Investigation on the effect of non-genotyped and low-density genotyped dams on the imputation accuracy

Maryam Kamaei¹, Mehdi Aminafshar*, Mahmood Honarvar*, Naser Emam Jomekashan², Mojtaba Hosseinpour Mashhadi*

¹Science and Research Branch, Islamic Azad University, Tehran, Iran
²Department of Animal Science, Science and Research Branch, Islamic Azad University, Tehran, Iran
³Department of Animal Science, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran
⁴Department of Animal Science, mashhad Branch, Islamic Azad University, Mashhad, Iran

Key words: Imputation, low-density genotyped, non-genotyped, offspring, dam

Article published on November 10, 2014

Abstract

Prediction of non-genotyped markers, called imputation, is a strategy that allows using the same low-density chips for all traits (and for different breeds). The difficulty for imputation lies in the fact that phenotyped individuals have no genotypes, thus information for imputation has to come from relatives. In this study to assess the accuracy of imputation of non-genotyped and low-density genotyped individuals, 3 different scenarios were considered on the basis of their relatives genotypes. The software Beagle was used for imputation. The first scenario had the most imputation accuracy and the third scenario had the least imputation accuracy.

*Corresponding Author: Mehdi Aminafshar aminafshar@gmail.com
Introduction
Genomic selection (Meuwissen et al., 2001) has become a new tool for genetic improvement in livestock species and plants thanks to the discovery of many thousands of single nucleotide polymorphisms (SNP) spreading across the whole genome and cost effective high throughput genotyping technology (Zhang and Druet, 2010). In the SNP genotype 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 deduce the missing genotypes and could be helpful in increasing the accuracy of genomic selection (Weng et al., 2013). Several software programs for imputation are available, some programs were designed for human populations and others for livestock populations (Pei et al., 2008). The performance of different imputation programs depends mostly on the data structure, density of SNP panels, size of the reference population, and whether related or unrelated individuals were genotyped (Johnston et al., 2011) (Ma et al., 2013). BEAGLE imputed missing genotypes of both animals with and without complete pedigree with high accuracy (Browning, 2010). Erbe et al., (2012) used the software BEAGLE without pedigree information to impute genotypes at 800 k SNPs from dairy bulls genotyped at 50 k and reported accuracies of imputation (defined as the proportion of correctly imputed genotypes) ranging from 0.96 to 0.98 in Jersey and Holstein cattle, respectively. The objective of this study was to investigate the effect of non-genotyped and low-density genotyped dams on the success of imputation.

Methods and materials
simulation
A genome which consisted of 5 chromosomes, each had 100 cM length and each chromosome involved 1000 SNP was stimulated, and mutation was considered $2.5 \times 10^{-8}$. A population consist of 510 individuals and 85 families was formed. Each family contains Maternal Grand Sire (MGS), dam, Sire 1, Sire 2, Offspring 1 and Offspring 2.

Scenarios
To assess the accuracy of imputation of non-genotyped and low-density genotyped individuals, 3 different scenarios were considered. In low-density genotyped individuals 50% and 95% genotypes with evenly spaced were missed. In the first scenario, dam in reference population is genotyped for all SNPs, one of the 2 half-sib offspring of this dam is completely genotyped and the second offspring is low-density genotyped. In the second scenario, dam is low-density genotyped, first offspring is completely genotyped and the second offspring is low-density genotyped. But dam is low-density genotyped, first dam’s genotype should be determined, and then using of genotype information from dam, second sire and first half-sibs offspring, genotypes of second offspring will be determined. Therefore, to determine dam’s genotype, genotype information of dam close relatives is needed, here only information of MGS and first offspring is available. In third scenario, dam is non-genotyped and the first offspring of this dam completely genotyped and the second offspring was low-density genotyped (Fig. 1).

Imputation
Beagle is able to impute missing genotypes of both groups of animals with and without completely pedigree. Imputation of Genotypes was carried out using the software Beagle. After imputation of genotypes in each scenario, in order to assess the accuracy of imputation, a comparison between imputed genotypes and real genotypes were carried out and the amount of genotypic correspondence for the means of all individuals was calculated.

Results and discussion
Table 1 shows the imputation accuracy and calculated SD of non-genotyped and low-density genotyped dams. The range of imputation accuracy varied from 0.76 to 0.91. In a condition with non-genotyped dam, the least amount of imputation accuracy and most SD were achieved. The most amount of imputation accuracy was related to genotyped dams with 50% determined genotype, as missing SNPs in dams increased, the imputation accuracy amount...
decreased. Table 2 provides a total comparison between different scenarios and presents percentage of correct and incorrect imputed genotypes in each scenario. In the third scenario which dam is non-genotyped and one of the offspring is low-density genotyped, the percentage of correct imputed genotypes in a condition that 50% and 5% of offspring’s genotypes are determined, were reported as 80.42%, and 75.16%. whereas, in the second scenario in a situation that 50% of dam’s genotype are determined, the percentage of correct imputed genotypes in a condition that 50% and 5% of offspring’s genotypes are determined were 87.5% and 77.61% and in 5% of determined genotypes were 79.9 and 76.8% orderly. in the first scenario that dam is completely genotyped there are the most amount of imputation accuracy and percentage of correct imputed genotypes. As missing SNPs in offspring increased, imputation accuracy reduces clearly. accuracy with non-genotyped dam decrease to 0.75, in conditions that genotype information of second offspring is not completely available. Pimentel et al., (2013) imputed non-genotyped individuals with one genotyped offspring, and reported that imputation accuracy ranged from 0.52 to 0.93 which is related to population structure and used method. Cleveland et al., (2011) investigated use of phenotypes from imputed animals in a simulated population. Their method used segregation analysis and information on haplotype frequencies, and they reported a success rate of 69% when dams were completely un-genotyped.

**Table 1.** Imputation accuracy (r) for low-density genotyped and non-genotyped dam.

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th>Sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>low-density genotyped</td>
<td>0.91</td>
<td>0.009</td>
</tr>
<tr>
<td>and non-genotyped dam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low-density 50%</td>
<td>0.96</td>
<td>0.09</td>
</tr>
<tr>
<td>low-density 5%</td>
<td>0.79</td>
<td>0.06</td>
</tr>
<tr>
<td>non-genotyped</td>
<td>0.76</td>
<td>0.07</td>
</tr>
</tbody>
</table>

1mean of imputation accuracy calculated as the correlation ^between true genotypes and Imputed genotype dosages, ^standard deviation.

**Table 2.** Imputation accuracy (r) and percentage of correct and incorrect imputed genotypes for offspring in each scenario.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Offspring</th>
<th>Dam</th>
<th>correct</th>
<th>incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario1</td>
<td>0.96</td>
<td>0.9</td>
<td>96.47%</td>
<td>3.53%</td>
</tr>
<tr>
<td>Scenario2, 50%</td>
<td>0.87</td>
<td>0.78</td>
<td>87.2%</td>
<td>12.8%</td>
</tr>
<tr>
<td>Scenario2, 5%</td>
<td>0.8</td>
<td>0.77</td>
<td>79.9%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Scenario3</td>
<td>0.8</td>
<td>0.75</td>
<td>80.42%</td>
<td>19.58%</td>
</tr>
</tbody>
</table>

1dam with low-density 50%, 2 dam with low-density 5%.

![Fig. 1. Assumed Family members with available genotypic information (black) used for imputing a low-density genotyped (green) or non-genotyped (red) individual.](image-url)
References

http://dx.doi.org/10.1016/j.ajhg.2009.01.005

http://dx.doi.org/10.1186/1753-6561-5-S3-S6

http://dx.doi.org/10.3168/jds.2011-5019


http://dx.doi.org/10.3168/jds.2012-6316


http://dx.doi.org/10.1186/1297-9686-45-12
