*J. Agr. Sci. Tech. (2021) Vol. 23(1): 27-40*

# Impact of Imputation, Reference Population Structure, and Single Nucleotide Polymorphism Panel Density on Accuracy of Genomic Evaluation in Purebred and Crossbred Populations

S. Barjasteh[1], G. R. Dashab[1*], M. Rokouei[1], M. M. Shariati[2], and M. Vafaye Valleh[1]

## ABSTRACT

**The objective of this study was to compare the accuracy of genomic breeding values prediction with different marker densities before and after the imputation in the simulated purebred and crossbred populations based on different scenarios of reference population and methods of marker effects estimation. The simulated populations included two purebred populations (lines A and B) and two crossbred populations (Cross and Backcross). Three different scenarios on selection of animals in the reference set including: (1) A high relationship with validation population, (2) Random, and (3) High inbreeding rate, were evaluated for imputation of validation population with the densities of 5 and 50K to 777K single marker polymorphism. Then, the accuracy of breeding values estimation in the validation population before and after the imputation was calculated by ABLUP, GBLUP, and SSGBLUP methods in two heritability levels of 0.25 and 0.5. The results showed that the highest accuracy of breeding values prediction in the purebred populations was obtained by GBLUP method and in the scenario of related reference population with validation set. However, in the crossbred population for the trait with low heritability ($h^2$= 0.25), the highest accuracy of breeding values prediction in the weighting mechanism was equal to ($\lambda$=0.2). Also, results showed that in the scenario of related reference population selection when 50K panel was used for genotype imputation to 777K SNPs, the prediction accuracy of genomic breeding values increased. But, in most scenarios of random and inbred reference set selection, there was no significant difference in the accuracy of genomic breeding values prediction between 5K and 50K SNPs after genotype imputation to 777K.**

**Keywords**: Genomic selection, Genotype imputation, Marker density, Prediction accuracy.

## INTRODUCTION

Recent developments in genotyping technologies have led to more knowledge on animal differences even in single nucleotide sequences. Next generation sequencing (NGS) technology is able to sequence millions of SNPs throughout the genome. In genomic selection area, density of SNP chips can affect the accuracy of prediction, because with increasing density, the Linkage Disequilibrium (LD) between markers and Quantitative Trait Loci (QTLs) increases and results in capturing QTL effects more accurately ( Wang *et al*., 2017; Chang *et al*., 2018). It is also possible to partition SNP effects into direct, indirect, and total SNP effects (Momen *et al*., 2018). Imputation from single-nucleotide polymorphic chips with low density to high-density panels is an important step before starting a genomic selection, since high-density panels can show more reliable genomic predictions (Júnior*et al*., 2017). Imputation is a powerful tool for increasing the power of genome-

[1]Department of Animal Science, Faculty of Agriculture, University of Zabol, Zabol, Islamic Republic of Iran.
*Corresponding author, e-mail: dashab@uoz.ac.ir
[2]Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Islamic Republic of Iran.





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

related studies, which allows to include genotyped animals with low-density panels in genomic evaluations without genotyping them with more expensive high-density panels. Additionally, this process is used to predict missing genotypes in breeding programs with purebred or crossbred population, which can lead to an increase in the accuracy of the genomic selection (Weigel *et al*., 2010; Zhang and Druet, 2010). Weigel *et al*. (2010) showed that accurate imputation of high-density genotypes from inexpensive low- or medium-density platforms could greatly enhance the efficiency of genome selection programs in dairy cattle. The accuracy of imputation and, consequently, the GEBV depends on a number of factors including population structure, panel density, the level of LD within a population, genetic structure of the trait, the number of individuals in reference populations genotyped by high-density panels, and the relationship between the reference population and the selection candidates (Van Binsbergen *et al*., 2014;Weigel*et al*., 2010).

Chen *et al*. (2014) showed that performance of both GBLUP and Bayesian methods was influenced by imputation errors. They demonstrated that for traits influenced by a few QTL with large effect, the Bayesian method resulted in a greater reduction in the accuracy than GBLUP,as imputation mademore errors and resulted in lower accuracy of genomic prediction in very low-density panels. Larmer *et al*. (2017) in studying animals from a variety of beef and dairy breeds reported that the combination of reference population from different breeds with low relationships, or the reference with sharing less haplotypes with those in the imputation population, reduced the accuracy of imputation. Poorly imputed individuals may also have a significant deleterious effect on the accuracy of whole-genome detection results and genomic prediction steps. It has been shown that the error rate of imputation depends on the relationship between the animals of the imputation targets and the reference populations, and the inclusion of the closest ancestors in the reference population with a high-density chip could help to reduce the errors (Schrooten *et al*., 2014).

So far, different statistical methods have been proposed to estimate breeding values using genomic information (De Los Campos *et al*., 2013;Meuwissen *et al*., 2001). However, GBLUP, which is a linear mixed model integrating a marker-based Genomic relationship matrix (G), is generally preferred for genomic evaluations more than the other methods, e.g. Bayesian, because of low computational demand (VanRaden, 2008). This method can provide GEBV with high levels of accuracy in many economically important traits, especially the traits with high or moderate heritability; these GEBVs can be achieved at an early age resulted an early selection even before having any phenotypic information for candidate animals (Schaeffer, 2006). Over the past 10 years, genomic selection has been extensively introduced in several major livestock species for its high accuracy, short generation intervals, and recently low breeding costs (Georges *et al*., 2019). The genomic information can be included in models to estimate breeding value along the pedigree information, e.g. single step genomic BLUP. Due to abundance of pedigree information, phenotypic records, and completeness of genotyping, combining both information, pedigree and genomic, can be helpful and increase accuracy of prediction (Gray *et al*., 2012). Therefore, the purpose of this study was to compare the accuracy of genomic predictions using low-, moderate- and high-density marker panels by considering imputation through three combinations of marker and pedigree information including pedigree-based BLUP (ABLUP), GBLUP and combining genomic and pedigree information in SSGBLUP models. We also aimed to investigate the effect of reference population structure on accuracy of genomic evaluations under two levels of heritability: 0.25 and 0.5.





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

## MATERIALS AND METHODS

### Population and Genome Simulations

The QMSim software was used to simulate historical and recent population structures (Sargolzaei and Schenkel, 2009). A genome consisting of 29 autosomal chromosomes with a similar length to the cattle chromosomes was simulated. A total of 777,026 balletic Single Nucleotide Polymorphisms (SNPs) consisting of 725 QTLs with equal frequency in the first generation were designed. SNPs were uniformly distributed over the whole genome. Distribution of QTLs across genome was random. To study the effects of trait heritability, both levels of 0.25 and 0.5 and phenotypic variance 1 were considered. In the historical population, to create the initial LD between marker and QTL and establish mutation-drift equilibrium, at the first generation, we constituted 500 animals (250 males and 250 females). Then, 1,000 randomly mated generations without changing numbers and 1,000 gradually expanded to 4,000 offspring were simulated. The number of males in the last generation in the historical population was considered 50 animals. To create the first purebred population (line A), 20 males and 200 females were selected from this generation and based on positive assortative mated through 10 generations with two progenies per dam. To create the second purebred population (line B), 20 males and 200 females were selected again from the last generation of the historical population and based on positive assortative mated through 10 generations. To utilize the maximum heterosis properties, the selection of animals was carried out based on the high breeding value (line A) and low breeding value (line B). For lines A and B, the genotype, phenotype, and pedigree information related to the generations 8, 9, and 10 were registered and the generation 10 was considered as the validation set and the generations 8 and 9 as the reference training set. In the next step, the hybrid populations (cross and backcross) were simulated. The cross population was created by mating 20 randomly selected males from the generation 10 in line A and 200 randomly selected females from generation 10 in line B. The backcross population was generated by mating 20 randomly selected males from generation 10 in line A and 200 randomly selected females from cross population through one generation of random mating (Table 1). The genotype, phenotype, and pedigree information that were related to the backcross population was registered. In this study, the animals of backcross population were considered as validation set (imputation) and animals of generation 10 from line A and animals of cross population were considered as reference set.

After genomic simulation, the quality control was performed and SNPs with Minor Allele Frequency (MAF) less than 0.01 and the monomorphic loci were excluded from the genotype data and, finally, 407935 SNPs were left for analysis. For 5K and 50K SNPs, we sampled from the remaining SNPs (400K) and the reduced genotype file was used for imputation populations.

### Imputation

A haplotype-base algorithm that was programmed in FImpute software was implemented to impute from low- (5K) and moderate-density (50K) panels to a high-density panel (777K) (Sargolzaei *et al*., 2014). This software uses the pedigree information (if known) and searches for long to short haplotypes representing close to far relationships, respectively. In comparison to most of population imputation software, FImpute assumes that all animals are related and uses Overlapping Sliding Window to find the haplotype fragments that have associated with common ancestor between individuals. FImpute uses an Overlapping Sliding Window approach to efficiently exploit relationships or haplotype similarities between target and reference





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

**Table 1**. Parameters for the stimulation of populations and genome.

| Population structure | Information of population simulation |
|---|---|
| **Step 1: Creating base population** | |
| Number of animals (Number of generations) | 500[0]500[1000]4000[2000] |
| Number of males in the last generation of base population | 50 |
| **Step 2: Recent (undergoing selection) population** | |
| **population of line A** | |
| Number of males from base population | 20 |
| Number of females from base population | 200 |
| Number of generations | 10 |
| Mating system | TBV/h positive assortative |
| Number of iterations | 10 |
| Heritability | 0.25,0.5 |
| Phenotype variance | 1 |
| **Population of line B** | |
| Number of males from base population | 20 |
| Number of females from base population | 200 |
| Number of generations | 10 |
| Mating system | TBV/l positive assortative |
| Number of iterations | 10 |
| Heritability | 0.25,0.5 |
| Phenotype variance | 1 |
| **Cross population** | |
| Number of males from the last generation of line A | 20 |
| Number of females from the last generation of line B | 200 |
| Male sex ratio | 0.5 |
| Number of generation | 1 |
| Mating system | Random |
| **Backcross population** | |
| Number of males from the last generation of line A | 20 |
| Number of females from cross population | 200 |
| Male sex ratio | 0.5 |
| Number of generation | 1 |
| Mating system | Random |
| Genome structure | Genome simulation information |
| Number of chromosomes | 29 |
| Number of markers (For each chromosome) | 26794 |
| Distribution of Markers | Evenly |
| Number of QTL (For each chromosome) | 25 |
| QTL distribution | Random |

individuals. The process starts with long windows to capture haplotype similarity between close relatives. After each chromosome sweep, the window size is shrunk by a constant factor allowing for shorter haplotype similarity (arising from more distant relatives) to be taken into account. Because closer relatives usually share longer haplotypes while more distant relatives share shorter haplotypes, the algorithm simply assumes that all individuals are related to each other at different degrees.

### Reference Population Structure

Three scenarios were investigated to evaluate the effect of population structure on prediction accuracy as: (1) Animals having the highest relationship with the validation set, (2) The animals with the highest inbreeding, and (3) The randomly selected animals.





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

## Prediction Model

Variance components and genomic breeding values were computed by a mixed linear model. Generally, GBLUP model assumes that all markers contributed equally to genetic variation with no major genes. To do this, linear mixed models were used to estimate the animal effects. Three different relationship information were implemented in BLUP models as follow:

Pedigree-based BLUP (ABLUP) method: The numerator relationships matrix (A) is calculated based on the pedigree information using the individuals' relationship average. It is worthwhile to mention that the accuracy of these estimates can be affected by the pedigree's accuracy and quality (Calus, 2010). The Estimated Breeding Values (EBVs) were derived from a linear model as follows:

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

Where, **y** represents a vector of phenotype of interest, 1 is a vector of 1, μ is the average population, a and e are vectors of breeding values and residual effects, respectively, and Z is a design matrix for the random effects. In this model, it is assumed that $a \sim N(0, A\sigma_a^2)$ and $e \sim N(0, D\sigma_e^2)$. The mixed model equations to estimate the breeding value are as equation (2):

$$\begin{bmatrix} 1'1 & 1'Z \\ Z'1 & Z'Z + A^{-1}\alpha \end{bmatrix} \times \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'y \\ Z'y \end{bmatrix} \quad (2)$$

Where, $\boldsymbol{\alpha}$ is the ratio of error variance to additive variance and $A^{-1}$ shows the inversion of relationship matrix.

Genotype-based BLUP (GBLUP) method: The relationship matrix was calculated based on the genotype information, which resulted in a Genomic relationship matrix (G). The G tends to measure an actual section of the common alleles between individuals not an expected section such as pedigree-based relationship matrix. The individuals with the same genotype for a large number of markers are genetically more similar and have a large value in their corresponding location in the genomic matrix. This matrix was created and calculated based on the VanRaden's model (2008) as follows:

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

Where, M is the genotypes matrix with codes -1 and 1 for homozygotes and code 0 for heterozygotes, P is Minor Allelic Frequency (MAF) matrix and $p_i$ shows MAF for $i_{th}$ marker, and Q is a matrix that is obtained from subtraction of P and M.

Combining genomic and pedigree information in SSGBLUP: The genomic and pedigree information were used in a form of Kernel matrix (K). This matrix combines the pedigree information (A) and the marker information (G) as follows:

$$K = \lambda A + (1-\lambda)G \quad (4)$$

Where, **λ** is a limited parameter ranging between 0 and 1. In this study, we chose **λ** equal to 0 (GBLUP), 0.1, 0.2, 0.5, and 1 (ABLUP).

In all models, we evaluated the accuracy of GEBV using the different marker densities and various subsets of reference population under two levels of heritability: 0.25 and 0.5.

## Prediction Accuracy Access

The *predictive* accuracy was calculated through evaluating the Pearson correlation between the GEBV and TBVas:

$$\rho_{TBV,GEBV} = \frac{\sigma(TBV,GEBV)}{\sigma_{TBV}\sigma_{GEBV}} \quad (5)$$

The *Duncan's* multiple range test (with α= 0.01) was performed to compare the effects of different scenarios on the accuracy of GEBV including the reference subset selection, marker densities heritability, and statistical methods.

## RESULTS AND DISCUSSION

### Purebred Population (Line A)

The results related to the accuracy of GEBVs in the different marker densities before and after genotype imputation from the low- (5K) and moderate-density (50K)





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

**Table 2.** Accuracy of GEBVs (EBVs for λ equal one) in the different marker densities with or without genotype imputation to the high-density panel (777K) and using the different weight parameter (λ = 0, 0.2, 0.5 and 1) in the different scenarios of reference population selection for simulated traits with the heritabilities of 0.25 and 0.5 in the line A.[a]

| $h^2$ | Reference sub-setting method (SNP panel density) | Imputation status | λ | | | |
|---|---|---|---|---|---|---|
| | | | **0** | **0.2** | **0.5** | **1** |
| 0.25 | Inbreeding (5K) | No | 0.29$^{g}$ (0.022) | 0.22$^{f}$ (0.002) | 0.18$^{e}$ (0.005) | 0.14$^{d}$ (0.025) |
| | | Yes | 0.34$^{cd}$ (0.008) | 0.24$^{e}$ (0.001) | 0.21$^{cd}$ (0.007) | |
| | Inbreeding (50K) | No | 0.32$^{f}$ (0.016) | 0.24$^{e}$ (0.012) | 0.19$^{de}$ (0.004) | 0.15$^{cd}$ (0.024) |
| | | Yes | 0.36$^{ab}$ (0.012) | 0.26$^{cd}$ (0.014) | 0.21$^{cd}$ (0.007) | |
| | Random (5K) | No | 0.31$^{f}$ (0.01) | 0.25$^{de}$ (0.009) | 0.21$^{cd}$ (0.018) | 0.16$^{bc}$ (0.007) |
| | | Yes | 0.34$^{de}$ (0.006) | 0.25$^{de}$ (0.011) | 0.22$^{abc}$ (0.028) | |
| | Random (50K) | No | 0.33$^{ef}$ (0.01) | 0.24$^{ef}$ (0.029) | 0.20$^{de}$ (0.021) | 0.15$^{cd}$ (0.011) |
| | | Yes | 0.35$^{bc}$ (0.006) | 0.26$^{cd}$ (0.017) | 0.22$^{ab}$ (0.007) | |
| | Relatedness (5K) | No | 0.32$^{f}$ (0.009) | 0.27$^{bc}$ (0.004) | 0.20$^{cd}$ (0.012) | 0.17$^{abc}$ (0.005) |
| | | Yes | 0.35$^{bc}$ (0.006) | 0.29$^{b}$ (0.01) | 0.21$^{bcd}$ (0.014) | |
| | Relatedness (50K) | No | 0.34$^{cde}$ (0.006) | 0.29$^{b}$ (0.005) | 0.21$^{bcd}$ (0.011) | 0.18$^{a}$ (0.005) |
| | | Yes | 0.37$^{a}$(0.008) | 0.30$^{a}$(0.007) | 0.23$^{a}$(0.004) | |
| 0.5 | Inbreeding (5K) | No | 0.48$^{ab}$ (0.043) | 0.35$^{ab}$ (0.056) | 0.32$^{vc}$ (0.006) | 0.26$^{ce}$ (0.003) |
| | | Yes | 0.52$^{a}$ (0.07) | 0.39$^{ab}$ (0.057) | 0.34$^{bc}$ (0.007) | |
| | Inbreeding (50K) | No | 0.50$^{ab}$ (0.039) | 0.38$^{ab}$ (0.055) | 0.38$^{ab}$ (0.043) | 0.27$^{bcd}$ (0.008) |
| | | Yes | 0.51$^{ab}$ (0.065) | 0.38$^{ab}$ (0.067) | 0.36$^{bc}$ (0.058) | |
| | Random (5K) | No | 0.49$^{ab}$ (0.07) | 0.38$^{ab}$ (0.077) | 0.36$^{ab}$ (0.058) | 0.27$^{bce}$ (0.016) |
| | | Yes | 0.51$^{ab}$ (0.141) | 0.36$^{ab}$ (0.125) | 0.35$^{bc}$ (0.0916) | |
| | Random (50K) | No | 0.43$^{b}$ (0.11) | 0.30$^{b}$ (0.121) | 0.31$^{c}$ (0.09) | 0.25$^{e}$ (0.025) |
| | | Yes | 0.52$^{ab}$ (0.067) | 0.37$^{ab}$ (0.063) | 0.38$^{ab}$ (0.043) | |
| | Relatedness (5K) | No | 0.51$^{ab}$ (0.047) | 0.38$^{ab}$ (0.075) | 0.39$^{ab}$ (0.05) | 0.30$^{ab}$ (0.005) |
| | | Yes | 0.54$^{a}$ (0.062) | 0.38$^{ab}$ (0.075) | 0.41$^{ab}$ (0.066) | |
| | Relatedness (50K) | No | 0.51$^{ab}$ (0.043) | 0.37$^{ab}$ (0.066) | 0.40$^{ab}$ (0.039) | 0.31$^{a}$ (0.008) |
| | | Yes | 0.57$^{a}$ (0.06) | 0.41$^{a}$ (0.074) | 0.44$^{a}$ (0.043) | |

[a] Groups with the same heritability and different letters within each column are significant (P< 0.01).

panels to the high-density panel (777K) under a different weight parameter (λ= 0, 0.2, 0.5 and 1) are shown in Table 2. These accuracies were calculated in the different scenarios of reference population schemes of the simulated traits with 0.25 and 0.5 heritabilities in the purebred population line A.

In the present study, values of λ equal to 0, 0.2, 0.5 and 1 were studied. When λ= 0, only G matrix and when λ= 1, only A matrix were used. In λ= 0.2 and 0.5, different percentages of G and A matrices were used. Increasing the values of λ resulted in increasing the contribution of a matrix. When λ= 1 and only A matrix was used to calculate the breeding values, the values of





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

EBV and, in other states, the values of GEBV were obtained. Comparing prediction accuracy of GEBVs (EBVs) across different scenarios of reference sub sets with a low heritability ($h^2$= 0.25) revealed that the highest accuracy achieved when references were selected based on the highest relationship with the test population. Similar results were obtained when a trait with the high heritability ($h^2$= 0.5) was simulated. With increasing the contribution of pedigree information in the model through the reduction in λ to retrieve the EBVs in two heritability levels, the accuracies decreased significantly. The accuracies for the best sub-setting method, relatedness, with or without the imputation from low-density to the high-density panels in the high heritability (0.5) scenario were not significantly different. Therefore, in the purebred populations for the traits with a high heritability, genotyping with a low-density panel (5K) can obtain an accuracy similar to the higher densities (50K) and thus reducing the cost of genomic selection. In contrast, in the scenario with low heritability, the use of 50K panel compared to 5K panel resulted in a higher accuracy before and after genotype imputation. These various results for low and high heritabilities indicated that selecting a panel with appropriate density could be varied based on heritability levels of the trait studied. Compared to the other values of weighting parameters, the results of accuracy in the purebred population were generally higher when only G matrix was used (λ= 0). With increasing λ in the SSGBLUP model from 0.2 to 0.5 the accuracy of estimations reduced. In ABLUP (λ= 1) in which only matrix A was used, the accuracies of estimations were the lowest. The accuracy of genotype imputation in the cattle from SNP panels with a low density to the panels with 50 or 777K densities, especially in the breeds with a large reference population with the dense genotypes and a high level of linkage disequilibrium in the genome has been reported (Larmer *et al*., 2017; Sargolzaei *et al*., 2014). The range of genomic prediction accuracy in the dairy cattle in the developed countries has been reported for the traits with intermediate to high heritability such as milk production from 0.5 to 0.85 and for the traits with low heritability such as reproductive and survival traits from 0.2 to 0.5 (Weigel *et al*., 2010). While the accuracy of genomic predictions has been reported from low to intermediate and in the range of 0.21-0.6 (Mrode *et al*., 2019) that is consistent with the results of the present research. The lower accuracy of genomic breeding values in the developing countries can be due to less effective population size of reference set in industrial countries than developing ones, the lower accuracy of phenotypic data than the proven bulls in the developed countries, as well as lack of appropriate breeding programs in these countries (Mrode *et al*., 2019).

Boison, *et al.* (2017) investigated the impact of relations between the validation and training populations on the accuracy of genomic predictions. They showed that the increase of 0.1 in the average of genomic relationships between the reference and validation population (equal to adding selection candidate sire to the reference population) yields a high increase in prediction accuracy about 0.05, which was in agreement with our results. Additionally, as we showed in this study, increasing the density of marker panels causes increase in the LD between QTL and SNP and results in higher prediction ability and, consequently, the increase of accuracy of GEBVs. Although the use of 50K chips for predicting the genomic breeding values within breed is suitable and the desired results have been obtained (Boison, *et al*. 2017), the highest density chips such as 777K can certainly be available and increase LD between markers and QTL, consequently, the trend of reducing accuracy of GEBVs during generations may become slower.





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

**Table 3.** Accuracy of GEBVs (EBVs for λ equal one) in the different marker densities with or without genotype imputation to the high-density panel (777K) and using different weight parameters (λ= 0, 0.2, 0.5 and 1) in the different scenarios of reference population selection for simulated traits with the heritability leves of 0.25 and 0.5 in the crossbred population.[a]

| $h^2$ | Reference sub-setting method (SNP panel density) | Imputation status | λ | | | |
|---|---|---|---|---|---|---|
| | | | **0** | **0.2** | **0.5** | **1** |
| 0.25 | Inbreeding (5K) | No | $0.16^{d}$ (0.011) | $0.20^{g}$ (0.011) | $0.14^{h}$ (0.003) | $0.09^{f}$ (0.01) |
| | | Yes | $0.17^{d}$ (0.012) | $0.22^{efg}$ (0.004) | $0.14^{fgh}$ (0.003) | |
| | Inbreeding (50K) | No | $0.18^{cd}$ (0.015) | $0.21^{fg}$ (0.012) | $0.14^{gh}$ (0.003) | $0.10^{ef}$ (0.011) |
| | | Yes | $0.18^{cd}$ (0.019) | $0.23^{ef}$ (0.011) | $0.15^{fg}$ (0.004) | |
| | Random (5K) | No | $0.20^{bc}$ (0.015) | $0.23^{def}$ (0.015) | $0.16^{de}$ (0.007) | $0.11^{e}$ (0.015) |
| | | Yes | $0.21^{b}$ (0. 009) | $0.24^{cde}$ (0.019) | $0.17^{cd}$ (0.012) | |
| | Random (50K) | No | $0.20^{bc}$ (0.021) | $0.23^{ef}$ (0.023) | $0.15^{bf}$ (0.016) | $0.11^{ef}$ (0.021) |
| | | Yes | $0.22^{b}$ (0.004) | $0.25^{bc}$ (0.009) | $0.17^{cd}$ (0.008) | |
| | Relatedness (5K) | No | $0.21^{b}$ (0.02) | $0.25^{bcd}$ (0.012) | $0.18^{bc}$ (0.009) | $0.14^{bc}$ (0.01) |
| | | Yes | $0.21^{b}$ (0.019) | $0.24^{cde}$ (0.019) | $0.19^{a}$ (0.009) | |
| | Relatedness (50K) | No | $0.22^{ab}$ (0.025) | $0.25^{bc}$ (0.011) | $0.19^{ab}$ (0.003) | $0.16^{a}$ (0.012) |
| | | Yes | $0.24^{a}$ (0.026) | $0.28^{a}$ (0.019) | $0.20^{a}$ (0.007) | |
| 0.5 | Inbreeding (5K) | No | $0.35^{ab}$ (0.05) | $0.25^{c}$ (0.014) | $0.25^{f}$ (0.011) | $0.13^{ce}$ (0.003) |
| | | Yes | $0.39^{ab}$ (0.053) | $0.29^{bc}$ (0.017) | $0.28^{def}$ (0.011) | |
| | Inbreeding (50K) | No | $0.38^{ab}$ (0.05) | $0.28^{bc}$ (0.025) | $0.27^{ef}$ (0.025) | $0.14^{abc}$ (0.004) |
| | | Yes | $0.39^{ab}$ (0.061) | $0.29^{bc}$ (0.23) | $0.29^{cdef}$ (0.028) | |
| | Random (5K) | No | $0.38^{ab}$ (0.078) | $0.30^{bc}$ (0.045) | $0.29^{def}$ (0.048) | $0.13^{dc}$ (0.016) |
| | | Yes | $0.37^{ab}$ (0.126) | $0.31^{bc}$ (0.071) | $0.32^{bcde}$ (0.049) | |
| | Random (50K) | No | $0.31^{b}$ (0.111) | $0.27^{bc}$ (0.066) | $0.29^{def}$ (0.048) | $0.12^{e}$ (0.02) |
| | | Yes | $0.37^{ab}$(0.063) | $0.32^{ab}$(0.035) | $0.33^{abc}$ (0.028) | |
| | Relatedness (5K) | No | $0.38^{ab}$ (0.075) | $0.32^{ab}$ (0.055) | $0.33^{bcd}$ (0.043) | $0.15^{ab}$ (0.009) |
| | | Yes | $0.39^{ab}$ (0.081) | $0.34^{ab}$ (0.062) | $0.34^{ab}$ (0.05) | |
| | Relatedness (50K) | No | $0.37^{ab}$ (0.067) | $0.33^{ab}$ (0.054) | $0.34^{abc}$ (0.045) | $0.16^{a}$ (0.005) |
| | | Yes | $0.42^{a}$ (0.074) | $0.37^{a}$ (0.063) | $0.38^{a}$ (0.048) | |

[a] Groups with same heritability and different letters within each column are significant (P< 0.01).

## Crossbred Population

The results related to the accuracy of GEBVs and EBVsin simulated population with the different densities, imputation status, λ values (0, 0.2, 0.5 and 1) and sub-set selection of reference population for the traits with heritability levels of 0.25 and 0.5 in the crossbred population are shown in Table 3.

Interestingly, the results of this research showed that in the crossbred population for the trait with the low heritability, the combined pedigree information and marker information (SSGBLUP with λ=0.2) improved the accuracy of breeding values prediction, and the highest accuracy was achieved when references were selected based on relatedness and 50K panel density imputed to a high-density panel. Therefore, in the crossbred population with low heritability, due to more complex genetic architecture, using the pedigree information along with genomic information can result in a better estimation of GEBVs. However, in the high heritability using marker information can result in a higher accuracy. In other words, the heritability level could mainly influence effect of the combined pedigree information and marker information in





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

**Table 4.** Accuracy of GEBVs (EBVs for λ equal one) in the different marker densities with or without genotype imputation to the high-density panel (777K) and using different weight parameters (λ= 0, 0.2, 0.5 and 1) in the different scenarios of reference population selection for simulated traits with the heritability levels of 0.25 and 0.5 in the backcross population.[a]

| $h^2$ | Reference sub-setting method (SNP panel density) | Imputation status | λ | | | |
|---|---|---|---|---|---|---|
| | | | 0 | 0.2 | 0.5 | 1 |
| 0.25 | Inbreeding (5K) | No | $0.16^{g}$ (0.012) | $0.19^{h}$ (0.003) | $0.14^{h}$ (0.007) | $0.06^{f}$ (0.014) |
| | | Yes | $0.17^{fg}$ (0.011) | $0.20^{g}$ (0.005) | $0.15^{g}$ (0.008) | |
| | Inbreeding (50K) | No | $0.17^{efg}$ (0.016) | $0.20^{g}$ (0.004) | $0.15^{gh}$ (0.005) | $0.07^{ef}$ (0.014) |
| | | Yes | $0.19^{def}$ (0.025) | $0.21^{f}$ (0.004) | $0.16^{efg}$ (0.008) | |
| | Random (5K) | No | $0.20^{cde}$ (0.015) | $0.22^{e}$ (0.005) | $0.16^{fg}$ (0.007) | $0.09^{cd}$ (0.015) |
| | | Yes | $0.21^{bc}$ (0. 021) | $0.24^{bc}$ (0.011) | $0.17^{def}$ (0.013) | |
| | Random (50K) | No | $0.20^{cde}$ (0.019) | $0.23^{de}$ (0.008) | $0.15^{g}$ (0.021) | $0.10^{c}$(0.013) |
| | | Yes | $0.22^{b}$ (0.009) | $0.25^{bc}$ (0.003) | $0.17^{cde}$ (0.01) | |
| | Relatedness (5K) | No | $0.21^{bcd}$ (0.019) | $0.24^{cd}$ (0.009) | $0.18^{cd}$ (0.008) | $0.14^{b}$ (0.01) |
| | | Yes | $0.22^{bc}$ (0.019) | $0.25^{b}$ (0.011) | $0.19^{ab}$ (0.007) | |
| | Relatedness (50K) | No | $0.22^{bc}$ (0.024) | $0.25^{b}$ (0.006) | $0.18^{bc}$ (0.005) | $0.15^{a}$ (0.01) |
| | | Yes | $0.25^{a}$ (0.027) | $0.27^{a}$ (0.003) | $0.20^{a}$ (0.009) | |
| 0.5 | Inbreeding (5K) | No | $0.34^{ab}$ (0.06) | $0.31^{c}$ (0.004) | $0.26^{d}$ (0.012) | $0.13^{d}$ (0.015) |
| | | Yes | $0.37^{ab}$ (0.059) | $0.33^{bc}$ (0.004) | $0.30^{bcd}$ (0.016) | |
| | Inbreeding (50K) | No | $0.36^{ab}$ (0.058) | $0.32^{bc}$ (0.0009) | $0.28^{cd}$ (0.024) | $0.14^{cd}$ (0.015) |
| | | Yes | $0.37^{ab}$ (0.071) | $0.33^{bc}$ (0.004) | $0.30^{bcd}$ (0.023) | |
| | Random (5K) | No | $0.37^{ab}$(0.073) | $0.35^{bc}$(0.052) | $0.30^{bcd}$(0.036) | $0.14^{cd}$ (0.01) |
| | | Yes | $0.37^{ab}$(0.059) | $0.35^{bc}$(0.1) | $0.32^{bc}$ (0.054) | |
| | Random (50K) | No | $0.29^{b}$ (0.114) | $0.31^{c}$ (0.081) | $0.29^{bcd}$ (0.053) | $0.14^{cd}$ (0.015) |
| | | Yes | $0.36^{ab}$ (0.059) | $0.37^{ab}$ (0.04) | $0.33^{abc}$ (0.027) | |
| | Relatedness (5K) | No | $0.37^{ab}$ (0.068) | $0.37^{abc}$(0.045) | $0.33^{abc}$(0.049) | $0.18^{ab}$ (0.031) |
| | | Yes | $0.37^{ab}$ (0.071) | $0.39^{ab}$ (0.05) | $0.34^{ab}$ (0.06) | |
| | Relatedness (50K) | No | $0.36^{ab}$ (0.062) | $0.38^{ab}$ (0.036) | $0.34^{ab}$ (0.049) | $0.20^{a}$ (0.024) |
| | | Yes | $0.40^{a}$ (0.067) | $0.41^{a}$ (0.04) | $0.37^{a}$ (0.056) | |

[a] Groups with same heritability and different letters within each column are significant (P< 0.01).

models on accuracy in crossbred populations. Selecting of reference population based on relatedness was still an appropriate approach to reduce the number of reference set for this population structure.

The results of crossbred population also revealed that using a higher density panel (50K) may have a beneficial effect on accuracy of the GEBVs. In the simulated trait with heritability of 0.5, the highest accuracy of breeding values were achieved after imputation of 50K density panel to a high-density panel. These results suggest that in crossbred populations, using high-density panels or imputation from low- to high-density panels could improve accuracy of GEBVs.

**Backcross Population**

The results of the accuracy of GEBVs in the various marker densities with or without genotype imputation under a different weight parameter (λ= 0, 0.2, 0.5 and 1) are shown in Table 4. These accuracies were estimated in different selection methods for reference population sub-setting for simulated traits with the heritability levels of 0.25 and 0.5 in the backcross population.

Prediction accuracy results obtained from the backcross population was similar to the results of the crossbred population. In this population, similar to the crossbred for the trait simulated with the low heritability, combining pedigree and marker information





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

(SSGBLUP with λ= 0.2) resulted in improvement of the accuracy of breeding values prediction, and the highest accuracy was achieved when references were selected based on relatedness and 50K panel density imputed to a high-density panel. Therefore, in this population, due to the complexities of breeding structure, using the pedigree information along with a small weight for genomic information can result in a better estimation of GEBVs. The results showed that with increasing the level of heritability, genomic accuracy also increased. Selecting reference population based on relatedness showed the highest accuracy, which means that this method of sub-setting was an appropriate approach to reduce the number of reference set.

Silva *et al*. (2016) studied the relationship between the reference population and three sampled validation populations (random, young, and unrelated) by using the pedigree relationship matrix and its influence on the genomic prediction accuracy. In their study, the random population had the highest relationship between the reference and validation populations in which 2.14% of the animals had relationship coefficients between0.25 to 0.5 in both reference and validation data sets. The corresponding estimations for the young and unrelated validation populations were 1.87 and 0.53%, respectively. In their study, the average of genomic predictions accuracies was higher in the random dataset. In the purebred population, the accuracies of genomic breeding values prediction in GBLUP (λ= 0) method in all scenarios of reference population selection were considerably higher than the traditional method of ABLUP (λ=1) and changed from 0.25 for the trait with heritability of 0.25 to 0.57 for the trait with heritability 0.5, while in the traditional method of ABLUP the accuracies of breeding values estimation were in the range of 0.14 to 0.35. The reason for the higher accuracy of genomic evaluations in GBLUP method compared to ABLUP is the use of all the variances between and within family in the genomic evaluations. The genomic selection by means of markers makes it possible to estimate the Mendelian sampling variance with a high accuracy that will lead to a better differentiation within families and a stable genetic gain, while in the traditional method of ABLUP selection all full sibs without record have the same breeding value. Villumsen *et al*. (2009) conclude that using genomic relationship matrix is more efficient than using the predicted relationship matrix for calculating the breeding value accuracy, because the pedigree-based relationship matrix has no ability for registering the Mendelian sampling effects, while the relationship marker matrix is able to calculate this effect. These finding are consistent with our results.

In both traditional and genomic evaluations, the trait with high heritability (0.5) had higher prediction accuracies than the trait with lower heritability (0.25). The reason for the reduction of genomic breeding values prediction accuracy in the lower heritability's is to increase in the estimates of Mendelian sampling variance of marker effects along with the increase of environmental variance (Lopes *et al*., 2017; Meuwissen *et al*., 2001). In the purebred populations using different scenarios of reference population, when the reference population selection is based on a high relationship with the validation population, it could yield a significant improvement in the breeding values accuracy (Tables 2). The results revealed that the increase in relationship between the validation population and the reference population increased the accuracy of genomic breeding values, consistent with the results obtained by Hayes *et al*. (2009) and Clark *et al*. (2012). The Stronger relationship between the reference population and validation population increases the efficiency of using LD due to common blocks, which are established between the related animals resulting from linkage disequilibrium between markers and gene loci. Also, the higher relationships, due to sharing more haplotypes, play an important role in the results related to the accuracy of GEBVs. By





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

studying the accuracy of genotype imputation in the purebred and crossbred sheep populations and its effect on the accuracy of genomic predictions, Moghaddar *et al*. (2015) reported that the crossbred animals need larger reference populations that have genotypes for all related breeds. The accuracy of genotype imputation in the purebred and crossbred population is increased when the breed-specific haplotypes are available in the reference population.

The accuracy of breeding values estimation in the purebred populations was relatively higher than cross and backcross populations. In the purebred population with more identical by descent loci, more common haplotypes are shared and the genetic interval between haplotypes in the reference and validation populations becomes shorter. Thus, the accuracy of imputation and, consequently, the accuracy of genomic breeding values prediction in the purebred is higher than the crossbred populations, which is consistent with the results obtained by Moghaddar *et al*. (2015). By studying the different strategies for genotype imputation in the crossbred dairy cattle populations (Guernsey×Holstein), Oliveira Júnior *et al*. (2017) showed that the highest imputation accuracy was observed when crossbred animals entered the reference population, but using only Guernsey animals in the reference population resulted in a low imputation accuracy. Their results revealed that haplotypes segregation in the reference population had more effect on the accuracy compared to the purebred haplotypes, and the crossbred animals should be included in the reference population to obtain the best genotype imputation accuracies. Lopes *et al*. (2017) studied the genomic selection in the purebred and crossbred populations and reported that considering the allele breed origin of alleles and using a model that considers the breed- specific effects can improve the accuracy of genomic prediction. They also concluded that when the breed-specific effects are considered, the use of crossbred data in the reference population results in a genomic prediction accuracy higher than the purebred data.

## CONCLUSIONS

In the present research, the accuracy of GEBVs in a different population structure including purebred, crossbred, and backcross populations was studied based on different combinations of pedigree and marker information and various methods of reference set selection. A higher accuracy of breeding values prediction in the purebred populations compared to the crossbred populations reveals that the haplotypes segregated in the purebred populations had probably more influence on the imputation accuracy and, consequently, on the accuracy of genomic breeding values prediction. The higher breeding values accuracy in the state of selecting the reference population with a strong relationship with the validation population demonstrates the importance of sharing more haplotypes. Additionally, the results of this research revealed that in most scenarios studied, GBLUP method (using only G matrix) resulted in the highest accuracy of genomic breeding values prediction. However, in the crossbred and back crossbred populations for the trait of interest with low heritability, using the pedigree information along with a small weight for genomic information can result in a more accurate estimation of GEBVs.

## ACKNOWLEDGEMENTS

We appreciate help from Dr. Mehdi Sargolzaei for sharing the commercial version of FImpute software.

## REFERENCES

1. Boison, S. A., Utsunomiya, A. T. H., Santos, D. J. A., Neves, H. H. R., Carvalheiro, R. and Mészáros, G. 2017. Accuracy of Genomic Predictions in Gyr (Bosindicus) Dairy Cattle. *J. Dairy Sci.*, **100**, 1–12.





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

2. Chang, L., Toghiani, S., Ling, A., Aggrey, S. E. and Rekaya, R. 2018. High Density Marker Panels, SNPs Prioritizing and Accuracy of Genomic Selection. B*MC Genetics*, **19**: 4.
3. Calus, M. P. L. 2010. Genomic Breeding Value Prediction: Methods and Procedures. *Animal*, **4**: 157-164.
4. Chen, L., Li, C., Sargolzaei, M. and Schenkel, F. 2014. Impact of Genotype Imputation on the Performance of GBLUP and Bayesian Methods for Genomic Prediction. *PLoS One*, **9**: e101544.
5. Clark, S. A., Hickey, J. M., Daetwyler, H. D. and van der Werf, J. H. 2012. The Importance of Information on Relatives for the Prediction of Genomic Breeding Values and the Implications for the Makeup of Reference Data Sets in Livestock Breeding Schemes. *Genet. Select. Evol.*, **44**: 4.
6. De Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D. and Calus, M. P 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*, **193**: 327-345.
7. Georges, M., Charlier, C. and Hayes, B. 2019. Harnessing Genomic Information for Livestock Improvement. *Nat. Rev. Genet*., **20**: 135-156.
8. Gray, K. A., Cassady, J. P., Huang, Y. and Maltecca, C. 2012. Effectiveness of Genomic Prediction on Milk Flow Traits in Dairy Cattle. *Genet. Sel. Evol*., **44**:24.
9. 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.
10. Larmer, S., Sargolzaei, M., Brito, L., Ventura, R. and Schenkel, F. 2017. Novel Methods for Genotype Imputation to Whole-Genome Sequence and a Simple Linear Model to Predict Imputation Accuracy. *BMC Genet*., **18**:120.
11. Lopes, M. S., Bovenhuis, H., Hidalgo, A. M., Arendonk, J. A., Knol, E. F. and Bastiaansen, J. W. 2017. Genomic Selection for Crossbred Performance Accounting for Breed-Specific Effects. *Genet. Sel. Evol*., **49**:51.
12. Meuwissen, T. H., Hayes, B. J. and Goddard, M. E. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, **157**: 1819-1829.
13. Moghaddar, N., Gore, K. P., Daetwyler, H. D., Hayes, B. J. and van der Werf, J. H. J., 2015. Accuracy of Genotype Imputation Basedon Random and Selected Reference Sets in Purebred and Crossbredsheep Populations and Its Effect on Accuracy of Genomic Prediction. *Genet. Sel. Evol.,* **47**: 97.
14. Momen, M., Ayatollahi Mehrgardi, A., Amiri Roudbar, M., Kranis, A., Mercuri Pinto, R., Valente, B. D., Morota, G., Rosa, G. J .M. and Gianola, D. 2018. Including Phenotypic Causal Networks in Genome-Wide Association Studies Using Mixed Effects Structural Equation Models. *Front. Genet*., **9**:455.
15. Mrode, R., Ojango, J. M. K., Okeyo, A. M. and Mwacharo, J. M. 2019. Genomic Selection and Use of Molecular Tools in Breeding Programs for Indigenous and Crossbred Cattle in Developing Countries: Current Status and Future Prospects. *Front. Genet*., **9**: 694.
16. Oliveira Junior, G. A., Chud, T. C. S., Ventura, R.V., Garrick, D. J., Cole, J. B., Munari, D. P., Ferraz, J. B. S., Mullart, E., DeNise, S. and Smith, S. 2017. Genotype Imputation in a Tropical Crossbred Dairy Cattle Population. *J. Dairy Sci*., **100**: 1–12.
17. Sargolzaei, M. and Schenkel, F. 2009. QMSim: A Large-Scale Genome Simulator for Livestock. *Bioinformatics*, **25**, 680-681.
18. Sargolzaei, M., Chesnais, J. P. and Schenkel, F. S. 2014. A New Approach for Efficient Genotype Imputation Using Information from Relatives. *BMC Genomics*, **15**: 478.
19. Schaeffer, L.R. 2006. Strategy for Applying Genome-Wide Selection in Dairy Cattle. *J. Anim. Breed. Genet*., **123**: 218-223.
20. Schrooten, C., Dassonneville, R., Ducrocq, V., Brondum, R., Lund, M. and Chen, J. 2014. Error Rate for Imputation from the Illumina Bovine SNP50 Chip to the Illumina Bovine HD Chip. *Genet. Sel. Evol*., **46**:10.
21. Silva, R. M. O., Fragomeni, B. O., Lourenco, D. A. L., Magalhães, A. F. B., Irano, N. and Carvalheiro, R. 2016. Accuracies of Genomic Prediction of Feed Efficiency Traits Using Different Prediction and Validation Methods in an Experimental Nelore Cattle Population. *J. Anim. Sci.*, **94**, 3613–3623.
22. Van Binsbergen, R., Bink, M. C., Calus, M. P., Van Eeuwijk, F.A., Hayes, B. J., Hulsegge, I. and Veerkamp, R. F. 2014. Accuracy of Imputation to Whole-Genome





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

Sequence Data in Holstein Friesian Cattle. *Genet. Sel. Evol.*, **46**: 41.
23. VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.*, **91**: 4414-4423.
24. Villumsen, T. M., Janss, L. and Lund, M. S. 2009. The Importance of Haplotype Length and Heritability Using Genomic Selection in Dairy Cattle. *J. Anim. Breed. Genet.*, **126**: 3-13.
25. Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F. and Xiang, J. 2017. Effects of Marker Density and PopulationStructure on the Genomic Prediction Accuracy for Growth Trait in Pacific White Shrimp *Litopenaeus vannamei. BMC Genetics,* **18:** 45.
26. Weigel, K. A., Van Tassell, C. P., O'Connell, J. R., VanRaden, P. M. and Wiggans, G. R. 2010. Prediction of Unobserved Single Nucleotide Polymorphism Genotypes of Jersey Cattle Using Reference Panels and Population-Based Imputation Algorithms. *J. Dairy Sci.*, **93**: 2229–2238.
27. Zhang, Z., and Druet, T. 2010. Marker Imputation with Low-Density Marker Panels in Dutch Holstein Cattle. *J. Dairy Sci.*, **93(11):** 5487-5494.

## تاثیر استنباط ژنوتیپی، ساختار جمعیت مرجع و تراکم پنل SNP بر صحت ارزیابی ژنومی در جمعیت‌های خالص و آمیخته

**ش. برجسته، غ. ر. داشاب، م. رکوعی، م. م. شریعتی، و م. وفای واله**

### چکیده

هدف از این تحقیق، مقایسه صحت پیش‌بینی ارزش‌های اصلاحی ژنومیکی با تراکم‌های مختلف نشانگری قبل و بعد از ایمپیوت در جمعیت‌های شبیه‌سازی شده خالص و آمیخته بر اساس سناریوهای مختلف انتخاب جمعیت مرجع و روش‌های مختلف برآورد آثار نشانگری بود. جمعیت‌های شبیه‌سازی شده شامل دو جمعیت خالص (لاین A و B) و دو جمعیت آمیخته (کراس و بک‌کراس) بودند. سه سناریو مختلف در ارتباط با نحوه انتخاب دام‌ها در جمعیت مرجع شامل: ۱-رابطه خویشاوندی بالا با جمعیت تأیید ۲-تصادفی ۳-همخونی بالا، برای ایمپیوت حیوانات جمعیت تأیید با تراکم‌های ۵K و ۵۰K به تراکممارکری ۷۷K ارزیابی شدند. سپس صحت برآورد ارزش‌های اصلاحی در افراد جمعیت تأیید، قبل و بعد از امپیوت با روش‌های ABLUP، GBLUP وSSGBLUP در دو سطح وراثت‌پذیری ۰/۲۵ و ۰/۵ محاسبه گردید. نتایج نشان داد که حداکثر صحت پیش‌بینی ارزش‌های اصلاحی در جمعیت‌های خالص با روش GBLUP و در سناریو انتخاب جمعیت مرجع خویشاوند با حیوانات جمعیت ایمپیوت (تأیید) بود. همچنین نتایج نشان دادند در سناریو انتخاب جمعیت مرجع خویشاوند در زمان استفاده از پنل ۵۰K برای استنباط ژنوتیپی به تراکم ۷۷K صحت پیش‌بینی ارزش-های اصلاحی ژنومی افزایش یافت، ولی در اکثر سناریوهای انتخاب جمعیت مرجع همخون و تصادفی





[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]

تفاوت معنی‌داری در صحت پیش‌بینی ارزش‌های اصلاحی ژنومی بین تراکم‌های ۵K و ۵۰K بعد از استنباط ژنوتیپی به ۷۷K وجود نداشت.



[ DOR: 20.1001.1.16807073.2021.23.1.11.4 ]