Comparison of different methods to calculate genomic predictions

– results from SNP-BLUP, G-BLUP and one-step H-BLUP

Minna Koivula, Ismo Strandén,
Esa Mäntysaari

MTT Agrifood Research Finland, Biotechnology
and Food Research, Biometrical Genetics

Introduction

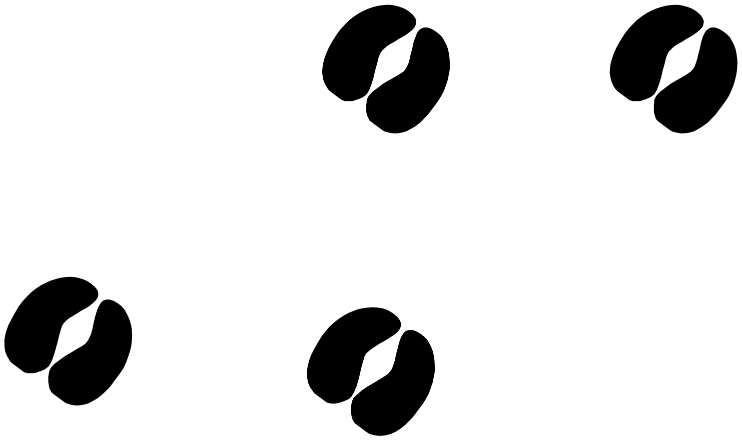
- Genomic selection refers to genetic improvement of animals through selection based on genomic breeding values (GEBVs)

- GEBVs predicted using a reference population of animals that have genotype as well as phenotypic information

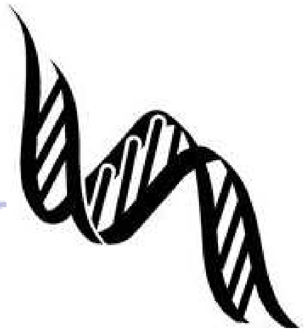
- Many statistical models proposed to predict GEBVs

- Objectives of this study were to apply
  - a simple marker model in SNP-BLUP
  - alternative genomic matrices in G-BLUP
  - one-step BLUP (H-BLUP)

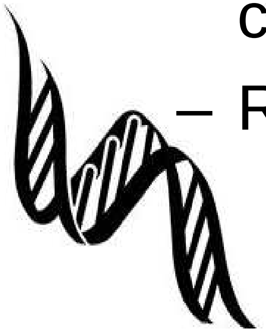
… analyses of production and mastitis traits in Nordic red cattle (RDC)

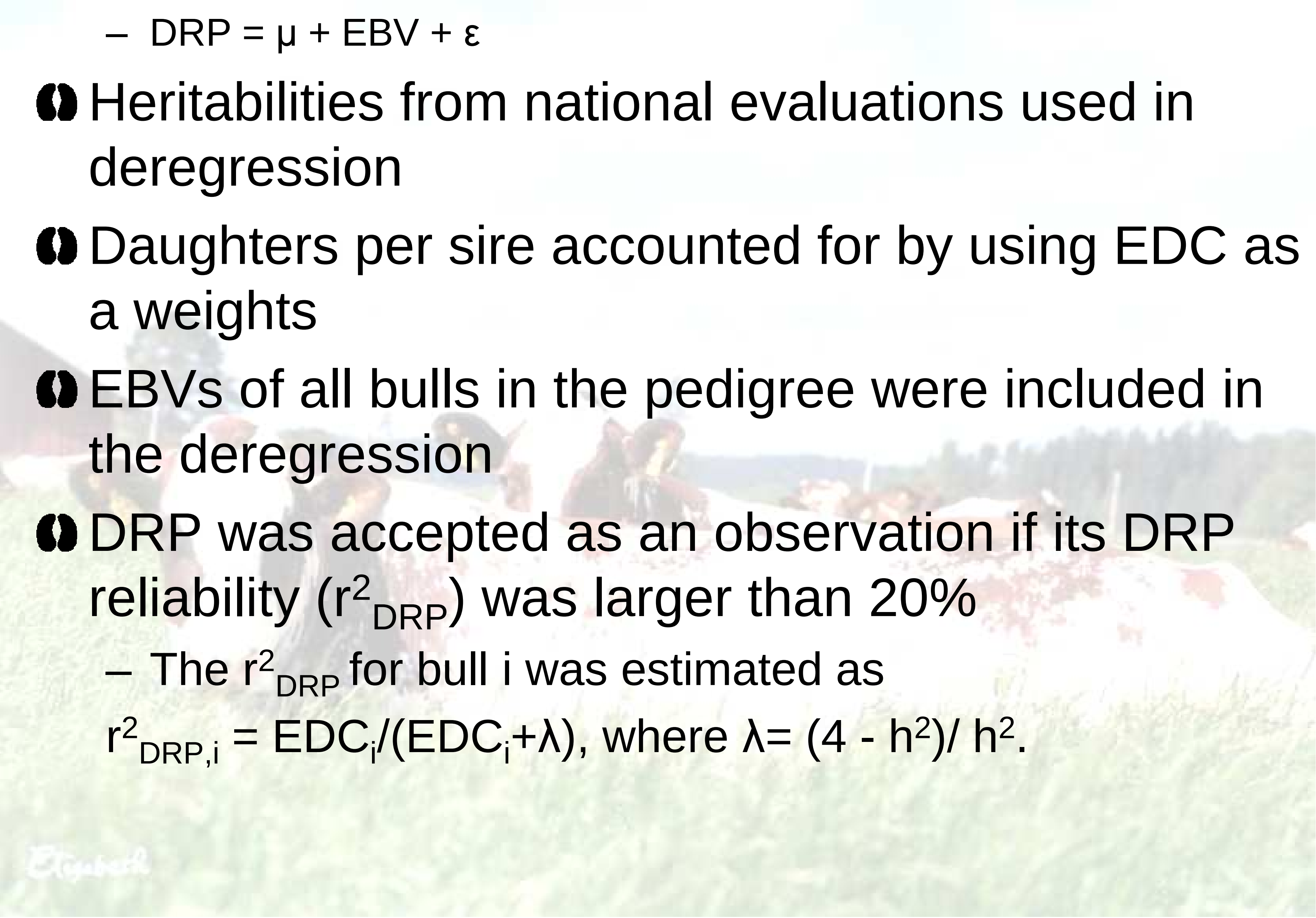


Materials and methods



- 6145 genotyped RDC bulls with 37996 snp’s
  - bulls genotyped using the Illumina Bovine SNP50 BeadChip (Illumina, San Diego, CA)
- Phenotypic data from official Nordic genetic evaluations for the RDC
  - Full data for bulls from March 2010 with EBVs, reliabilities (r^2EBV), and effective daughter contributions (EDC)
  - Reduced data with the same animals, but EBVs calculated using data until year 2005
  - Response variables deregressed proofs (DRP)





- Deregressions with procedure DeRegress in MiX99
  - $DRP = \mu + EBV + \varepsilon$
- Heritabilities from national evaluations used in deregression
- Daughters per sire accounted for by using EDC as a weights
- EBVs of all bulls in the pedigree were included in the deregression
- DRP was accepted as an observation if its DRP reliability (r^2_{DRP}) was larger than 20%
  - The r^2_{DRP} for bull i was estimated as

    $r^2_{DRP,i} = EDC_i/(EDC_i+\lambda)$, where $\lambda = (4 - h^2)/ h^2$.

Heritabilies (h^2), lambdas (λ), and the average r^2_{DRP} by trait

Trait	h^2	λ	r^2_{DRP} **Reference bulls**	r^2_{DRP} **Candidate bulls**
Milk	**0.39**	**9.26**	**0.96**	**0.94**
Protein	**0.31**	**11.90**	**0.95**	**0.93**
Fat	**0.36**	**10.11**	**0.96**	**0.93**
Mastitis	**0.04**	**99.00**	**0.88**	**0.80**

Number of bulls in different data sets

Trait	Full Data		Reference Bulls		Candidate bulls
	Bulls with DRP	Genotyped with DRP	Bulls with DRP*	Genotyped with DRP	
Milk	**6253**	**4145**	**5309**	**3330**	**809**
Protein	**6253**	**4145**	**5309**	**3330**	**809**
Fat	**6253**	**4145**	**5309**	**3330**	**809**
Mastitis	**6169**	**4431**	**5363**	**3649**	**780**

*For one-step method reference bulls include both genotyped and non-genotyped bulls

BLUP models

1. SNP-BLUP

y= **y=1μ + Mg + e**

- DGV estimated $\hat{\mathbf{a}} = \mathbf{1}\hat{\mu} + \mathbf{M}\hat{\mathbf{g}}$
- $\hat{\mathbf{g}}$ are the estimated marker effects from the SNP-BLUP

2. G-BLUP

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\mathbf{a} + \mathbf{e}$$

- Defined equivalent to the SNP-BLUP model but instead of including the SNP-effects to the model only the sum of the effects $\mathbf{a}=\mathbf{Z}\mathbf{g}$ included
- Uses either:
  - Unscaled G-matrix $\mathbf{G}_0 = \mathbf{Z}\mathbf{Z}'$
  - Scaled G-matrix $\mathbf{G}_k = \mathbf{Z}\mathbf{Z}'/k$
    - $k=2\Sigma p_i(1-p_i)$ is a scaling parameter

3. One-step H-BLUP

$$\mathbf{y_t} = \mathbf{1}\mu + \mathbf{Wa} + \mathbf{e}$$

- Uses G –matrix from genotyped animals and A_{11} relationship matrix from pedigree

- $\mathbf{H^{11}} = \mathbf{G_w}^{-1} - \mathbf{A_{11}}^{-1}$
  - $\mathbf{G_w} = w\, t\, \mathbf{G}_k + (1-w)\mathbf{A_{11}}$ with $t = \Sigma(\mathbf{A_{ii}}) / \Sigma(\mathbf{G_{k,ii}})$
  - t scales the sums of diagonals in $\mathbf{G}_k$ and **A** to be equal
  - w=0.90 which assumes that 10% of total genetic variance is due to the polygenic effect not described by the SNP markers

Analyses

- Variance of marker effects (σ_g^2) and residual variance (σ_e^2) estimated from the full data using a SNP marker genetic model with Bayesian method
- All weighted analysis with w=EDC/λ
  - $\lambda = (4 - h^2)/ h^2$
  - w accounted heterogeneous residual variances due to difference in reliabilities of DRP

Validation

- Regression of DRP 2010 to DGV or GEBV 2005
- Each R^2 value was divided by the average accuracy of DRP

$$R^2_{\text{validation}} = R^2_{\text{model}} / \overline{r^2_{\text{DRP}}}$$

- Parent Average BLUP
- Comparison of DGV
- Comparison of R^2

Results

- Computing time did not differ much with different BLUP models
  - The G_k-BLUP needed less iterations and less time than the other BLUP models.
  - In practice, differences between the models were negligible as the G_k-BLUP converged in about 50 seconds and SNP-BLUP in 5 G_0-BLUP in 9 and H-BLUP in 3 minutes (with 2.8 GHz)
- The most time consuming part in the G-BLUP and H-BLUP was the construction and inverse of the G-matrix and the H-matrix (~20 minutes)

Correlations of genomic predictions for candidate bulls

		G_k-BLUP	G_0-BLUP	H-BLUP
Milk	**SNP-BLUP**	0.9996	0.9998	0.9704
	G_k-BLUP		0.9997	0.9707
	G_0-BLUP			0.9709
Protein	**SNP-BLUP**	0.9997	0.9999	0.9616
	G_k-BLUP		0.9997	0.9618
	G_0-BLUP			0.9622
Fat	**SNP-BLUP**	0.9997	0.9999	0.9782
	G_k-BLUP		0.9997	0.9786
	G_0-BLUP			0.9785
Mastitis	**SNP-BLUP**	0.9997	0.9999	0.9837
	G_k-BLUP		0.9996	0.9832
	G_0-BLUP			0.9836

Reliability, candidates

	Milk		**Protein**		**Fat**		**Mastitis**	
	b_1	**R^2**	**b_1**	**R^2**	**b_1**	**R^2**	**b_1**	**R^2**
Parent Average	0.73	0.19	0.77	0.20	0.83	0.23	0.65	0.08
SNP-BLUP	0.76	0.30	0.77	0.31	0.85	0.40	0.76	0.17
G_k-BLUP	0.77	0.30	0.78	0.31	0.86	0.40	0.77	0.17
G_0-BLUP	0.76	0.30	0.77	0.31	0.85	0.40	0.76	0.17
H-BLUP	0.80	0.32	0.83	0.34	0.90	0.42	0.77	0.17

Conclusions

- Computing time did not differ much with different BLUP models

- SNP-BLUP and G_0-BLUP give same evaluations
- G_k-BLUP slightly different b1 values
- b_1 values suggests that DGVs overpredict the variation in DRPs

- Milk and Protein validation values were low (~0.31)
- Fat $R^2_{validation}$ slightly higher (~0.40)
- Mastitis validation very low (0.17)

- The results indicate that different genomic models give comparable results
  - For the candidate bulls, the SNP-BLUP and G-BLUP gave the same DGV's but there was a difference to those from H-BLUP

- In general, reliability of DGV was 45% higher than reliability of traditional PA, averaged over the production and mastitis traits

- H-BLUP had slightly higher validation reliabilities than the other models
  - Presumably because of 0.9 weight for H^{11} matrix

Acknowledged for providing the genotype data and the EBV data

Nordic Cattle Genetic Evaluation

VIKINGGENETICS®

faba

DET JORDBRUGSVIDENSKABELIGE FAKULTET
AARHUS UNIVERSITET