Research Article

Comparison of BLUP and Bayesian methods for different sizes of training population in genomic selection

Samet Hasan ABACI* ⓘ, **Hasan ÖNDER** ⓘ

Department of Animal Science, Faculty of Agriculture, Ondokuz Mayıs University, Samsun, Turkey

Abstract: This study aims to compare the accuracy of pedigree-based and genomic-based breeding value prediction for different training population sizes. In this study, Bayes (A, B, C, Cpi) and GBLUP methods for genomic selection and BLUP method for pedigree-based selection were used. Genomic and pedigree-based breeding values were estimated for partial milk yield (158 days) of Holstein cows (400 individuals) from a private enterprise in the USA. For this aim, populations were created for indirect breeding value estimates as training (322–360) and test (78–40) populations. In animals genotyped with a 54k SNP, the marker file was encoded as –10, 0, and 10 for AA, AB, and BB marker genotypes, respectively. Bayes and GBLUP methods were performed using GenSel 4.55 software. A total of 50,000 iterations were used, with the first 5000 excluded as the burn-in. Pedigree-based breeding values were estimated by REML using MTDFREML software employing an animal model. Correlations between partial milk yield and estimated breeding values were used to assess the predictive ability for methods. Bayes B method gave the highest accuracy for the indirect estimate of breeding value.

Key words: Bayes, best linear unbiased prediction, breeding value, genomic selection

1. Introduction

In order to obtain more and higher-quality yields from animal materials which are of great importance in nutrition, the environmental conditions and genetic structure of animals should be improved [1,2]. Since the level of the yield that can be achieved by improving environmental conditions is only to the extent that the genotype will allow, it is highly important to carry out breeding studies to increase the genotypic value provided that appropriate environmental conditions are met. There are many methods to estimate the genotypic values.

Today, best linear unbiased prediction (BLUP) method is the most useful method for breeding studies with pedigree information and phenotypic properties. In the 1950s, Henderson developed and introduced the BLUP method to estimate the breeding values of animals. However, this term began to be used after the 1960s. Thanks to the increasing computer technology, the actual use of the BLUP method has taken its final status since the 1980s with the addition of an animal model. Thanks to BLUP method, which allows simultaneous estimation of breeding value and fixed factors, an estimated breeding value is calculated [3], it is made possible to sort the animals according to their estimated genetic potentials, and faster genetic progression is achieved through generations with more accurate selection results [4].

Recent years, with the development of molecular genetics, have been dominated by the expectation (in terms of animal husbandry and product varieties) that the information at the DNA level will lead to a faster genetic gain than the information which is obtained based solely on phenotypic data [5].

The identification of an intermittent map of genetic markers has enabled the detection of some quantitative trait loci (QTL) [6]. The inclusion of marker information in the BLUP-derived breeder values was suggested by Fernando and Grossman [7] and was predicted to provide extra genetic gain by 8–38% [8]. However, the ideal method for estimating the breeding value from genomic data is to calculate the conditional average of the breeding value of the animal considering the genotype in each QTL. This conditional average can only be calculated using the predistribution of QTL effects. Therefore, it can only form part of the research to implement genomic selection. In practice, since more sequence and single nucleotide polymorphisms (SNP) data are obtained by using marker genotypes instead of QTL genotypes, it may be considered to approach the ideal result by genomic selection in the estimation of breeding value [9].

Genomic selection methods that emerged with the development of molecular genetics and statistics together

* Correspondence: samet.abaci@omu.edu.tr

with the presence of high-density panels of SNP markers and new perspectives for marker-assisted selection applications are being used widely in breeding studies [10]. Meuwissen and Goddard [8] have developed a way to combine large-scale DNA information available in animal model theory to estimate genomic breeding values [4].

With this theory, instead of using a limited number of marker information belonging to individuals, the breeding value obtained by using the information belonging to all the markers in the genome has been named as genomic estimated breeding value (GEBV) [5,11,12]. To calculate GEBV, a prediction equation based on SNP is first obtained. The animal genome from data estimated from the population whose phenotypes and genotypes are known is divided into small pieces. Thus, the effect levels of all loci contributing to the genetic variation in the investigated properties are obtained and used even if their individual effects are very small. In subsequent generations, individuals are genotyped to determine in which region of the chromosome they carry markers. Then, GEBVs are calculated by summing the effects of the regions in which the individuals carry the markers [13,14]. This improved technique allows selecting individual animals genomically.

The expression "genomic selection" is used for selection made according to the GEBVs of the animals [5]. In genomic selection, parameter estimates obtained by using a population in which genetic marker and phenotypic values are available (training population) are also used to estimate breeding values of individuals in the population with only marker information (test population) [5,12].

Genomic breeding value estimates can be obtained in two ways: directly and indirectly. In direct methods, genomic breeding value estimations can be obtained in one step with mixed equation models of individuals whose phenotypes and genotypes are known. In indirect methods, firstly, marker effects are estimated with the help of a training population, and breeding values estimates can be obtained only with marker effects in the population with genotype information and the population in which selection will be made [12,13,15]. Three methods are used to estimate genomic breeding value: ordinary least square (OLS), BLUP, and Bayesian. Due to the low amount of data and the high number of markers, OLS method is not preferred. Of these methods, the most commonly used are BLUP (GBLUP) and Bayes (A, B, C, Cpi). According to the literature review conducted, the majority of the studies are based on data obtained by simulation, and more studies with real data are needed.

The aim of this study was to compare the marker effects obtained from training populations of different sample sizes with the estimation of genomic breeding values of the test population for various methods that are preferred the most (Bayes A, Bayes B, Bayes C, Bayes Cpi, and GBLUP). Thus, the study aimed to determine which one or more of the methods mentioned in different sample sizes can make more reliable genomic breeding value estimation with real data.

2. Materials and methods

2.1. Materials
In this study, the total of partial milk yield (PMY) records (phenotype) obtained daily up to 158th day, pedigree, and 50K single nucleotide polymorphism (54609 SNP) genotype records from 400 Holstein dairy cattle reared in a private enterprise in the United States were used.

2.2. Methods
In order to make breeding value estimates with BLUP method, it was determined that of the existing animals, 78 had mother PMY records. Therefore, it was decided to use 78 animals as test population. In addition, to determine the effect of the sample size of the training population on the accuracy of the methods, 40 animals were chosen out of 78 animals randomly, and 360 training - 40 test population and 322 training - 78 test population were formed. In the BLUP method, analyses were performed assuming that there were no PMY records (missing values) of the animals in the test population. SPSS statistical program was used to determine the environmental factors (lactation order, calving season, calving year, and milking period-covariate) that affect PMY. Bonferroni multiple comparison test was used to compare the subgroups of the factors that were found statistically significant. The calving season was taken into consideration in the calendar season. In determining lactation order, 1st, 2nd and 3rd lactations were taken individually, and 4th and later lactations were included in the 4th lactation due to the small number of observations. Variance components of PMY, genetic parameters, and estimation of phenotypic breeding value were analyzed according to the individual animal model using the MTDFREML package [16] program.

The model used to investigate the effect of environmental factors is given below:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + b(X_{ijkl} - \bar{X}) + e_{ijkl}$$

In equality: Y_{ijkl}; observed value of the PMY, μ; population means, α_i; i. effect of lactation order, β_j; j. effect of calving season, γ_k; k. effect of calving year, b; constant regression coefficient for days in milk, X_{ijkl}; in the ijk subgroup, l. cow milking time, $\bar{X}$; average milking time of population, e_{ijkl}; random error.

The model used to estimate variance components and breeding values is given below:

$$Y_{ijklm} = F_{ijkl} + a_m + e_{ijklm}$$

In equality: Y_{ijklm}; observed value of the PMY, F_{ijkl}; constant environmental factors ($\alpha_i + \beta_j + \gamma_k + b_1 A$), $b_1 A$; direct effect of milking time on milk yield, a_m; additive gene effect of the animal: a ~ N (0, σ^2_α), e_{ijklm}; random error: e ~ N (0, σ^2_e).

The following general model presented by Gianola et al. [17] was used to calculate the marker effects and genomic breeding values using the GBLUP (Bayes C0) and Bayesian methods (Bayes A, Bayes B, Bayes C, and Bayes Cpi).

$$y = Xs + C\beta + W\alpha + e$$

In equation: y; phenotypic vector, X; incidence matrix for constant effects (in the simplest case, the overall mean is reduced to a vector with elements 1), s; constant effects vector, C; covariate design matrix, β; covariate effect vector, W; a known matrix of numerical genotype scores for each marker (–10, 0, 10 for AA, AB, BB respectively), α; marker additive effects vector. e; random error vector: $e \sim N(0, I\sigma^2_e)$.

In another and more detailed illustration, the statistical model for the marker-based methods with polygenic effects is as follows:

$$y = \mu 1_n + Xs + C\beta + \sum_j X_j \alpha_j \delta_j + Zu + e$$

Here y is an $N \times 1$ vector of phenotypes with N being the numbers of individuals, μ is the overall mean, 1_n is a vector of ones of length n, X is an incidence matrix for constant effects (lactation order, calving season, calving year), s is a constant effects vector, C is a covariate design matrix, β is a covariate effect vector, X_j is an $N \times 1$ vector of genotypes at SNP j, coded (-10, 0, 10), α_j is the random allele substitution effect for SNP j, δ_j is a 0/1-indicator variable which equals 1 if SNP j is included in the model and zero otherwise, Z is the associated design matrix, u is a vector with random polygenic effects of all individuals with $Var(u) = A\sigma^2_u$, (A is the numerator relationship matrix, and σ^2_u is the polygenic variance), and e is a vector of random residuals $e \sim N(0, I\sigma^2_e)$.

In the Bayes A method, all $\delta_j = 1$ so that all markers fit in the model. The prior distribution of marker substitution effect α_j is normal N $(0, \sigma^2_{\alpha_j})$, and the prior distribution for marker variance $\sigma^2_{\alpha_j}$ is a scaled inverse chi-square distribution. The prior distribution of the error variance is σ^2_e.

The distribution of SNP follows a Student's t-distribution. This allows for a higher probability of moderate to large SNP effects than a normal distribution.

In reality, the distribution of genetic variances across loci is such that there are many loci with no genetic variance (not segregating) and a few with genetic variance. However, the prior density of method Bayes A does not have a density peak at $\sigma^2_{g_j} = 0$, which is infinitesimal. Method Bayes B, therefore, uses a prior that has a high density π at $\sigma^2_{g_j} = 0$ and has an inverted chi-square distribution for $\sigma^2_{g_j} > 0$ [5].

In Bayes A and Bayes B, there is only one additional degree of freedom compared with its prior, and so the shrinkage of SNP effects is largely dependent on the scale parameter, S. To overcome this limitation, proposed method is Bayes C, which involves estimating a single variance that is common to all SNPs, thereby reducing the influence of the scale parameter. In Bayes C, π is treated as an unknown, and it is assumed that it has Uniform distribution with mean = 0 and variance = 1.

In Bayes Cpi, marker effects on phenotypic traits were sampled from a mixture of null and normal distributions. The markers in the model shared a common variance σ^2_α and the probability π that markers do not have a genetic effect. In Bayes C, there is the implicit assumption that the probability, $\pi > 0$, i.e. a SNP has zero effect, is regarded as known. The shrinkage of SNP effects is affected by π and should be estimated from the data and proposed Bayes Cpi, which incorporates this estimation step. Thus, compared to Bayes C, the additional feature of Bayes Cpi is estimating π from the data [10].

Method G-BLUP fitted all SNPs in the model, assuming that every SNP explained an equal proportion of the total genetic variance. This method can be named as Bayes C0 for executive simplicity. It is the same as the Bayes C when pi = 0 [18].

In order to determine the accuracy of methods for estimating genomic breeding value indirectly, marker effects were determined by using phenotypic and genotypic recordings of 322 and 360 animals (training population) out of 400 animals. Next, estimated genomic breeding values were obtained by using only the genotypic recordings of the 78 and 40 animals in the first lactation, by using the marker effects of the training population.

Given the estimates of the marker effects and the marker genotypes, genomic estimated breeding values (GEBV) for the individuals in the test population set are predicted as:

$$GEBV = \sum_{j=1}^{k} z_{ij} \widehat{\alpha}_j$$

where GEBV is the GEBV for individual i in the test population dataset, k is the number of marker (54609), z_{ij} is the marker genotype of individual i for marker j, and $\widehat{\alpha}_j$ is the posterior mean effect of marker j.

In simulation studies, the correlation between direct breeding value (DBV) and true breeding values (TBV) is used to represent the accuracy of the DBV. However, TBV is not available in the field data, and response variable (phenotype recordings, estimated breeding value, deregressed estimated breeding value, etc.) is generally used to obtain DBV and the accuracy of DBV [19]. The relationship between genomic and pedigree-based breeding values obtained using different methods was determined by Pearson correlation. In addition, the accuracy of the methods was calculated by Pearson correlation between PMY and genomic and pedigree-based breeding values. The deviation coefficients of the methods were found as

linear regression coefficient of genomic and pedigree-based breeding values on PMY. In the implementation of Bayesian methods (Bayes A, Bayes B, Bayes C, Bayes Cpi, and Bayes C0 (GBLUP)), the GenSel package program running online under the Cy-Verse cyber infrastructure web interface Discovery Environment was used[1]. In the study, Markov chain was run for 50,000 iterations of Gibbs sampling, the optimization was achieved successfully, and it was thought that the deviation of the first 5000 iterations was burn; thus, they were ignored and excluded from the experiment [20]. Due to the monomorphic structure in the genes, some markers (6497 and 6435) for different population sizes (322 and 360) were excluded from the analysis.

For correlation and regression analysis, SPSS package program was used. Mantel test was used to determine whether there was a difference between similarity matrices obtained using different methods and different sample sizes and analysis was performed in XLSTAT package program. The difference between the obtained correlation coefficients was tested online by Fisher Z transformation[2].

3. Results and discussion

This section includes results obtained related to indirect genomic estimated breeding values. The animals with a training population of 400 were divided into two different groups as the training population (n = 322 and 360) and the test population (n = 78 and 40). While obtaining the pedigree-based breeding values, PMY records of the animals in the test population were considered as missing observations (0.0), and the estimated breeding values were obtained from related pedigree records. In genomic selection, indirect estimation of the breeding value was made in two steps. In the first step, genomic estimated breeding values were obtained as indirect breeding value estimation, while marker values of the training population were determined by correlating them with PMY records. In the second step, the marker values obtained from the training population and the marker values of the animals in the test population were correlated, and the breeding value estimates were obtained without considering PMY records. Method accuracy was determined by checking Pearson correlation coefficient ($r_{(f,\hat{y})}$) between the estimated breeding value (BV) and PMY. In addition, the linear regression coefficient of the breeding value on PMY was calculated, and the deviations of the methods ($b_{(f,\hat{y})}$) were determined.

Descriptive statistics of PMYs according to lactation order for different training population sizes are given in Table 1. According to the results of variance analysis for training population sizes 322 and 360, the effect of

lactation order on PMY was significant. The highest PMY was observed in animals in lactation 4 and above, whereas the lowest PMY was found in animals in lactation 1.

Descriptive statistics of PMYs of animals of different size training populations according to their calving years are given in Table 2. According to the results of the variance analysis, the effect of the calving years on PMY was not significant for the sizes of different training populations (P > 0.05). It can be said that the determined values are reliable in terms of variation.

Descriptive statistics of PMYs of animals of different size training populations according to calving season are given in Table 3.

According to the results of the variance analysis, the effect of calving season on PMY was not significant for the sizes of different training populations (P > 0.05). It can be said that the determined values are reliable in terms of variation.

Genetic variance, error variance, total variance, heritability, and calculation times related to pedigree and genomic breeding values obtained using different methods for different numbers of training populations are presented in Table 4. In the estimation of genomic breeding values, for initial values to be used for genetic variance and error variance, the results estimated from pedigree-based breeding values were used. Since monomorphic structure was observed in 6497 markers of the group 360 with a training population size of 322 and 6435 markers of the group with the estimation of genomic breeding values with a training population size of 360, analyses were performed on 48,112 and 48,174 markers.

When Table 4 is examined, it can be observed that, in terms of heritability, for both training populations, the highest degree of heritability was found with BLUP, Bayes A, Bayes B, and Bayes C methods, while the lowest degree of heritability was found with Bayes Cpi method. It may result from the fact that Bayes Cpi is estimating π from the data. Degree of heritability obtained from training population size 322 was found to be higher when compared with that of training population size 360. When the literature was reviewed, estimates of heritability for milk yield in Holstein Friesian cattle were estimated as 0.07 by Abubakar et al. [21] and as 0.13 by Kim et al. [22]. Accordingly, they were found to be lower than the highest heritability obtained. Some studies have found values of heritability similar to the one obtained, and reported degree of heritability was in parallel with what the following names reported: Cañón et al. [23] 0.17, Saatçi et al. [24] 0.16, and Ertuğrul et al. [25] 0.16. It was found that in most of the other studies, the degree of heritability for milk yield was estimated to be between 0.19 and 0.45

Table 1. Descriptive statistics of PMYs according to lactation order.

Lactation Order	n	322		n	360	
		$\bar{X} \pm S_{\bar{X}}$	CV (%)		$\bar{X} \pm S_{\bar{X}}$	CV (%)
1	148	5812.0 ± 66.26[c]	13.87	186	5770.1 ± 59.85[c]	14.15
2	83	7142.9 ± 116.80[b]	14.90	83	7142.9 ± 116.80[b]	14.90
3	49	7609.9 ± 148.53[ab]	13.66	49	7609.9 ± 148.53[ab]	13.66
≥4	42	7794.2 ± 138.93[a]	11.55	42	7794.2 ± 138.93[a]	11.55
P		<0.001			<0.001	
b		40.033 (P < 0.001)			40.596 (P < 0.001)	

b: Correction coefficient and significance level for different milking time; P: Significance level; [a,b,c]: There is a difference between the means indicated by different letters in the same column; PMY: Partial milk yield; $\bar{X}$: Mean; $S_{\bar{X}}$: Standard error; CV(%): Coefficient of variation; n: Number of animals

Table 2. Descriptive statistics of PMYs according to years of calving.

Calving years	n	322		n	360	
		$\bar{X} \pm S_{\bar{X}}$	CV (%)		$\bar{X} \pm S_{\bar{X}}$	CV%)
2008	143	7028.0 ± 67.41	17.39	148	6980.5 ± 102.74	17.91
2009	139	6524.7 ± 90.26	19.32	165	6382.3 ± 96.64	19.45
2010	40	6033.5 ± 164.39	14.17	47	5960.7 ± 123.34	14.19
P		0.430			0.257	
b		40.033 (P < 0.001)			40.596 (P < 0.001)	

b: Correction coefficient and significance level for different milking time; P:Significance level; PMY: Partial milk yield; $\bar{X}$: Mean; $S_{\bar{X}}$: Standard error; CV(%): Coefficient of variation; n: Number of animals

Table 3. Descriptive statistics of PMYs according to calving season.

Calving season	n	322		n	360	
		$\bar{X} \pm S_{\bar{X}}$	CV (%)		$\bar{X} \pm S_{\bar{X}}$	CV(%)
Autumn	76	6920.5 ± 143.18	18.04	83	6818.2 ± 138.28	18.48
Winter	81	6406.6 ± 126.82	17.82	94	6327.5 ± 114.03	17.47
Spring	81	6566.4 ± 150.80	20.67	85	6513.4 ± 146.16	20.69
Summer	84	6863.1 ± 128.19	17.12	98	6653.1 ± 127.90	19.03
P		0.167			0.147	
b		40.033 (P < 0.001)			40.596 (P < 0.001)	

b: Correction coefficient and significance level for different milking times; P:Significance level; PMY: Partial milk yield; $\bar{X}$: Mean; $S_{\bar{X}}$: Standard Error; CV(%): Coefficient of variation; n: Number of animals

[26–30]. When the calculation times of marker effects are examined, it can be seen that they differ between 1079 s and 3177 s; and while calculation time lasted the longest with Bayes A method, it was found to last the shortest with Bayes C method.

Correlations between breeding values calculated using different methods for different test population sizes (78: lower diagonal; 40: upper diagonal) are given in Table 5. As a result of the analysis, it was determined that breeding values obtained by BLUP and Bayes C methods had higher

Table 4. Results of variance elements.

n	Variance elements	BLUP	GBLUP (Bayes C0)	Bayes A	Bayes B	Bayes C	Bayes Cpi
322	Genetic variance	129,095	108,384	128,532	128,855	123,150	61,085
	Error variance	665,456	677,066	658,366	657,822	665,423	722,846
	Phenotypic variance	794,551	785,450	786,898	786,677	788,573	783,931
	Heritability	0.16	0.14	0.16	0.16	0.16	0.08
	Pi	-	0.00	0.00	0.95	0.95	0.36
	Calculation time (Sec)	-	2255	3177	1357	1322	1794
360	Genetic variance	95,352	88,759	94,947	94,876	88,531	38,147
	Error variance	688,976	691,169	684,246	684,058	692,074	738,020
	Phenotypic variance	784,329	779,928	779,193	778,934	780,605	776,166
	Heritability	0.12	0.11	0.12	0.12	0.11	0.05
	Pi	-	0.00	0.00	0.95	0.95	0.47
	Calculation time (s)	-	2395	2673	1229	1079	1805

BLUP: Best linear unbiased prediction; GBLUP: Genomic best linear unbiased prediction; n: Number of animals

correlation than the other methods, while the lowest correlation was found between breeding values obtained by BLUP and Bayes Cpi methods. A high degree of correlation was determined between the Bayesian methods, and the correlations were found to be statistically significant (P < 0,01).

The accuracy $(r_{(f,\hat{y})})$ and deviations $(b_{(f,\hat{y})})$ of the methods for different test population sizes were calculated, and they are shown in Table 6. As a result of the analysis, it was found that the accuracy (correlation) between the breeding values obtained from Bayes B and Bayes C methods, and PMY was higher for the test population size 78 when compared with the other methods. The lowest accuracy was obtained by the BLUP method. While the lowest deviation was found in the equation obtained by BLUP method, it was found that the highest deviation was in the estimation of PMY with breeding values obtained from Bayes Cpi method.

Breeding values obtained from Bayes B method for test population 40 were found to have higher accuracy (correlation) than the other methods. The second highest accuracy was determined by GBLUP and Bayes A methods. However, it was found that the deviation in the estimation was the lowest in the BLUP method, and the highest deviation was obtained from the Bayes Cpi method.

According to the results of Fisher Z test analysis conducted to determine the difference between correlation coefficients, there is no statistically significant difference between the correlation coefficients obtained for test population sizes 78 and 40, and PMY (P > 0.05). Correlation matrices can be said to be similar.

According to the results of the study, there is no significant difference between Bayesian methods in obtaining indirectly estimated genomic breeding values (P > 0.05). It was determined that higher accuracy can be obtained in breeding values obtained with the contribution of genomic information compared to pedigree-based breeding values estimated with pedigree information. Among the Bayesian methods, the reliability of Bayes B and Bayes A methods was found to be higher when compared with the other methods.

Ding et al. [31] calculated the mean accuracy in the Holstein population of China to be 0.380 for the Bayes B method for milk yield. They reported that Bayes B method predicted better than GBLUP. Rolf et al. [32] reported that high direct breeding accuracy was obtained consistently for all traits in models using Bayes A method in mixed breed commercial feeder cattle. Karaman et al. [18] compared genomic prediction methods, namely GBLUP, Bayes B, and Bayes C, in the data of human length genome project. They reported that when the training population was small (n < 6000 individuals), Bayes B and Bayes C applied to the 30M genome for human-size variable selection methods were not superior to GBLUP, but that they were superior when more samples were included in the training population.

Habier et al. [33] compared the Bayes Cpi and Bayes Dpi methods they developed with Bayes A and Bayes B methods for both milk yield, fat yield, protein yield, and somatic cell score, both through simulation and in North American Holstein bulls. They stated that the accuracy they obtained with Bayes Cpi and Bayes Dpi methods were similar and that Bayes A method was a good choice to

Table 5. Correlations between breeding values calculated using different methods for different test population sizes (Lower diagonal (78), Upper diagonal (40)).

Methods	n	BLUP	GBLUP (Bayes C0)	Bayes A	Bayes B	Bayes C	Bayes CPi
n			40	40	40	40	40
BLUP		-	0.840**	0.837**	0.836**	0.837**	0.822**
GBLUP (Bayes C0)	78	0.802**	-	0.999**	0.999**	0.999**	0.997**
Bayes A	78	0.805**	0.999**	-	0.999**	0.999**	0.998**
Bayes B	78	0.806**	0.999**	0.999**	-	0.999**	0.998**
Bayes C	78	0.810**	0.999**	0.999**	0.999**	-	0.997**
Bayes Cpi	78	0.790**	0.999**	0.998**	0.998**	0.997**	-

**P < 0.01 BLUP: Best linear unbiased prediction; GBLUP: Genomic best linear unbiased prediction; n: number of animals

Table 6. Accuracy and deviations of methods for different test population sizes.

Method	$r_{(f,\hat{y})}$		Fizher Z	$b_{(f,\hat{y})}$	
	n: 78	n: 40	P	n: 78	n: 40
BLUP	0.060	0.145	0.334	0.436	1.161
GBLUP (Bayes C0)	0.070	0.248	0.181	0.532	2.108
Bayes A	0.071	0.247	0.184	0.463	1.921
Bayes B	0.074	0.250	0.183	0.480	1.945
Bayes C	0.074	0.245	0.191	0.516	2.091
Bayes Cpi	0.069	0.234	0.200	0.874	4.115

BLUP: Best linear unbiased prediction; GBLUP: Genomic best linear unbiased prediction; n: Number of animals; P: Significance level; $r_{(f,\hat{y})}$: Pearson correlation coefficient; $b_{(f,\hat{y})}$ Linear regression coefficient; f: Phenotypic value; ŷ: Estimated breeding value

estimate the genomic estimated breeding value with actual data. They stated that the calculation time of Bayesian Cpi method was shorter than that of Bayesian Dpi method and that Bayes A method had the longest application time.

4. Conclusion

The use of genomic breeding values allows the selection of animals at an earlier age. In the studies conducted, accuracy of selection based on genomic breeding value estimation was found to be quite high compared to traditional methods based on pedigree information [34]. Genomic selection methods, which have a significant impact on animal breeding programs, provide an important accumulation of knowledge, especially in terms of the decisions to be made about which features will be improved.

The correlation results between indirect genomic breeding values and PMY show no significant difference between Bayesian methods. However, Bayes A and Bayes B methods were found to give more reliable results when compared with the other methods. Therefore, it can be said that Bayes A and Bayes B methods can be used in indirect estimation. The increase in the number of animals in the training population increased the accuracy of the estimates. In addition, it was observed that the accuracy of the other methods also increased as the number of animals in the training population increased. Meuwissen [35] reported that simulation studies may provide more accurate estimates of breeding values when individuals of the training and test population are close relatives.

As a result, it can be said that in estimating breeding value indirectly, the accuracy rate of the methods increases as the number of animals in the training population increases, and Bayes B method makes more accurate estimates without a big difference.

Since most of the studies on genomic selection have been obtained by simulation, their accuracy has not been proven yet. For this reason, comparison of real genomic

values with different Bayesian methods which were not used in this study and by using a bigger sample size will contribute to the determination of sensitivity of these methods against real values. In practice, genomic selection can be used to select candidate sires indirectly using train data sets of the flock.

Conflict of interest
The authors declare that they have no conflict of interest.

Data availability
The data were obtained from a private business in the USA. Restrictions apply to the availability of these data, which were used with permission for this study.

References

1. Kızılkaya K. Application of different methods in the estimation of breeding values in dairy cattle. MSc, Ankara University, Ankara, Turkey, 1993 (in Turkish with an abstract in English).

2. Ünalan A. Prediction of genetic parameters related to milk yields of Holstein Friesian population reared in the Ceylanpınar Agricultural State Farm by bayesian and REML methods. PhD, Çukurova University, Adana, Turkey, 2002 (in Turkish with an abstract in English).

3. Mrode RA. Linear Models for the Prediction of Animal Breeding Values. Edinburgh, UK: CAB International; 1996.

4. Oldenbroek K, van der Waaij L. Textbook Animal Breeding: Animal Breeding and Genetics for BSc Students. the Netherlands: Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre, 2014.

5. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics 2001; 157 (4): 1819-1829.

6. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R et al. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 1995; 139 (2): 907-920.

7. Fernando RL, Grossman M. Marker assisted selection using best linear unbiased prediction. Genetics Selection Evolution 1989; 21 (4): 467-477. doi: 10.1186/1297-9686-21-4-467

8. Meuwissen THE, Goddard ME. The use of marker haplotypes in animal breeding schemes. Genetics Selection Evolution 1996; 28 (2): 161-176. doi: 10.1186/1297-9686-28-2-161

9. Goddard ME, Hayes BJ. Genomic selection. Journal of Animal Breeding and Genetics 2007; 124 (6): 323-330. doi: 10.1111/j.1439-0388.2007.00702.x

10. Neves HH, Carvalheiro R, Queiroz SA. A comparison of statistical methods for genomic selection in a mice population. BMC Genetics 2012; 13 (100): 1-17. doi: 10.1186/1471-2156-13-100

11. Fan B, Du ZQ, Gorbach DM, Rothschild MF. Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Australasian Journal of Animal Sciences 2010; 23 (7): 833-847. doi: 10.5713/ajas.2010.r.03

12. Yılmaz O, Öner Y, Karaman E, Cemal İ. Genome wide association studies and genomic selection in livestock. In: 9th National Animal Science Congress; Konya, Turkey; 2015. pp. 394-401 (in Turkish with an abstract in English).

13. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 2009; 92 (2): 433-443. doi: 10.1186/1471-2105-12-186

14. Özbeyaz C, Kocakaya A. Genomic evaluation in dairy cattle (A review). Lalahan Hayvancılık Araştırma Enstitüsü Dergisi 2011; 51 (2): 93-104. (in Turkish with an abstract in English)

15. Calus MPL, Meuwissen THE, de Roos APW, Veerkamp RF. Accuracy of genomic selection using different methods to define haplotypes. Genetics 2008; 178 (1): 553-561. doi: 10.1534/genetics.107.080838

16. Boldman KG, Kriese L, Van Vleck L, Van Tassell C, Kachman S. A manual for use of MTDFREML. A set of programs to obtain estimates of variance and covariances (Draft) USDA, ARS, Lincoln, NE, 1995.

17. Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. Additive genetic variability and the bayesian alphabet. Genetics 2009; 183 (1): 347-363. doi: 10.1534/genetics.109.103952

18. Karaman E, Cheng H, Fırat MZ, Garrick DJ, Fernando RL. An upper bound for accuracy of prediction using GBLUP. Plos One 2016; 11 (8): 1-18. doi: 10.1371/journal.pone.0161054

19. Saatchi M, Schnabel RD, Rolf MM, Taylor JF, Garrick DJ. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. Genetics Selection Evolution 2012; 44 (38): 1-10. doi: 10.1186/1297-9686-44-38

20. Fernando RL, Garrick DJ. GenSel-User manual for a portfolio of genomic selection related analyses. 2009.

21. Abubakar BY, McDowell RE, Van Vleck LD. Genetic evaluation of Holsteins in Columbia. Journal of Dairy Science 1986; 69 (4): 1081-1086. doi: 10.3168/jds.S0022-0302(86)80504-5

22. Kim JS, Park KD, Jeong HY, Ahn BS, Lee KJ. Estimation of regional genetic trends for milk and fat yields in the Korean Holstein population. Korean Journal of Animal Sciences 1999; 41: 11-14.

23. Cañón J, Berger PJ, Gutiérrez JP, Muñoz A. Estimate of (co) variance components for milk and fat yield in Spanish holstein population using REML. Archivos De Zootecnia 1989; 38 (142): 249-259.

24. Saatçi MR, Ulutaş Z, Dewi IA, Akkuş İ. Environmental effects, variance components and estimated breeding values of milk yield for Holstein cows in Dalaman state farm. Atatürk Üniversitesi Ziraat Fakültesi Dergisi 2000; 31 (2): 97-101.

25. Ertuğrul O, Orman MN, Güneren G. Some genetic parameters of milk production in the Holstein breed. Turkish Journal of Veterinary and Animal Sciences 2002; 26 (3): 463-469 (in Turkish with an abstract in English).

26. Campos MS, Wilcox CJ, Becerril CM, Diz A. Genetic parameters for yield and reproductive traits of Holstein and Jersey cattle in Florida. Journal of Dairy Science 1994; 77 (3): 867-873. doi: 10.3168/jds.S0022-0302(94)77021-1

27. Dimov G, Albuquerque LG, Keown JF, Van Vleck LD, Norman HD. Variance of interaction effects of sire and herd for yield traits of Holsteins in California, New York, and Pennsylvania with an animal model. Journal of Dairy Science 1995; 78 (4): 939-946. doi: 10.3168/jds.S0022-0302(95)76709-1

28. Gebreyohannes G, Koonawootrittriron S, Elzo MA, Suwanasopee T. Variance components and genetic parameters for milk production and lactation pattern in an ethiopian multibreed dairy cattle population. Asian-Australasian Journal of Animal Sciences 2013; 26 (9): 1237-1246. doi: 10.5713/ ajas.2013.13040

29. Gengler N, Tijani A, Wiggans GR, Van Tassell CP, Philpot JC. Estimation of (co)variances of test day yields for first lactation Holsteins in the United States. Journal of Dairy Science 1999; 82 (1): 225.e1-e14. doi: 10.3168/jds.S0022-0302(99)75227-6

30. Muir BL, Fatehi J, Schaeffer LR. Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holsteins. Journal of Dairy Science 2004; 87 (9): 3029-3037. doi: 10.3168/jds.S0022-0302(04)73435-9

31. Ding X, Zhang Z, Li X, Wang S, Wu X et al. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. Journal of Dairy Science 2013; 96 (8): 5315-5323. doi: 10.3168/jds.2012-6194

32. Rolf MM, Garrick DJ, Fountain T, Ramey HR, Weaber RL et al. Comparison of bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. Genetics Selection Evolution 2015; 47 (23): 1-14. doi: 10.1186/s12711-015-0106-8

33. Habier D, Fernando RL, Kızılkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 2011; 12 (186): 1-12. doi: 10.1186/1471-2105-12-186

34. Pryce JE, Daetwyler HD. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Animal Production Science 2012; 52 (3): 107-114. doi: 10.1071/AN11098

35. Meuwissen THE. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genetics Selection Evolution 2009; 41 (35): 1-9. doi: 10.1186/1297-9686-41-35