Imputing from BovineSNP50 v1 to v2 chip

Hatem Alkhoder[1,2], Z. Liu[1], F. Reinhardt[1] ,H. H. Swalve[2], and R. Reents[1]

[1]Vereinigte Informationssystem Tierhaltung **vit**, Heideweg 1, 27283 Verden (Aller), Germany

[2]Institute of Agricultural and Nutritional Sciences Martin-Luther-University Halle-Wittenberg,Theodor-Lieser-Str. 11, D-06120 Halle/Saale Germany

Conclusion

Findhap was 20-70 times faster and required much less memory than Beagle for imputing with 20,055 animals in a simulation study

The error rate for imputing with Beagle 0.91% and much lower than Findhap (Δ=0.41%)

Subsetting validation population decreased allele error rate about 36% (from 1.41%to 0.91%)

When the validation population very large (150% from the reference population), Findhap was more accurate than Beagle

The error rate found in the simulation study should be higher than in routine imputation with all genotyped animals.

Switching to v2 from v1 chip increased variance of direct genomic breeding values of candidates between 0.3% and 7.6%

direct genomic breeding values of candidates estimated using v1 and v2 chips were very highly correlated (>0.99)

Objectives

Impute genotypes of all animals with the old version Illumina BovineSNP50 chip (v1) to new version Illumina BovineSNP50 chip (v2)

Estimate SNP effects and direct genomic value using (imputed) genotypes of the new version chip v2

Data and Materials

20,055 animals (v2) in reference population

33,304 animals (v1) in validation population

Imputing softwares: Findhap & Beagle

52,340 common SNPs, 1661 v1 (removed), 2269 only in v2

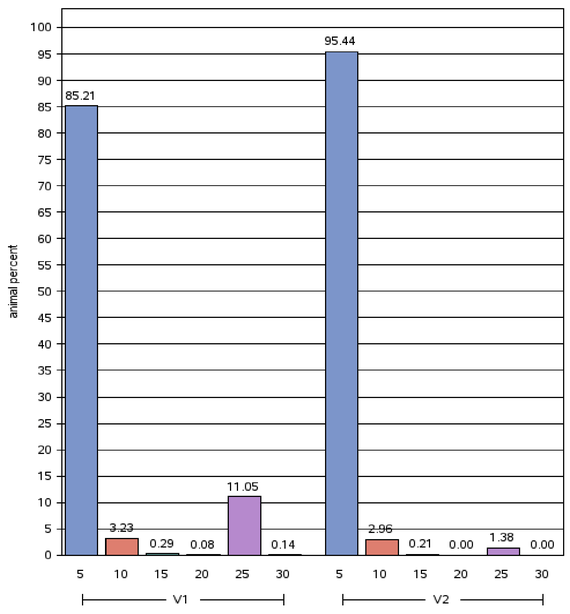
Final selected 45,613 SNPs with call rate (>=90%) and MAF (0.01)

1439 new SNPs on to impute

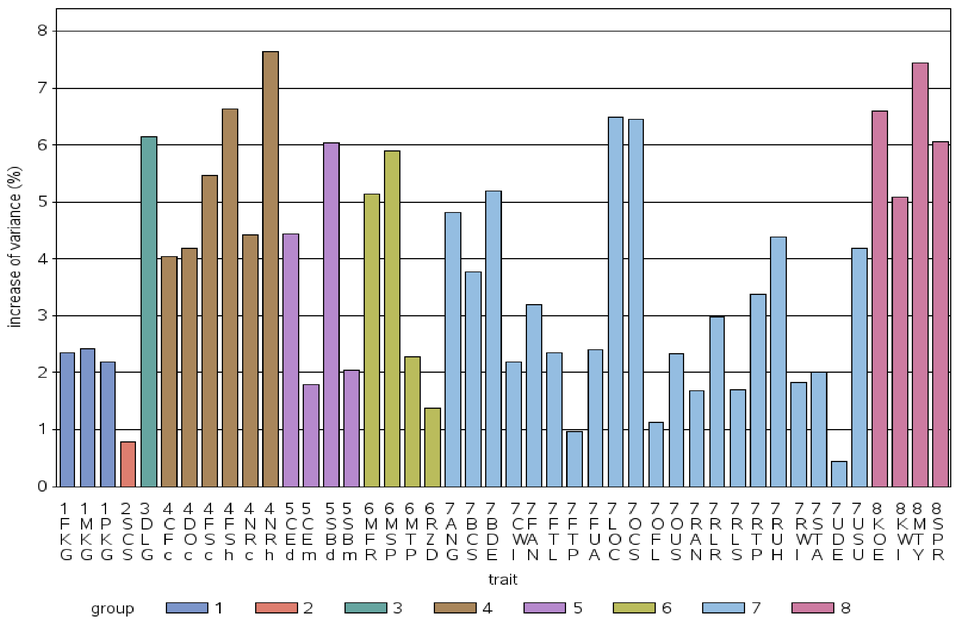
SNPs on sex chromosome were imputed only with Findhap

Validation animals (v1) were divided into 15 groups and imputed separately

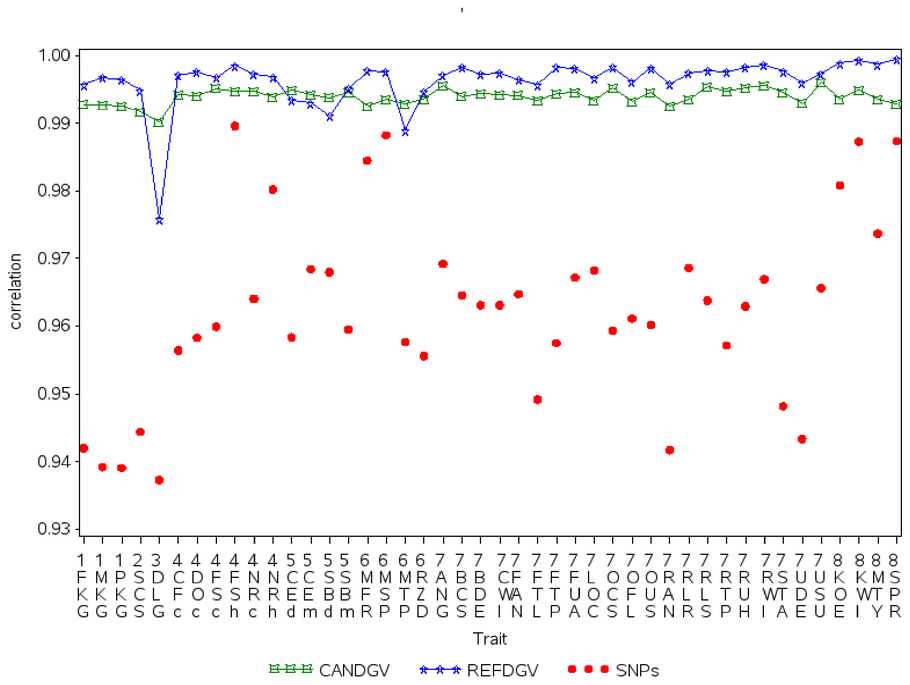
Results



Frequency of animals by missing genotype rate (%)



Increase of variance in direct genomic breeding values of candidates for all traits when v2 was used instead of v1



Correlations of SNP effect estimates or direct genomic values between two evaluations using v2 and v1 chips

Simulation for investigating imputation accuracy from v1 to v2

All 20,055 v2 animals were devided in two groups:

8,055 reference animals

12,000 validation animals

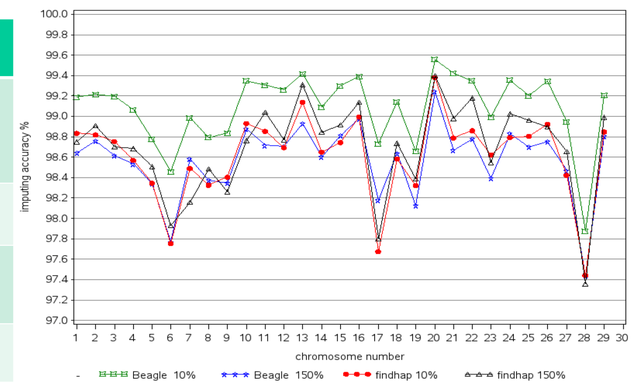
Validation animals were imputed in three scenrios :

One large group (1 x 12,000)

Three medium groups (3 x 4000)

15 small groups (15 x 800)

Data subset Groups	Beagle		Findhap	
	Error rate (%)	Time (hr)	Error rate (%)	Time (hr)
Small	0.91	305	1.38	15
Medium	1.12	85	1.32	6
Large	1.41	239	1.30	3



Imputation accuracy=1- allele error rate

Acknowledgements

The financial support by FBF Förderverein Biotechnologieforschung e.V. is appreciated.