Journal of Livestock Science and Technologies, **2019, 7 (1): 45-53** ***http://lst.uk.ac.ir***

DOI: 10.22103/jlst.2019.13608.1261

Estimation of genotype imputation accuracy using reference populations with varying degrees of relationship and marker density

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

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

[2]Department of Animal Science, College of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran.

*Corresponding author, E-mail address: dashab@uoz.ac.ir

Abstract Genotype imputation from low-density to high-density (SNP) chips is an important step before applying genomic selection, because denser chips can provide more reliable genomic predictions. In the current research, the accuracy of genotype imputation from low and moderate-density panels (5K and 50K) to high-density panels in the purebred and crossbred populations was assessed. The simulated populations included two purebred populations (lines A and B) and two crossbred populations (cross and backcross). Three scenarios were assessed for selecting the subset of the references that used to impute un-genotyped loci of animals in the validation set, where: 1) high relationship with validation set, 2) randomly, and 3) high inbreeding selecting. Imputing the individuals of validation set 5K and 50K to marker density 777K using the various combinations of reference set was performed by FImpute software. The imputation accuracies were calculated using two methods including Pearson correlation coefficient (PCC) and concordance rate (CR). The results showed that imputation accuracy in the purebred populations lines A and B was higher than the cross and backcross populations. When the reference set has been selected based on high relationships, the genotype accuracy in lines A and B was the highest, and there was less difference between imputation from 5K and 50K density to 777K compared to the other subset selection methods. In the crossbred population with imputation from 50K to 777K, the imputation accuracy was the highest in the state of the randomly selected of the reference population (0.98 and 0.97 for PCC and CR, respectively). In the backcross population, the imputation accuracy was the lowest when the reference set selected according to the high inbreeding, which it could be resulting from the lower homozygosis in these populations.

Keywords: genotype imputation, single nucleotide polymorphism, accuracy

Received: 16 Feb. 2019, *accepted:* 22 May. 2019, *published online:* 14 Oct. 2019

Introduction

Recent advances in DNA sequencing technology and availability of high-density single nucleotide polymorphism (SNP) genotyping platforms have provided unique opportunities to promote the breeding programs in livestock, poultry and plant species to better understand the genetic basis of complex traits. More accurate breeding values were obtained by such genomic information (Momen et al., 2018a). The superiority of genomic selection is possible only when high densities of SNP panels are used to track genes and the effects of QTLs on the traits. Unfortunately, even with decreasing continuously the genotyping costs, only a small fraction of the population was genotyped by these high-density panels. In order to reduce the genomic selection costs, often a larger fraction of population is genotyped by low-density and low cost SNP panels and then imputed to a higher density. Imputing from a low-density panel to a high-density SNP panel has been recently used a common and operational method in the various species of genome breeding programs (Pimentel et al., 2013).

To estimate genomic breeding values, dense marker panels are needed to use the linkage disequilibrium (LD) between quantitative trait loci (QTL) and markers which they can control traits by partitioning them into direct, indirect, and total SNP effects (Hayes et al., 2009; Momen et al., 2018b). A dense marker map is also prerequisite for appropriate mapping in order to precisely locate QTL (Meuwissen et al., 2001). Although high-density genotyping is possible for dairy cattle, genotyping thousands of individuals with high-density panel is still expensive. To reduce the genotyping costs, the reference population can be genotyped by a high-

density panel, while the other animals are genotyped by a low-density panel in which markers are spaced evenly. Then, by using information resulted from the reference population, genotypes for missed loci can be inferred for individuals genotyped with the low-density panel (Habier et al., 2009).

Many factors have been suggested as influencing imputation accuracy, including reference population's structure, the number of animals in reference population, allele frequency, the position of SNPs on the chromosome and the density of SNP used for the reference population. (Bolormaa et al., 2015; Calus et al., 2014). Jattawa et al. (2016) studied the imputation accuracy from low-density chips to intermediate-density chips in Thailand's multi-generational dairy cattle. They showed that the imputation accuracies varied from 76.79% through 93.94% depending on which algorithm implemented, and the combined family and population–based algorithm has a higher imputation accuracy than the other algorithms. Studying the imputation accuracy on Holstein, Guernsey and Ayrshire population by use of FImpute and Beagle software, Larmer et al. (2014) reported that there was very little difference in imputation accuracy between these two methods when the Holstein reference population was used; however, FImpute algorithm outperformed when the size of reference population in Guernsey and Ayrshire population was reduced.

By studying of two dairy cattle populations of France and Scandinavia, Dassonneville et al. (2012) reported that with increasing the number of reference population animals the markers imputation error rate reduced up to 5.5% and 3.9% in the Scandinavian Peninsula and France, respectively. According to them, the lower error rate of imputation between France's dairy cattle resulted from using more markers due to the way of markers edition and having denser genome. By assessing the imputation accuracy using different densities of SNP panel in the sheep multi-bred population of New Zealand, Ventura et al. (2016) demonstrated that imputation from 5K panel to high-density (600K) was slightly better (0.6%) than imputation from 5K to 50K. Two step imputation from 5K to 50K and then from 50K to high-density (600K) outperformed imputation from 5K to 600K. Also, they reported a slight loss in imputation accuracy when a large fixed reference population was used compared to a smaller within-bred reference population. Calus et al. (2014) assessed the different measures of genotype imputation accuracy and showed that correlation between the true and imputed genotypes is the most beneficial and unbiased method for calculating the imputation accuracy.

There is a few research about the level of imputation accuracy in the purebred and crossbred populations with different relationship levels. Therefore, the present research was carried out in order to study the different scenarios of selecting livestock composition in the reference populations and also the effects of 5K and 50K panels promotion to 777K panel density in the purebred and crossbred populations.

Materials and methods

Populations (Purebred and crossbred lines)

Populations were simulated by using the QMSim software (Sargolzaei et al., 2009). In the first simulation step, to create the initial LD between marker and QTL and establish mutation - drift equilibrium a historical population constituted of 500 animals (250 males and 250 females) was created. Then 1000 randomly mated generations without changing numbers and 1000 mated generations with a gradual increase of numbers to 4000 animals were simulated. The number of males in the last base population generation was considered 50 animals, then to create the first purebred population (line A) 20 males and 200 females were selected from the last base population generation. In this step 10 generations were simulated in which there was two progenies per dam. To create the second purebred population (line B), 20 males and 200 females were selected from the last generation of the historical population and 10 mating generations were produced. For lines A and B the genotype, phenotype and pedigree information related to the generations 8, 9 and 10 was registered. In these populations the generation 10 was selected as a validation set and the generations 8 and 9 were selected as the reference set. In the next step, the hybrid populations (cross and backcross) were simulated. The cross population was created from mating 20 males from generation 10 of line A and 200 females from generation 10 of line B. The backcross population was generated from mating 20 males from generation 10 of line A and 200 females from cross population. One randomly mating generation was produced, as illustrated in Figure1. 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 considered as reference set.

Genome and genotypes

In the genome simulation section, a genome consisting of 29 pairs of autosomes with different length and simi-

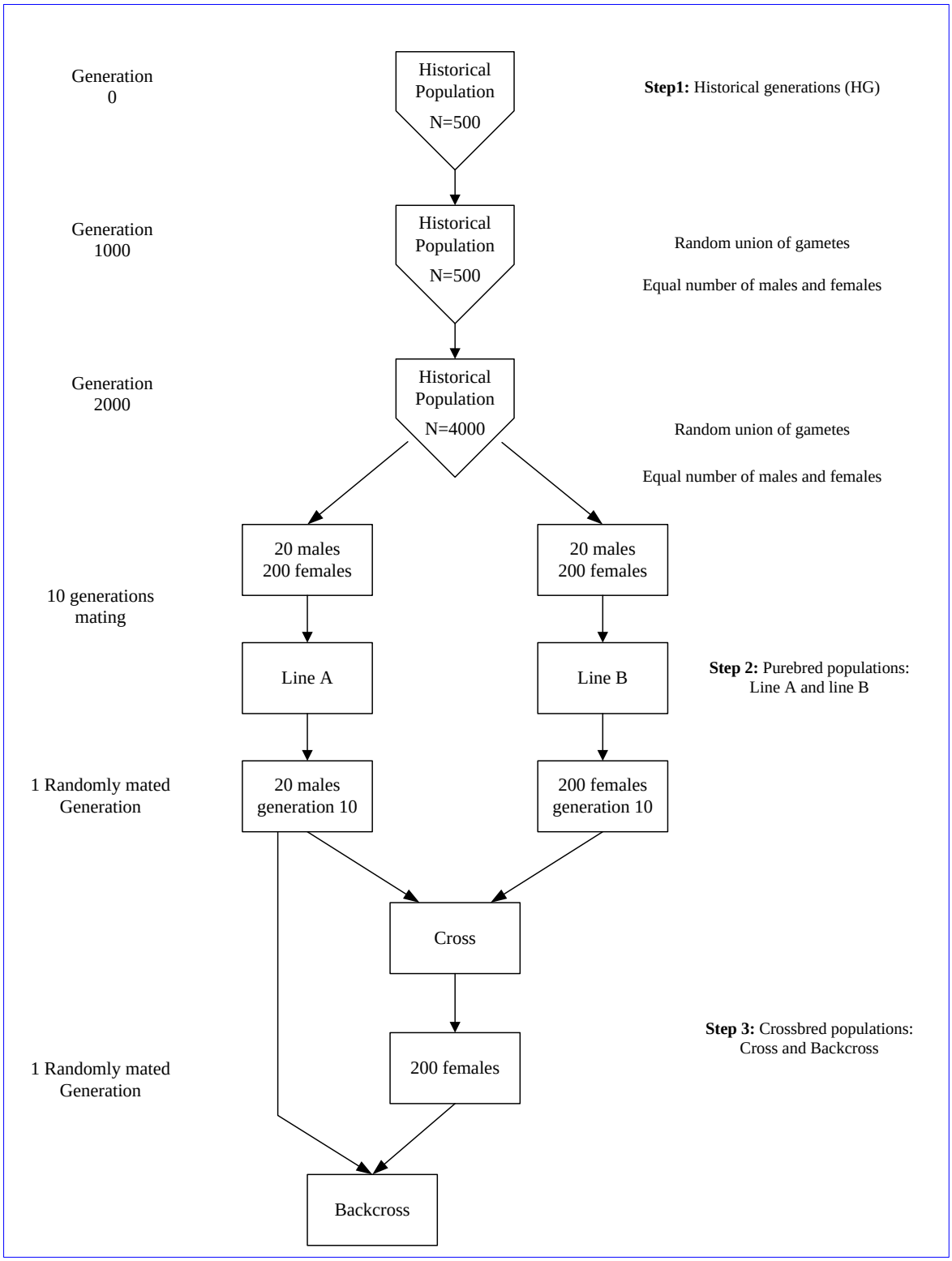

Figure 1. Schematic representation of the simulation steps. The pure breeding schemes (Line A, Line B) started in step 2 and consisted of 20 males and 200 females that were selected from the last historical population generation. The crossbreeding scheme started in step 3 and consisted of 20 males from generation 10 of line A and 200 females from generation 10 of line B (Cross). The backcross population consisted of 20 males from generation 10 line A and 200 females cross from population.

lar to the bovine chromosomes size was simulated. Then, the number of 777026 bi-allelic markers and 725 bi-allelic QTLs with the initial allele frequency identical to evenly distribution were randomly positioned on the genome. QTLs allele effects were sampled from a Gamma distribution with a shape parameter of 0.4. Table 1 shows the parameters that were used for simulation.

After simulating data for 777K density, sampling for 5K and 50K densities was performed and the 5K and 50K genome files for imputation populations (validation set) were established. In the next step, the quality control was performed and SNPs with minor allele frequency (MAF) less than 0.01 and the monomorphic loci were deleted in which numbers of 369091 SNPs were deleted and 407935 SNPs with known loci were left for analysis. After sampling, the 5K and 50K panels constituted of 4975 SNPs and 50992 SNPs, respectively. Actually 4975 and 50992 SNPs from 5K and 50K panels corresponded with 777K panel.

Imputation scenarios

Four scenarios were considered for imputing missing genotypes. In all of these scenarios the animals of reference population were genotyped by the high-density panel (777K). While the other animals in imputation population were genotyped with low-density (5K) or intermediate-density (50K) panels that using the information that were obtained from reference population the genotypes of missed loci for the genotyped individuals with low-density panel were inferred.

In scenarios 1 and 2, which were related to the purebred populations' imputation of lines A and B, the animals of generations 9 and 10 of lines A and B were selected as reference population, which had genotypes

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

Population structure	Information of population simulation
Step 1:Creating base population	
Number of generations of base population	500[0]500[1000]4000[2000]
Number of males in the last generation of base population	50
Step 2: 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
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
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

from 777K panel. To impute the 5K and 50K, the generation 10 within lines was used which included 400 animals and were considered as the validation population. In scenario 3, the cross population animals were selected as the validation population and animals of generation 10 from line A and generation 10 from line B were selected as the reference population and used to impute the cross population. In scenario 4, the backcross animals were selected as the validation population and the animals of generation 10 from line A and cross animals were selected as the reference population and used to impute the animals of backcross population. In all scenarios of imputation, the number of reference population animals was 800 of which 400 animals were selected as the reference population by 3 methods including: 1) animals which had the highest relationship level with the reference population, 2) animals which had been selected randomly, and 3) animals which had a high inbreeding.

Genotype imputation was performed by FImpute software (Sargolzaei et al., 2011). FImpute uses the pedigree imputation algorithm and then uses the population imputation steps based on an Overlapping Sliding Window (OSW) method. In comparison to most of population imputation software, FImpute assumes that all animals are related and uses OSW to find the haplotype fragments, which have associated with common ancestor between individuals. Initially, imputation is conducted in the family individuals if pedigree information is available and then the close relatives are studied by searching for the large haplotype regions in reference population by OSW and going to search in smaller haplotype regions in the distant relatives and this process is frequently repeated.

Imputation accuracy

Concordance rate (CR) and Pearson correlation coefficient (PCC) as imputation accuracy criteria for each animal were individually calculated with 10 repetitions. Concordance rate was expressed as follow:

$$CR=M/G \quad (1)$$

where, M is the proportion of matching genotypes; and G represent the total proportion of genotypes. Several reports show that PCC does not relate to the allele frequency (Calus et al., 2014) and the relationship between them is as follows:

$$PCC_{anim}=\frac{\sum_{j=1}^{L_k}(g_{ij}-\bar{g})(\hat{g}-\bar{\hat{g}})}{\sum_{j=1}^{L_k}(g_{ij}-\bar{g})^2\sum_{j=1}^{L_k}(\hat{g}_{ij}-\bar{\hat{g}})^2} \quad (2)$$

where L shows the total number of markers to be imputed. g_{ij} and $\hat{g}_{ij}$ show the observed and imputed genotypes for SNP_j of individual i; $\bar{g}_{ij}$ and $\bar{\hat{g}}_{ij}$ show the average values of observed and imputed genotypes, respectively.

Results and discussion

The results of imputation accuracies estimated based on CR and PCC methods in different combinations of the densities and the subset selection scenarios in the reference population are shown in Figure 2. Our results were in agreement with the previous study conducted by Ma et al. (2013). They obtained the accuracy results similar to ours by imputing genotypes from 5K to HD panel in Swedish and Finnish red cattle. We showed that the imputation accuracy in the purebred populations (lines A and B) were higher compared with the other populations (crossbred and backcross populations). These results confirmed that imputation accuracy depends on the relevant haplotypes number in the reference set. This can be resulted most likely due to a more relationship to each other through common ancestor in purebred populations, which they share more haplotypes and the genetic distance between haplotypes in the reference set and the training set. Therefore, imputation accuracy in the purebred populations is higher than hybrid populations (Moghaddar et al., 2015).

In the purebred populations under different states of selecting reference population, a significant improvement in imputation accuracy were observed when the individuals of reference population were selected based on high relationship with validation set in compared to subset selection based on inbreeding and random. These results show that with increasing of relationship between the imputed and reference animals can increase the imputation accuracy. This was in agreement with the results of Ma et al. (2013) and Hickey et al. (2012).

In the purebred populations, the imputation accuracy from 50K density to 777K density in all scenarios was significantly higher compared to the imputation accuracy from 5K to 777K density ($P<0.01$). This demonstrates that due to sharing more haplotypes, a higher relationship has more importance in the results of imputation accuracy, and panel density is more important when there is a lower genomic relationship between validation set and the animals of reference set. Ventura et al. (2014) reported that the imputation accuracy is from 5K to 50K in multi-breed beef population with concordance rate static. Results showed that imputation accuracy is in the highest level when there is a close relationship between reference set and validation set.

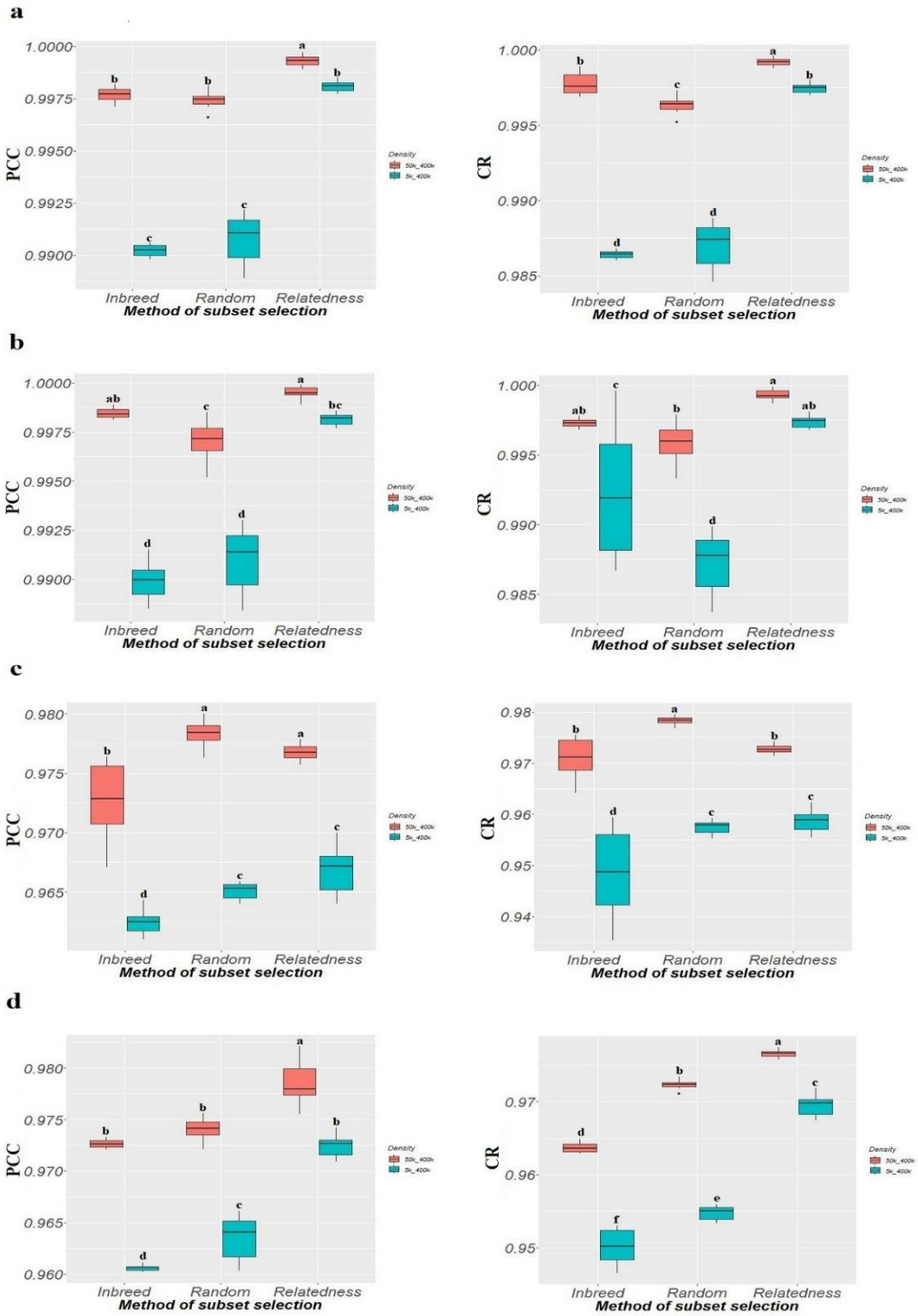

Figure 2. Accuracy of genotype imputation based on subset selection in the reference set (high relationship with validation set, random and high inbreeding) using low-density (5K) and intermediate-density (50K) panels to high-density (777K) panel based on Pearson correlation coefficient (PCC) and concordance rate (CR). Accuracies estimated in different population structure including: purebreds (line A (a) and line B (b)), cross (c) and backcross (d). The PCC and CR with different letters on their boxes indicate significant differences (empirical P< 0.01).

In crossbred population when imputation was performed from 50K density to 777K density panel, the imputation accuracy in the randomly selected of reference set was high (with PCC of 97.13%), while the accuracy increased in the highly related reference population (97.17%) and inbred reference set (96.78%). Higher imputation accuracy in the crossbred population in the state of randomly selected reference set can be a result of complication of breed composition in animal populations. Consequently, the common haplotype blocks-based relationships (Identity by descent) are used when the close relationships exist and haplotypes are small to maintain the relation between different generations. But in the crossbred populations, due to fracture in the haplotype blocks, lower accuracy is observed and in such situations the allele frequency-based replacement can have more precision (Moghaddar et al., 2015). This means that the number and size of common haplotype fragments among the animals in crossbred populations is probably fewer and smaller than haplotype fragments in purebred populations. Therefore, the lower accuracy that were obtained can be a result from a smaller number and shorter fragments of common haplotype among animals in the reference and validation sets (Ventura et al., 2016).

By studying different strategies to impute genotype in crossbred Girolando dairy cattle (Gyr and Holstein), Oliveira Junior et al. (2017) showed that the highest imputation accuracy is when Girolando were used in reference population; however, using Guernsey animals solely cause to reduce the imputation accuracy. Their results showed that segregating haplotypes in Girolando population in compared to purebred haplotypes had more effect on accuracy and crossbred Girolando, which could be included in reference population to create the best imputation accuracies. In the backcross population, the imputation accuracy in the state of the highly related of reference population was significantly higher ($P<0.01$). When the reference population was selected randomly and highly inbreeded, the accuracy reduced significantly ($P<0.01$). Lower imputation accuracy in the state of selecting the reference according to inbreeding in the crossbred and backcross populations can be resulted from lower importance of inbreeding in the crossbred populations due to lower homozygosity in these populations which was announced by the previous study (Ventura et al., 2016).

These results show that imputation accuracy depends on panel density, population structure and the way of selecting reference population. It means that there is no single method for providing a higher imputation accuracy for all scenarios. Nevertheless, an important advantage of imputation is reducing the genomic selection costs and increasing the accuracy of genome breeding values prediction. In general, performance and accuracy of imputation depend on several factors.

Conclusions

Higher imputation accuracy in the purebred populations in comparison to crossbred populations showed that haplotypes segregated in the purebred sets had a greater influence on the imputation accuracy than the crossbred haplotypes. Higher imputation accuracy in purebred populations based on the selection of individuals from the reference population, most closely resembling with validation set population, reveals the importance of the same haplotypes shared between individuals by common ancestors and transitions from one generation to the next. But, in the crossbred populations, due to segregation and recombination of haplotype during generations, the accuracy was lower. Also, the results of current study provide the useful information on reducing of cost of genotyping in pure and crossbred populations for future genomic selection strategy.

References

Bolormaa, S., Gore, K., van der Werf, J.H.J., Hayes, B.J., Daetwyler, H.D. 2015. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Animal Genetics* 46, 544–556.

Calus, M.P., Bouwman, A.C., Hickey, J.M., Veerkamp, R.F., Mulder, H.A. 2014. Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. *Animal* 8, 1743-1753.

Dassonneville, R., Baur, A., Fritz, S., Boichard, D., Ducrocq, V. 2012. Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. *Genetics Selection Evolution* 44, 40.

Habier, D., Fernando, R.L., Dekkers, J.C.M. 2009. Genomic selection using low-density marker panels. *Genetics* 182, 343–353.

Hayes, B., Bowman, P., Chamberlain, A., Verbyla, K., Goddard, M. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution* 41, 51.

Hickey, J.M., Kinghorn, B.P., Tier, B., van der Werf, J.H., Cleveland, M.A. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genetics Selection Evolution* 44, 9.

Jattawa, D., Elzo, M.A., Koonawootrittriron, S., Suwanasopee, T. 2016. Imputation accuracy from low to moderate-density single nucleotide polymorphism chips in a Thai multi-breed dairy cattle population. *Asian-Australasian Journal of Animal Sciences* 29, 464–470.

Larmer, S.G., Sargolzaei, M., Schenkel, F.S. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *Journal of Dairy Science* 97, 3128–3141.

Ma, P.P., Brøndum, R. F., Zhang, Q., Lund, M.S., Su, G. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish red Cattle. *Journal of Dairy Science* 96, 4666–4677.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Moghaddar, N., Gore, K.P., Daetwyler, H.D., Hayes, B.J., van der Werf, J.H.J. 2015. Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. *Genetics Selection Evolution* 47, 97.

Momen, M., Ayatollahi Mehrgardi, A., Sheikhi, A., Kranis, A., Tusell, L., Morota, G., Rosa, G., Gianola, D. 2018a. Predictive ability of genome-assisted statistical models under various forms of gene action. *Scientific Reports* 8, 12309.

Momen, M., Ayatollahi Mehrgardi, A., Amiri Roudbar, M., Kranis, A., Mercuri Pinto, R., Valente, B.D., Morota, G., Rosa, G., Gianola, D. 2018b. Including phenotypic causal networks in genome-wide association studies using mixed effects structural equation models. *Frontiers in Genetics* 9, 455.

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., Smith, S. 2017. Genotype imputation in a tropical crossbred dairy cattle population. *Journal of Dairy Science* 100, 9623-9634.

Pimentel, E.C., Wensch- Dorendorf, M., Konig, S., Swalve, H.H. 2013. Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. *Genetics Selection Evolution* 45, 12.

Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2011. FImpute-An efficient imputation algorithm for dairy cattle populations. *Journal of Dairy Science* 94 (E-Suppl. 1), 421. (Abstract)

Sargolzaei, M., Schenkel, F.S. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25, 680-681.

Ventura, R. V., Miller, S.P, Dodds K.G., Auvray,B., Lee, M., Bixley, M., Clarke, S.M., McEwan, J.C. 2016. Assessing accuracy of imputation using different SNP panel densities in a multibreed sheep population. *Genetics Selection Evolution* 48, 71.

Ventura, R.V., Schenkel, D., Lu, F.S., Wang, Z., Li, C., Miller, S.P. 2014. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. *Journal of Animal Science* 92, 1433–1444.

Communicating editor:

ارزیابی صحت ایمپیوت ژنوتیپی با استفاده از درجات مختلف خویشاوندی جمعیت مرجع و تراکم نشانگری

ش. برجسته[1]، غ. ر. داشاب[1*]، م. رکوعی[1]، م. م. شریعتی[2] و م. وفای واله[1]

[1]گروه علوم دامی، دانشکده کشاورزی، دانشگاه زابل، زابل، ایران.

[2]گروه علوم دامی، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران.

*نویسنده مسئول، پست الکترونیک: dashab@uoz.ac.ir

چکیده استنباط ژنوتیپی از تراشه‌های چندشکلی تک‌نوکلئوتیدی با تراکم پائین به تراشه‌ها با تراکم بالا گام مهمی پیش از انجام انتخاب ژنومی می‌باشد، زیرا تراشه‌های متراکم‌تر می توانند پیش‌بینی‌های ژنومی قابل اطمینان‌تری را ارائه می‌دهند. در این تحقیق صحت ایمپیوت ژنوتیپی از پنل‌های با تراکم پائین (5K) و متوسط (50K) به تراکم بالا (777K) در جمعیت‌های خالص و آمیخته مورد ارزیابی قرار گرفت. جمعیت‌های شبیه‌سازی شده شامل دو جمعیت خالص (لاین A و B) و دو جمعیت آمیخته (کراس و بک‌کراس) بودند. سه سناریو برای انتخاب دام‌ها در جمعیت مرجع مورد ارزیابی قرار گرفت و برای ایمپیوت جایگاه‌های تعیین ژنوتیپ نشده حیوانات در جمعیت تأیید استفاده شد، سناریوها بر اساس: ۱- رابطه خویشاوندی بالا با جمعیت تأیید ۲- تصادفی ۳- همخونی بالا انتخاب شدند. ایمپیوت افراد جمعیت تأیید 5K و 50K به تراکم نشانگری 777K با استفاده از ترکیب‌های مختلف جمعیت مرجع با استفاده از نرم افزار FImpute انجام شد. صحت‌های ایمپیوت با دو روش ضریب همبستگی پیرسون (PCC) و نرخ مطابقت (CR) محاسبه شد. نتایج نشان داد که صحت ایمپیوت در جمعیت‌های خالص لاین‌های A و B بالاتر از جمعیت‌های آمیخته کراس و بک‌کراس بود. در لاین‌های A و B صحت ایمپیوت ژنوتیپی در حالتی که جمعیت مرجع بر اساس روابط خویشاوندی بالا انتخاب شده بود بالاترین بود و تفاوت کمتری بین ایمپیوت از تراکم 5K و 50K به 777K در مقایسه با روش‌های دیگر انتخاب جمعیت مرجع بود. در جمعیت کراس زمانی که ایمپیوت از 50K به 777K صورت گرفت صحت ایمپیوت در حالت انتخاب جمعیت مرجع تصادفی بالاترین بود (۰/۹۸ و ۰/۹۷ به ترتیب بر اساس PCC و CR). در جمعیت بک‌کراس صحت ایمپیوت در حالت انتخاب جمعیت مرجع همخون در پائین‌ترین حد قرار داشت که می‌تواند به دلیل اهمیت پائین‌تر همخونی به دلیل هموزیگوسیتی کمتر در این جمعیت باشد.