

# Accuracy of genomic breeding values estimated from simulation of the dairy cattle population of northern Thailand

Sirijanya Aryuman<sup>1\*</sup>, Nattaphon Chongkasikit<sup>1</sup>, Chirawath Phatsara<sup>1</sup>, Saowaluck Yammuen-art<sup>1</sup>, Krasit Vasupen<sup>2</sup>, Chot Rachwicha<sup>1</sup> and Chalernporn Jirapanyalert<sup>3</sup>

<sup>1</sup>Department of Animal and Aquatic Sciences, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50100, Thailand

<sup>2</sup>Department of Animal Science, Faculty of Agricultural Innovation and Technology, Rajamangala University of Technology Isan, Nakhon Ratchasima 30210, Thailand

<sup>3</sup>Northern Institute of Vocational Education in Agriculture, Tak 63180, Thailand

\*Corresponding author: [Sirijanya\\_aryuman@cmu.ac.th](mailto:Sirijanya_aryuman@cmu.ac.th)

Received: November 5, 2023. Revised: December 19, 2023. Accepted: December 20, 2023.

## ABSTRACT

Accuracy is of significant importance in animal breeding, as it directly impacts the response to selection. The objectives of this study were to estimate the accuracy of genomic breeding values (GEBV) from a simulated population of the dairy cattle population of northern Thailand, which has exhibited an increasing trend from the past to the present. Data were simulated using a calibration set (CS) of 2,000 and 3,000 animals, heritability ( $h^2$ ) ranging from 0.05 to 0.50, and the number of SNPs at 20K and 40K. The GEBV was estimated using BLUP under animal model, and the accuracy was estimated by the correlation between GEBV and TBV from the simulation. The accuracy of GEBV ranged from 0.0870 to 0.8761. The CS of 3,000 animals was higher than the CS of 2,000 animals. Additionally, it was observed that the accuracy of the low  $h^2$  trait was unstable and lower than the high  $h^2$  trait, and the accuracy between 20K and 40K of SNPs was similar, with the highest values being 0.8761 and 0.8189, respectively. This study showed the CS of 3,000 animals and SNPs 40K would be appropriate for estimating GEBV.

**Keywords:** accuracy, dairy cattle, GEBV, calibration set, simulated

## INTRODUCTION

Dairy cattle breeding aims to improve the efficiency and productivity of dairy cattle by selecting animals with desirable traits. Estimated Breeding Value (EBV) was used to predict the genetic value of a trait in an individual animal based on the Best Linear Unbiased Prediction Method (BLUP; Henderson, 1975). Genomic Selection (GS; Meuwissen et al., 2001) was developed for use in dairy cattle breeding by using genetic markers distributed throughout the genome to estimate genomic breeding value (GEBV) (Guarini et al., 2019). The GEBV is the predicted genetic value of an individual animal based on genomic information and is calculated using statistical models that incorporate the animal's genotypic data as well as information on the inheritance and expression of the trait of interest (Hayes et al., 2010). The GEBV can be used in breeding programs to make selection decisions for individual animals. The animals with the highest GEBV for these traits can be selected as sires and dams for the next generation to improve the overall genetic potential of the population (de los Campos et al., 2013; García-Ruiz et al., 2022).

The accuracy of GEBV depends on many factors, such as the historical population, CS,  $h^2$  of the

trait, the number of single nucleotide polymorphisms (SNPs), linkage disequilibrium (LD), minor allele frequency (MAF), quantitative trait loci (QTL), and the relatedness between the CS and validation set (VS) (VanRaden et al., 2008; Goddard, 2009; Daetwyler et al., 2010; Wientjes et al., 2013). Corbin et al. (2010) reported that historical population data reveal genetic diversity and patterns of genetic variation in populations, which can help in selection planning. Hayes et al. (2009) and Hickey et al. (2011) reported that large CS led to greater genetic diversity. This leads to a more accurate assessment of the influence of markers, especially for traits with a complex genetic structure. Similarly, Meuwissen et al. (2016) reviewed the literature on GS in animal breeding and highlighted the importance of combining genotypic and phenotypic information to improve the accuracy of genomic predictions. The authors also emphasized the need to account for the genetic architecture of the trait being predicted and the effects of population structure and environmental factors. The group of animals consists of a group of animals with a known phenotype (calibration set; CS) and an unknown phenotype or progeny group (validation set; VS). Data simulation is a prevalent method. It can be used to study the methods for estimating GEBV, the factors affecting GEBV, and

the accuracy of GEBV in different populations. It can apply the results to real data by adjusting environmental factors or other fixed factors to obtain GEBV and accuracy. However, there are many factors affecting GEBV and accuracy estimation. The number of CS and SNPs is important for different populations and traits ( $h^2$ ). Therefore, the objective of this study was to estimate the accuracy and study the trend of changes in accuracy within the dairy cattle population of northern Thailand, considering different levels of CS,  $h^2$ , and SNPs. This study will provide information on the selection and planning of dairy cattle breeding for the future.

## MATERIALS AND METHODS

### Data Simulation

This study uses dairy cattle population data collected from the dairy cattle data of northern Thailand from 1964 to 2020 to be used as a historical population (location of Chiang Mai, the latitude and longitude were 18° N and 98° E, respectively). It simulated the data from two groups of animals: a reference group of animals with known phenotypes (calibration set; CS) of 2 levels, 2,000 and 3,000 animals, and a group of animals with unknown phenotypes (validation set; VS). In the VS group are the offspring born from the CS group. The QMSim package (Sargolzaei and Schenkel, 2009; Scheper, 2016; Mehdi and Flavio, 2019) was used to simulate all animals at two levels of SNPs, 20,000 (20K) and 40,000 (40K), at MAF of 0.05 and 10 levels of  $h^2$  of 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, and 0.50, as shown in Table 1. Phenotypic and genotypic data, SNPs, and true breeding value (TBV) from the data simulation were determined by the  $h^2$  of the trait without environmental factors or fixed effects.

### Estimation of Genomic Breeding Values (GEBV)

The SNPs obtained from the simulation are used to create the genomic relationship matrix (G matrix) to estimate the GEBV of the trait determined by  $h^2$  in the CS and VS using BLUP (Best Linear Unbiased Prediction) under the Animal model.

$$y = 1\mu + Zu + e$$

Where  $y$  = vector of observation

$\mu$  = the overall mean

$u$  = vector of random animal effects

$Z$  = design matrix relates records to random animal effects

$e$  = vector of random residual errors

Generally, the mixed model equation (MME) is as follows:

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

Where  $\alpha = \frac{\sigma_e^2}{\sigma_u^2}$

$G$  = genomic relationship matrix

**Table 1.** Base parameters used in the simulations

Parameters	
<b>Historical population</b>	
No. of animals in the 0 generation	1,252
No. of animals in the 10 <sup>th</sup> generation	35,844
No. of animals in the 20 <sup>th</sup> generation	70,724
<b>Current population</b>	
Replacement ratio for sires	0.50
Replacement ratio for dams	0.25
Criteria for selection/culling	Age
Heritability ( $h^2$ )	0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50
<b>Genome</b>	
No. of chromosome	30
number of markers (1 K = 1,000 SNPs)	20K and 40K
Minor allele frequency (MAF)	0.05

### Accuracy of Genomic Breeding Values (GEBV)

The accuracy of GEBV was calculated from the correlation between GEBV in the CS and VS and True Breeding Value (TBV) (Takeda et al., 2021).

## RESULTS

### Accuracy of genomic breeding values (GEBV)

This study showed the accuracy of GEBV under simulated data from the dairy cattle population

of northern Thailand, as shown in Table 2. It was found that the accuracy ranged from 0.0870 to 0.8761. The accuracy of CS and VS ranges from 0.0870 to 0.8761 and 0.1278 to 0.7015, respectively.

**Table 2.** Accuracy of genomic breeding values (GEBV)

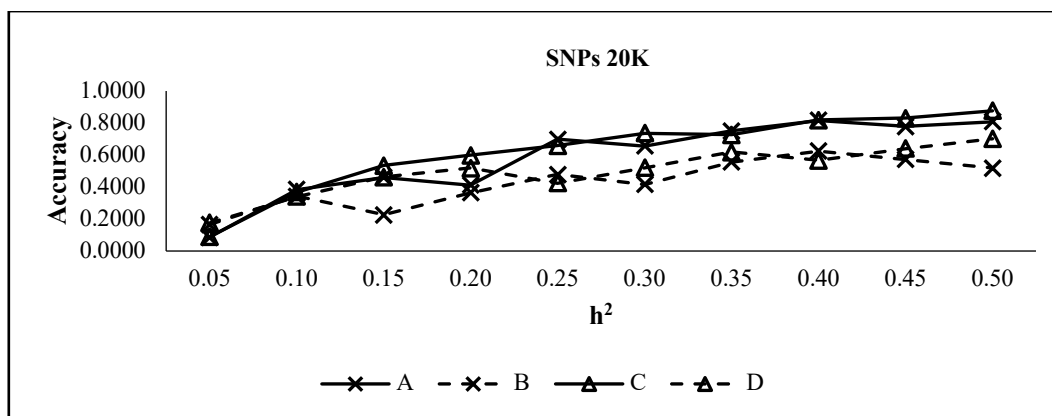
SNP	h <sup>2</sup>	Rep.	2,000 animals				3,000 animals			
			Calibration set		Validation set		Calibration set		Validation set	
			$\bar{X}$	SE	$\bar{X}$	SE	$\bar{X}$	SE	$\bar{X}$	SE
20K	0.05	10	0.0870	0.1064	0.1656	0.0669	0.0890	0.0259	0.1752	0.0304
	0.10	10	0.3849	0.0846	0.3426	0.0367	0.3676	0.0475	0.3408	0.0589
	0.15	10	0.4596	0.0323	0.2267	0.1726	0.5339	0.0794	0.4625	0.0713
	0.20	10	0.4116	0.1185	0.3637	0.0336	0.5986	0.0081	0.5181	0.0170
	0.25	10	0.6990	0.0647	0.4790	0.0665	0.6589	0.0248	0.4241	0.0620
	0.30	10	0.6560	0.0401	0.4169	0.0182	0.7356	0.0331	0.5200	0.0591
	0.35	10	0.7512	0.0379	0.5560	0.0351	0.7256	0.0685	0.6189	0.0438
	0.40	10	0.8154	0.0264	0.6238	0.0554	0.8185	0.0345	0.5659	0.0316
	0.45	10	0.7781	0.0530	0.5729	0.0956	0.8305	0.0269	0.6398	0.0491
	0.50	10	0.8084	0.0865	0.5179	0.2223	0.8761	0.0082	0.7015	0.0273
40K	0.05	10	0.1837	0.0889	0.1278	0.0652	0.1883	0.1051	0.2282	0.1275
	0.10	10	0.3480	0.1088	0.2924	0.0631	0.3014	0.0567	0.1917	0.0415
	0.15	10	0.3892	0.0799	0.3141	0.0296	0.6573	0.0398	0.4956	0.0371
	0.20	10	0.5796	0.0761	0.4036	0.0913	0.6808	0.0277	0.5532	0.0445
	0.25	10	0.6038	0.0761	0.4693	0.0428	0.7218	0.0333	0.5433	0.0455
	0.30	10	0.7029	0.0500	0.4407	0.1540	0.7579	0.0318	0.5765	0.0704
	0.35	10	0.7931	0.0165	0.5600	0.0368	0.8189	0.0092	0.6483	0.0271
	0.40	10	0.7460	0.0511	0.5963	0.0563	0.7589	0.0231	0.5660	0.0293
	0.45	10	0.8359	0.0179	0.6209	0.0300	0.8274	0.0125	0.6048	0.0337
	0.50	10	0.8306	0.0220	0.5761	0.0449	0.8039	0.0301	0.5805	0.0488

SNPs = single nucleotide polymorphism, h<sup>2</sup> = heritability, Rep. = replication.

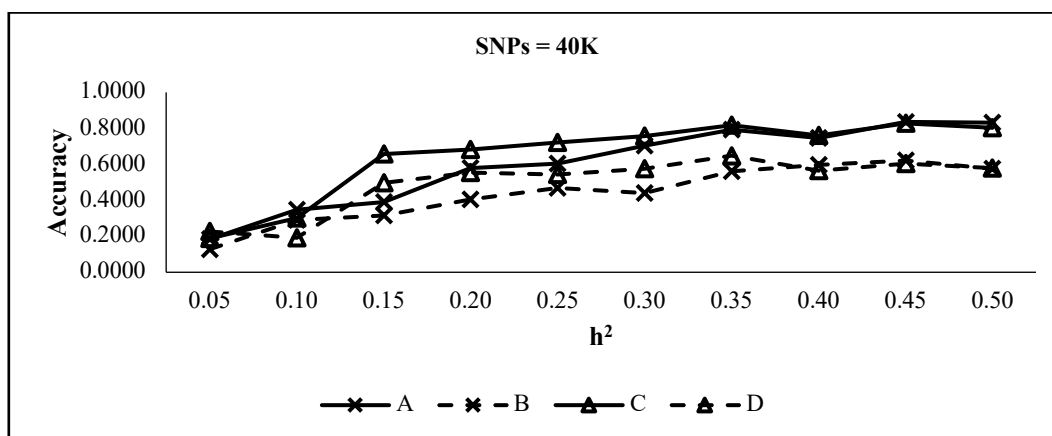
### Effect of Calibration set (CS) on the accuracy of genomic breeding values (GEBV)

The accuracy of GEBV under the CS in 2,000 and 3,000 animals with 20K SNPs was similar in both sizes. The accuracy of the CS of 3,000 animals and the VS from the CS of 3,000 animals had the highest values of 0.8761 and 0.7015, respectively. The accuracy of the CS of 2,000 animals and the VS from the CS of 2,000 animals had the highest values

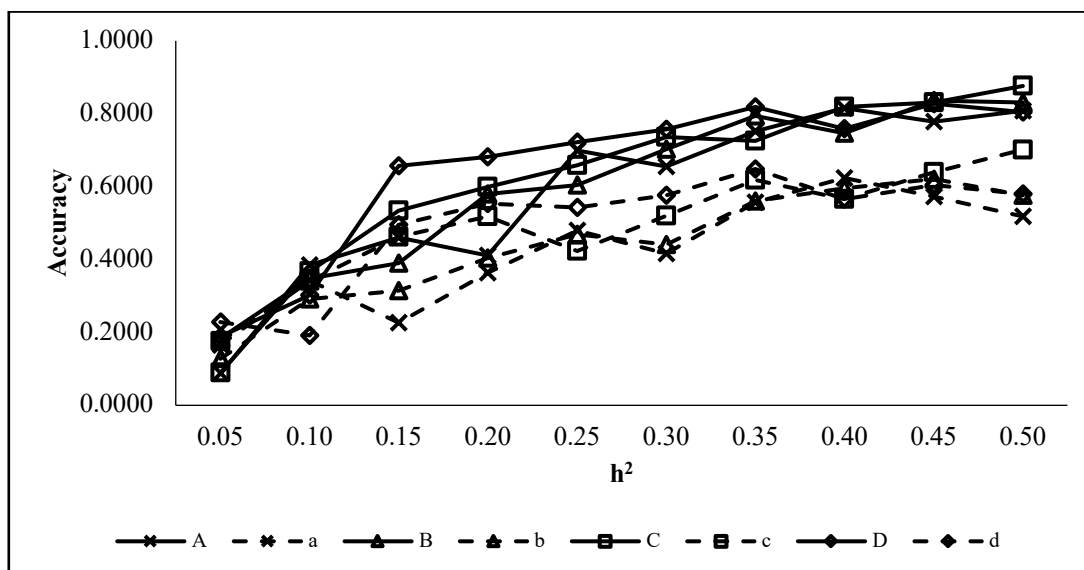
of 0.8154 and 0.6238, respectively (Figure 1). Similarly, with the number of SNPs at 40K, the accuracy for both sizes was similar. The accuracy of the CS of 3,000 animals and the VS from the CS of 3,000 animals had the highest values of 0.8189 and 0.6483, respectively. The CS of 2,000 animals and the VS from the CS of 2,000 animals had the highest values of 0.7931 and 0.5963, respectively, as shown in Figure 2.



**Figure 1.** Accuracy of genomic breeding values (GEBV) of CS of 2,000 and 3,000 animals, and number of SNPs 20K. (A = accuracy of CS of 2,000 animals, B = accuracy of VS from CS of 2,000 animals, C = accuracy of CS of 3,000 animals, D = accuracy of VS from CS of 3,000 animals).



**Figure 2.** Accuracy of genomic breeding value (GEBV) of CS of 2,000 and 3,000 animals and number of SNPs 40K. (A = accuracy of CS of 2,000 animals, B = accuracy of VS from CS of 2,000 animals, C = accuracy of CS of 3,000 animals, D = accuracy of VS from CS of 3,000 animals).



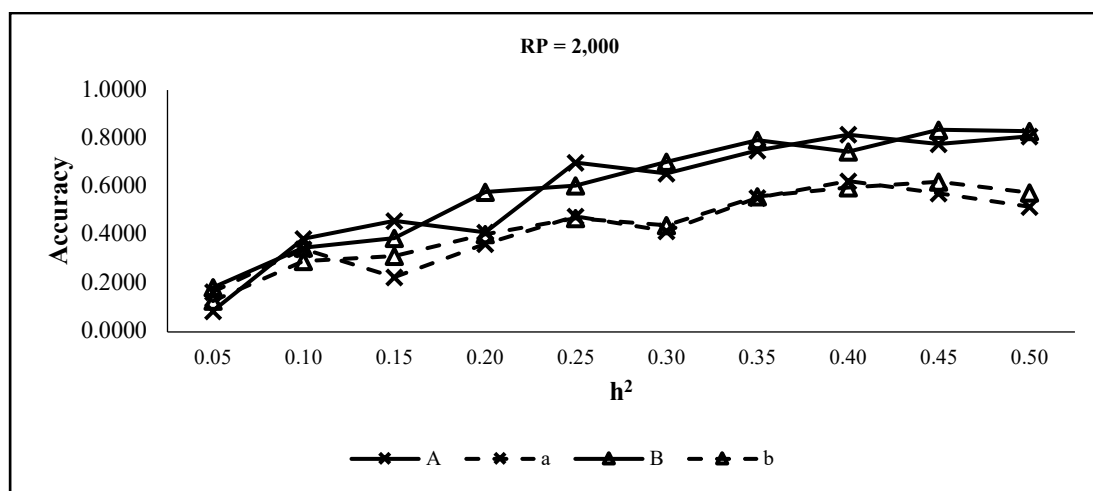
**Figure 3.** Accuracy of genomic breeding value (GEBV) at different levels of heritability (h²). (A = accuracy of CS of 2,000 animals and 20K SNPs, a = accuracy of VS from CS of 2,000 animals and 20K SNPs, B = accuracy of CS of 2,000 animals and 40K SNPs, b = accuracy of VS from CS of 2,000 animals and 40K SNPs, C = accuracy of CS of 3,000 animals and 20K SNPs, c = accuracy of VS from CS of 3,000 animals and 20K SNPs, D = accuracy of CS of 3,000 animals and 40K SNPs, d = accuracy of VS from CS of 3,000 animals and 40K SNPs).

**Effect of heritability ( $h^2$ ) on the accuracy of genomic breeding values (GEBV)**

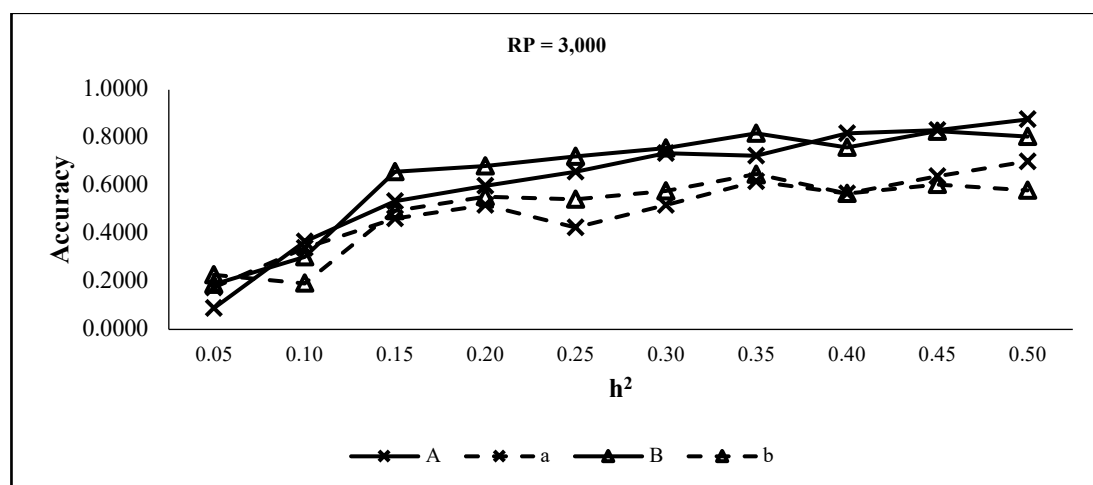
The effect of  $h^2$  on the accuracy of GEBV was found to increase and remain constant when  $h^2$  increased (Figure 3). The accuracy of GEBV in the CS of 3,000 animals and the VS from the CS of 3,000 animals were highest in the  $h^2$  at 0.45 and 0.35, respectively; and with a group of 2,000 animals, it had the highest  $h^2$  of 0.45 for both CS and VS.

**Effect of number of SNPs on the accuracy of genomic breeding values (GEBV)**

The effect of the number of SNPs on the accuracy of GEBV, it was found that the number of SNPs 20K and 40K in the CS of 2,000 animals had the highest values of 0.8154 and 0.7931, and in the VS from the CS of 2,000 animals, the values were 0.6238 and 0.5963, respectively. And, at the CS of 3,000 animals, the highest values were 0.8761 and 0.8189, and in the VS from the CS of 3,000 animals, the highest values were 0.7015 and 0.6483 at the number of SNPs 20K and 40K, respectively (Figure 4 and 5).



**Figure 4.** Accuracy of GEBV at the number 20K and 40K of SNPs of CS of 2,000 animals. (A = accuracy of CS 2,000 animals on 20K SNPs, a = accuracy of VS from CS 2,000 animals on 20K SNPs, B = accuracy of CS 2,000 animals on 40K SNPs, b = accuracy of VS from CS 2,000 animals on 40K SNPs).



**Figure 5.** Accuracy of GEBV at the number 20K and 40K of SNPs in 3,000 CS. (A = accuracy of CS 2,000 animals on 20K SNPs, a = accuracy of VS from CS 2,000 animals on 20K SNPs, B = accuracy of CS 2,000 animals on 40K SNPs, b = accuracy of VS from CS 2,000 animals on 40K SNPs).

### The standard error (SE) of accuracy

The standard error (SE) from simulation found that the SE values of the CS of 2,000 animals with SNPs 20K and 40K ranged from 0.1064 to 0.0865 and 0.0889 to 0.0220, respectively. The CS of 3,000 animals with SNPs 20K and 40K ranged from 0.0259 to 0.0082 and 0.1051 to 0.0301, respectively, as shown in Table 2.

## DISCUSSION

A study of the historical population of the dairy cattle population of Northern Thailand from 1964 to 2020 found that the size of the dairy cattle population has increased. This is consistent with a study by Gerdsook (2016), who reported that the dairy cattle population of Northern Thailand from 1952 to 2014 has increased and is on the trend of increasing. Numerous studies have reported the influence of historical populations on accuracy. For example, a study by Nwogwugwu et al. (2020) reported the accuracy of a cattle population in Korea by simulating the historical population with a constant size of 1,000 animals up to a generation of 1,000, which decreased to 200 animals in the next 95 generations. The accuracy was found to be 0.563, 0.735, and 0.808, with  $h^2$  values of 0.1, 0.3, and 0.5, respectively. Consistent with the Atefi et al. (2016) study, the historical population was simulated. The size was constant at 1,000 animals until 100 generations and gradually decreased until the remaining 500 animals in the 200<sup>th</sup> generation. The accuracy was found to be 0.57, 0.70, and 0.77, with  $h^2$  0.15, 0.30, and 0.45, respectively. The large historical population sizes indicate greater genetic diversity, which is helpful for accurately estimating genetic parameters and genetic markers associated with various traits. However, other factors, such as the quality and quantity of phenotypic data, the genetic structure of the trait, and the statistical methods used in the study, should also be taken into account to influence the accuracy of the assessment of GEBV.

The accuracy of GEBV by using simulation and data from the dairy cattle population of northern Thailand. The results showed that the accuracy of the CS was higher than that of the VS. Consistent with Bouwman et al. (2014), they estimated the accuracy of GEBV from simulations using genotyping data from relatives. They found that using genotype data from relatives and offspring can increase accuracy, especially in the VS. Similar to Calus et al. (2013), they used genomic technology to estimate genetics and evaluate the accuracy of the VS by predicting from the CS data that it has good genetic and

phenotypic information. The results showed highly accurate breeding values. Boison et al. (2017) evaluated the accuracy of GEBV in Gyr (*Bos indicus*) dairy cattle using different VS sizes. It was found that the accuracy of the production traits ranged from 0.28 to 0.49. However, the VS is a population that does not have direct phenotypic data but is derived from predicting animal data in the population with known phenotypes. As a result, the accuracy is lower than that of animals with known phenotypes.

When considering the effect of CS on the accuracy of GEBV, it was found that increasing the CS can improve accuracy. Several studies have reported that CS affects accuracy. The accuracy of GEBV is higher as the CS increases. Increasing the CS will increase genetic diversity, resulting in high accuracy. However, an appropriate CS must be considered due to the high cost of genome selection. This study shows that the CS of 2,000 animals is sufficient to achieve an accuracy of more than 0.4 for the dairy cattle population of northern Thailand with  $h^2$  of 0.15. Similarly, Takeda et al. (2021) reported that the CS of 4,000 animals was sufficient to achieve a GEBV accuracy of carcass traits in Japanese black cattle of more than 0.4. Zhang et al. (2023) reported that using the CS of 6,000 animals could increase the accuracy of GEBV for the production traits of Chinese Holstein cattle by greater than 0.40. It can be seen that the accuracy of GEBV is different when using different CS. This is because animals in each population have different relationships. Falconer et al. (1997) reported that increasing the number of offspring (progeny) per sire in the VS can reduce differences in accuracy estimation methods. However, the CS studied this time was similar, resulting in similar accuracy values. In the next study, a larger CS should be added to observe the accuracy trend further.

It was considering the influence of different levels of  $h^2$  on the accuracy of GEBV. At low  $h^2$ , the accuracy was relatively low. But while the  $h^2$  is high, the accuracy of GEBV tends to increase. This is consistent with the study by Buaban et al. (2021), which reported that the accuracy values of milk yield traits of dairy cattle in Thailand ranged from 0.27 to 0.37. Similarly, Yan et al. (2022) reported that the accuracy ( $r$ ) of GEBV by simulation in goats with  $h^2$  values of 0.11 and 0.34 was 0.465 and 0.604, respectively. While Togashi et al. (2019) reported the reliability values ( $r^2$ ) of the dairy cow population with  $h^2$  of 0.1, 0.3, and 0.5, which were equal to 0.0769, 0.2000, and 0.2941, respectively, the accuracy values were lower than this study. However, it can be seen that when the  $h^2$  value increases, the accuracy tends to increase. This is because accuracy and

$h^2$  are important factors affecting the response to selection. Traits with high  $h^2$  are more responsive to selection than those with low  $h^2$ , resulting in rapid genetic progress.

The effect of SNPs of 20K and 40K on the accuracy of GEBV, it was found that the accuracy values for both CS estimated from 40K SNPs were only slightly higher than those from 20K SNPs. The more informative SNPs, the higher the accuracy of GEBVs. However, using a larger number of SNPs can improve the accuracy of GEBV, but there may be little or no increase. Consistent with the study of Hayes et al. (2019), they reported that increasing the number of SNPs from 24K to 728K resulted in a slight increase in the accuracy of GEBV in three cattle breeds, with values of 0.22 and 0.24, respectively. They also reported that using many SNPs results in higher breeding costs. However, some studies reported that using a high-density SNP resulted in higher GEBV accuracy than using a low-density SNP (Bolormaa et al., 2015). This study showed that increasing the accuracy of GEBV in the dairy cattle population in northern Thailand can use low-density SNPs for selection. Using low-density SNPs results in lower selection costs. Consistent with the study of Cole et al. (2016), they studied the efficiency of genomic selection in a multi-breed dairy population using 777,962 SNPs. The study found that 50K SNPs were sufficient to increase the accuracy of genetic selection. Using a larger number of SNPs did not result in higher accuracy. Lopes et al. (2020) evaluated the accuracy of GEBV in US Holstein cattle using 777,962 SNPs. Increasing the number of SNPs improves the accuracy of GEBV. However, the accuracy tends to increase only slightly when the number of SNPs increases over 50K.

In the study of the standard error (SE) of the accuracy of GEBV, it was found that increasing the CS will result in a lower SE, especially for traits with low  $h^2$ . Consistent with the study of Gerdsook (2016), the SE decreased from 0.10 to 0.06 for the trait with  $h^2$  of 0.10 when the CS increased from 100 to 1,500 animals. Similarly, Yan et al. (2022) reported the SE of the trait with a  $h^2$  of 0.11. The SE decreased from 0.0862 to 0.0210 when the CS increased from 500 to 3,000 animals.

Each research (historical population) produced different results, as evidenced by the above information. Therefore, further research is important by using other factors and different levels of CS and SNPs for more results.

## CONCLUSIONS

Accuracy of GEBV from simulation using dairy cattle population data from northern Thailand. It was found that the accuracy of the CS tended to be higher than that of the VS. In contrast, the accuracy of the CS of 3,000 animals and the VS from the CS of 3,000 animals tended to be slightly higher than that of the CS of 2,000 animals and the VS from the CS of 2,000 animals. In addition, we found that the accuracy of traits with low  $h^2$  tended to be unstable. This is because the SE is higher than traits with a high  $h^2$ . Moreover, traits with low  $h^2$  have low accuracy, and SE is relatively high. While traits with high  $h^2$  have high accuracy and relatively low SE, the accuracy of traits with low  $h^2$  can be increased by increasing CS. Moreover, it was found that the accuracy for the number of SNPs 20K and 40K was similar, but the SE of the number of SNPs 40K was lower than 20K. From the results of the study, it was found that using the CS of 3,000 animals and SNPs 40K was appropriate for estimating GEBV in the dairy cattle population in northern Thailand. This is because the accuracy of GEBV is high, and the SE is low. Additionally, for low  $h^2$  traits, increasing the CS size can reduce the SE of the accuracy.

## REFERENCES

- Atefi, A., Shadparvar, A.A., and Hossein-Zadeh, N.G. 2016. Comparison of whole genome prediction accuracy across generations using parametric and semi parametric methods. *Acta Scientiarum*. 38: 447–453.
- Boison, S.A., Utsunomiya, A.T.H., Santos, D.J.A., Neves, H.H.R., Carvalheiro, R., Mészáros, G., Utsunomiya, Y.T., do Carmo, A.S., Verneque, R.S., Machado, M.A., Panetto, J.C.C., Garcia, J.F., Sölkner, J., and da Silva, M.V.G.B. 2017. Accuracy of genomic predictions in Gyr (*Bos indicus*) dairy cattle. *J. Dairy Sci.* 100: 5479–5490.
- Bolormaa, S., Pryce, J.E., Zhang, Y., Reverter, A., Barendse, W., Hayes, B.J., and Goddard, M.E. 2015. Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. *Genetics Selection Evolution*. 47: 1–12.
- Bouwman, A.C., Hickey, J.M., Calus, M.P.L., and Veerkamp, F.R. 2014. Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle. *Genetics Selection Evaluation*. 46: 1–11.
- Buaban, S., Prempre, S., Sumreddee, P., Duangjinda, M., and Masuda, Y. 2021. Genomic prediction of milk production traits and somatic cell score using single-step genomic best linear unbiased predictor with random regression test-day model in Thai dairy cattle. *J. Dairy Sci.* 104: 12713–12723.
- Calus, M.P.L., Haas, Y.D., Pszczola, M., and Veerkamp, R.F. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. *Anim.* 7: 183–191.
- Cole, J.B., and da Silva, M.V.G.B. 2016. Genomic selection in multi-breed dairy cattle populations. *R. Bras. Zootec.* 45: 195–202.

- Corbin, L.J., Blott, S.C., Swinburne, J.E., Vaudin, M., Bishop, S.C., and Woolliams, J.A. 2010. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* 41: 8–15.
- Daetwyler, H.D., Pong-Wong, R., Villanueva, B., and Woolliams, J.A. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics.* 185: 1021–1031.
- 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.
- Falconer, D.S., and Mackay, T.C. 1997. *Introduction to quantitative genetics*, 4<sup>th</sup> ed. John Wiley & Sons, Hoboken, NJ.
- García-Ruiz, I., Quiñones, A., and Taborsky, M. 2022. The evolution of cooperative breeding by direct and indirect fitness effects. *Science Advance.* 8: 1–10.
- Gerdsook, N. 2016. Relationship between reference population size and accuracy of genomic breeding values of dairy cattle population in upper northern Thailand. Master's thesis, Chiang Mai University, Chiang Mai, Thailand.
- Goddard, M.E. 2009. Genomic selection: prediction of accuracy and maximization of long term response. *Genetica.* 136: 245–257.
- Guarini, A.R., Lourenco, D.A.L., Brito, L.F., Sargolzaei, M., Baes, C.F., Miglior, F., Tsuruta, S., Misztal, I., and Schenkel, F.S. 2019. Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. *J. Dairy Sci.* 102: 8175–8183.
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J., and Goddard, M.E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92: 433–443.
- Hayes, B.J., and Goddard, M.E. 2010. Genome-wide association and genomic selection in animal breeding. *Genome.* 53: 876–883.
- Hayes, B.J., Corbet, N.J., Allen, J.M., Laing, A.R., Fordyce, G., Lyons, R., McGowan, M.R., and Burns, B.M. 2019. Towards multi-breed genomic evaluations for female fertility of tropical beef cattle. *J. Anim. Sci.* 97: 55–62.
- Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics.* 31: 423–447.
- Hickey, J.M., Kinghorn, B.P., Tier, B., and Wilson, J.F. 2011. The impact of genetic architecture on genome-wide evaluation methods. *Genetics Research.* 93: 343–355.
- Lopes, M.S., Silva, F.F., Meuwissen, T.H., and Guimarães, S.E. 2020. Optimal number of SNP markers for genomic prediction under the genetic architecture of complex traits. *Bioinformatics.* 21: 1–17.
- Mehdi, S., and Flavio, S.S. 2019. Qmsim version 1.10. User's guide. Centre for Genetic Improvement of Livestock Department of Animal and Poultry Science University of Guelph, Canada.
- 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.
- Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. 2016. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers.* 6: 6–14.
- Nwogwugwu, C.P., Kim, Y., Choi, H., Lee, J.H., and Lee, S. 2020. Assessment of genomic prediction accuracy using different selection and evaluation approaches in a simulated Korean beef cattle population. *Asian-Australas J Anim Sci.* 33: 1912–1921.
- Sargolzaei, M. and Schenkel, F.S. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics.* 25: 680–681.
- Scheper, C. 2016. Technical note for QUALSim (usage in combination with QMSim). Department of Animal Breeding, University of Kassel, Germany.
- Takeda, M., Inoue, K., Oyama, H., Uchiyama, K., Yoshinari, K., Sasago, N., Kojima, T., Kashima, M., Suzuki, H., Kamata, T., Kumagai, M., Takasugi, W., Aonuma, T., Soma, Y., Konno, S., Saito, T., Ishida, M., Muraki, E., Inoue, Y., Takayama, M., Nariai, S., Hideshima, R., Nakamura, R., Nishikawa, S., and Uemoto, Y. 2021. Exploring the size of reference population for expected accuracy of genomic prediction using simulated and real data in Japanese Black cattle. *BMC Genomics.* 22: 1–11.
- Togashi, K., Adachi, K., Kurogi, K., Yasumori, T., Tokunaka, K., Ogino, A., Miyazaki, Y., Watanabe, T., Takahashi, T., and Moribe, K. 2019. Effects of preselection of genotyped animals on reliability and bias of genomic prediction in dairy cattle. *Asian-Australas J Anim Sci.* 32: 159–169.
- VanRaden, P.M., VanTassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., and Schenkel, F.S. 2008. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92: 16–24.
- Wientjes, Y.C.J., Veerkamp, R.F. and Calus, M.P.L. 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics.* 193: 621–631.
- Yan, X., Zhang, T., Liu, L., Yu, Y., Yang, G., Han, Y., Gong, G., Wang, F., Zhang, L., Li, W., Yan, X., Mao, H., Li, Y., Du, C., Li, J., Zhang, Y., Wang, R., Lv, Q., Wang, Z., Zhang, J., Liu, Z., Wang, Z., and Su, R. 2022. Accuracy of genomic selection for important economic traits of cashmere and meat goats assessed by simulation study. *Frontiers in Veterinary Science.* 9: 1–13.
- Zhang, Z., Shi, S., Zhang, Q., Aamand, G.P., Lund, M.S., Su, G., and Ding, X. 2023. Improving genomic prediction accuracy in the Chinese Holstein population by combining with the Nordic Holstein reference population. *Animals.* 13: 1–12.