Imputation of non-genotyped individuals based on genotyped relatives: a real case scenario

A.C. Bouwman[1], J.M. Hickey[2], M.P.L. Calus[1], R.F. Veerkamp[1]

[1]Animal Breeding and Genomics Centre, P.O. Box 135, 6700 AC Wageningen, the Netherlands, [2]International Maize and Wheat Improvement Center (CIMMYT), Apdo. 06600, Mexico D.F., Mexico; aniek.bouwman@wur.nl

Imputing genotypes for non-genotyped individuals is attractive because it enables inclusion of historic datasets with valuable phenotypes (e.g., feed intake) to a training set, and it might help to reduce genotyping cost of breeding programs. The objective of this study was to see if, and how accurate non-genotyped individuals can be imputed from genotyped relatives. This study was based on a real dataset for feed intake of dairy cows with 1,021 cows phenotyped and genotyped; 1,344 cows were only phenotyped and thus needed to be imputed; and 3,076 relatives with genotypes only. Genotypes were simulated for all individuals in the pedigree. Subsequently genotypes were set to missing in different scenarios: the real situation, adding sire and maternal grandsire information, and adding information from 1, 2 or 4 offspring. AlphaImpute was used to impute missing genotypes based on pedigree information. Accuracy of imputation was assessed per individual using correlations between true and imputed genotype dosage, both corrected for mean gene content. As expected, imputation accuracy increased when more close relatives were genotyped. Most interesting is the increase in accuracy without phasing from 0.59 (0 offspring) to 0.73, 0.82 and 0.92 by adding 1, 2, and 4 genotyped offspring, respectively. With genotyped offspring, imputation accuracy appeared to be higher than the expected accuracy based on selection index theory. This is because the imputation method can make use of correlations between markers due to linkage and linkage disequilibrium. In these situations a two-step approach, where imputed genotypes are used in further analyses, will therefore give better results than an one-step approach using for instance a H-matrix. In conclusion, imputation of non-genotyped individuals was possible with acceptable accuracy when multiple offspring were genotyped.