing the SNPs number, also the accuracy would increase. The accuracy of breeding values in 777K density were higher than 50K density and in 50K density were higher than 5K density. In the other words, the marker interval reduction between them would lead to the accuracy increasing. This can be more possible because of an increase of LD level between the markers and QTL or their fewer intervals that was improved for the QTL effects capturing. As it was expected, the estimated accuracy in the low heritability (0.25) was significantly less in comparison with the accuracy in the high heritability (0.5) ($P < 0.05$).

Consequently, the accuracies of GEBV and EBV will increase in the traits with a higher heritability, and as a result, more genetic gain will be obtained. The GBLUP model accuracy (where only the G matrix was used and λ was 0) was higher than the other models with the more than zero λ . Due to the fact that even small errors in the pedigree can significantly affect the breeding value accuracy and consequently the extent of genetic gain (Hayes *et al.* 2008), we demonstrated that using genomic relationship matrix, which was derived from the marker information, could be considered as more efficient in comparison with the methods that just used pedigree information. This could be mainly as a result of the inability of pedigree-based relationship matrix to register the Mendelian sampling effects.

The reference population with the highest relatedness

The results of the accuracy that was achieved by the use of the subset of the highly related reference populations in different density of SNP panels and also with various λ and heritability are presented in Table 3. Similar to using all individuals in the reference population, also 777K density panel indicated the highest accuracy in comparison with other panels with lower density, the patterns of high marker density (777K), due to providing higher LD between SNP markers and QTL, yield more reliability in genomic predictions. This increase in reliability may be due to the specific relations that do not provide enough accuracy when using lower density patterns (5K and 50K). In this scenario there were no significant differences in the accuracies of breeding values prediction in 5K and 50K densities. This demonstrates the more efficiency of related reference population scenario which can obtain accuracy similar to the higher densities (50K) through reducing costs of genotyping by using the lower densities (5K).

Table 1 Characteristics and parameters of the simulated population

Population structure	Information of population simulation
Step 1	
Creating base population	
Number of base population generations	500[0]500[1000]4000[2000]
Number of males in the last base population generation	50
Step 2	
Number of males from base population	20
Number of females from base population	200
Number of generations	10
Number of iterations	10
Heritability	0.25, 0.5
Phenotype variance	1
Genome structure	Genome simulation information
Number of chromosomes	29
Number of markers (for each chromosome)	26794
Distribution of markers	Evenly
Number of QTL (for each chromosome)	25
QTL distribution	Random

Table 2 The genomic estimated breeding value (GEBV) and estimated breeding value (EBV) accuracies using all individuals in the reference population, with the different single nucleotide polymorphism (SNP) panel densities, the various λ and the heritability

Heritability	Density	λ				
		0	0.1	0.3	0.5	1
0.25	777K	0.2567 ^a (0.01)	0.2519 ^a (0.002)	0.2518 ^a (0.005)	0.2420 ^a (0.011)	0.2043 ^b (0.037)
	50K	0.2433 ^a (0.007)	0.2433 ^a (0.009)	0.2430 ^a (0.007)	0.2381 ^a (0.01)	0.1979 ^b (0.041)
	5K	0.2384 ^a (0.001)	0.2376 ^a (0.001)	0.2373 ^a (0.002)	0.2333 ^a (0.006)	0.1916 ^b (0.194)
0.5	777K	0.4358 ^{ab} (0.066)	0.4339 ^{bc} (0.061)	0.4269 ^{bcd} (0.102)	0.4166 ^{bc} (0.067)	0.3527 ^c (0.07)
	50K	0.4216 ^{bc} (0.049)	0.4211 ^a (0.08)	0.4133 ^{bc} (0.062)	0.4033 ^{bcd} (0.07)	0.340 ^{cde} (0.076)
	5K	0.3999 ^{bcd} (0.055)	0.3995 ^{cde} (0.10)	0.3945 ^{bc} (0.055)	0.3843 ^{bcd} (0.046)	0.3290 ^{de} (0.073)

The means within the same column with at least one common letter, do not have significant difference ($P > 0.05$).

Table 3 The genomic estimated breeding value (GEBV) and estimated breeding value (EBV) accuracies using the 400 highly related individuals in the reference population, with the different single nucleotide polymorphism (SNP) panel densities, the various λ and the heritability

Heritability	Density	λ				
		0	0.1	0.3	0.5	1
0.25	777K	0.2274 ^a (0.019)	0.2268 ^{ab} (0.016)	0.2263 ^{abc} (0.028)	0.2220 ^{ab} (0.018)	0.1978 ^d (0.031)
	50K	0.2189 ^{abc} (0.014)	0.2180 ^{ab} (0.016)	0.2167 ^{abc} (0.016)	0.2144 ^{abc} (0.019)	0.1877 ^{bcd} (0.033)
	5K	0.2210 ^{abcd} (0.024)	0.2190 ^{abcd} (0.044)	0.2195 ^a (0.025)	0.2183 ^{abc} (0.023)	0.1928 ^{cd} (0.042)
0.5	777K	0.4056 ^a (0.074)	0.3999 ^a (0.074)	0.3939 ^a (0.115)	0.3919 ^a (0.074)	0.3554 ^a (0.067)
	50K	0.3771 ^a (0.055)	0.3756 ^a (0.072)	0.3719 ^a (0.069)	0.3675 ^a (0.077)	0.3335 ^a (0.072)
	5K	0.3801 ^a (0.059)	0.3790 ^a (0.107)	0.3781 ^a (0.057)	0.3727 ^a (0.053)	0.3349 ^a (0.072)

The means within the same column with at least one common letter, do not have significant difference ($P > 0.05$).

Using **G** or **A** matrix in traits with lower heritability can result in a significant difference between methods. For instance, ABLUP showed significant lower accuracy compared to those methods that were used genomic information (GBLUP and BLUP|GA). In high heritability, it seems that using **A** matrix is sufficient for reaching to proper accuracy.

Selecting of a subset of the reference population with the highest inbreeding

Table 4 shows the results of estimating the accuracy of breeding values in different densities of 5K, 50K and 777K with the various weighting coefficients of ($\lambda=0, 0.1, 0.3, 0.5, 1$) in the selection scenario of inbreeding reference population.

Similar to other methods of reference set selecting, Table 4 indicates that **G** matrix can make a better accuracy, especially in a low heritability. Selection of the reference population in terms of the inbreeding, indicated a significant accuracy reduction in comparison with the subset selection of the reference population, regarding the relatedness ($P<0.05$). This demonstrated that the selection of a reference population by the use of inbreeding was not an appropriate approach for decreasing the genotyped animals' number.

Randomly selected reference population

The accuracy results of the randomly selected reference population under different scenarios are presented in Table 5. The accuracy in this reference population subset was significantly lower than those references that were selected using relatedness ($P<0.05$). However, there was observed no significant difference between inbred and randomly selected reference population.

This research results indicated that the prediction accuracy of the studied models (GBLUP, ABLUP and BLUP|GA) had a similar pattern in the all scenarios, as the accuracies in the GBLUP model, it means that when the matrix **G** was just used ($\lambda=0$), were higher than the λ more than zero. In the BLUP|GA model, by the λ increasing, that is, the increase of using pedigree information in order to create the relationship matrix, the breeding value accuracies decreased, especially when λ in the ABLUP model was equal to one.

The main advantage of GBLUP method over the pedigree-based evaluation (ABLUP) is that the pedigree information (pedigree depth and quality) is one of the key factors for the breeding values estimation in the classic evaluation. While in the GBLUP method, the genomic information can be applied in order to determine the exact genetic relationships between animals (Silva *et al.* 2014).

The range of errors in the pedigree registering has earlier been reported between 5 to 22%, but in recent years, be-

cause of using modern registering systems of relatives' information, it has been reduced to 10%, approximately.

This error can decrease the genetic gain between 2 to 12%, and there might be an increase in hidden inbreeding caused by some errors in the pedigree (Silva *et al.* 2014). GBLUP uses the genomic relationship matrix instead of the classic relationship matrix that was obtained from the molecular information. In the genomic relationship matrix, the individuals with the same genotype are genetically more similar for a large number of markers, and have a larger value in their correspondent locus in the genomic matrix. In general, in the genomic selection methods (GBLUP), the accuracy of genomic is higher by considering the ability of Mendelian sampling estimation in comparison with the ABLUP method (Van Raden, 2008).

In a simulation study, Villumsen *et al.* (2009) performed that using genomic relationship matrix is more efficient than using the predicted relationship matrix for calculating the breeding value accuracy, because the pedigree-based relationship matrix has no ability for registering the Mendelian sampling effects, while the relationship marker matrix is able to calculate this effect. These findings are consistent with this study results.

By including 20 individuals from the close relatives of training population in the reference population, Clark *et al.* (2012) attained the prediction accuracy as 0.57 by the use of the genomic relationship matrix (**G**) for the trait with the 0.3 heritability. Pedigree-based prediction accuracy for the close relatives was less (accuracy of 0.42 with 10 pedigree generations) in comparison with the genomic relations. There was a lower genomic prediction accuracy for those animals with intermediate association (0.41), but BLUP method showed much lower accuracy (0.21 with 10 pedigree generations) by the use of the pedigree. For the unrelated group, the pedigree method presented a very low and close to zero accuracy (accuracy of 0.04 with 10 pedigree generations), but GBLUP still showed an acceptable accuracy (0.34).

Su *et al.* (2012) demonstrated that the levels of breeding value accuracy were 35.8, 45.4 and 36.6% for the traits of milk production, fat and protein percent in GBLUP method, respectively, and also were 19.4, 25.1 and 19.9% in the pedigree-based traditional method, respectively. Therefore, GBLUP accuracy level was about 0.1 higher than the traditional method, and this was in agreement with this research results.

The range of genomic prediction accuracy in the dairy cattle in the developed countries has been reported from 0.5 to 0.85 for the traits with intermediate to high heritability such as milk production and for the traits with low heritability such as reproductive and survival traits from 0.2 to 0.5 (Wiggans *et al.* 2017).

Table 4 The genomic estimated breeding value (GEBV) and estimated breeding value (EBV) accuracies using the 400 individuals with the highest inbreeding in the reference population, with the different single nucleotide polymorphism (SNP) panel densities, the various λ and the heritability

Heritability	Density	λ				
		0	0.1	0.3	0.5	1
0.25	777K	0.1695 ^a (0.018)	0.1682 ^a (0.017)	0.1657 ^a (0.031)	0.1598 ^a (0.02)	0.1206 ^b (0.036)
	50K	0.1699 ^a (0.01)	0.1675 ^a (0.014)	0.1662 ^a (0.015)	0.1610 ^a (0.02)	0.1232 ^b (0.039)
	5K	0.1656 ^a (0.016)	0.1655 ^a (0.03)	0.1596 ^a (0.019)	0.1506 ^a (0.022)	0.1007 ^b (0.053)
0.5	777K	0.3712 ^a (0.69)	0.3681 ^{abcd} (0.066)	0.3644 ^{abcd} (0.115)	0.3550 ^a (0.078)	0.2961 ^{bd} (0.103)
	50K	0.3758 ^{abc} (0.052)	0.3726 ^{abc} (0.07)	0.3692 ^{abcd} (0.07)	0.3608 ^{abcd} (0.082)	0.3035 ^{abcd} (0.12)
	5K	0.3194 ^{abcd} (0.059)	0.3158 ^{abcd} (0.111)	0.3126 ^{ab} (0.063)	0.301 ^{abcd} (0.062)	0.2341 ^{acd} (0.119)

The means within the same column with at least one common letter, do not have significant difference ($P>0.05$).

Table 5 The genomic estimated breeding value (GEBV) and estimated breeding value (EBV) accuracies using the 400 randomly selected individuals from the reference population, with the different single nucleotide polymorphism (SNP) panel densities, various λ and the heritability

Heritability	Density	λ				
		0	0.1	0.3	0.5	1
0.25	777K	0.2099 ^a (0.01)	0.2082 ^a (0.009)	0.2102 ^a (0.01)	0.2079 ^a (0.012)	0.1774 ^{cd} (0.054)
	50K	0.2092 ^a (0.002)	0.2085 ^a (0.012)	0.2073 ^a (0.016)	0.207 ^{ab} (0.013)	0.1797 ^{abcd} (0.059)
	5K	0.2093 ^{abcd} (0.015)	0.2084 ^{abcd} (0.031)	0.2075 ^a (0.016)	0.2072 ^{ab} (0.017)	0.1865 ^{bd} (0.057)
0.5	777K	0.3635 ^a (0.083)	0.3622 ^{ab} (0.074)	0.3797 ^{ab} (0.125)	0.3535 ^a (0.081)	0.3100 ^{ab} (0.086)
	50K	0.3698 ^{ab} (0.063)	0.3676 ^{ab} (0.082)	0.3562 ^{ab} (0.078)	0.3587 ^{ab} (0.088)	0.3111 ^{ab} (0.088)
	5K	0.378 ^{ab} (0.061)	0.3679 ^{ab} (0.113)	0.3760 ^{ab} (0.062)	0.3711 ^{ab} (0.057)	0.3112 ^{ab} (0.085)

The means within the same column with at least one common letter, do not have significant difference ($P>0.05$).

While the accuracy of genomic predictions in the developing countries has been reported from low to intermediate and in the range of 0.21-0.6 (Mrode *et al.* 2019), that is consistent with the results of present research. The lower accuracy of genomic breeding values in the developing countries can be due to the smaller size of reference populations and a lower accuracy of phenotypic data than the proven bulls in the developed countries as well as lack of appropriate breeding programs in these countries (Mrode *et al.* 2019).

Heritability is considered as one of the effective factors on the accuracy of genomic breeding values estimations, and it was known as a factor that researcher is not able to control it. Since the trait heritability reduction causes dramatically reduction in the accuracy of genomic breeding values prediction, one can resolve this deficiency by increasing the animals' number in the reference population or along with the marker density increasing. The high heritability of a trait demonstrates that the environmental factors in creating diversity are less effective, in comparison with the genetic factors.

Reducing the environmental factors effects on the trait's phenotypic value would result in decreasing the error variance of model and consequently increasing the accuracy of genomic breeding values prediction. It has been indicated that there is a higher breeding value prediction accuracy and genetic gain in the traits with a high heritability (De los Campos *et al.* 2013). The comparison between estimated accuracies shown in Tables 2 to 5, indicated that in all scenarios, the accuracy of genomic breeding values estimation in the higher heritability (0.5) was more than the lower heritability (0.25).

The number of animals in the reference population is one of the key tools for the genomic selection. Because this factor can affect the accuracy of allelic effects estimations, and consequently the genetic gain level. In this study, the prediction accuracy was the highest at the time that all available animals were considered as the reference population.

Perez-Cabal *et al.* (2012) described that increasing the relationships between the reference population and validation population would cause the increase of the genomic selec-

tion accuracy, which is in agreement with this study results. By reducing the number of the reference population, selecting the subset according to the highest relatedness indicated the highest accuracy, in comparison with the subset selection according to inbreeding and random.

Solberg *et al.* (2008) investigated the effect of marker density on the genomic selection accuracy. They reported that the increase of marker density could cause increasing in the accuracy of genomic selection, which was consistent with this study results. Since the higher density of markers leads to the genotyping costs increase, breeders are searching for the cost-effective methods for genetic improvement; low-density panels could be considered as more effective if they give the accuracy similar to the high-density panels. This research results showed that there was not a significant difference between high and intermediate panels when all possible animals in reference population were used. Therefore, it is suggested to using an intermediate panel for the traits with heritability ranged from 0.25 to 0.5.

CONCLUSION

In this research, the accuracy of genomic breeding values was studied in the different marker densities and different combinations of G and A matrix, in terms of the various subsets of the reference populations. The results indicated that the use of G matrix played an important role in increasing the accuracy of breeding value estimation, when higher weights were assigned to the marker-based relationship matrix, the prediction accuracies would increase in all scenarios. These results also indicated that the reference set composition plays an important role in the prediction accuracy. The larger reference sets and an increase in the markers density can lead to a higher accuracy of breeding values. The highest accuracy of breeding values prediction was obtained when all individuals of reference population were used and after that in the related reference population the accuracies were higher. In the scenario of inbred reference population, the accuracy of genomic breeding values in the different densities was estimated lower than the other scenarios, which it shows less efficiency of inbred reference population than the other scenarios. This study demonstrated that when there is a high relationship between the reference population and the validation population; the markers could be used with a lower density in the genomic evaluation, and as a result achieving the genetic gain of interest. In the scenario of related reference population, the more relationship between the reference population and validation population increases the efficiency of using LD due to common haplotype blocks which are established between the related animals resulting from LD between markers and genes loci. Also, the higher relationships, due to sharing more haplotypes, play an important role in the

results related to the accuracy of breeding values prediction. In conclusion, using the intermediate marker density and the reference individuals' subset with the highest relationship, and consequently, applying a higher weighting on marker information for creating the relationship matrix can result in better genomic selection efficiency.

ACKNOWLEDGEMENT

We thank the Bioinformatics Center of Ferdowsi University of Mashhad and Shahid Bahonar University of Kerman for providing computer facilities for analyzing of large data.

REFERENCES

- Amiri Roudbar M., Abdollahi-Arpanahi R., Mehrgardi A.A., Mohammadabadi M., TaheriYeganeh A. and Rosa G.J.M. (2018). Estimation of the variance due to parent-of-origin effects for productive and reproductive traits in Lori-Bakhtiari sheep. *Small Rumin. Res.* **160**, 95-102.
- Amiri Roudbar M., Mohammadabadi M., Mehrgardi A.A. and Abdollahi-Arpanahi R. (2017). Estimates of variance components due to parent of origin effects for body weight in Iran-Black sheep. *Small Rumin. Res.* **149**, 1-5.
- Boichard D., Ducrocq V., Croiseau P. and Fritz S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *C.R. Biol.* **339**, 274-277.
- Boison S.A., Utsunomiya A.T.H., Santos D.J.A., Neves H.H.R., Carvalheiro R. and Mészáros G. (2017). Accuracy of genomic predictions in Gyr (*Bos indicus*) dairy cattle. *J. Dairy Sci.* **100**, 1-12.
- Calus M.P.L. (2010). Genomic breeding value prediction: Methods and procedures. *Animal.* **4**, 157-164.
- Calus M.P.L., Meuwissen T.H.E., de Roos A.P.W. and Veerkamp R.F. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics.* **178**, 553-561.
- Clark S.A., Hickey J.M., Daetwyler H.D. and Van der Werf J.H. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* **44**, 4-14.
- 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**, e3395.
- De los Campos G., Hickey J.M., Pong-Wong R., Daetwyler H.D. and Calus M.P.L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics.* **193**, 327-345.
- Habier D., Fernando R.L. and Dekkers J.C.M. (2007). The impact of genetic relationship information on genome-assisted Breeding values. *Genetics.* **177**, 2389-2397.
- Hayes B.J. and Goddard M.E. (2008). Technical note: Prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* **86**, 2089-2092.
- Hayes B.J., Bowman P.J., Chamberlain A.C. and Goddard M.E. (2009a). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* **92**, 433-443.

- Hayes B.J., Bowman P.J., Chamberlain A.C., Verbyla K. and Goddard M.E. (2009b). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet. Sel. Evol.* **24**, 41-51.
- Henderson C.R. (1984). Applications of Linear Model in Animal Breeding. University of Guelph, Guelph, Ontario, Canada.
- Lourenco D.A.L., Fragomeni B.O., Tsuruta S., Aguilar I., Zumbach B., Hawken R.J., Legarra A. and Misztal I. (2015). Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. *Genet. Sel. Evol.* **47**, 56-66.
- Lund M.S., De Roos A.P.W., De Vries A.G., Druet T., Ducrocq V., Fritz S. and Schrooten C. (2010). Improving genomic prediction by Euro Genomics collaboration. Pp. 24-29 in 10th World Conf. Genet. Appl. Livest. Prod., Leipzig, Germany.
- Meuwissen T.H.E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* **41**, 35-44.
- Momen M., AyatollahiMehrgardi A., AmiriRoudbar M., Kranis A., Mercuri Pinto R., Valente B.D., Morota G., Rosa G.J.M. and Gianola D. (2018). Including phenotypic causal networks in genome-wide association studies using mixed effects structural equation models. *Front. Genet.* **9**, 455-463.
- Mrode R., Ojango J.M.K., Okeyo A.M. and Mwacharo J.M. (2019). Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. *Front. Genet.* **9**, 694-702.
- Pérez-Cabal M.A., Vazquez A.I., Gianola D., Rosa G.J.M. and Weigel K.A. (2012). Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. *Front. Genet.* **3**, 27-35.
- Sargolzaei M. and Schenkel F. (2009). QMSim: A large-scale genome simulator for livestock. *Bioinformatics.* **25**, 680-681.
- Silva M.V., dos Santos D.J.A., Solomon A.B., Utsunomiya A.T.H., Carmo A.S., Sonstegard T.S., Cole J.B. and Van Tassell C.P. (2014). The development of genomics applied to dairy breeding. *Livest. Sci.* **166**, 66-75.
- Solberg T.R., Sonesson A.K., Woolliams J.A. and Meuwissen T.H. (2008). Genomic selection using different marker types and densities. *J. Anim. Sci.* **86**, 2447-2454.
- Su G., Madsen P., Nielsen U.S., Mäntysaari E.A., Aamand G.P., Christensen O.F. and Lund M.S. (2012). Genomic prediction for Nordic Red cattle using one-step and selection index blending. *J. Dairy Sci.* **95**, 909-917.
- Teimourian M., Shariati M.M. and Aslaminejad A. (2015). Comparing various statistical methods in genomic selection of Holstein populations. *Anim. Prod. Res.* **7**, 198-203.
- Van Raden P.M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414-4423.
- Van Raden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F. and Schenkel F.S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **92**, 16-24.
- Villumsen T.M., Janss L. and Lund M.S. (2009). The importance of haplotype length and heritability using genomic selection in dairy cattle. *J. Anim. Breed. Genet.* **126**, 3-13.
- Vitezica Z.G., Aguilar I., Misztal I. and Legarra A. (2011). Bias in genomic predictions for populations under selection. *Genet. Res.* **93**, 357-366.
- Wang Q., Yu Y., Yuan J., Zhang X., Huang H., Li F. and Xiang J. (2017). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *BMC Genet.* **18**, 45-51.
- Wiggans G.R., Cole J.B., Hubbard S.M. and Sonstegard T.S. (2017). Genomic selection in dairy cattle: The USDA experience. *Annu. Rev. Anim. Biosci.* **5**, 309-322.