

A multi breed reference improves genotype imputation accuracy in Nordic Red cattle

Rasmus Froberg Brøndum^{1*}, Peipei Ma^{1,2}, Mogens Sandø Lund¹, Guosheng Su^{1*}

¹ Centre for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark, ² Department of Animal Genetics, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. *Corresponding authors: rasmusf.brondum@agrsci.dk & guosheng.su@agrsci.dk

BACKGROUND

- High density genotypes might offer higher reliability for genomic predictions.
- To avoid re-genotyping all animals, accurate imputation methods and strategies are needed
- The Nordic Red populations have close relationships, and some Danish animals also have Holstein Genetics

HYPOTHESIS

- Combining reference populations for Nordic Red populations improves the accuracy of imputation from 54k to HD
- Adding Holstein to the reference further improves the imputation accuracy
- Imputation accuracy depends on the allele frequency
- Using imputed HD data increases the reliability of genomic prediction

CONCLUSIONS

- Increasing the reference population increases the imputation accuracy when target animals are related to added animals
- Imputation accuracy is lower for rare alleles. However, this is not seen when calculating it as percent correct alleles since this does not factor in the allele frequency
- Using imputed HD data for genomic prediction only shows small gains in reliability

METHODS

- Different combinations of references for Danish (RDM), Swedish (SRB), Finnish (FAY), all Nordic Red (RDC) and Holstein (HOL) were tried for imputation.
- All imputations were done using BEAGLE v 3.3 with default parameters
- Accuracy was estimated by deleting markers not on the 54K chip for validation animals and imputing to HD. It was calculated as the average percentage of correct alleles or correlation between true and imputed genotypes (R)
- All Nordic Red animals were imputed from 54K to HD data using the RDC reference, and DGVs were predicted using GBLUP or a Bayesian mixture model (JDS 95:4657-4665).
- Reliability was calculated as the squared correlation between DRP and DGV divided by the reliability of DRP

Reference	N_ref	Validation	N_val	Correct rate (SD)	R
RDM	197	RDM	50	97.36 (2.43)	0.925
SRB	160	SRB	50	98.61 (1.42)	0.958
FAY	199	FAY	50	99.13 (0.58)	0.973
RDC	556	RDM	50	98.25 (1.84)	0.950
		SRB	50	99.41 (0.29)	0.982
		FAY	50	99.46 (0.30)	0.983
SRB+FAY	359	RDC	150	99.04 (1.22)	0.971
		SRB	50	99.04 (1.31)	0.971
		FAY	50	99.46 (0.31)	0.983
HOL	457	HOL	100	99.23 (0.61)	0.975
HOL+RDC	1013	HOL	100	99.32 (0.52)	0.977
		RDM	50	98.83 (1.35)	0.967
		SRB	50	99.40 (0.29)	0.982
		FAY	50	99.45 (0.28)	0.983
		RDC	150	99.23 (0.85)	0.977

	N ref	N val	GBLUP		Bayesian Mixture ($\pi=0.2$)	
			54K	HD	54K	HD
Protein	3,421	924	0.346	0.358	0.346	0.359
Fertility	3,377	941	0.297	0.304	0.299	0.307
Mastitis	3,421	979	0.244	0.257	0.243	0.259
Average	-----	-----	0.296	0.306	0.296	0.308

