

Imputation of non-genotyped individuals based on genotyped relatives

Aniek Bouwman, John Hickey, Mario Calus and Roel Veerkamp

EAAP Nantes, 2013



Introduction

Historic datasets with valuable phenotypes

- Lack genotypes, no DNA material available

Imputation of non-genotyped individuals as solution

(Grand)sires and offspring might be available for genotyping

Objectives

How well can we impute non-genotyped individuals?

Which relatives should be genotyped to improve the imputation accuracy?



Data

Real dataset of 2,365 phenotyped dairy cows

- Half of them non-genotyped (1,344)
 - 805 without offspring in the data
- The other half genotyped (1,021)
- 3,076 other individuals in pedigree genotyped

Simulated genotypes of all individuals in pedigree

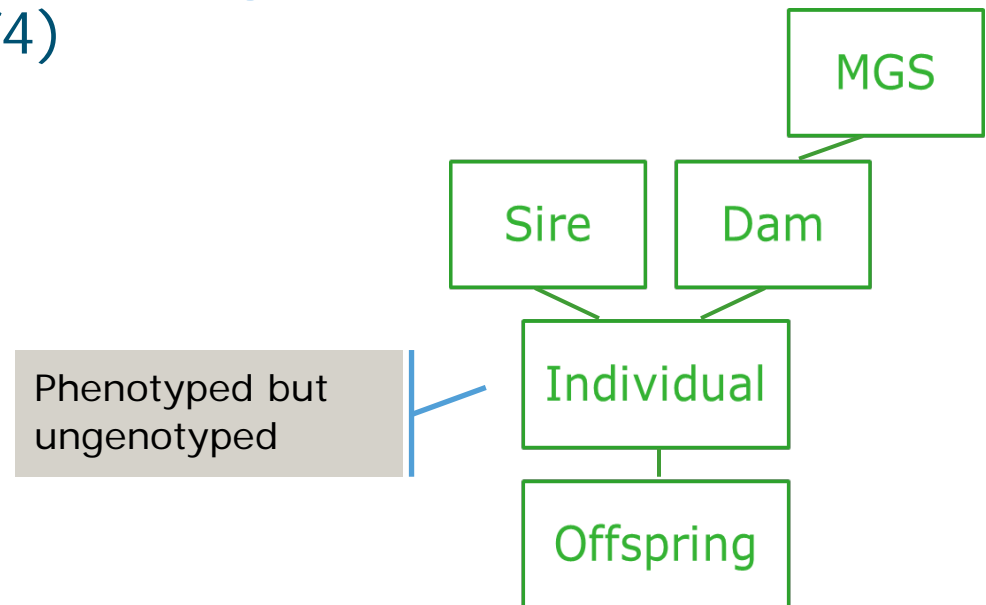
- Real pedigree + 4 offspring (18,053)
- 10 replicates

Scenario's

Real situation (Real)

Adding sires and maternal grandsires (SireMGS)

Adding genotypes of offspring
(Off0, Off1, Off2, Off4)



Imputation

Imputation based on genotyped relatives through pedigree information

AlphaImpute

- using basic inheritance rules and segregation analysis
- probabilities

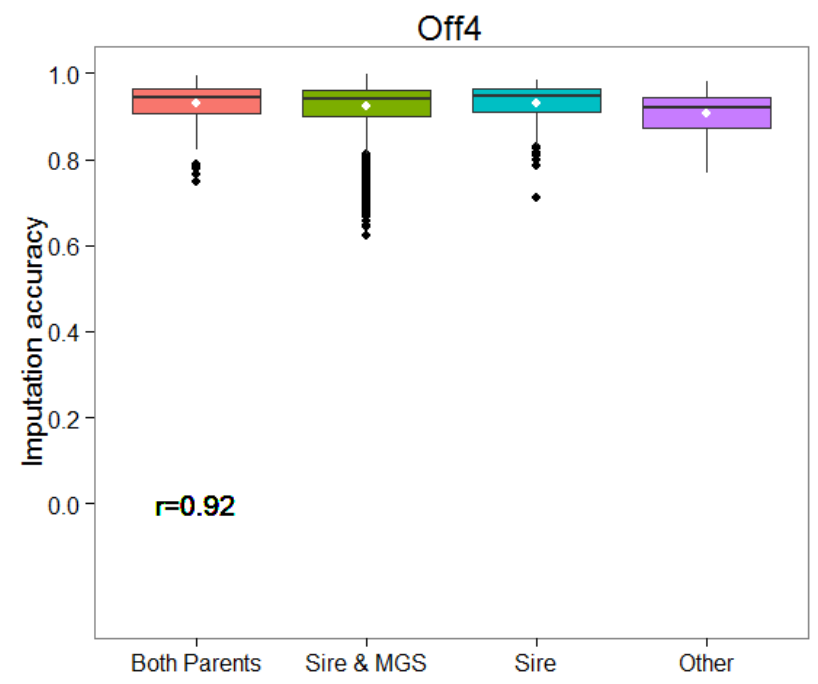
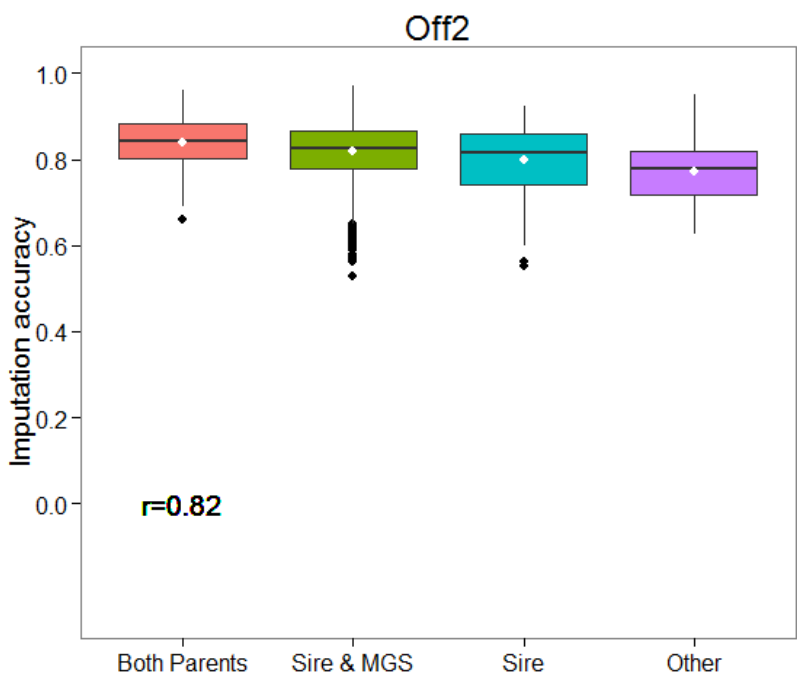
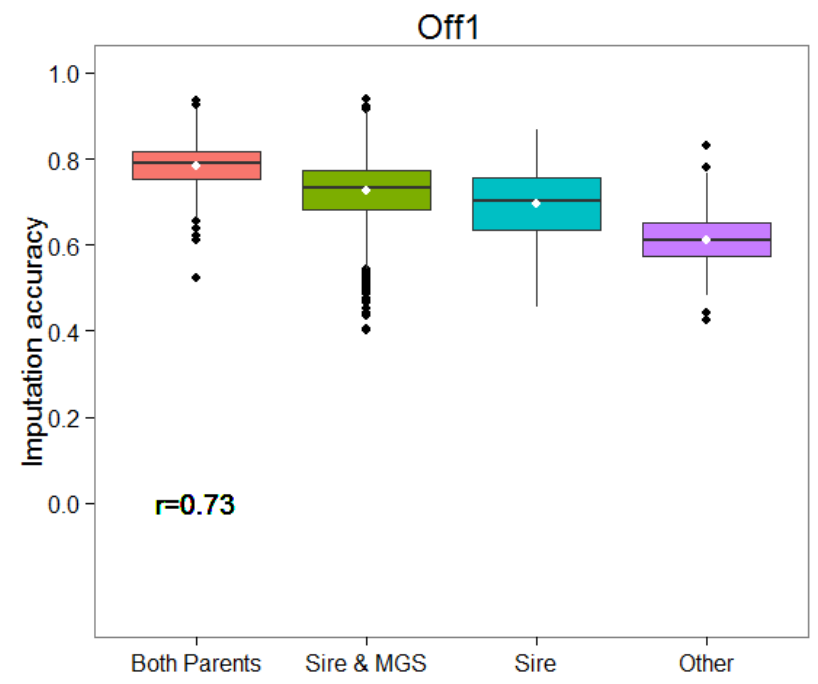
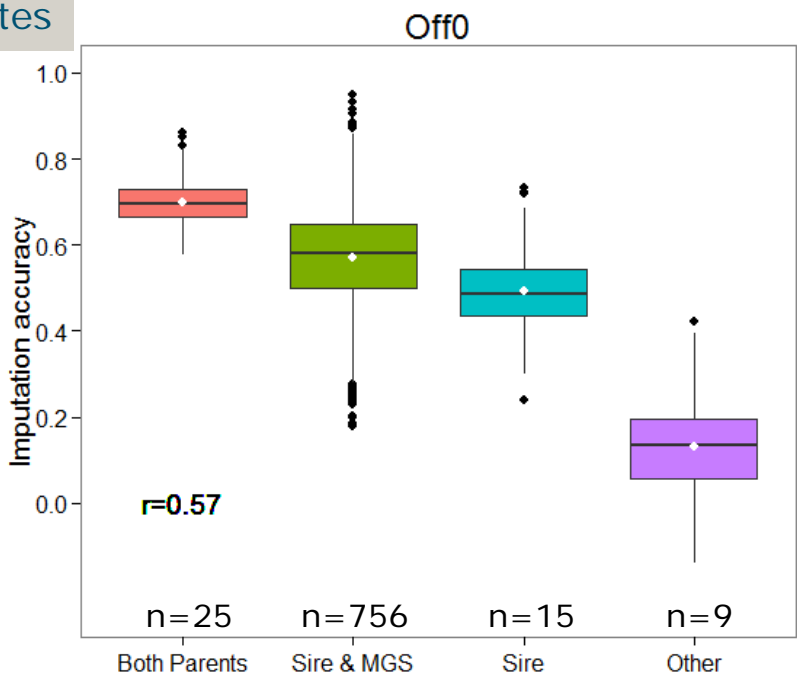
Animal-specific imputation accuracy

Correlation between true genotype and imputed genotype after subtracting the mean genotype ($2p$) of each SNP

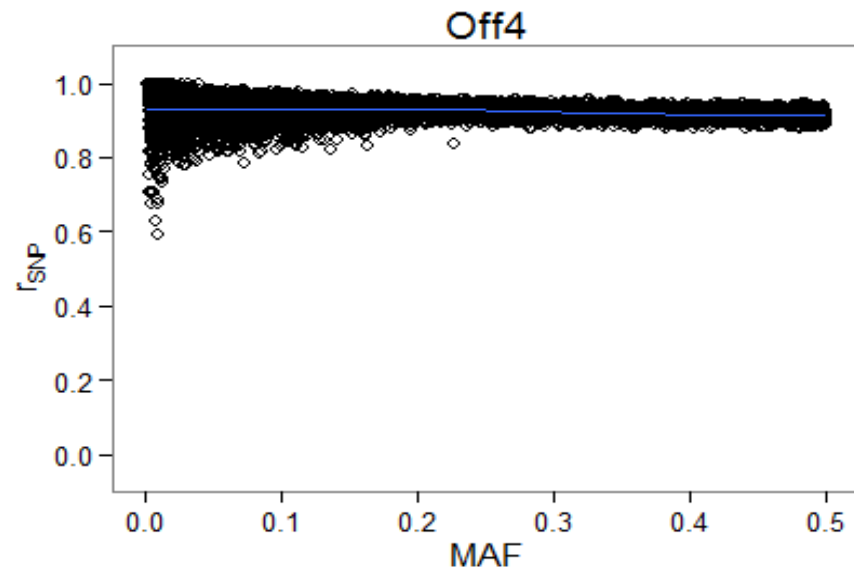
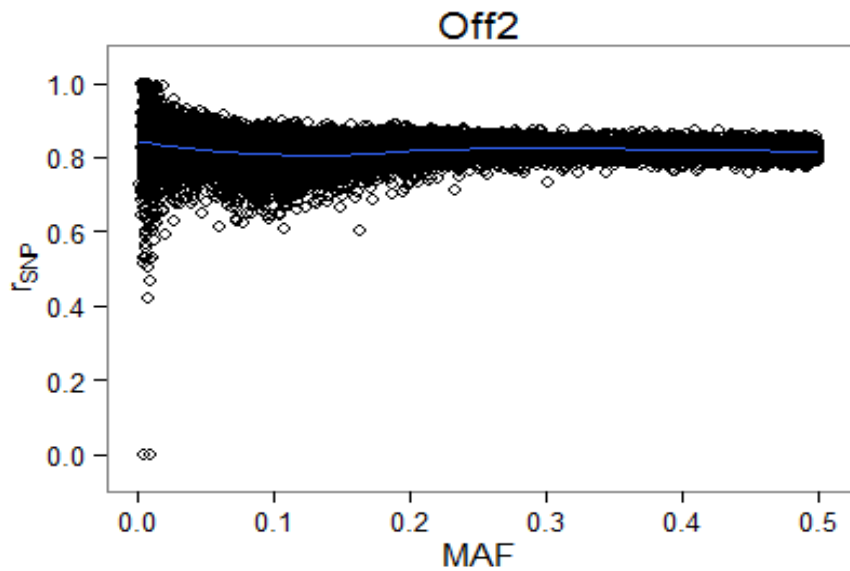
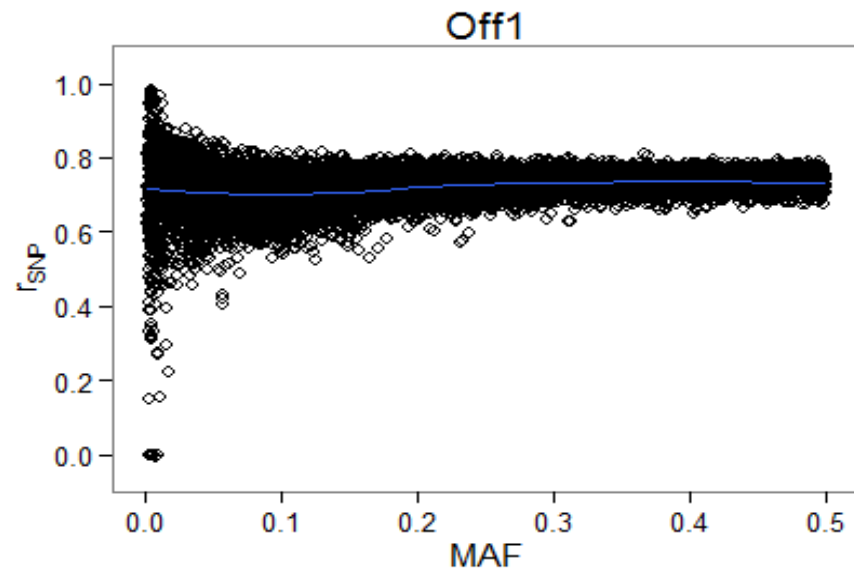
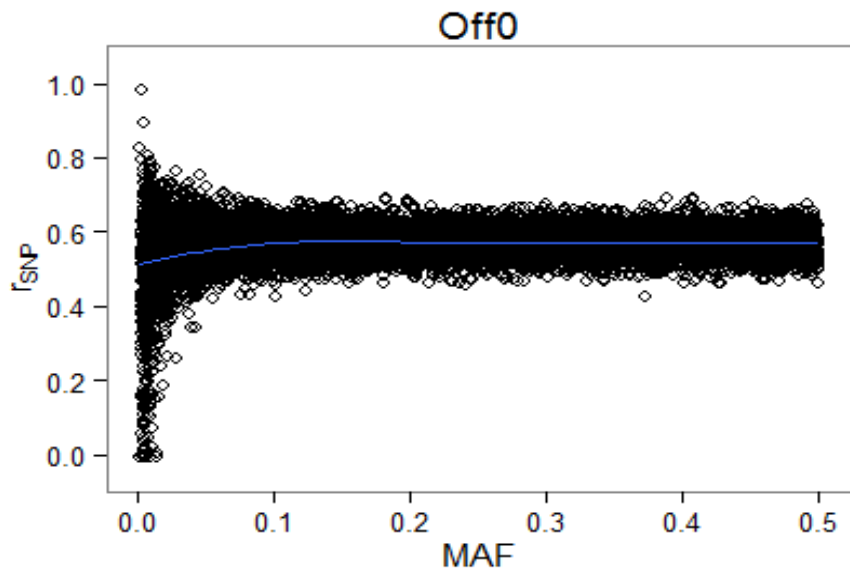
Reduces bias due to allele frequencies

Comparable to selection index

- Imputation of parent average genotype:
 - expect close to $\sqrt{0.5} = 0.71$
 - $r_{\text{corrected}} = 0.70$
 - $r_{\text{uncorrected}} = 0.87$



Accuracy per SNP versus MAF



Imputation accuracy \geq predicted with selection index (SI)

Both parents genotyped (n=25)

Scen	imputation	SI
Off0	0.70	0.71
Off1	0.78	0.76
Off2	0.84	0.79
Off4	0.93	0.84

Sire+MGS genotyped (n=756)

Scen	imputation	SI
Off0	0.57	0.56
Off1	0.73	0.66
Off2	0.82	0.73
Off4	0.92	0.80

Conclusions

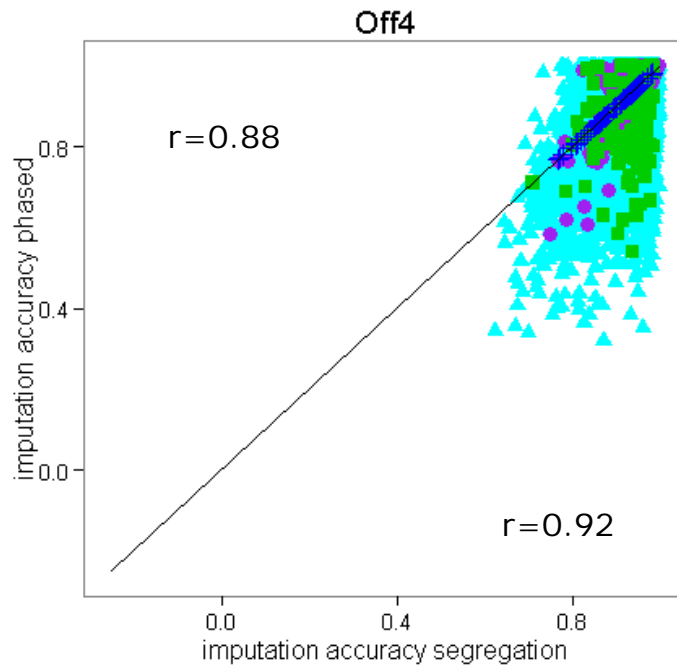
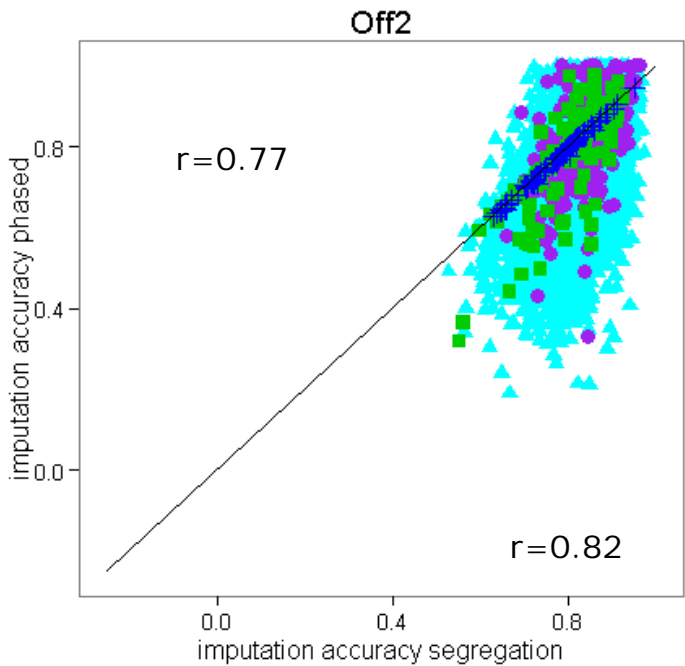
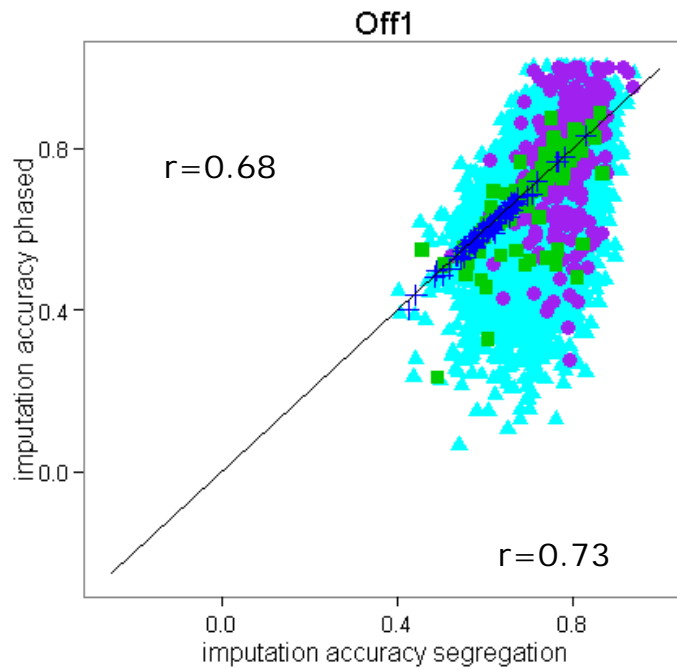
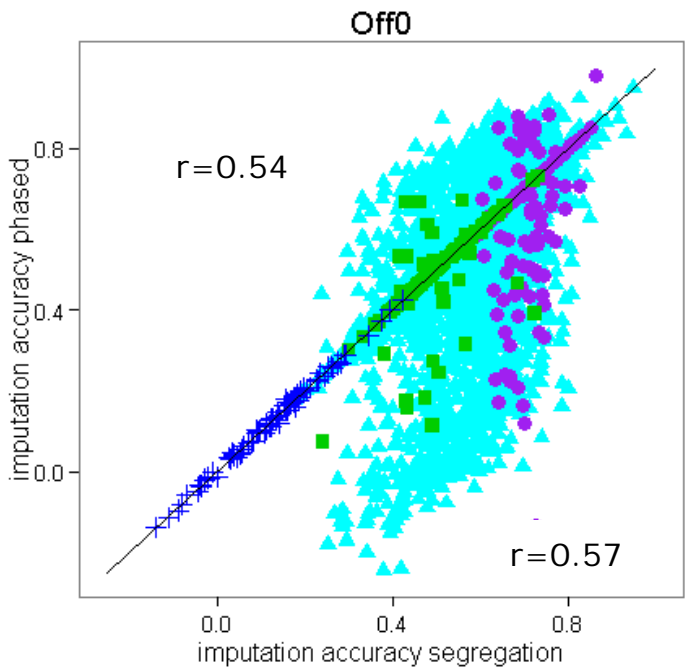


Imputation of ungenotyped individuals is possible

Need at least 2 but preferably 4 genotyped offspring for accurate imputation

With genotyped offspring

- higher accuracy than selection index
- high imputation accuracy for low MAF SNP
- ancestor information less relevant



10 replicates

- 2 parents (25)
- ▲ sireMGS (756)
- sire (15)
- + other (9)