Journal of Livestock Science and Technologies, 2017, 5 (2): 43-52 *http://lst.uk.ac.ir*
DOI: 10.22103/jlst.2017.9332.1164

Imputation of parent-offspring trios and their effect on accuracy of genomic prediction using Bayesian method

M. Kamaei[1*], M. Honarvar[2], M. Aminafshar[1] and R. Abdollahi-Arpanahi[3]

[1]Department of Animal Science, Science and Research Branch, Islamic Azad University, Tehran, Iran.
[2]Department of Animal Science, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran.
[3]Department of Animal Science, College of Abouraihan, University of Tehran, Tehran, Iran.
*Corresponding author, E-mail address: kamaei_62@yahoo.com

Abstract The objective of this study was to evaluate the imputation accuracy of parent-offspring trios under different scenarios. By using simulated datasets, the performance Bayesian LASSO in genomic prediction was also examined. The genome consisted of 5 chromosomes and each chromosome was set as 1 Morgan length. The number of SNPs per chromosome was 10000. One hundred QTLs were randomly distributed across chromosomes. Three low density SNP panels, containing 0.5k, 1k and 5k SNPs, were generated from the 10k panel. Six scenarios were evaluated, each containing two trios (dam, sire and offspring) and sire of each dam for parent-offspring pair data. These scenarios were compared from completely genotyped offspring to low-density genotyped and dams that were completely genotyped, low density genotyped and non-genotyped. It was assumed that the genotypes of the offspring's sires were available. The Beagle 3.3.2 program was used for imputation of parent-offspring trios. The Bayesian LASSO were used to estimate the marker effects using the R package of "BLR". The results showed that accuracy of both imputation and genomic evaluation was influenced by imputation errors. Imputation accuracy ranged from 0.67 to 0.96 for genotyped individuals. Genotype imputation accuracy increased with increasing marker density of low-density genotyping platform and with dams having high-density genotypes. Results showed that imputation accuracies decreased significantly ($P < 0.05$) when dam was non-genotyped and both of offspring were low-density genotyped. In case of factors affecting imputation accuracy, the imputation accuracy of SNPs with low MAF increased considerably when a dam was completely genotyped. Imputation of non-genotyped individuals can help to include valuable phenotypes for genome-wide association studies or for genomic prediction, especially when the non-genotyped individuals have genotyped offspring.

Keywords: low-density, genotype imputation, genotyped individual, non-genotyped

Received: 20 Oct. 2016, *accepted:* 05 Nov. 2017, *published online:* 25 Dec. 2017

Introduction

Breeding values can be predicted with high accuracy using genomic information (Meuwissen et al., 2001). A successful genetic improvement program requires accurate genetic parameter estimates (Molaei Moghbeli et al., 2013). Moreover, due to recent advances in genotyping technologies, the amount of genomic information available for genomic selection (GS) has increased from a few thousands (Sargolzaei et al., 2008), to 50K (Pimentel et al., 2011) and 800K (Erbe et al., 2012). Nowadays, genomic evaluation programs tend towards whole-genome sequence (Ober et al., 2012). Genomic selection combines information on genotypes, phenotypes and pedigree to increase the accuracy of the estimated breeding values (EVBs) (Weigel et al., 2010). In SNPs genotyping data obtained from the SNP chip technique, missing genotype information is a common phenomenon that leads to a low call rate for some SNPs and for some animals. Imputation can be used to predict the missing genotypes and could be helpful in increasing the accuracy of genomic selection. A major challenge in implementing genomic selection in most species is the cost of genotyping (Boichard et al., 2012). Genotype imputation can help reduce genotyping costs particularly for implementation of genomic selection (Sargolzaei et al., 2014). If a relevant genotyping strategy can be chosen such that imputation accuracy becomes sufficiently high, imputation of non-genotyped animals might also be of interest for breeding programs to reduce

genotyping costs (Williams et al., 2012). Genotype imputation is an important process of predicting unknown genotypes, which uses reference population with dense genotypes to predict missing genotypes for both human and animal genetic variations at a low cost (Boichard., 2012). Phasing and imputation methods can be broadly divided into family-based methods (which use linkage information from close relatives) and population-based methods, which use population linkage disequilibrium information (Sargolzaei, 2014). A "trio" data consists of genotypes from father–mother–child triplets and some phasing algorithms are adapted to be used in this type of data (Lu and Cantor, 2014). These conditions make it possible to infer the genotypes of a non-genotyped individual using genomic information from its family members (Pimentel et al., 2013). Often sires and grandsires of these non-genotyped individuals are genotyped. Imputation methods can be divided into family-based methods (which use linkage information from close relatives) and population-based methods, which use population linkage disequilibrium information (Sargolzaei et al., 2014). The accuracy of imputation depends on several factors, such as the number of SNPs in the low density panel, the relationship between the animals genotyped, the effective population size, and the method used (Wellmann et al., 2013).

There are many software programs for imputation which are fast- PHASE (Scheet and Stephens, 2006), MACH (Willer et al., 2008) and Beagle (Browning and Browning, 2009). Some programs are designed for human and livestock populations and others have been used to infer missing genotypes based on known information derived from flanking markers for livestock populations (Sargolzaei et al., 2014). Erbe et al. (2012) used the Beagle software (version 3.3.2) without pedigree information to impute genotypes at 800 k SNPs from dairy bulls genotyped at 50k and reported accuracies of imputation ranging from 0.96 to 0.98 in Jersey and Holstein cattle, respectively. The Beagle software uses population information for imputation. Therefore it is expected that program be able to impute genotypes of animals with incomplete pedigree (Johnston et al., 2011). The Beagle program imputes missing genotypes of animals with and without complete pedigree with high accuracy (Browning and Browning, 2009).

Meuwissen and Goddard (2010) applied a method for imputing whole sequence genotypes on individuals genotyped at a low density panel and reported that 10% of the missing genotypes were erroneously imputed.

Under a SNPs markers whole-genome scans approach, many markers are likely to be located in regions that are not involved in the determination of traits of interest. On the other hand, some markers may be in linkage disequilibrium with some QTL, or in regions harboring genes involved in the infinitesimal component of the trait. This suggests that differential shrinkage of marker effects should be a feature of the model, then an alternative is the use the LASSO (Least Absolute Shrinkage and Selection Operator) regression, which provides good features of subset selection (i.e., variable selection) with the shrinkage theory. De los Campos et al. (2010) proposed a Bayesian approach of LASSO regression in genome-wide selection (GWS), and the validity of this methodology has been reported (Silva et al., 2011). If a large number of markers are included in a regression model, marker-specific shrinkage of regression coefficients may be needed. For this reason, the Bayesian least absolute shrinkage and selection operator (LASSO) appears to be an interesting approach for fitting marker effects in a regression model (De los Campos et al., 2009).

The objectives of this study were to investigate the accuracy of imputation for low-density genotyped offspring of parent-offspring trios and to evaluate the performance of the Bayesian LASSO method when imputed genotypes were used for genomic prediction. To evaluate the factors affecting imputation accuracy, minor allele frequency (MAF) was examined.

Materials and methods

Simulation

Genomic data were simulated using the statistical software package R (R Development Core Team, 2014). The R package of Hypred (Technow, 2015) was used for simulating the genomic data.

The genome consisted of 5 chromosomes and each chromosome was set as one Morgan length. The number of SNPs per chromosome were 10000 and the recombination rate per chromosome was performed using hypredRecombine function obtained from the Hypred. One hundred QTLs were randomly distributed across chromosomes. Gene substitution effects for each QTL were assigned randomly from a standard normal distribution, a ~ N(0,1). Marker allele frequencies in the first historical generation were set equal to 0.5 (Villumsen et al., 2009). An historic population consisted of 100 individuals (50 males,50 females) that were randomly mated during 50 generations using mutation rate of 2.5×10^{8} per site. To reach at a mutation-drift balance, the method of Villumsen et al. (2009) was used. The reference population was generated from the historic population by mating parent groups. The parent groups were randomly selected from the last generation of the

historic population. This structure was followed by 50 generations of random mating. The paternal and maternal haplotypes for each individual were generated based on Haldane mapping function to generate recombinant haplotypes. Eighty-five families were randomly selected from the reference population. Each family contained the dam, its sire (MGS), two offspring (Offspring 1 and Offspring 2) and the offspring's sires (Sire 1, Sire 2). Each family had two parent-offspring trios.

Scenarios

To assess the effect of imputation on the accuracy of estimation, 6 different scenarios were considered (Figure 1). In all scenarios, the missing SNP genotypes rates were 50, 90 and 95 percent under a random missing pattern genotypes. 3 low-density panels (5K, 1K and 0.5K SNPs) were created based on a high-density panel (10K SNP). Our first goal was to determine the accuracy of imputation in low-density offspring, when the dam and an offspring were genotyped (S1). Our second goal was to determine the accuracy of imputation in both low-density offspring, when the dam was genotyped (S2). The third and fourth goals were to assess the accuracy when low-density dams had one and two low-density offspring respectively (S3 and S4). In the scenarios four and six we intended to determine the accuracy of imputation in one and two low-density offspring, respectively, when dam was non-genotyped. In the fourth and sixth scenarios, imputation on low-density and non-genotyped dams was performed using known genotypes from the sire of each dam (MGS) as parent-offspring pair data (Figure 1).

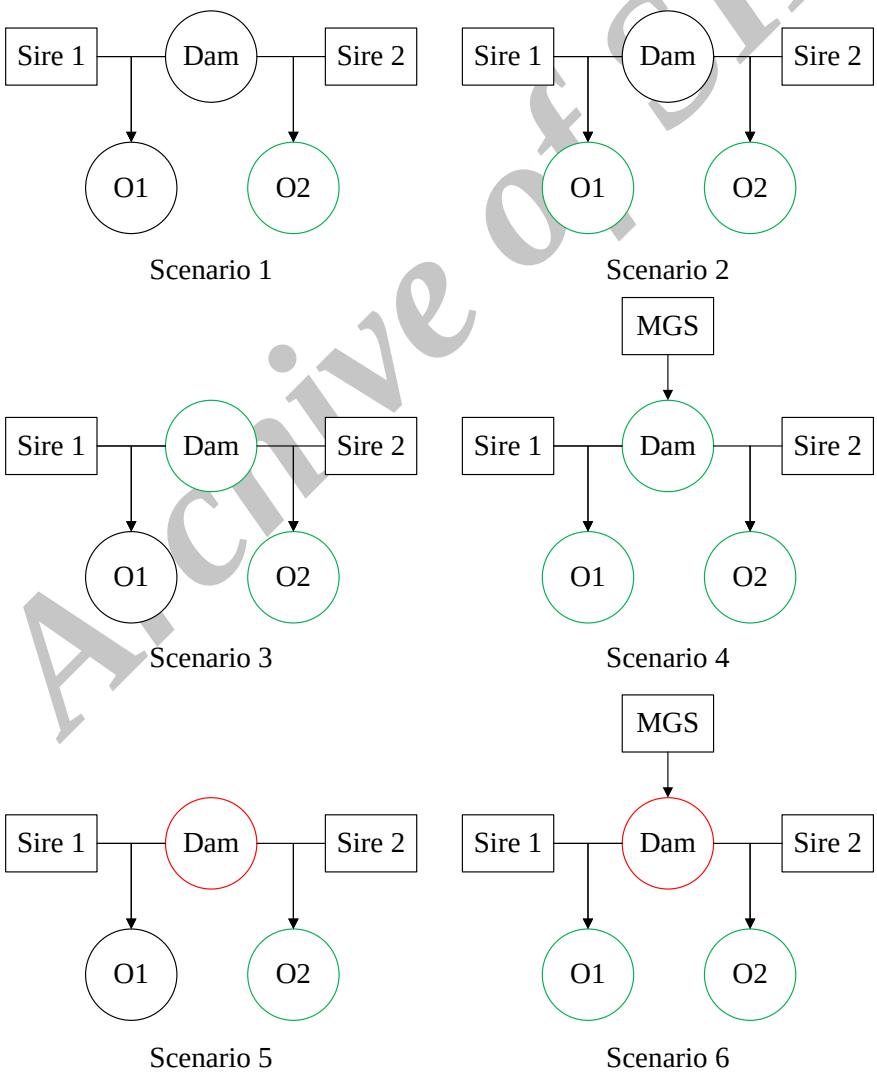

Figure 1. Assumed family members with available genotypic information (black) used for imputing a low-density genotyped (green) or non-genotyped (red) individual. O1 is offspring 1 and O2 is offspring 2

Imputation

The Beagle software (version 3.3.2) has special options allowing the user to provide genotypes from parent-offspring trios and parent-offspring pairs for phasing. Using the haplotype phasing and imputation program, it is possible to impute genotypes from low-to-high density. We used the Beagle software (v.3.3.2) to impute parent-offspring trios and parent-offspring paired data.

Assessing imputation accuracy

After imputation of genotypes in each scenario, the imputation accuracy was calculated as the correlation between imputed and real genotypes. For each scenario, the mean of imputation accuracy, standard deviation, and the percentage of the correct and incorrect imputed SNP genotypes of each individual with 10 replicates were calculated.

Assessing genomic prediction accuracy

A Bayesian implementation of the Lasso method with BLR package in R (De los Campos et al., 2010) was used to estimate marker effects. Genomic breeding value accuracy was defined as the correlation between GEBVs and true breeding values.

Results

SNP-specific Imputation accuracy of non-genotyped and low-density genotyped dams

Table 1 shows the imputation accuracy and corresponding standard deviation (SD) of non-genotyped and low-density genotyped dams. Imputation accuracies ranged from 0.76 to 0.91. The least amount of imputation accuracy and highest SD were achieved for non-genotyped dams. The imputation accuracy for the 5k genotyped dam was the highest. The imputation accuracy decreased with the increase of missing SNPs. The results of each scenario indicated that imputation of low-density genotyped and non-genotyped individuals based on parent-offspring trios and parent-offspring paired is possible.

Table 1. Average imputation accuracy (r) for low-density genotyped and non-genotyped dams

Genotyping	r[a]	SD[b]
5k genotyped	0.91	0.009
1k genotyped	0.82	0.04
0.5k genotyped	0.79	0.06
non-genotyped	0.76	0.07

[a]Mean of imputation accuracy calculated as the correlation between true genotypes and imputed genotype dosages. Values are means across 10 replicates. [b]Standard deviation

SNP-specific Imputation accuracy of offspring

Table 2 shows that average of imputation accuracy and percentage of correctly imputed genotypes increased with genotyped dam and offspring. In S1 (dam and one offspring were genotyped completely) the imputation accuracy and percentage of correctly imputed genotypes were the highest and in S6 (dam was not genotyped and both of offspring were low-density genotyped) were lowest. Genotyping density also affected the accuracy and percentage of correctly imputed genotypes. For example in S6 average of imputation accuracy for 5K genotyped offspring was 0.73 and for 0.5K genotyped offspring was 0.67. In this scenario the average of imputation accuracy was 0.73 for the 5K genotyped offspring.

Animal-specific imputation accuracy

In genomic selection, it is important to know the imputation accuracy per individual, because there is a direct relation with the accuracy of genomic prediction (Mulder et al., 2012) and therefore the response to selection. Imputation accuracy of individuals ranged from 0.70 to 0.85 and was 0.85 for full genotyped dam. The lowest of animal-specific imputation accuracy was for non-genotyped dam (Figure 2)**.**

The effect of MAF on the imputation accuracy

The SNP imputation accuracies increased as the number of offspring genotyped increased (Figure 3). More interesting was the fact that imputation accuracy of SNPs with high MAF decreased considerably when a dam was non-genotyped but imputation accuracy depended less on MAF when a dam and one of the offspring were genotyped (S1). The imputation accuracy of SNPs with low MAF increased considerably when the dam was completely genotyped (Figure 3). In this study, there were no typed SNPs thus only a few SNPs had imputation accuracy equal to 1.

Genomic breeding value accuracy

Three populations were simulated for this section based on the scenarios used. The reference population (P_r): in P_r most individuals were genotyped for all SNPs in each scenario. Validation 1 population (P_1) and Validation 2 population (P_2) had the same individuals: only in P_1 these individuals had true genotypes and in P_2 had imputed genotypes (Table 3).

Table 2. Average imputation accuracy (r) and percentage of correct imputed genotypes for offspring in several each scenarios

	r			correct		
Scenarios	[a]5k	[b]1k	[c] 0.5k	[a]5k	[b]1k	[c] 0.5k
Scenario1	0.96	0.92	0.9	96.17	91.76	91.47
Scenario2	0.88	0.84	0.8	87.64	85.29	80.29
Scenario3						
[d]dam (5k)	0.87	0.83	0.78	87.5	83.12	77.61
[e]dam (1k)	0.81	0.79	0.77	80.78	79.41	77.14
[f]dam (0.5k)	0.8	0.77	0.77	79.9	76.9	76.8
Scenario4						
[d]dam (5k)	0.83	0.78	0.75	83.05	77.51	74.85
[e]dam (1k)	0.75	0.73	0.73	74.7	73.14	72.91
[f]dam (0.5k)	0.74	0.72	0.72	74.21	72.41	72.35
Scenario5	0.8	0.77	0.75	80.42	76.8	75.16
Scenario 6	0.73	0.69	0.67	73.10	68.88	67.34

S1 (Full genotyped dam with one low-density genotyped offspring), S2 (Full genotyped dam with two low-density genotyped offspring), S3 (low-density genotyped dam with one low-density genotyped offspring), S4 (low-density genotyped dam with two low-density genotyped offspring), S5 (Non-genotyped dam with one low-density genotyped offspring), S6 (Non-genotyped dam with two low-density genotyped offspring). [a]5k genotyped offspring, [b]1k genotyped offspring, [c]0.5k genotyped offspring,[d]5k genotyped dam, [e]1k genotyped dam, [f]0.5k genotyped dam.

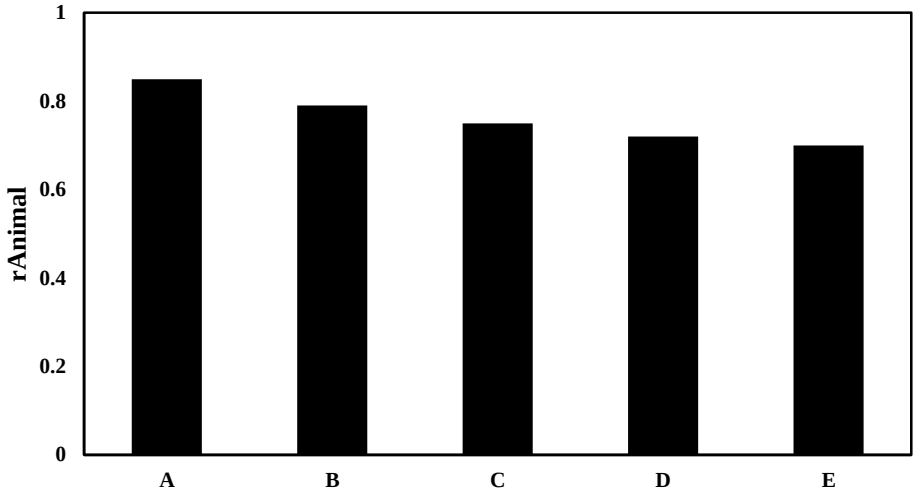

Figure 2. Individual imputation accuracy with full, low-density and non-genotyped dam. A: full genotyped dam, B: 5k genotyped dam, C: 1k genotyped dam, D: 0.5k genotyped dam, E: non-genotyped dam

The accuracies for each scenario and alternative low-density genotyping strategies are shown in Table 4. The GEBV accuracy for P_r was higher than two other populations. The P_2 had lowest accuracy because in this population genotypes were imputed. As expected maximum accuracies for three population were obtained at S1 where the dam was completely genotyped and one offspring was low-density genotyped. In S1, the accuracy for P2 with 5k SNP was 0.78, while for S2 (completely genotyped dam with two low-density genotyped offspring) was 0.73. Accuracies for S3 where the dam and one offspring were genotyped with low density panels were lower than S1 and S2. Also the accuracy for S4 where the dam and both offspring were low density genotyped was lower than S3. In S5 and S6, the accuracy decreased compared to S4. This is due to the presence of non- genotyped dams in these scenarios. In S5, where only one offspring was genotyped with low density panel, the accuracy was 0.60 for the 5k SNP and 0.56 for S6 with two low-density genotyped offspring.

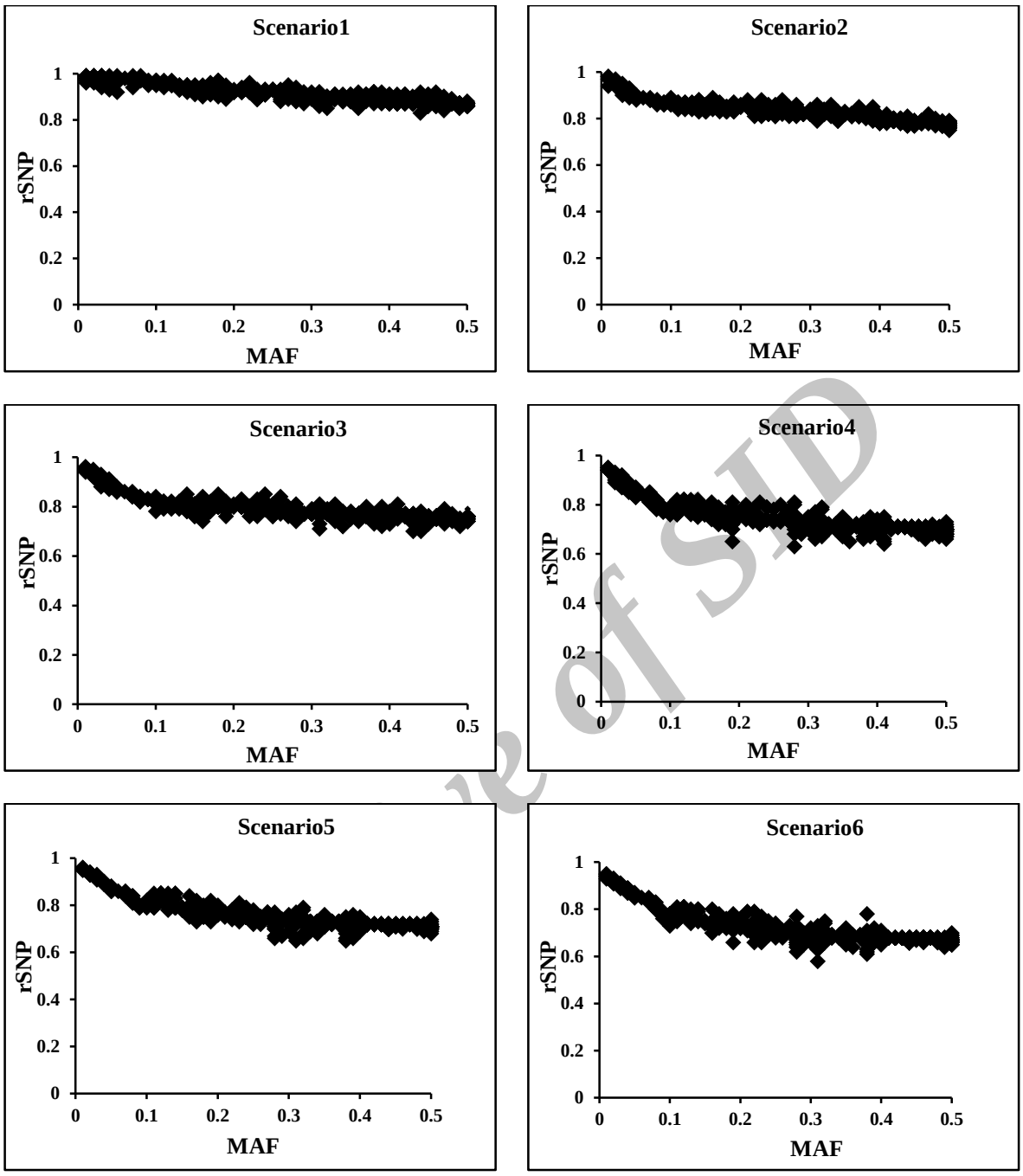

Figure 3. Imputation accuracy by SNP (rSNP) plotted against the minor allele frequency (MAF) in conditions that the offspring in all scenarios and the dam in third and fourth scenarios had 0.5K SNPs.

Table 3. Reference and validation populations for several scenarios

Scenario	Reference (P_r)	Validation 1 and 2 (P_1 and P_2)
1	Sire 1, Sire 2, Dam, Offspring a	Offspring b
2	Sire 1, Sire 2, Dam	Offspring a and b
3	Sire 1, Sire 2, MGS, Offspring a	Dam, Offspring b
4	Sire 1, Sire 2, MGS	Dam and Offspring a and b
5	Sire 1, Sire 2, MGS, Offspring a	Dam, Offspring b
6	Sire 1, Sire 2, MGS	Dam and Offspring a and b

P_1 has true genotype and P_2 has imputed genotype

Table 4. Accuracy of genomic breeding values in reference and validation populations for several scenarios

Scenarios	Density for offspring	Reference (P_r)	Validation1 (P_1)	Validation2 (P_2)	P_2/P_1
Scenario1	5k			0.06 ± 0.78	0.96
Full genotyped dam with one low-density offspring	1k	0.01 ± 0.94	0.06 ± 0.81	0.08 ± 0.73	0.91
	0.5k			0.10 ± 0.72	0.89
Scenario2	5k			0.10 ± 0.73	0.91
Full genotyped dam with two low-density offspring	1k	0.02 ± 0.91	0.06 ± 0.80	0.13 ± 0.69	0.86
	0.5k			0.15 ± 0.67	0.84
Scenario3	5k			0.11 ± 0.72	0.88
5k genotyped dam with one low-density offspring	1k	0.01 ± 0.92	0.05 ± 0.83	0.15 ± 0.67	0.81
	0.5k			0.15 ± 0.65	0.78
Scenario3	5k			0.16 ± 0.67	0.81
1k genotyped dam with one low-density offspring	1k	0.01 ± 0.92	0.05 ± 0.83	0.19 ± 0.63	0.76
	0.5k			0.20 ± 0.62	0.75
Scenario3	5k			0.17 ± 0.66	0.82
0.5k genotyped dam with one low-density offspring	1k	0.01 ± 0.92	0.05 ± 0.80	0.21 ±0.61	0.76
	0.5k			0.21 ± 0.60	0.75
Scenario4	5k			0.13 ± 0.68	0.86
5k genotyped dam with two low-density offspring	1k	0.03 ± 0.91	0.05 ± 0.79	0.19 ± 0.63	0.80
	0.5k			0.20 ± 0.62	0.78
Scenario4	5k			0.19 ± 0.64	0.81
1k genotyped dam with two low-density offspring	1k	0.03 ± 0.91	0.05 ± 0.79	0.22 ± 0.58	0.73
	0.5k			0.22 ± 0.56	0.71
Scenario4	5k			0.20 ± 0.63	0.79
0.5k genotyped dam with two low-density offspring	1k	0.03 ± 0.91	0.05 ± 0.79	0.23 ± 0.55	0.69
	0.5k			0.24 ± 0.54	0.68
Scenario5	5k			0.21 ± 0.60	0.72
Non-genotyped dam with one low-density offspring	1k	0.01 ± 0.92	0.05 ± 0.83	0.26 ± 0.53	0.64
	0.5k			0.26 ± 0.53	0.64
Scenario6	5k			0.22 ± 0.56	0.71
Non-genotyped dam with two low-density offspring	1k	0.03 ± 0.91	0.05 ± 0.79	0.30 ± 0.50	0.63
	0.5k			0.31 ± 0.49	0.62

P_r and P_1 were genotyped completely and P_2 was imputed with three low density panels.

Discussion

This study investigated the accuracy of imputation of low-density genotyped offspring of completely genotyped, low-density genotyped and non-genotyped dams. The results showed sufficient accuracy could be obtained when a dam is genotyped. The average of imputation accuracy in 5k genotyped offspring with completely genotyped dam was 0.96 and with 5k genotyped dam, it decreased to 0.87. Genotype imputation accuracy increased with increasing marker density of low-density genotyping platform and with close relatives having high-density genotypes. An important question is whether the use of phenotypes from imputed animals is advantageous, for example, in GWAS or genomic prediction. This question is not specifically addressed in the simulations presented here, but has received some attention in the literature. For example, in human GWAS studies, inclusion of predicted genotypes for individuals increased the power of GWAS when close relatives were genotyped (Chen et al., 2012).

This enables re-use of valuable phenotypes from historical datasets for, e.g. GWAS or genomic prediction. Usually, datasets with valuable phenotypes are small and in such cases, adding phenotypes with imputed genotypes can have a relatively larger impact on the power of GWAS or on the improvement of the accuracies of genomic prediction. Chen et al. (2014) showed the accuracy of genotypes that were imputed from various low density panels to the 50k SNP panel under different scenarios. The imputation accuracy was the highest (0.98) when all bulls in the training set were genotyped with 50 k panel and bulls in the validation set were genotyped on the 6k panel. In this study, genotypes on 0.5k, 1k, 5k panels were simulated from 10k genotypes. In reality, there are more genotyping errors in 0.5k genotypes than 1k or 5k. Using the Bayesian method, the 5k SNP panel performed better than the 1k and 0.5k panels. Imputations from lower density panels were more prone to errors and resulted in lower accuracy of genomic predic-

tion. But for individuals with both parents genotyped, genotype imputation achieves a relatively high accuracy. Chen et al. (2014) investigated the impact of genotype imputation on the performance of GBLUP and Bayesian methods; their results showed that performance of both methods was influenced by imputation errors. Boichard et al. (2012) showed there were more genotyping errors in lower density panel. Therefore, lower density panels performed worse due to more inaccurate imputation of SNP genotypes.

In the literature, several definitions of imputation accuracy are used. As pointed out by Hickey et al. (2012) and empirically shown by Brøndum et al. (2012), the widely used percentage of correctly imputed SNPs depends on the MAF, and the correlation between the true genotype and the imputed genotype (or dosage) is a better measure of the quality of imputation. However, for the animal-specific imputation accuracy, different SNPs have different MAF, and thus also a distribution with a different mean, while a Pearson correlation assumes that the correlated variables are bivariate normally distributed. Therefore, we calculated imputation accuracy as the correlation between imputed genotypes and real genotypes. Callus et al. (2014) calculated imputation accuracy as the correlation between true and imputed alleles because this definition is in line with the definition of the accuracy of breeding values, which is commonly used in the context of animal breeding.

In the future, more animals might be genotyped on low-density panels. One might have to decide whether to include these animals in the validation population to derive genomic prediction equations. From Table 4, the accuracy of genomic prediction was consistently reduced when more animals in the validation population were imputed when the density of the SNP panel was lower than 5 k. Currently, nearly all males used for breeding are genotyped or regenotyped on panels with a density of 5 k or higher and results from this study justified the application of the 5 k panel. The trend that the accuracy changed with the density of SNP panel agreed with results of Weigel et al. (2010).

Conclusion

Non-genotyped individuals could be imputed with an imputation accuracy, ranging from 0.67 to 0.96. Imputed genotypes are calculated for use in genomic evaluation but the accuracy of breeding values will depend on the level of genotyping in close relatives. Phenotypes with imputed genotypes can have a relatively larger impact on the power of GWAS or on the improvement of the accuracies of genomic prediction. Imputations from lower density panels were more prone to errors and resulted in less accurate genomic prediction. But for individuals with both parents genotyped, genotype imputation achieve a relatively high accuracy. The accuracy of genomic prediction was reduced when more animals in the validation population were imputed when the density of the SNP panel was less than 5k.

References

Boichard, D., Chung, H., Dassonneville, R., David, X., Eggen, A., Fritz, S., 2012. Design of a bovine low-density SNP array optimized for imputation. *PLoS ONE* 7.e34130 doi: 10.1371/journal.pone.0034130.

Bouwman, A.C., Hickey, J.M., Calus, M.P., Veerkamp, R.F., 2014. Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle. *Genetics Selection Evolution* 46, 6-10 doi: 10.1186/1297-9686-46-6.

Brøndum, R.F., Ma, P., Lund, M.S., Su, G., 2012. Short communication: genotypeimputation within and across Nordic cattle breeds. *Journal of Dairy Science* 95, 6795–6800.

Browning, B.L., Browning, S.R., 2009. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics* 84,210-223 doi: 10.1016/j.ajhg.2009.01.005.

Calus, M.P.L., 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 doi:10.1017/S1751731114001803

Chen, L., Li, C., Sargolzaei, M., Schenkel, F., 2014. Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. *Plos ONE* 9, e101544.

Chen, M.H., Huang, J., Chen, W.M., Larson, M.G., Fox, C.S., Vasan, R.S., Seshadri, S., O'Donnell, C.J., Yang, Q., 2012. Using family-based imputation in genome-wide association studies with large complex pedigrees: the Framingham heart study. *PLoS ONE* 7:e51589.

De los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K., Cotes, J.M., 2009 Predicting Quantitative Traits With Regression models for Dense Molecular Markers and pedigree. *Genetics* 182, 375-385.

De los Campos, G., Pérez, P., 2010. BLR: Bayesian Linear Regression. R package version 1.1.

Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., Mason, B.A., Goddard, M.E., 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science* 95, 4114–4129 doi: 10.3168/jds.2011-5019.

Hickey, J.M., Crossa, J., Babu, R., de losCampos, G., 2012. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Science* 52, 654–663 doi: 10.2135/cropsci2011.07.0358.

Johnston, J., Kistemaker, G., Sullivan, P.G., 2011. Comparison of different imputation methods. *Interbull Bulletin* 44, 25–33.

Lu, A.T., Cantor, R.M., 2014. Identifying rare-variant associations in parent-child trios using a Gaussian support vector machine. *BMC Proceedings* 8, S98 doi: 10.1186/1753-6561-8-S1-S98.

Meuwissen, T.H.E., Goddard, M.E., 2010. The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole genome sequence density genotypic data. *Genetics* 185, 1441–1449 doi: 10.1534/genetics.110.113936.

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.

Molaei Moghbeli, S., Barazandeh, S., Vatankhah, M., Mohammadabadi, M., 2013. Genetics and non-genetics parameters of body weight for post-weaning traits in Raini Cashmere goats. *Tropical Animal Health and Production* 45, 1519-1524 doi: 10.1007/s11250-013-0393-4

Mulder, H.A., Calus, M.P.L., Druet, T., Schrooten, C., 2012. Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *Journal of Dairy Science* 95,876-889 doi: 10.3168/jds.2011-4490.

Ober, U., Ayroles, J.F., Stone, E.A., Richards, S., Zhu, D., Gibbs, R.A., Stricker, C., Gianola, D., Schlather, M., Mackay, T.F.C., Simianer, H., 2012. Using whole-genome sequence data to predict quantitative trait phenotypes in Drosophila melanogaster. *PloS Genetics* 8:e1002685 doi: 10.137/journal.pgen.1002685.

Pimentel, E.C.G., Erbe, M., König, S., Simianer, H., 2011. Genome partitioning of genetic variation for milk production and composition traits in Holstein cattle. *Frontires in Genetics* 2, 19-25.

Pimentel, E.C.G., 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, 45-12 doi: 10.1186/1297-9686-45-12.

R Development Core Team. R: a language and environment for statistical computing, Vienna. 2014. Available at: http://www.r-project.org/.

Sargolzaei, M., Jansen, G.B., Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15,478 doi: 10.1186/1471-2164-15-478.

Sargolzaei, M., Schenkel, F.S., Jansen, G.B., Schaeffer, L.R., 2008. Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science* 91, 2106–2117 doi: 10.3168/jds.2007-0553.

Scheet, P., Stephens, M., 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics* 78, 629–44.

Silva, F.F., Rose, G., Guimaräes, S., Lopes, P.S., Campos, G., 2011. Tree-step Bayesian factor analysis applied to QTL detection in crosses between outbred pig populations. *Livestock Science* 142, 210-215.

Technow, A.F., 2015. Hypred, simulation of genomic data in applied genetics. R package version 0.5. Available at: http://cran.rproject.org//web/packages/hypred

Villumsen, T.M., Janss, L., Lund, M.S., 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. *Journal of Animal Breeding and Genetics* 126, 3-13.

Weigel, K.A., Van Tassell, C.P., O'Connell, J.R., VanRaden, P.M., Wiggans, G.R., 2010. Prediction of unobserved single nucleotide poly-morphism genotypes of Jersey cattle using reference panels and population based imputation algorithms. *Journal of Dairy Science* 93, 2229-2238 doi: 10.3168/jds.2009-2849.

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, 28 doi: 10.1186/1297-9686-45-28.

Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., 2008. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genetics* 40, 161–9.

Williams, A.L., Patterson, N., Glessner, J., Hakonarson, H., Reich, D., 2012. Phasing of many thousands of genotyped samples. *The American Journal of Human Genetics* 91, 238–251 doi: 10.1016/j.ajhg.2012.06.013.

Communicating editor: *Ali K. Esmailizadeh*

تخمین ژنوتایپی از خانواده‌های سه تایی (والدین–فرزند) و اثر آن بر پیش‌بینی‌های ژنومیک با استفاده از روش Bayesian

م. کمایی[۱]، م. هنرور[۲]، م. امین افشار[۳] و ر. عبدالهی آرپناهی[۴]

[۱]دانش آموخته دکتری دانشگاه آزاد اسلامی، واحد علوم وتحقیقات تهران، دانشکده کشاورزی و منابع طبیعی، گروه علوم دامی، تهران، ایران.

[۲]استادیار دانشگاه آزاد اسلامی، واحد شهر قدس، دانشکده علوم دامی، تهران، ایران.

[۳]استادیار دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، دانشکده کشاورزی و منابع طبیعی، گروه علوم دامی، تهران، ایران.

[۴]استادیاردانشگاه تهران، پردیس ابوریحان، دانشکده علوم دامی، تهران، ایران.

*نویسنده مسئول، پست الکترونیک: kamaei_62@yahoo.com

چکیده هدف از این مطالعه ارزیابی صحت تخمین ژنوتیپی با استفاده از استراتژی‌های متفاوت از خانواده‌های سه تایی (والدین–فرزندان) است. با استفاده از داده‌های شبیه‌سازی شده، عملکرد روش Bayesian LASSO در پیش‌بینی ژنومیک مورد استفاده قرار گرفت. در این تحقیق ژنومی متشکل از پنج کروموزوم هر یک با طول یک مورگان شبیه‌سازی گردید. تعداد ۱۰۰۰۰ نشانگر SNPبر روی هر کروموزوم شبیه‌سازی شد .برای بررسی اثر QTLبر صحت برآورد ارزشهای اصلاحی ژنومی تعداد ۱۰۰ مکان صفت کمی که با استفاده از توزیع یکنواخت بر روی کروموزوم پراکنده شده‌اند شبیه‌سازی شد. سه سطح SNP با تراکم پایین۰/۵ k، ۱k، ۵k) مورد استفاده قرار گرفت. شش استراتژی که هر کدام شامل دو خانواده سه تایی (پدر، مادر و فرزندان) و پدربزرگ مادری برای داده‌های دوتایی (والد–فرزند) است شبیه سازی گردید. در این استراتژی‌ها مادرها در سه وضعیت کاملاً ژنوتایپ شده، ژنوتایپ شده با تراکم پایین و یا ژنوتایپ نشده و فرزندان در دو وضعیت ژنوتایپ شده بطور کامل و ژنوتایپ شده با تراکم پایین هستند. از برنامه Beagle (۳,۳,۲) برای تخمین ژنوتایپی و از نرم‌افزار R پکیج BLR با استفاده از روش Bayesian LASSO برای پیش‌بینی اثرات مارکرها استفاده شد. نتایج نشان داد که صحت تخمین ژنوتایپی و پیش‌بینی ژنومیک تحت تاثیر خطاهای تخمین ژنوتایپی است. صحت تخمین ژنوتایپی رنجی از ۰/۶۷ تا ۰/۹۶ را دارد که به تراکم ژنوتایپی در فرزندان و مادر بستگی دارد. در استراتژی که مادر ژنوتایپ نشده است و هر دو فرزند با تراکم پایین ژنوتایپ شده‌اند، صحت تخمین ژنوتایپی کاهش می‌یابد. صحت تخمین ژنوتایپی در SNPهایی با فراوانی آللی پایین وقتی که یک مادر بطور کامل ژنوتایپ شده است بطور قابل ملاحظه افزایش می‌یابد. صحت تخمین ژنوتایپی از افراد ژنوتایپ نشده موجب ایجاد فنوتیپ‌های با ارزش برای پیش‌بینی ژنومیک میشود بویژه وقتی افراد ژنوتایپ نشده فرزند ژنوتایپ شده دارند.