The Contribution of Linkage and Linkage Disequilibrium Information to the Accuracy of Genomic Selection

Tu Luan*, John A. Woolliams *,† Jørgen Ødegård*,‡ and Theo H.E. Meuwissen*

*Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences

†The Roslin Intitute (Edinburgh), University of Edinburgh, Roslin, UK

‡Nofima Marin, Ås, N-1432, Norway

Introduction

- Genomic Selection (GS) refers to selection decisions based on genome-wide breeding values (GW-EBV) predicted through the use of dense markers covering the whole genome.
- The prediction of GW-EBV is generally thought to be achieved by using linkage disequilibrium (LD) between the markers and QTL.
- Habier et al. (2007) showed that GS implicitly also uses genetic relationships.

Objective

To quantify to what extent the reliability of GW-EBV prediction is due to linkage analysis information and how much due to LD that already existed in the founders of the pedigree.

Material

- **All Data**
  - 255 British Holstein bulls
  - Average number of daughters is 200
  - Genotypic data – 45,888 SNP markers
  - Phenotypic data – DYDs for milk yield, fat yield and protein yield

- **Training Data**
  - randomly mask 51 bulls each time
  - 5 non-overlapping training data sets
  - each of all 255 bulls masked once
  - 6 replicates

Provision of LA and LD

- **LA information**

A genomic identity-by-descent (IBD) matrix, $\boldsymbol{G}_{IBD}$, containing identity-by-descent probabilities within the known pedigree, depicts LA information.

- **LD information**

LD information in the founders of the pedigree is provided by a genomic identity-by-state (IBS) matrix , $\boldsymbol{G}_{IBS}$.

Set up of LA based G_{IBD} matrix

$\boldsymbol{G}_{IBD}$ construction:

- based on Fernando and Grossman's method (1989)
- averaged over maker positions

To investigate the effect of the number of generations of pedigree used on the accuracy of the GW-EBV prediction, we set up $\boldsymbol{G}_{IBD}$ matrix using 1, 2, 3, 4 and 5 generations of pedigree.

Model

- GW-EBV prediction with LA information

$$y = 1\mu + Zu + e$$

- GW-EBV prediction with LD information

$$y = 1\mu + Za + e$$

- GW-EBV prediction with LD+LA information

$$y = 1\mu + Za + Zu + e$$

$$\boldsymbol{u} \sim N(0, \boldsymbol{G}_{IBD}\sigma_u^2) \quad \boldsymbol{a} \sim N(0, \boldsymbol{G}_{IBS}\sigma_a^2)$$

Prediction of GW-EBV

- In total 30 analyses (5 training data sets X 6 replicates)
- Only converged analyses were used
- Mean accuracy for converged analyses

Number of converged calculations

Method	fat yield	milk yield	protein yield
IBD	8	30	25
IBS	30	30	30
IBD+IBS	8	30	29

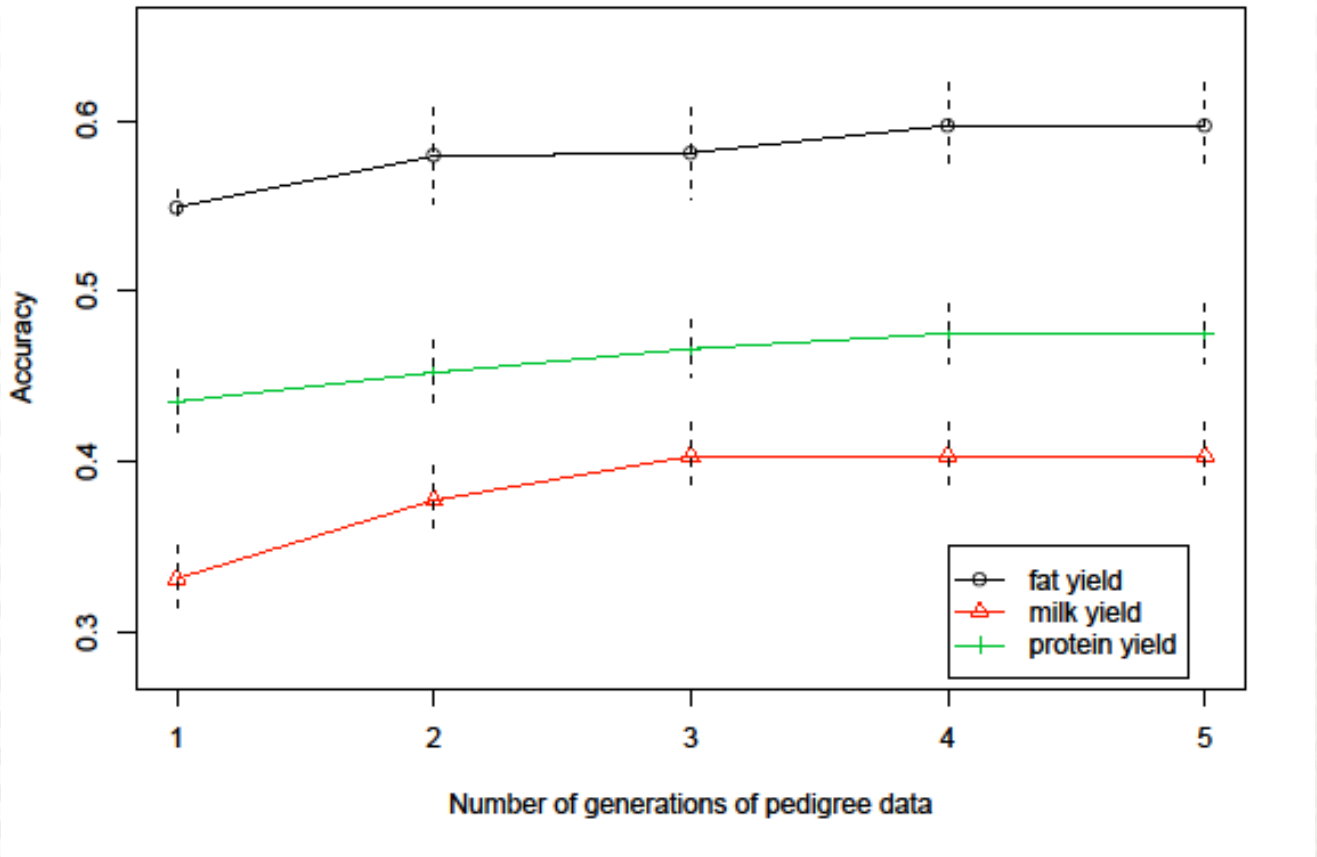
Results

Reliability of GW-EBV (±SE) obtained using IBD, IBS and IBD+IBS genomic relationship matrices

Methods	Fat yield	Milk yield	Protein yield
IBD	0.3364 ± 0.0010	0.1624 ± 0.0004	0.2228 ± 0.0004
IBS	0.3158 ± 0.0002	0.1584 ± 0.0004	0.2209 ± 0.0003
IBD+IBS	0.3272 ± 0.0015	0.1673 ± 0.0004	0.1989 ± 0.0003

Results

Accuracy of GW-EBV (±SE) predicted with IBD information using different number of generations of pedigree data



Conclusions

- Results show genomic relationships between known relatives (IBD relationships) are responsible for nearly all the accuracy of GS.
- Although GS in principle does not require the availability of pedigree data, it does use available pedigree structure.
- $\boldsymbol{G}_{IBD}$ does not use much information more than 5 generations ago.

Acknowledgement

This research was supported by Cogent and the SABRE project that has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2006-016250.

Thanks !