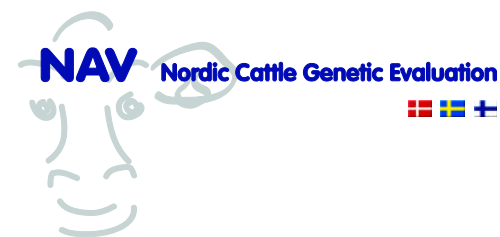
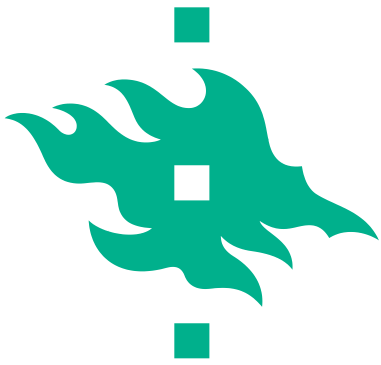


Single step evaluations using haplotype segments

M. L. Makgahlela, T. Knürr, G. P. Aamand, I. Strandén
& E. A. Mäntysaari



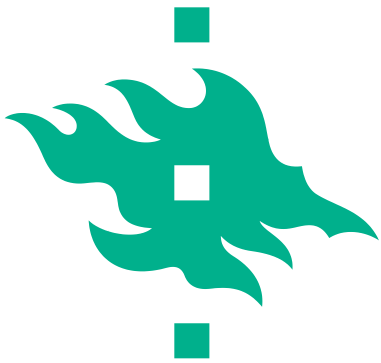


Introduction



- Genomic evaluations, as originally proposed, were based on regression on **haplotype segments, which are;**
 - **closely located allele combinations that tend to be jointly inherited**

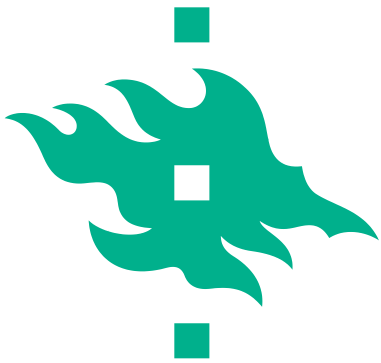
- Many current evaluations however, are based on regression on a large number of individual markers in models that are;
 - **simplified and less computationally demanding**



Introduction



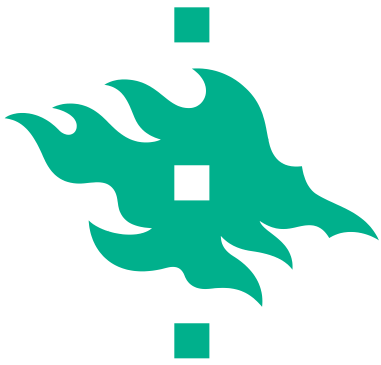
- If the observed reliabilities are low, haplo-block models may be an alternative to improve evaluations
 - 1) They were found to be more reliable than individual markers
 - Because ancestral haplotype segments capture greater linkage disequilibrium (LD) with the QTL than single markers
 - 2) They could greatly reduce the number of markers for genomic evaluations
- Many free and reliable haplotyping software are available



Objectives



- To examine the use of genomic relationship matrix (**G**) constructed using haplotype segments in **single step evaluations**
 - applied on the Nordic Red dairy cattle (RDC)
- Compare the haplo-block model with standard single-step GBLUP



Data were provided by NAV

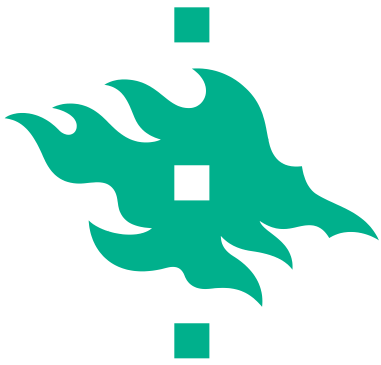
Genotypes

- After editing, there were 38,194 informative SNPs available for 4,727 bulls; born between 1971-2008

Phenotypes

- **Deregressed Proofs (DRP)** of cows for milk and protein
 - Full data (DRP_F) → 3,633,481 cows
 - Reduced data (DRP_R) i.e., discard cows born after > 2005 → 3,146,448 cows

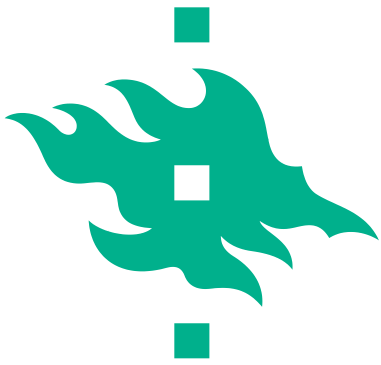
 Full RDC pedigree (n=4,873,703)



For validation



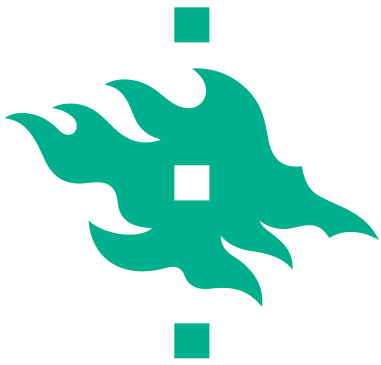
- ApaX program in MiX99 was used for calculating EDCs
- 2 runs of animal model were used to solve deregressed bull EBVs as follows;
 - 1) full run → with DRP_F → generate DRP for 519 validation bulls born between 2002-2008 with $EDC \geq 20$
 - 2) reduced run → with DRP_R → daughters of 4,208 training bulls born between 1971-2005



Construction of haplotype blocks

Details in Knurr et al. EAAP 2013

- 1) BayesB joint estimation of SNP effects in a multilocus model
- 2) Rank SNPs by the absolute effect size $\hat{\beta}_g$
- 3) Haplotype (phase) genotyped using Beagle v3.3
- 4) Construct **5-SNP** haplotypes (i.e., 2 SNPs before and after the one with the highest absolute $\hat{\beta}_g$)
- 5) Estimate haplotype variances
- 6) Tested 750 and 1500 haplotype segments

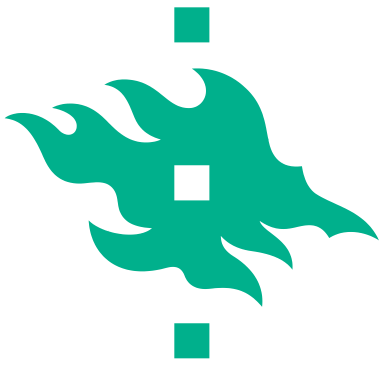


Single step model

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}w^{-1} - \mathbf{A}^{-1}_{22} \end{bmatrix}, \text{ where}$$

- \mathbf{A}^{-1} includes all animals and \mathbf{A}_{22}^{-1} is a sub-matrix for genotyped bulls
- $\mathbf{G}w = (1-w)\mathbf{G}k + w\mathbf{A}_{22}$

$$\checkmark k = \frac{\text{trace}(\mathbf{A}_{22_{ii}})}{\text{trace}(\mathbf{G}_{ii})} ; w \text{ was varied at } 0.10, 0.20 \text{ or } 0.40$$



Single step model

- **Haplo-block \mathbf{G}**

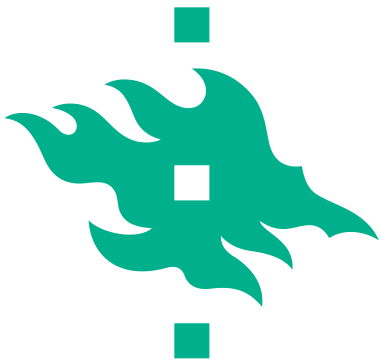
$$\mathbf{G} = \mathbf{ZDZ}' ; \mathbf{z}_{i,j} \leftarrow 0,1,2$$

- 0,1 or 2 is the number of copies of j^{th} haplo-block
- \mathbf{D} is a diagonal of haplo-block variances
- Haplo-block \mathbf{G} was constructed with 750 segments (HAP750) and 1500 (HAP1500)

- **Individual SNP-based \mathbf{G}**

$$\mathbf{G} = \mathbf{Z}\mathbf{Z}'/2\sum pq ; \mathbf{z}_{i,j} \leftarrow (0-2p_j); (1-2p_j); (2-2p_j),$$

- p_j is the frequency for the 2nd allele



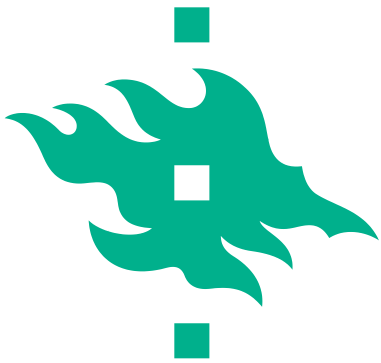
GEBV evaluation



$$DRP_{R_{cow}} = \mathbf{1}_n \mu + Wa + e,$$

➤ where :

- $\text{var}(\mathbf{a}) = \mathbf{H}\sigma_a^2$ with variances from NAV routine evaluations
- $DRP_{R_{cow}}$ are the deregressed proofs of all cows in reduced data (incl. Daughs of bulls in the reduced data)
- Effective record number of the cow was used as weight



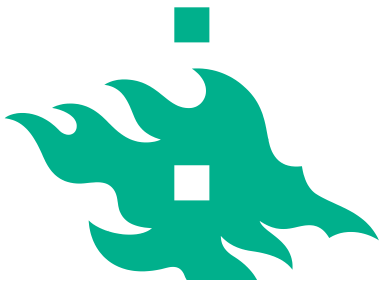
GEBV validation



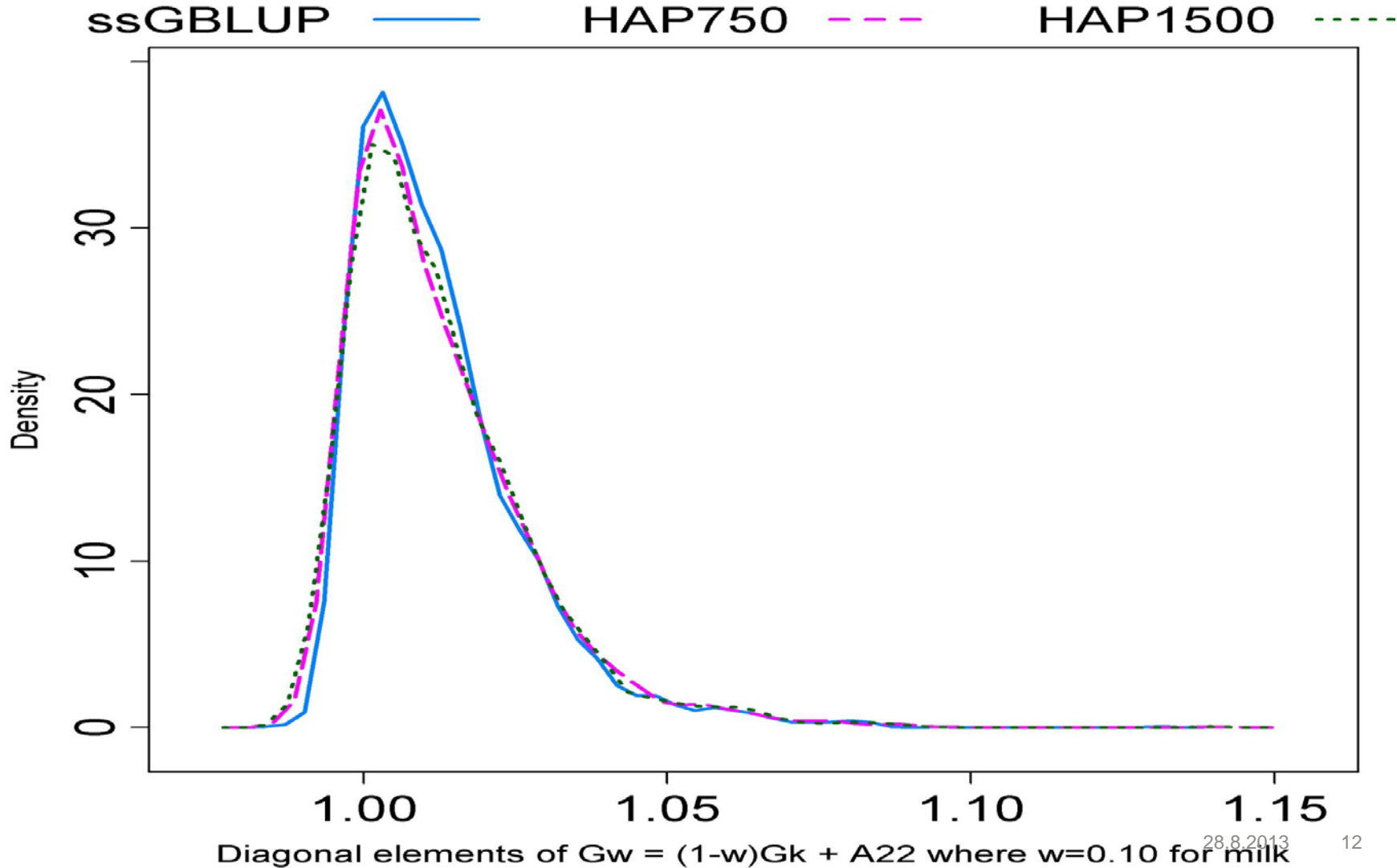
$$DRP_{F_{bull}} = b_0 + b_1 GEBV + e,$$

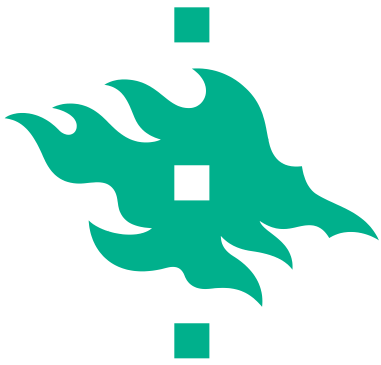
➤ where :

- $DRP_{F_{bull}}$ is the deregressed proof of the candidate bull from the animal model with full data
- EDC of the bull was used as weight

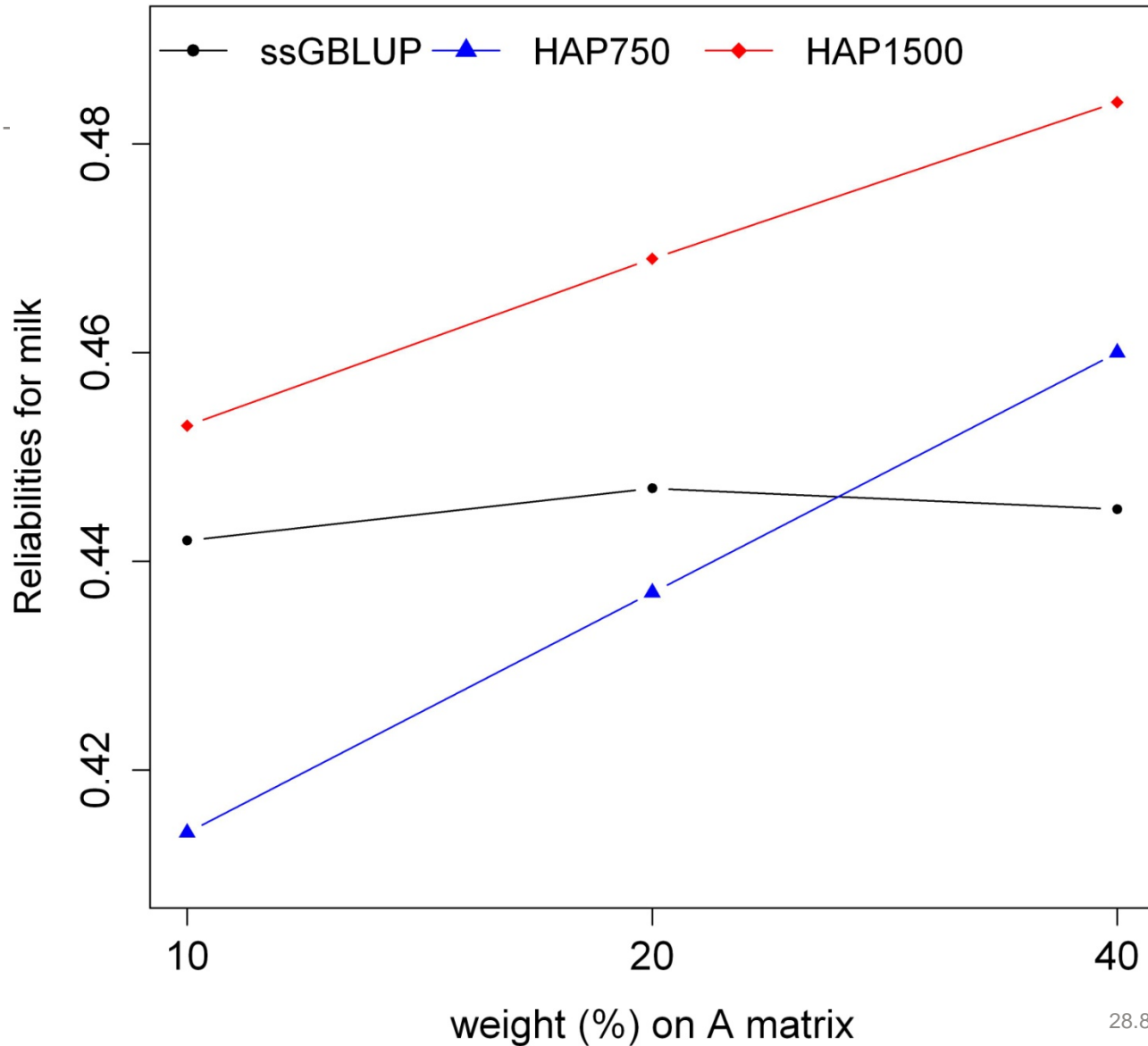


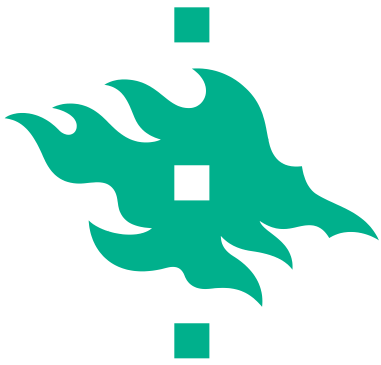
Diagonal elements from ssGBLUP, HAP750 and HAP1500



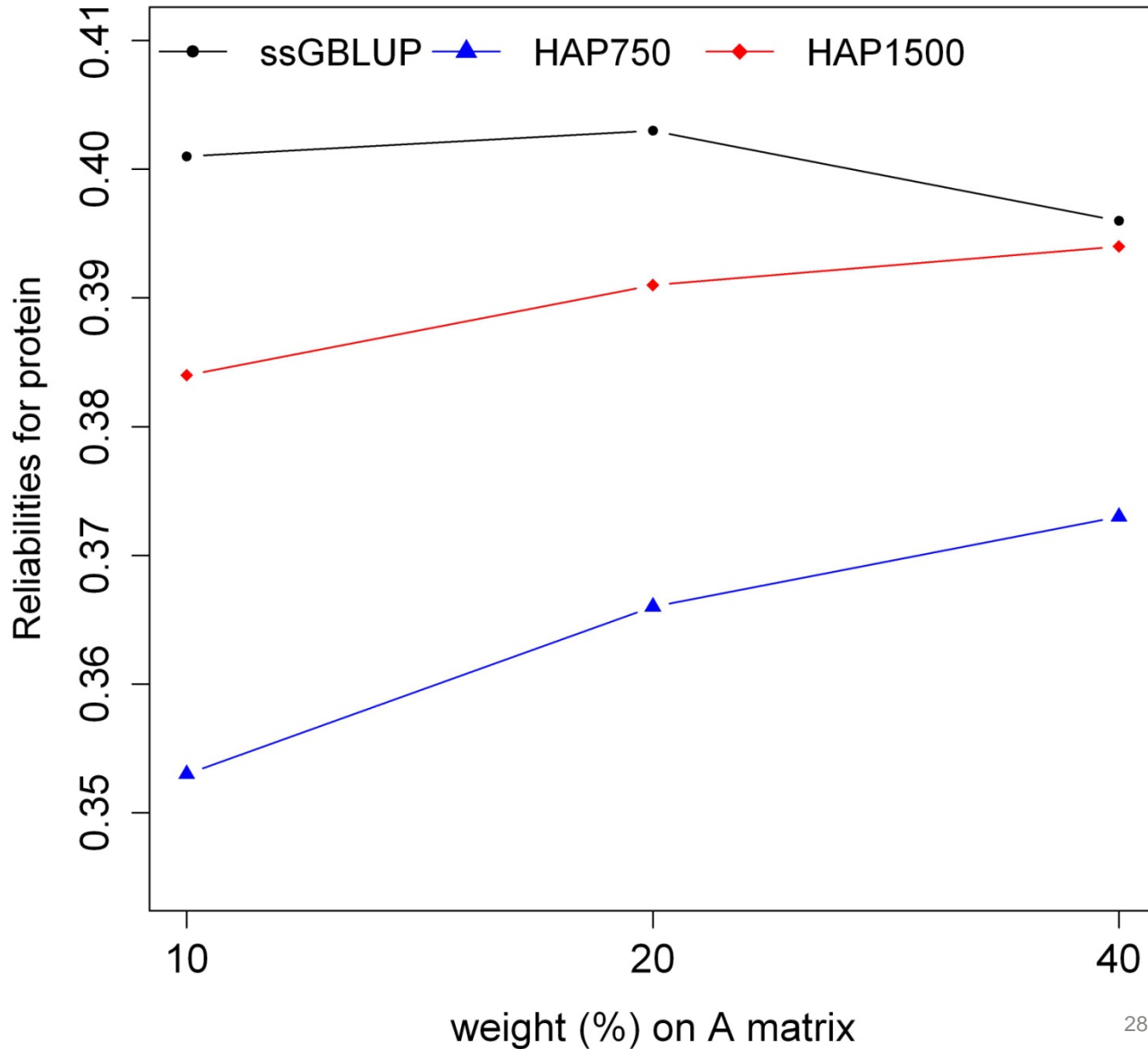


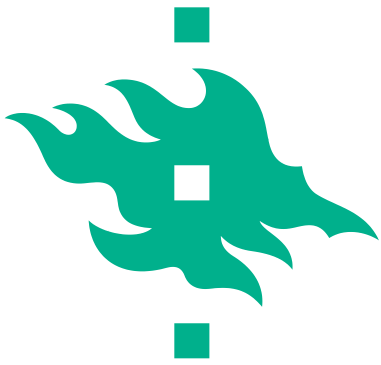
Validation reliabilities for milk



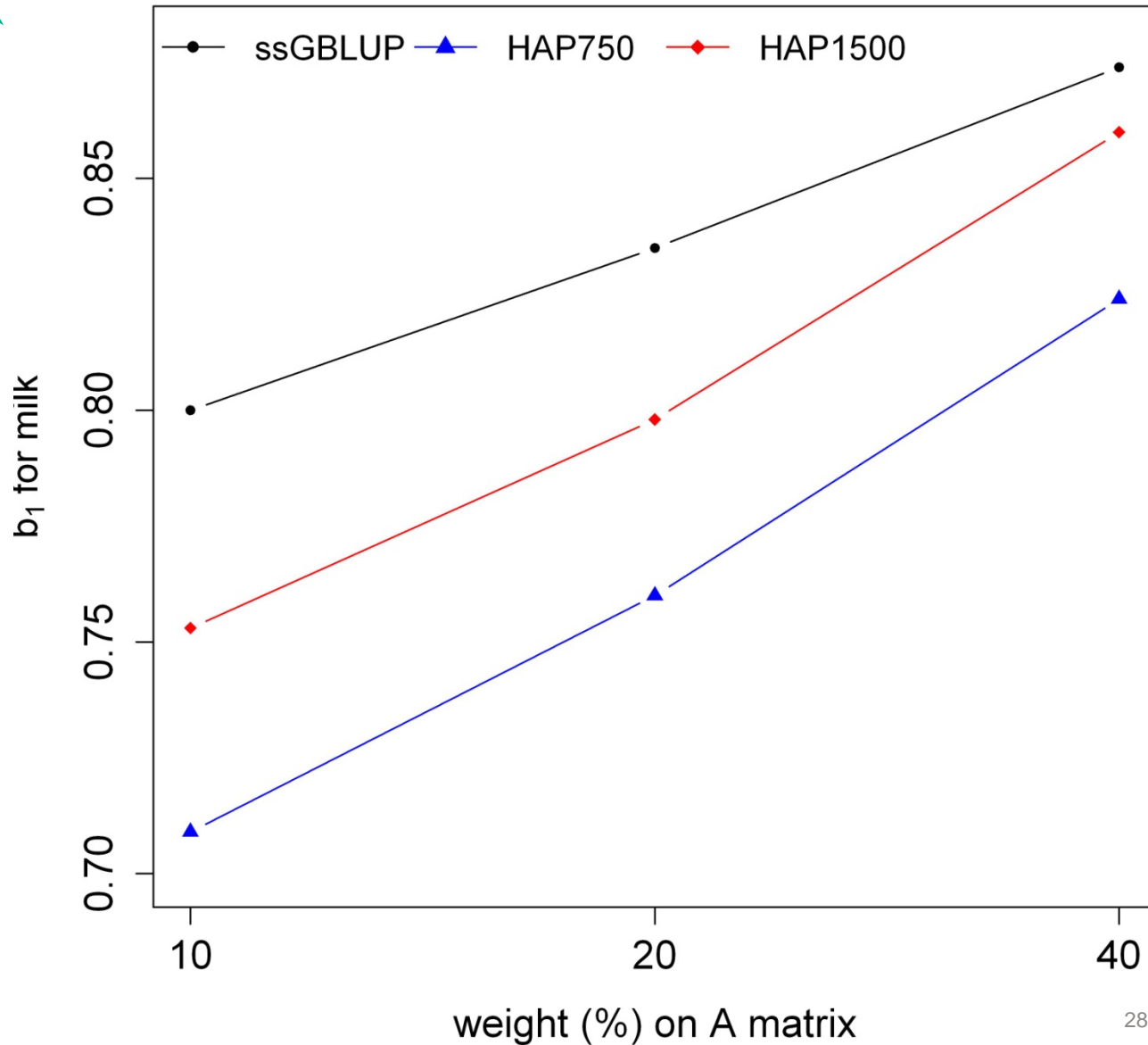


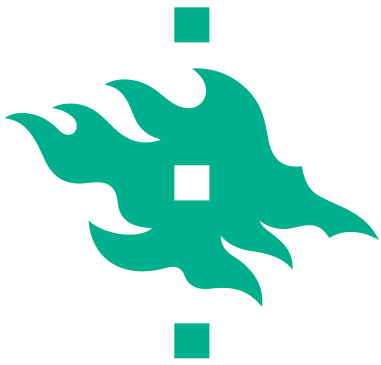
Validation reliabilities for protein



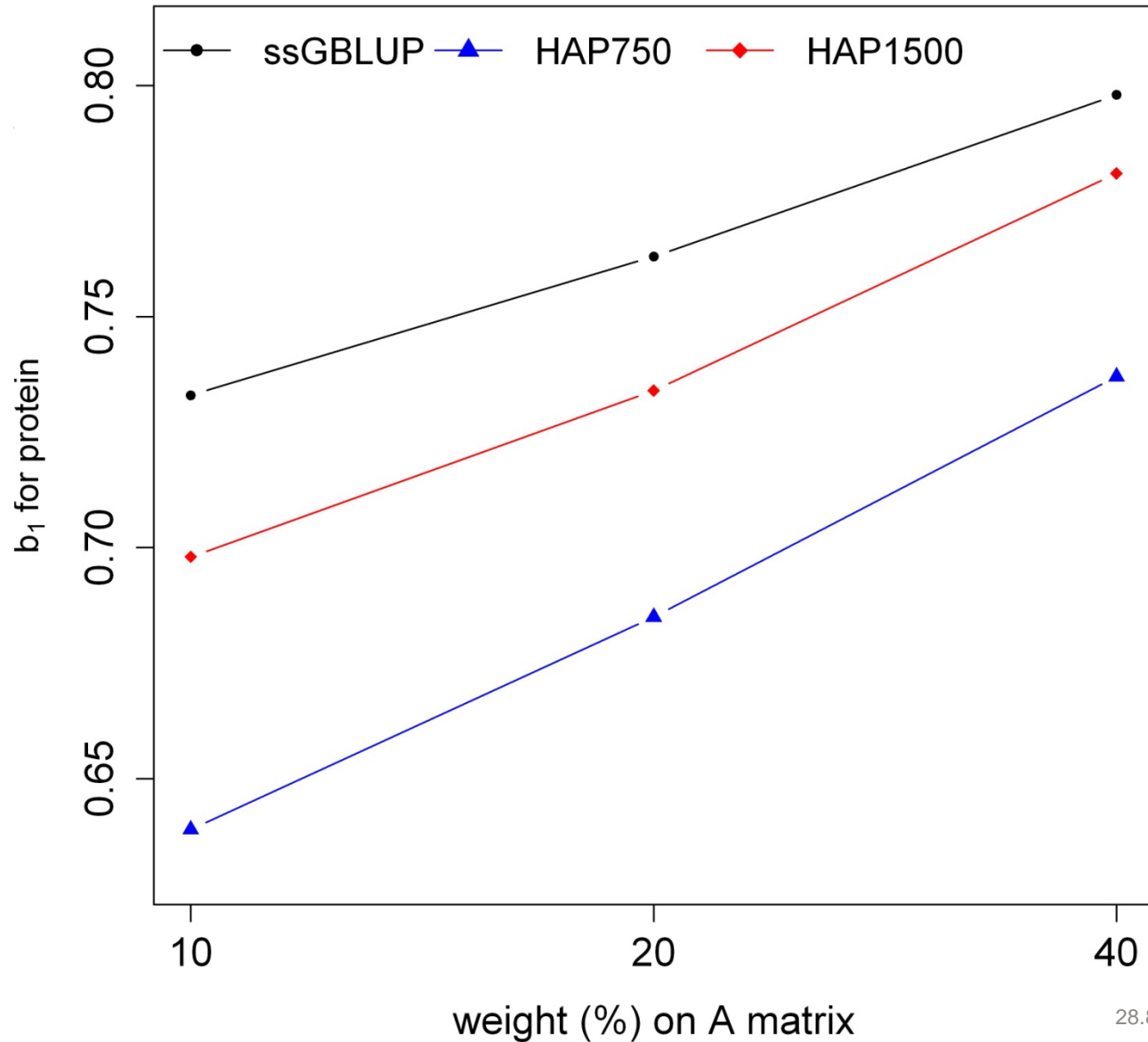


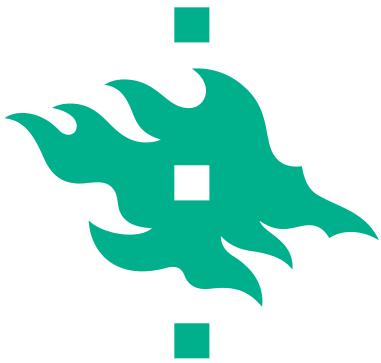
Inflation factors b_1 : MILK





Inflation factors b_1 : PROTEIN

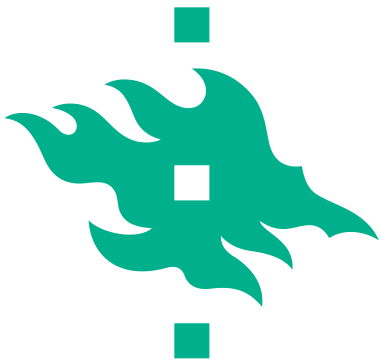




Conclusions



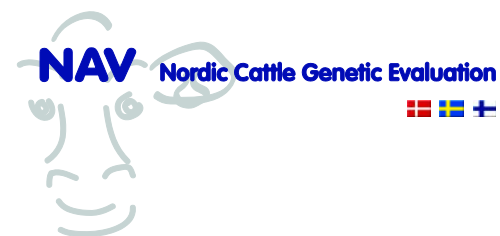
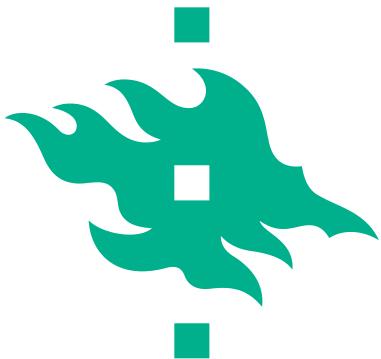
- The use of haplotype segments appeared to be promising, and highlights a need for balance between the number of haplo-blocks, and the optimal scaling with **A**
- Evaluations for milk were improved using more haplotype segments i.e., **HAP1500 than HAP750**
- Reliabilities of haplo-block models tended to increase with increasing weight on **A**



Conclusions



- Inflation of GEBV was greater with haplo-block models than standard single step
 - Inflation with haplo-block models was reduced when the weight on **A** increased



THANK YOU !!!