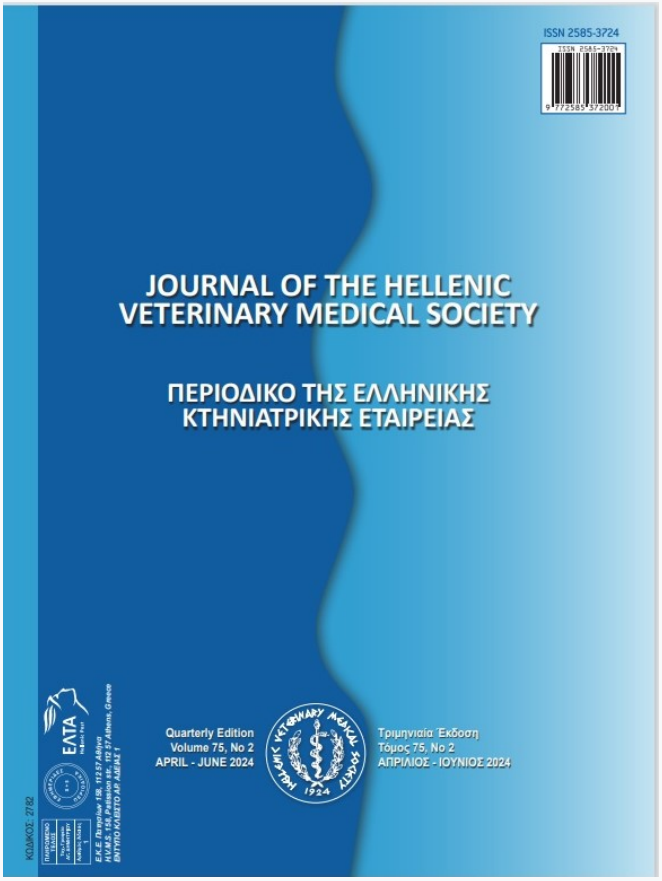
Journal of the Hellenic Veterinary Medical Society

Vol 75, No 2 (2024)



Use genetic modification indicator to increase the precision and accuracy of genomic analysis

M Safarzadeh, MH Kalani, M Nezafat , M Ouni, P Ghadermarzi, H Aghazadeh, S Rajabi; J Pish Jang Aghajeri

doi: 10.12681/jhvms.35220

To cite this article:

Safarzadeh, M., Kalani, M., Nezafat , M., Ouni, M., Ghadermarzi, P., Aghazadeh, H., Rajabi, S., & Pish Jang Aghajeri, J. (2024). Use genetic modification indicator to increase the precision and accuracy of genomic analysis. *Journal of the Hellenic Veterinary Medical Society*, *75*(2), 7537–7550. https://doi.org/10.12681/jhvms.35220 (Original work published July 16, 2024)

J HELLENIC VET MED SOC 2024, 75 (2): 7531-7544
ΠΕΚΕ 2024, 75 (2): 7531-7544

Research article
Ερευνητικό άρθρο

Use genetic modification indicator to increase the precision and accuracy of genomic analysis

M. Safarzadeh[1], M.H. Kalani[2], M. Nezafat[3], M. Ouni[4], P. Ghadermarzi[5], H. Aghazadeh[6,7], S. Rajabi[8*], J. Pish Jang Aghajeri[9*]

[1]*Department of Medicine, Tabriz Branch, Islamic Azad University, Tabriz, Iran*

[2]*Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran*

[3]*Department of Biology, Central Tehran Branch, Islamic Azad University, Tehran, Iran*

[4]*Department of Chemical and Polymer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran*

[5]*Department of New Technologies, University of Zabol, Zabol, Iran*

[6]*Tabriz University of Medical Sciences, Tabriz, Iran*

[7]*Young Researchers and Elites Club, Islamic Azad University of Tabriz Branch, Tabriz, Iran*

[8]*Department of Natural Sciences, University of Tabriz, Tabriz, Iran*

[9]*Department of Animal Science, Maragheh Branch, Islamic Azad University, Maragheh, Iran*

ABSTRACT: The high expense of genotype determination and the low accuracy of the evaluation in a small number of genotyped samples when using genomic data and genetic markers to perform genomic evaluations are two major problems. The effect of levels of genetic markers in an F2 population obtained from a Two-way crossing of native Iran chicken with a low growth rate and Arian meat strain with a high growth rate was investigated in this study to isolate SNPs with higher effect and use these markers in genomic evaluation as a suitable method of screening SNPs to increase the accuracy of the evaluation and reduce genotyping costs. In this study, we examined the prediction accuracy of correction values in five marker groups with various minor allele frequencies (MAFs). We also demonstrated the superiority of the ssGBLUP approach over the BLUP method using the 5-fold cross-validation (CV) method in the single-step assessment strategy. We found that markers with an MAF of 0.4-0.5 were the most suitable for genomic evaluations of the growth trait. Specifically, using SNPs with an MAF of 0.4-0.5 in the second to seventh weeks resulted in higher predictive accuracy compared to using information from all SNPs. Additionally, using SNPs with an MAF of 0.4-0.5 and developing low-density SNP chips with markers possessing these properties can be a reliable method for evaluating individuals based on genetic merit.

***Keywords*:** Genomic evaluation; Genetic markers; BLUP

Corresponding Authors:
Jafar Pish Jang Aghajeri, Department of Animal Sciences, Maragheh Branch, Islamic Azad University, Maragheh, Iran
E-mail address: ja.pja@iau.ac.ir & j.pishjang@gmail.com

Soroush Rajabi
Department of Natural Sciences, University of Tabriz, Tabriz, Iran
E-mail address: Rajabisoroush3@gmail.com

Date of initial submission: 11-08-2023
Date of acceptance: 21-11-2023

INTRODUCTION

The high cost of genotype determination and the poor evaluation accuracy in a small number of samples are challenges when using genomic data and genetic markers to conduct genomic evaluations at the farm level. In comparison to a kinship matrix based purely on pedigree information, accuracy predictions employing a mixed matrix of pedigree information and genomic information are predicted to perform better. The successful genetic enhancement of important features in cattle and poultry species can be achieved by developing an accurate and unbiased genomic prediction system (Mrode et al., 2018).

Numerous genomic evaluation studies in cattle and poultry species have examined and compared various approaches to genomic prediction (Cardoso-Silva et al., 2014; Neves et al., 2014; Costa et al., 2019). For example, study 6 confirmed that the ssGBLUP methods improved prediction accuracy for residual feed intake (RFI) and feed conversion ratio (FCR) in Nellore cattle compared to BayesCπ and GBLUP techniques. The ssGBLUP methods are based on the infinitesimal model of polygenic trait control (Karaman et al., 2018; Karaman et al., 2016), which assumes a common effect and variance across all indicators. The ssGBLUP method is considered superior to traditional pedigree-based BLUP and GBLUP methods (Gao et al., 2012; Koivula et al., 2015; Christensen et al., 2012; Li et al., 2014; Guo et al., 2015; Song et al., 2017) because it incorporates both genomic and pedigree information.

The prediction accuracy of the BLUP, GBLUP, BayesC, ssGBLUP, and BayesCπ algorithms was assessed in an average genotyped size of the Canadian pig population in a study by Salek Ardestani et al., (2021) to compare their predictive capacities. The ssGBLUP approach had the highest prediction accuracy in the majority of the study's situations (Salek et al., 2021). Additionally, Yan et al., (2018) revealed that ssGBLUP EBVs were more accurate and had less skewness than BLUP by investigating a pure line of laying hens.

Using the 5-fold cross-validation (CV) method, the prediction accuracy obtained from a traditional pedigree-based method (BLUP), a genomic BLUP (GBLUP), and a One-step genomic BLUP (ssGBLUP) was compared in a study that was conducted for the genomic evaluation of traits related to body weight in a Yorkshire population consisting of 592 pigs by the Illumina PorcineSNP80 panel. As a whole, the GBLUP approach has worse prediction accuracy than the BLUP and ssGBLUP methods, according to the study's findings. For single-step techniques, ssGBLUP's accuracy for features associated with body weight in pigs ranged from 0.54 to 0.78 and its regression coefficients ranged from 0.83 to 1.02 respectively. The prediction accuracy of ssGBLUP in this investigation, which involved a population of pigs, was about 1% greater than that of classical BLUP. Because of the few genotyped animals and shallow pedigrees, the improvement in ssGBLUP's prediction accuracy was smaller than anticipated. The findings demonstrated that the prediction accuracy of GBLUP was low in comparison to other approaches with various genotyped reference population sizes and that this accuracy increased with the genotyped reference population size (Song et al., 2018).

Using a 50K chip, Rolf et al. revealed that a G-matrix and precise EBV forecasts for feed yield can be created using chips with 2,500 to 10,000 SNP markers dispersed across the genome. Angus cattle can utilize it (Rolf et al., 2010). Salvian et al., (2020) study demonstrated that at least 10% of high-density chip SNPs can be exploited for genomic evaluation. Additionally, a study on the population of dairy cows revealed that there were no appreciable changes in the precision of genomic prediction when the number of markers was increased from a chip with medium density (54,000K) to a chip with high density (777,000) (Su et al., 2012). In addition to the aforementioned, decreasing SNP density has been widely reported in numerous research (Habier et al., 2009; Wellmann et al., 2013) as a way to lower genotyping costs.

MATERIALS AND METHODS

Study population

The fast-growing Arian meat line and the local breed of Iran's East Azarbaijan region (Tabriz) were crossed bilaterally to create the F2 population used in this study. In February 2018, the parents of these hens were moved to Tabriz University of Medical Sciences, where they were later crossed. These crosses produced the F1 population, and the mating of the F1s produced the F2 population of chickens. For 11 weeks, these chickens were grown in identical environmental and dietary conditions in separate cages with access to water and food. F1 and F2 birds were not given any antibiotics or immunizations throughout the breeding season. However, the required steps were made to provide a clean and healthy environment in the hall.

Genotypic data

The salting out technique was used to extract DNA from the blood samples of 312 birds from the F2 generation. The commercial Illumina 60k SNP Chip (Illumina 60k SNP Chip) donated by Cobb and Aarhus University in Denmark was used to genotype blood samples (Zhang et al., 2014). Each sample was genotyped using 55.329 SNP markers. After initial editing, which included removing duplicate SNPs and SNPs whose chromosome numbers were unknown, 54.338 SNP markers were left in the genetic map.

Phenotypic data

The fast-growing commercial line Arian (A) and the slow-growing native chickens of Tabriz, Iran (N) were crossed to produce an F2 population. F1 chicks were born as a result of the mating of ♀A ♂ × N and ♀ N ♂ × A birds. Four to eight females from different families were mated with each F1 male from the cross. Finally, 5 separate hatches resulted in the production of 488 F2 chickens, including 312 genotyped and 176 genotype-free birds, from 8 Tatney families. F2 chicks were weighed at one day old, and they were raised outside for seven days at a temperature of 33°C with 24-hour exposure. On the sixth day, this temperature dropped to 30 degrees Celsius. The birds were weighed on the seventh day and moved to separate cages with temperatures of 30 °C, which gradually dropped until the final temperature reached 22 °C. During the experimental period, the exposure cycle was 22 hours of light and 2 hours of darkness. Vaccinations weren't given to chickens while they were being raised. Chickens were given unlimited access to food and water, and throughout various weeks, variations in body weight were monitored and examined.

Bioinformatics software and operating system

Windows and Linux operating systems were used for this research. Also, various stages of evaluation were done with the help of PLINK, BLUPF90, GCTA, GAPIT, TASSEL, R, RRBLUP, ASREML, MIXBLUP, STRUCTURE, and other related software.

Data quality control

Type 1 and type 2 errors in the test findings are eventually caused by the biases that are introduced during the experiment's design as well as the faults of the genotyping procedure, which result in systemic errors in genomic investigations. Therefore, by using proper sampling and precise laboratory techniques, many of these inaccuracies can be decreased. Even if the experiment is properly designed, it is still necessary to thoroughly review the data quality, and identify and exclude any markers or samples that do not meet the required standards. Type I and II mistakes are reduced when these samples and markers are removed (Anderson et al., 2010). Using the PLINK software version 9.1 (Zhernakova et al., 2011), a data quality check was carried out. Samples with genotyping rates lower than 90%, minor allele frequencies (MAF) lower than 0.05%, Hardy-Weinberg equilibrium test p-values lower than 6-10, and markers in which fewer than 5% of samples were genotyped for the target marker (geno>0.5) were all disregarded from the study.

Samples with a call rate of less than 90% indicate that less than 90% of the markers have been genotyped for the intended bird; samples with a geno of more than 5% indicate that more samples than 5% have not been genotyped for the intended marker. The frequency of the minor marker allele is also known as the MAF rate. In actuality, the minor allele of the marker is an allele that is less common in society than the other allele, called the major. Hardy-Weinberg equilibrium was used to remove markers whose observed and predicted heterozygous genotype frequencies differed by more than 0.15. Excessive heterozygosity in markers can lead to type 1 and type 2 mistakes in genomic investigations. The population's selection process or genotyping errors may be to blame for this disparity of more than 0.15 (Clayton et al., 2005; Salanti et al., 2005). After performing data quality checks, 308 samples (140 male and 138 female birds) and 48379 SNP markers were chosen for additional analysis. Four samples with low genotyping rates were eliminated from the research. The synbreed program (Papanicolaou et al., 2016) was used to assess the distribution of SNPs before and after quality control as well as the average distance between adjacent SNPs in each chromosome. The results are shown in Table 1.

Examining the genetic structure of the population

The identification and consideration of population structure is a crucial aspect of genetic analysis. For clarity in watching the findings up till the results are appropriately controlled, population structure is classified as a genetic influence. Three techniques are frequently employed to mitigate the impacts of population structure in genomic investigations. These techniques include principal component analysis (PCA), structural association (SA), and genomic control (GC) (Zhang et al., 2015). That the PCA meth-

Table 1: Distribution of SNPs before and after quality control and average distance between adjacent SNPs in each chromosome

Chromosome	No. of SNP Markers after quality control	No. of SNP in chip	Average distance (kb)
1	7546	8303	26.5
2	5762	6355	26.7
3	4340	4739	26.3
4	3553	3872	26.5
5	2303	2542	27.1
6	1815	1995	19.6
7	1907	2089	20.1
8	1502	1636	20.1
9	1269	1366	18.8
10	1378	1553	16.1
11	1329	1531	16.4
12	1356	1559	14.4
13	1251	1371	14.6
14	1081	1179	14.3
15	1094	1222	11.8
16	20	24	21.7
17	898	994	11.8
18	930	1048	11.9
19	878	973	11.3
20	1587	1815	8.8
21	805	901	8.5
22	313	432	12.6
23	631	724	9.3
24	763	853	8.5
25	177	211	11.5
26	685	776	7.4
27	518	576	9.4
28	582	708	7.6
29	118	142	7.7
30	4	7	6.9
Z	1984	2842	37.5
Total	48379	54338	15.8

od will lessen the effects of population classification (Gu et al., 2011). Drawing two-dimensional graphs from two components of MDS (Yamaguchi-Kabata et al., 2008) or PCA (Lee et al., 2012) can demonstrate population structure based on genetic markers. In this study, MDS analysis was used to determine and assess the genetic distance, where each sample is placed in a graph, and compared to the results of PCA and MDS, which are generally similar [Wang et al., 2012]. The real genetic distance between the samples was revealed after population classification (Wang et al., 2012).

In addition to multidimensional scaling (MDS) analysis (Zhang et al., 2015), the method Neighbor-joining tree (Wang et al., 2020) was also used to achieve the precise structure of the F2 chicken population and understand the relationship between and within populations at the genome level. PLINK software (version 1.09) was used to assess population structure using MDS analysis (Zhang et al., 2015). As recommended by Wang et al. (2009), independent SNP markers were obtained for all autosomes using the independence-pairwise option, with a window size of 30 SNP markers and a threshold of r2 = 0.2. All individuals' identity distances or (identity-by-state) pairs were calculated using independent SNP markers. Additionally, MDS components were obtained using the IBS matrix-based MDS-plot option (Sun et al., 2013). The TASSEL software was used to perform a neighbor-joining tree-based cluster analysis of all genotypes based on genetic distance.

Formation of SNP subgroups with different MAFs

To study the relationship between allelic frequency and prediction ability, using Plink software, GCTA

(Chang et al., 2015), R3.2.2, and Python, the number of 48379 SNP markers that passed the quality control stage were divided and separated into 5 different MAF subgroups (MAF 0.05-0.1 with 6731 markers, MAF 0.1-0.2 with 8884 markers, MAF 0.2-0.3 with 10148 markers, MAF 0.3-0.4 with 11128 markers, and MAF 0.4-0.5 with 11488 markers) (Yang et al., 2013).

Statistical analysis

The AIREMLF90 (v1.61) module of the Blupf90 program was used to estimate the breeding values of each animal using Model 1 (Misztal et al., 2002).

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

where y is the adjusted phenotype vector, is the overall mean, X is the incidence matrix of fixed effects (sex, hatch, population structure, and age), b is the vector of fixed effects, Z is the incidence matrix of random effects to relate phenotypes to additive genetic effects, a is a vector of additive genetic effects assuming N (0, Aa2) distribution, A is the pedigree-based relationship matrix, a2 is the variance of additive genetic effects, and e is the vector of random residual effects with N (0, Ie2) distribution, where I is the identification matrix and e2 is the residual variance.

Additionally, the total of the EBV values and the remaining animals were used to obtain the corrected phenotypes (Misztal et al., 2020). Using the AIREMLf90 and Pridictf90 modules of the Blupf90 program and fitting the raw phenotypic in model 1, EBV and residual were calculated for each animal.

Model 2 to estimate one-step genomic modification values using AIREMLF90 (v1.61) (VanRaden et al., 2007) with all SNPs (48379 SNPs), as well as 5 subsets of SNPs with different MAFs (0.05-0.1) (6731 SNP markers), 0.0-1.2 (8884 SNP markers), 0.2-0.3 (10148 SNP markers), 0.0-3.4 (11128 SNP markers), and -0.5- 0.4 (11488 SNP markers), were used:

$$y = 1\mu + Xb + Zg + e \quad (2)$$

That y, X, b, and e are similar to model 1. Z is a design matrix for random additive genetic effects, and g is a vector of random additive genetic effects assuming N (0, H-g2), where H is a combination of the genomic relationship matrix (G) and the pedigree-based relationship matrix (A). Is. The inverse of the H matrix used in this study was created as follows:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & t(\alpha G+\beta A_{22})^{-1}-\omega A_{22}^{-1} \end{bmatrix} \quad (3)$$

Where the matrix's subset A22 is AIREMLF90 (v1.61)'s scaling factors, t and, are both set to one by default and A stands for genotyped animals. The combined variables were set to 0.95 and 0.05, respectively, to prevent singularity issues and enhance predictions (VanRaden et al., 2007; Lourenco et al., 2014).

The correlation between the corrected phenotypes of the birds in the validation population and the correction values (GEBVs/EBVs) was used to determine the assessment's correctness. Salek Ardestani (2021) used the formula below to compute the standard error of prediction accuracy (Salek et al., 2021):

$$\text{Standard error} = \frac{1\text{-accuracy}^2}{\sqrt{(\text{number of individuals-1})}} \quad (4)$$

The improvement in assessment accuracy was calculated using the following equation (Salek et al., 2021):

$$\text{Improvement accuracy} = (\text{EBV accuracy}) \times 100 \quad (5)$$

Bias and prediction errors were calculated as regression coefficients (r) of GEBVs on the corrected phenotype using the lm function in the R 4.0.2 program.

Cross-validation to evaluate the model

The 5-fold cross-validation (CV) approach was used to assess the performance of several prediction models. A total of 308 birds were used; 40 were chosen at random as the validation population and the remaining 268 were used as the reference population. Five times were used to complete this. The ssGBLUP method was used to estimate the GEBVs in the validation sets. The BLUP approach was also used to determine traditional correction values for various age groups. To examine the ability of various scenarios to predict outcomes, accuracy, and bias of GEBVs/EBVs were employed. In the end, the levels of MAF with the best predictive power were determined by comparing the accuracy values of GEBV evaluation connected to the information of all SNPs with the accuracy of GEBV evaluation of each subgroup of MAF separately.

RESULTS

The genotype rate of the samples was judged to be 99.9% after data quality control, and four birds with low genotype rates were disregarded from the analysis. The remaining 308 birds were divided between 138 chickens and 170 roosters. After quality control, the QQ-plot in R was used to establish the data's normality (Figure 1). The expected and actual P-values of 10 are represented on the x and y axes, respectively.

Population structure

The distribution of SNPs before and after quality control, along with the average distance of adjacent SNPs in each chromosome using synbreed software and R3.2.2, are summarized in Table 1. Multidimensional scaling (MDS) analysis of SNP markers with a threshold of r2 = 0.2 using the first MDS component showed that chickens branched from 8 stepfamilies (Figure 2).

The population structure of the samples utilized in this study was identified using the first and second components of MDS analysis, and its graph was created using R software. It was found that the current samples are made up of 8 closely related population groups. The first MDS component primarily captured the genetic structure of the data or its variance. Additionally, a neighbor-joining tree was created using 48379 SNP markers to calculate the genetic distance between all genotypes. This study outcome supported the existence of 8 subpopulations and was compatible with the population structure analysis (Figure 3).

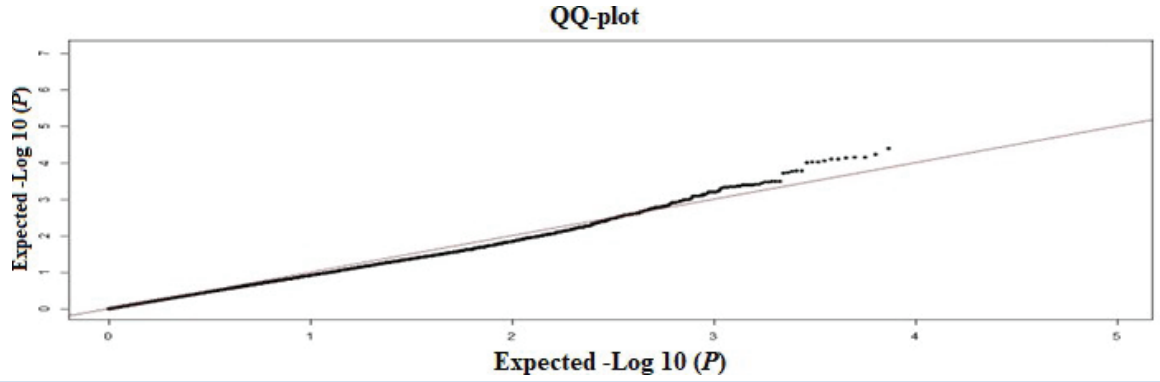


Figure 1: QQ-plot of 48379 SNP markers for body weight trait in chicken

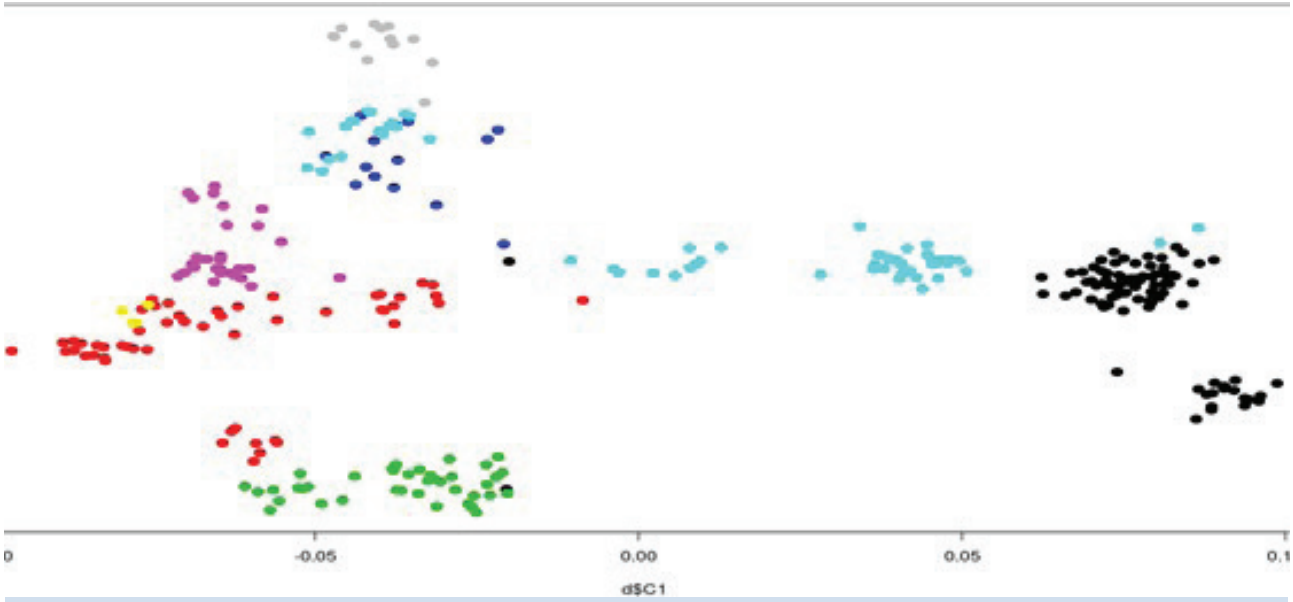


Figure 2: Identification of population structure using multidimensional scaling (MDS) analysis. Fullsib families are shown in one color (HSF = half-sibling family)

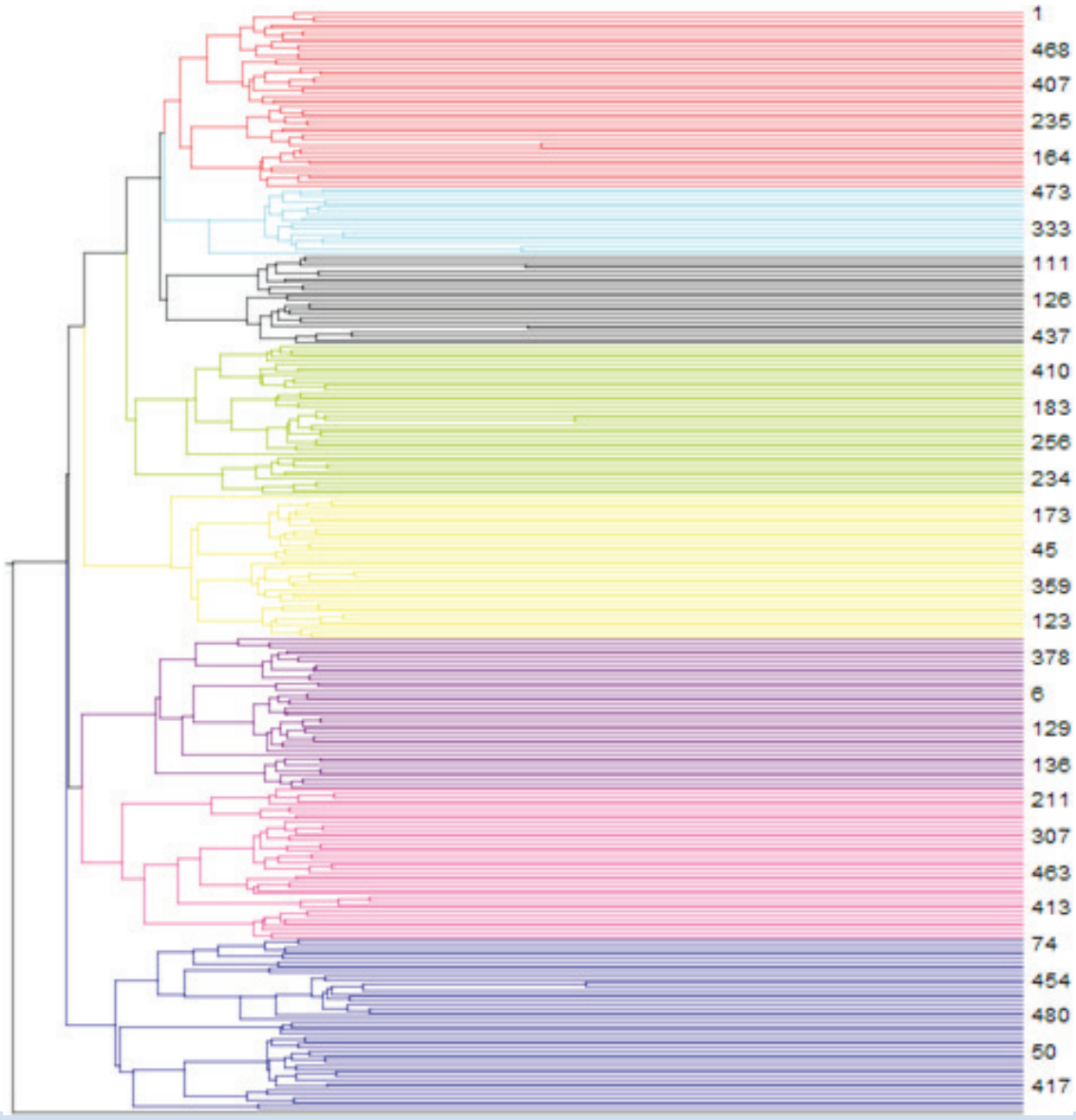


Figure 3: Genetic relationships among 8 subpopulations constructed using common allelic distance in a phylogenetic tree based on 48,379 single nucleotide polymorphisms (SNPs)

BLUP and ssGBLUP evaluations

By calculating the correlation between EBV in each validation group and Adjust phenotype after estimating EBV and Adjust phenotype for the validation groups, the estimated values of EBV prediction accuracy for the second, third, fourth, fifth, sixth, and seventh weeks are 0.116, 0.054, 0.215, 0.178, 0.219 and 0.146. Additionally, by evaluating the correlation between the GEBV in each group and the Adjust phenotype after estimating the GEBV for validation groups, the anticipated accuracy values of GEBV for each of these weeks are 0.264, 0.173, and 0.216, respectively. The estimates were 0.188, 0.22 and 0.15. For the second to seventh weeks, the regression coefficients for genomic prediction to assess the bias of estimates using the ssGBLUP technique are 1.3, 0.89, 0.74, 0.72, 0.71 and 0.73, respectively, and the improvement values. Each of these weeks' improvement accuracy was 59.03%, 220.37%, 0.46%, 5.61%, 0.45% and 2.73%, respectively (Table 2).

In various weeks, Figure 4 compares the forecast accuracy of the BLUP and ssGBLUP.

Genomic evaluation using different MAF sub-groups

For each of the groupings MAF 0.1-0.05, MAF 0.1-0.2, MAF 0.2-0.3, MAF 0.3-0.4 and MAF 0.4-

0.5 in the second week, the accuracy of the genomic evaluation was evaluated at 0.265, 0.273, and 0.259. These five subgroups' regression coefficients were computed as 1.3, 1.08, 1.8, 1.6 and 1.6, respectively. For the MAF 0.1-0.05, MAF 0.1-0.2, and MAF 0.4-0.5 subgroups, the genome prediction improvement values were 0.6%, 5.42% and 0.6%, respectively (Table 3).

In the second week, Figure 5 compares the evaluation accuracy of each subgroup of markers with the evaluation accuracy of data about all markers.

For each of the subgroups in the third week, the accuracy of the genomic evaluation was judged to be 0.149, 0.170, 0.159, 0.170 and 0.182, respectively. These five subgroups' regression coefficients were calculated to be 0.81, 0.92, 0.86, 0.86 and 0.90, respectively. When compared to using the data from all SNPs, genomic prediction for MAF subgroup markers 0.4-0.5 was better by 16.66% (Table 4).

In the third week, Figure 6 compares the evaluation accuracy of each subgroup of markers with the evaluation accuracy of data about all markers.

For each MAF subgroup in the fourth week, the accuracy of the genomic evaluation was judged to be 0.179, 0.199, 0.188, 0.234 and 0.229, respectively. These five subgroups' regression coefficients were

Table 2: Accuracy and skewness of BLUP and ssGBLUP prediction for the body weight trait of broilers in different weeks using the 5-fold cross-validation method

Weeks	Accuracy / BLUP	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Regression coefficient / ssGBLUP
2	0.166 ± 0.042	0.264 ± 0.044	59.03	1.3
3	0.054 ± 0.045	0.173 ± 0.043	220.37	0.89
4	0.215 ± 0.043	0.216 ± 0.043	0.46	0.74
5	0.178 ± 0.043	0.188 ± 0.043	5.61	0.72
6	0.219 ± 0.043	0.220 ± 0.043	0.45	0.71
7	0.146 ± 0.044	0.150 ± 0.044	2.73	0.73

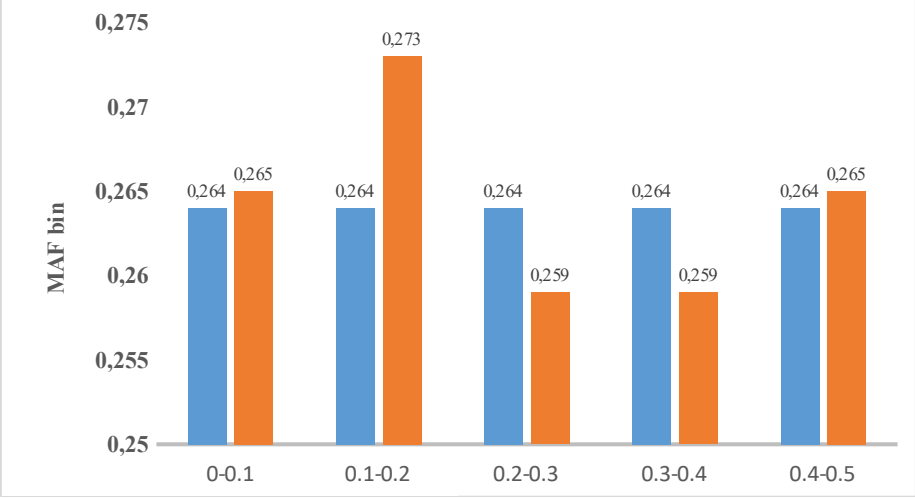


Figure 4: Comparison of accuracy of BLUP and ssGBLUP assessment in the second to seventh weeks for body weight trait in F2 chicks

Table 3: Accuracy and bias of genomic prediction of body weight trait using different MAFs at the age of two weeks

MAF	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Improvements for each MAF %	Regression coefficient / ssGBLUP
0.1-0.05	0.265 ± 0.042	59.63	0.6	1.32
0.1-0.2	0.273 ± 0.041	64.45	5.42	1.08
0.2-0.3	0.259 ± 0.042	56.02	-3.01	1.8
0.3-0.4	0.259 ± 0.042	56.02	-3.01	1.6
0.4-0.5	0.265 ± 0.042	59.63	0.6	1.6

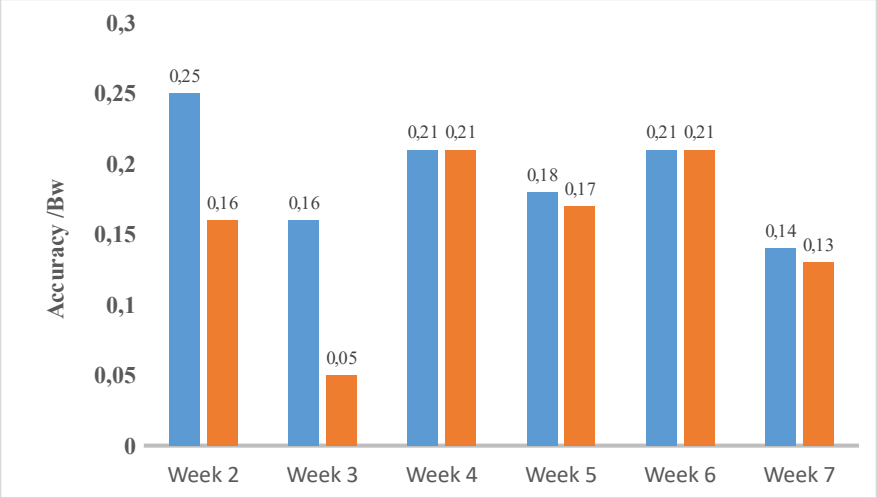


Figure 5: Comparison of the accuracy of evaluation of each MAF subgroup with the accuracy of evaluation of information related to one hundred percent of markers in the second week

Table 4: Accuracy and bias of genomic prediction of body weight trait using different MAFs at three weeks of age

MAF	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Improvements for each MAF %	Regression coefficient / ssGBLUP
0.1-0.05	0.149 ± 0.044	175.92	-44.45	0.81
0.1-0.2	0.170 ± 0.044	214.81	-5.56	0.92
0.2-0.3	0.159 ± 0.044	194.44	-25.93	0.86
0.3-0.4	0.170 ± 0.044	214.81	-5.56	0.86
0.4-0.5	0.182 ± 0.043	237.03	16.66	0.90

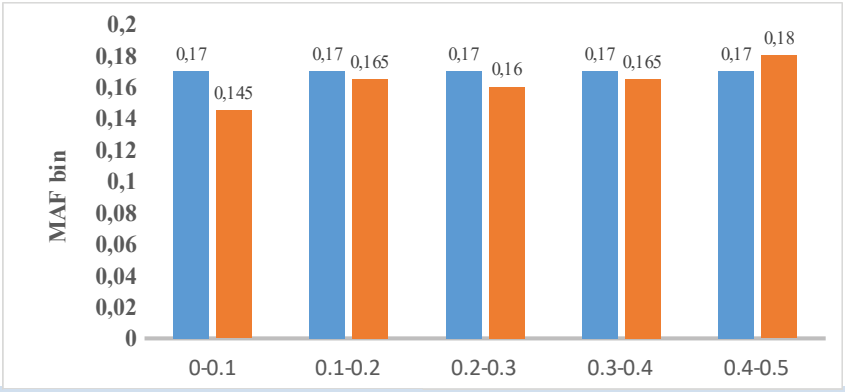


Figure 6: Comparison of the accuracy of evaluation of each MAF subgroup with the accuracy of evaluation of information related to one hundred percent of markers in the third week

calculated to be 0.64, 0.73, 0.7, 0.77 and 0.79, respectively. In the fourth week, Figure 7 compares the evaluation accuracy of each subgroup of markers with the evaluation accuracy of data about all markers.

In comparison to using the data from all SNPs this week, the genomic prediction for the MAF 0.3-0.4 and MAF 0.4-0.5 subgroups improved by 8.37% and 6.05%, respectively (Table 5).

In the fifth week, the accuracy of the genomic evaluation was judged to be 0.151, 0.184, 0.180, 0.195 and 0.196 for each of the MAF subgroups. These five subgroups' regression coefficients were calculated to be 0.65, 0.74, 0.76, 0.70 and 0.75, respectively. The accuracy of genomic prediction was enhanced for subgroups MAF 0.3-0.4 and MAF 0.4-0.5, respectively, when compared to the data of all SNPs this week (Table 6).

In the fifth week, Figure 8 compares the evaluation accuracy of each subgroup of markers with the evaluation accuracy of data about all markers.

For MAF subgroups, the accuracy of the genomic evaluation was judged to be 0.163, 0.236, 0.191, 0.219

and 0.229, respectively, in the sixth week. These five subgroups' regression coefficients were calculated to be 0.64, 0.77, 0.71, 0.71 and 0.75, respectively. The information about all markers this week is displayed by comparing the assessment accuracy of each subgroup of markers with the evaluation accuracy (Figure 9).

In the sixth week, the genome prediction improvement values for the MAF 0.1-0.2 and MAF 0.4-0.5

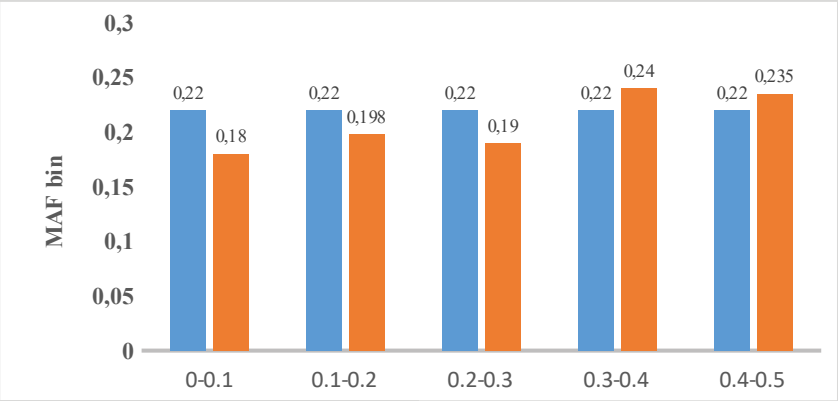


Figure 7: Comparison of the accuracy of assessment of each MAF subgroup with the accuracy of assessment of the information of one hundred percent of the markers in the fourth week

Table 5. Accuracy and bias of genomic prediction of body weight trait using different MAFs at the age of four weeks

MAF	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Improvements for each MAF %	Regression coefficient / ssGBLUP
0.1-0.05	0.179 ± 0.043	-16.74	-17.2	0.64
0.1-0.2	0.199 ± 0.043	-7.44	-7.9	0.73
0.2-0.3	0.188 ± 0.043	-12.55	-13.01	0.70
0.3-0.4	0.234 ± 0.042	8.83	8.37	0.77
0.4-0.5	0.229 ± 0.042	6.51	6.05	0.79

Table 6: Accuracy and bias of genomic prediction of body weight trait using different MAFs at the age of five weeks

MAF	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Improvements for each MAF %	Regression coefficient / ssGBLUP
0.1-0.05	0.151 ± 0.044	-15.16	-20.77	0.65
0.1-0.2	0.184 ± 0.043	3.37	-2.24	0.74
0.2-0.3	0.180 ± 0.043	1.12	-4.49	0.76
0.3-0.4	0.195 ± 0.043	9.55	3.94	0.70
0.4-0.5	0.196 ± 0.043	10.11	4.5	0.75

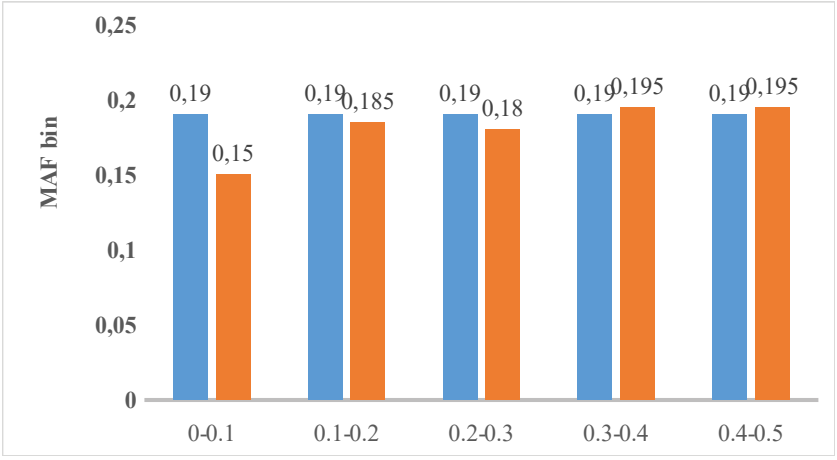


Figure 8: Comparison of the accuracy of assessment of each MAF subgroup with the accuracy of assessment of information related to one hundred percent of markers in the fifth week

subgroups were 7.31% and 4.11%, respectively (Table 7).

Additionally, the accuracy of genomic evaluation in the seventh week was 0.1, 0.145, and 116 for each of the subgroups MAF 0.1-0.05, MAF 0.1-0.2, MAF 0.2-0.3, MAF 0.3-0.4 and MAF 0.4-0.5. The calculated values were 0.0, 0.161 and 0.163. These five categories' regression coefficients were calculated to be 0.53, 0.64, 0.61, 0.59 and 0.58, respectively. For the MAF 0.3-0.4 and MAF 0.4-0.5 subgroups, the values of genomic prediction improvement this week compared to using the knowledge of all SNPs were 7.54% and 8.91%, respectively (Table 8).

In the seventh week, Figure 10 compares the evaluation accuracy of each subgroup of markers with the evaluation accuracy of data about all markers.

The advantage of markers with allelic frequencies of 0.4-0.5 across all weeks was validated by a comparison of the enhancement of genomic prediction by various MAFs (Figure 11).

Using data from 100% markers and various MAF groups, the values of standard error and improvement of genomic prediction were calculated using the equations presented in the statistical analysis section for each week and were then entered into the appropriate tables.

DISCUSSION

The current results demonstrate a significant difference in enhancing the accuracy of genomic prediction when the decreased number of SNPs (markers with particular MAFs) is utilized, in addition to establishing the superiority of the

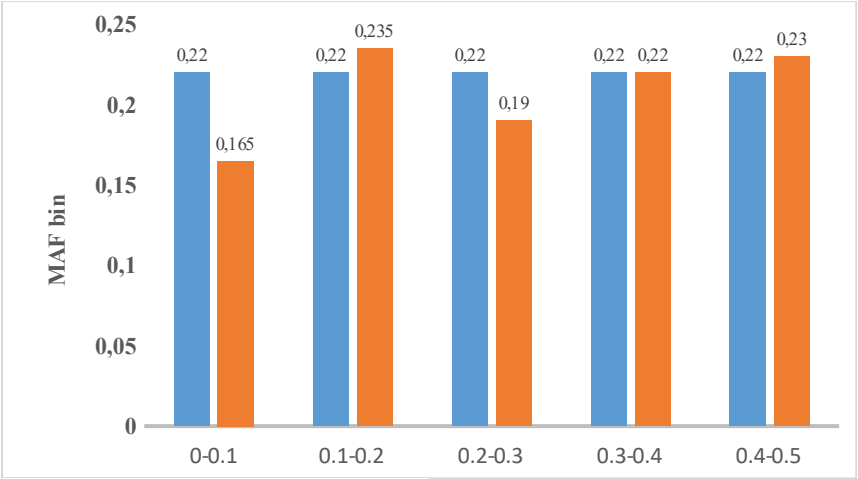


Figure 9: Comparison of the accuracy of evaluation of each MAF subgroup with the accuracy of evaluation of information related to one hundred percent of markers in the sixth week

Table 7: Accuracy and bias of genomic prediction of body weight trait using different MAFs at the age of six weeks

MAF	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Improvements for each MAF %	Regression coefficient / ssGBLUP
0.1-0.05	0.163 ± 0.044	-25.57	-26.02	0.64
0.1-0.2	0.236 ± 0.042	7.76	7.31	0.77
0.2-0.3	0.191 ± 0.043	-12.78	-13.23	0.71
0.3-0.4	0.219 ± 0.043	0	0	0.71

Table 8: Accuracy and bias of genomic prediction of body weight trait using different MAFs at the age of seven weeks

MAF	Accuracy / ssGBLUP	Improvement accuracy% / ssGBLUP	Improvements for each MAF %	Regression coefficient / ssGBLUP
0.1-0.05	0.100 ± 0.044	-31.50	-34.23	0.53
0.1-0.2	0.145 ± 0.044	-0.68	-3.41	0.64
0.2-0.3	0.116 ± 0.044	-20.54	-23.27	0.61
0.3-0.4	0.161 ± 0.044	10.27	7.54	0.59
0.4-0.5	0.163 ± 0.044	11.64	8.91	0.58

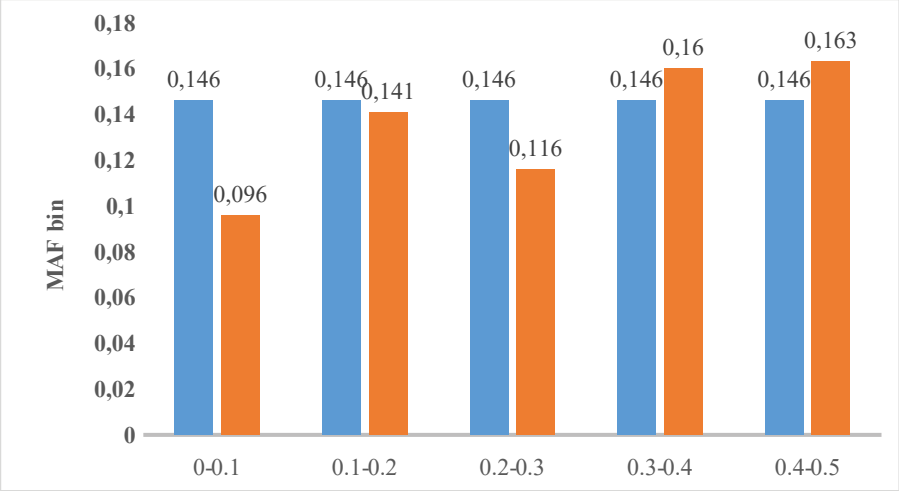


Figure 10: Comparison of the accuracy of evaluation of each MAF subgroup with the accuracy of evaluation of information related to one hundred percent of markers in the seventh week

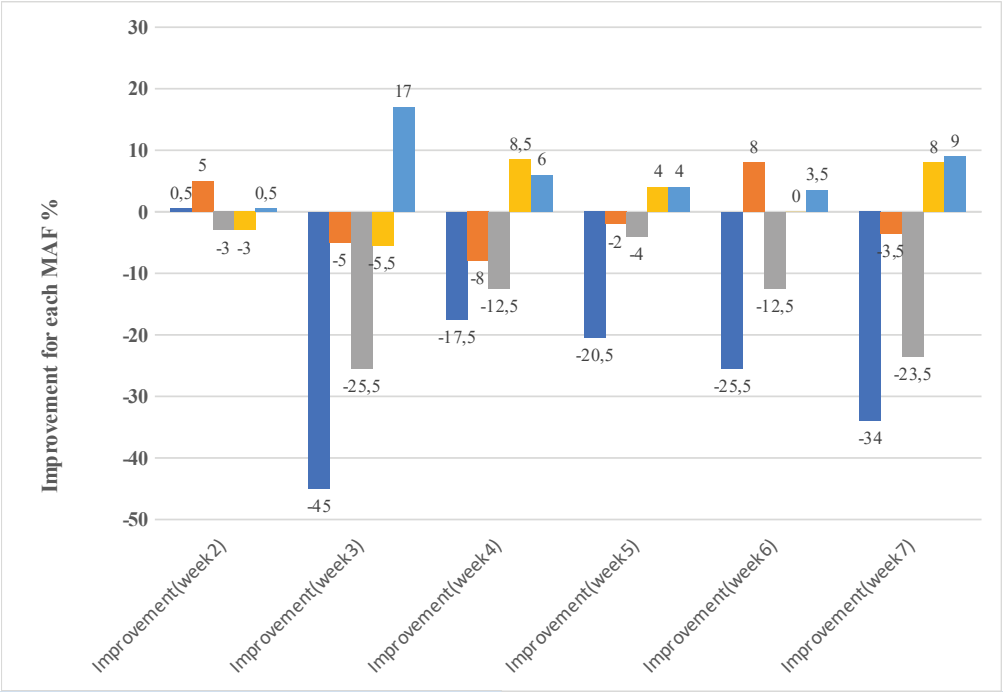


Figure 11: Improvement for each MAF group compared to 100% SNPs data in weeks 2 to 7

ssGBLUP technique over the BLUP method for the body weight trait in various weeks. It shows that it was utilized to create a matrix of genomic kinship ties in broiler chickens. It is clear from the current study that the ssGBLUP approach was the most accurate during the entire six-week period. Although only the F2 generation and 308 birds were genotyped, the accuracy of genomic prediction was estimated to be less than 30% in various weeks (Song et al., 2018). However, the higher prediction accuracy of the ssGBLUP method than the BLUP method in various weeks may be attributable to the combination of pedigree information and genotypic data used to predict GEBV values (Mrode et al., 2018; Silva et al., 2016).

The current investigation supported the findings of Salek Salek Ardestani (Salek et al., 2021). The improvement in genomic prediction was stronger in the third and second weeks than in the other weeks, respectively, even though we got a higher prediction accuracy for ssGBLUP than the other technique in all six weeks. In general, it is anticipated that employing high-density panels will allow a significant amount of genetic variance to be explained. However, the majority of the markers in these chips have causative

mutations and are in incomplete LD. The capacity to forecast is decreased when incomplete LD with causative mutations is present (Al Kalaldeh et al., 2019).

Consequently, using specific markers can help to increase the precision of genomic prediction. To look into the genomic prediction of the body weight characteristic in broiler chickens, SNPs were divided into five subgroups based on allelic frequency. From 0.1 in the allelic frequency of 0.05-0.1 in the seventh week to 0.273 in the allelic frequency of 0.1-0.2 in the second week, the accuracy of genomic evaluation for body weight trait by various SNP subsets ranged. The findings demonstrated that using SNPs with allelic frequencies of 0.05-0.1 in the second week, 0.1-0.2 in the second and sixth weeks, 0.3-0.4 in the fourth, fifth and seventh weeks, as well as 0.4-0.5 in every six weeks, can estimate the accuracy of the evaluation much more accurately than the information provided by all SNPs for the body weight trait.

The implementation of genomic evaluation using MAFs 0.0-0.1, 0.0-1.2 and 0.0-4.5 in similar populations at the age of two weeks can produce positive results, as evidenced by the higher prediction accuracy in the second week for all five MAFs (0.259) compared to the accuracy of various MAF groups in other weeks (0.236). These findings demonstrate that, despite having a lower cost, utilizing a panel with a lower density can nevertheless produce results that are superior to those obtained by using data from all SNPs (Habier et al., 2009). To support this, Ogawa et al. (2014) reported that the use of at least 4000 SNP markers is sufficient for the genetic prediction of body weight and carcass attributes in a study on a population of Japanese cattle. According to the results of the current study, Liang et al. (2018) also found comparable findings in pigs, demonstrating that an increase in marker density will not improve the accuracy of genomic prediction.

CONCLUSION

This study uses the 5-fold cross-validation (CV) method in a single-stage evaluation strategy to confirm the superiority of the ssGBLUP method over the BLUP method in an F2 population and investigate the accuracy of the correction values prediction using five groups of markers with various MAFs. This group of markers (MAF 0.4-0.5) is introduced as the best level of allelic frequency to perform genomic evaluations for the growth trait after the results showed that using SNPs with an allelic frequency of 0.4-0.5 in each of the second to seventh weeks shows a higher predictive accuracy than the information of all SNPs. The effectiveness of the poultry industry will be increased by decreasing the cost associated with genotype and other management costs, as well as increasing the accuracy of assessment for this crucial trait (growth), by applying this study, isolating and using selected markers, and even creating a genetic chip with low marker density in the following steps (markers with an allelic frequency of 0.4-0.5). In addition to imposing low genotyping costs on the industry, using SNPs with an allelic frequency of 0.4-0.5 and developing low-density SNP chips with markers with the aforementioned features can be utilized to accurately evaluate individuals based on genetic merit.

COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Al Kalaldeh, M., Gibson, J., Lee, S. H., Gondro, C., Van Der Werf, J. H. 2019. Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. Genetics Selection Evolution. 51, 1-18. https://doi.org/10.1186/s12711-019-0479-1

Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P., Zondervan, K. T. 2010. Data quality control in genetic case-control association studies. Nature Protocols. 5(9), 1564-1573. https://doi.org/10.1038/nprot.2010.116

Cardoso-Silva, C. B., Costa, E. A., Mancini, M. C., Balsalobre, T. W. A., Canesin, L. E. C., Pinto, L. R., …Vicentini, R. 2014. De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PloS one. 9(2), e88462. https://doi.org/10.1371/journal.pone.0088462

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., Lee, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4(1), s13742-015. https://doi.org/10.1186/s13742-015-0047-8

Christensen, K. G., Fenger-Grøn, M., Flarup, K. R., Vedsted, P. 2012. Use of general practice, diagnostic investigations, and hospital services before and after cancer diagnosis-a population-based nationwide registry study of 127,000 incident adult cancer patients. BMC Health Services Research. 12, 1. https://doi.org/10.1186/1472-6963-12-224

Clayton, D. G., Walker, N. M., Smyth, D. J., Pask, R., Cooper, J. D., Maier, L. M., ... Todd, J. A. 2005. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nature Genetics. 37(11), 1243-1246. https://doi.org/10.1038/ng1653

Costa, M. S., Gonçalves, Y. G., Teixeira, S. C., de Oliveira Nunes, D. C., Lopes, D. S., da Silva, C. V., ... Yoneyama, K. A. G. 2019. Increased ROS generation causes apoptosis-like death: mechanistic insights into the anti-Leishmania activity of a potent ruthenium (II) complex. Journal of Inorganic Biochemistry. 195, 1-12. https://doi.org/10.1016/j.jinorgbio.2019.03.005

Gao, Q., Zeng, X. J., Feng, G., Wang, Y., Qiu, J. 2012. T-S-fuzzy-model-based

approximation and controller design for general nonlinear systems. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). *42*(4), 1143-1154. https://doi.org/10.1109/TSMCB.2012.2187442

Gu, Q., Li, Z., Han, J. 2011. Joint feature selection and subspace learning. In Twenty-second international joint conference on artificial intelligence. Pp,1294-1299.

Guo, G. C., Wang, D., Wei, X. L., Zhang, Q., Liu, H., Lau, W. M., Liu, L. M. 2015. First-principles study of phosphorene and graphene heterostructure as anode materials for rechargeable Li batteries. The journal of physical chemistry letters. *6*(24), 5002-5008. https://doi.org/10.1021/acs.jpclett.5b02513

Habier, D., Fernando, R. L., Dekkers, J. C. 2009. Genomic selection using low-density marker panels. Genetics. 182(1), 343-353. https://doi.org/10.1534/genetics.108.100289

Karaman, T., Karaman, S., Dogru, S., Tapar, H., Sahin, A., Suren, M., ... Kaya, Z. 2016. Evaluating the efficacy of lavender aromatherapy on peripheral venous cannulation pain and anxiety: a prospective, randomized study. Complementary therapies in clinical practice. 23, 64-68. https://doi.org/10.1016/j.ctcp.2016.03.008

Karaman, E., Lund, M. S., Anche, M. T., Janss, L., Su, G. 2018. Genomic prediction using multi-trait weighted GBLUP accounting for heterogeneous variances and covariances across the genome. G3: Genes, Genomes, Genetics. 8(11), 3549-3558. https://doi.org/10.1534/g3.118.200673

Koivula, M., Strandén, I., Pösö, J., Aamand, G. P., Mäntysaari, E. A. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. Journal of Dairy Science. 98(4), 2775-2784. https://doi.org/10.3168/jds.2014-8975

Lee, Y. J., Yeh, Y. R., Wang, Y. C. F. 2012. Anomaly detection via online oversampling principal component analysis. IEEE transactions on knowledge and data engineering. 25(7), 1460-1470. https://doi.org/10.1109/TKDE.2012.99

Li, G., West, A. J., Densmore, A. L., Jin, Z., Parker, R. N., Hilton, R. G. 2014. Seismic mountain building: Landslides associated with the 2008 Wenchuan earthquake in the context of a generalized model for earthquake volume balance. Geochemistry, Geophysics, Geosystems. 15(4), 833-844. https://doi.org/10.1002/2013GC005067

Liang, Z., Gupta, S. K., Yeh, C. T., Zhang, Y., Ngu, D. W., Kumar, R., ... Schnabel, J. C. 2018. Phenotypic data from inbred parents can improve genomic prediction in pearl millet hybrids. G3: Genes, Genomes, Genetics. 8(7), 2513-2522. https://doi.org/10.1534/g3.118.200242

Lourenco, D. A. L., Misztal, I., Tsuruta, S., Aguilar, I., Lawlor, T. J., Forni, S., Weller, J. I. 2014. Are evaluations on young genotyped animals benefiting from the past generations. Journal of Dairy Science. 97(6), 3930-3942. https://doi.org/10.3168/jds.2013-7769

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D. H. 2002. BLUPF90 and related programs (BGF90). In Proceedings of the 7th world congress on genetics applied to livestock production. 28, (07).

Misztal, I., Lourenco, D., Legarra, A. 2020. Current status of genomic evaluation. Journal of Animal Science. 98(4), skaa101. https://doi.org/10.1093/jas/skaa101

Mrode, R., Ojango, J., Mwai, O. 2018. 522 Developing innovative digital technology and genomic approaches to livestock genetic improvement in developing countries. Journal of Animal Science. 96(Suppl 3), 507. https://doi.org/10.1093/jas/sky404.1109

Neves, H. H., Carvalheiro, R., O'brien, A. M. P., Utsunomiya, Y. T., Do Carmo, A. S., Schenkel, F. S., ... Garcia, J. F. 2014. Accuracy of genomic predictions in Bos indicus (Nellore) cattle. Genetics Selection Evolution, 46, 1-13. https://doi.org/10.1186/1297-9686-46-17

Ogawa, S., Matsuda, H., Taniguchi, Y., Watanabe, T., Nishimura, S., Sugimoto, Y., Iwaisaki, H. 2014. Effects of single nucleotide polymorphism marker density on degree of genetic variance explained and genomic evaluation for carcass traits in Japanese Black beef cattle. BMC Genetics. 15, 1-13. https://doi.org/10.1186/1471-2156-15-15

Papanicolaou, A., Schetelig, M. F., Arensburger, P., Atkinson, P. W., Benoit, J. B., Bourtzis, K., ... Handler, A. M. 2016. The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome biology. 17, 1-31. https://doi.org/10.1186/s13059-017-1155-9

Rolf, M. M., Taylor, J. F., Schnabel, R. D., McKay, S. D., McClure, M. C., Northcutt, S. L., ... Weaber, R. L. 2010. Use of bovine SNP50 data for feed efficiency selection decisions in Angus cattle. Pg, 103-117. BIF, Columbia, MO.

Salanti, G., Amountza, G., Ntzani, E. E., Ioannidis, J. 2005. Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. European journal of human genetics. 13(7), 840-848. https://doi.org/10.1038/sj.ejhg.5201410

Salek Ardestani, S., Jafarikia, M., Sargolzaei, M., Sullivan, B., Miar, Y. 2021. Genomic prediction of average daily gain, back-fat thickness, and loin muscle depth using different genomic tools in Canadian swine populations. *Frontiers in Genetics*. *12*, 665344. https://doi.org/10.3389/fgene.2021.665344

Silva, R. M. O., Fragomeni, B. O., Lourenco, D. A. L., Magalhães, A. F. B., Irano, N., Carvalheiro, R., ... Albuquerque, L. G. 2016. Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. Journal of Animal Science. 94(9), 3613-3623. https://doi.org/10.2527/jas.2016-0401

Salvian, M. 2020. Enrichment of genotyping panels for the genomic selection of special traits in broiler chicken. Doctoral dissertation, Universidade de São Paulo. https://doi.org/10.11606/T.11.2020.tde-12082020-153615

Song, H., Zhang, J., Jiang, Y., Gao, H., Tang, S., Mi, S., ... Ding, X. 2017. Genomic prediction for growth and reproduction traits in pig using an admixed reference population. Journal of Animal Science. 95(8), 3415-3424. https://doi.org/10.2527/jas.2017.1656

Song, H., Zhang, J., Zhang, Q., Ding, X. 2019. Using different single-step strategies to improve the efficiency of genomic prediction on body measurement traits in pig. Frontiers in genetics. *9*, 730. https://doi.org/10.3389/fgene.2018.00730

Su, G., Brøndum, R. F., Ma, P., Guldbrandtsen, B., Aamand, G. P., Lund, M. S. 2012. Comparison of genomic predictions using medium-density (~ 54,000) and high-density (~ 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. Journal of Dairy Science. 95(8), 4657-4665. https://doi.org/10.3168/jds.2012-5379

Sun, Y., Zhao, G., Liu, R., Zheng, M., Hu, Y., Wu, D., ... Wen, J. 2013. The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. BMC Genomics. 14(1), 1-11. https://doi.org/10.1186/1471-2164-14-458.

VanRaden, P. M., Tooker, M. E., Cole, J. B., Wiggans, G. R., Megonigal Jr, J. H. 2007. Genetic evaluations for mixed-breed populations. Journal of Dairy Science. 90(5), 2434-2441. https://doi.org/10.3168/jds.2006-704

Wang, C., Zöllner, S., & Rosenberg, N. A. 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. Plos genetics. 8(8), e1002886. https://doi.org/10.1371/journal.pgen.1002886

Wang, Y., Cao, X., Luo, C., Sheng, Z., Zhang, C., Bian, C., ... Li, N. 2020. Multiple ancestral haplotypes harboring regulatory mutations cumulatively contribute to a QTL affecting chicken growth traits. Communications Biology. 3(1), 472. https://doi.org/10.1038/s42003-020-01199-3

Wellmann, R., Preuß, S., Tholen, E., Heinkel, J., Wimmers, K., Bennewitz, J. 2013. Genomic selection using low density marker panels with application to a sire line in pigs. Genetics Selection Evolution. 45(1), 1-11. https://doi.org/10.1186/1297-9686-45-28

Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M., ... Kamatani, N. 2008. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. The American Journal of Human Genetics. 83(4), 445-456. https://doi.org/10.1016/j.ajhg.2008.08.019

Yan, Y., Wu, G., Liu, A., Sun, C., Han, W., Li, G., Yang, N. 2018. Genomic prediction in a nuclear population of layers using single-step models. Poultry science. 97(2), 397-402. https://doi.org/10.3382/ps/pex320

Yang, J., Lee, S. H., Goddard, M. E., Visscher, P. M. 2013. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. Genome-wide association studies and genomic prediction. 215-236. https://doi.org/10.1007/978-1-62703-447-0_9

Zhang, Y., Yang, R., Burwinkel, B., Breitling, L. P., Brenner, H. 2014. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. Environmental health perspectives. 122(2), 131-137. https://doi.org/10.1289/ehp.1306937

Zhang, G. X., Fan, Q. C., Wang, J. Y., Zhang, T., Xue, Q., Shi, H. Q. 2015. Genome-wide association study on reproductive traits in Jinghai Yellow Chicken. Animal reproduction science. 163, 30-34. https://doi.org/10.1016/j.anireprosci.2015.09.011

Zhang, W., Zhang, Q., Yu, B., Zhao, L. 2015. Knowledge map of creativity research based on keywords network and co-word analysis, 1992-2011. Quality & Quantity. 49, 1023-1038. https://doi.org/10.1007/s11135-014-0032-9

Zhernakova, A., Stahl, E. A., Trynka, G., Raychaudhuri, S., Festen, E. A., Franke, L., ... Plenge, R. M. 2011. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS genetics. 7(2), e1002004. https://doi.org/10.1371/journal.pgen.1002004