Comparative assessment of methods for estimating genomic relationships and their use in predictions in an admixed population

M.L. Makgahlela[1,2], I. Strandén[1,2], U.S. Nielsen[3] , M.J. Sillanpää[1,4], J. Juga[1] & E.A. Mäntysaari[2]

[1]University of Helsinki, Finland

[2]MTT Agrifood Research Finland, 31600 Jokioinen

[3]*Danish Agricultural Advisory Service, Udkaersvej 15, Denmark*

[4]University of Oulu, Finland

Introduction

- Accurate estimation of relationships between animals is an important step in any routine genetic evaluations
- Relationships were previously based on pedigree information only
- Conversely, most current evaluations use both marker-derived relationship matrix (**G**) and pedigree-based relationships (**A**)
- **G** estimators are more accurate than **A** because they have more variation between closely related individuals

Introduction

- The accuracy of **G** estimators may be even higher
  - If founder population allele frequencies were available
- In the absence, current population allele frequencies are used to make **G** and that defines the founder population
- The use of observed allele frequencies in structured populations however, may lead to biased estimation of **G**

Objectives

- To estimate **A** and **G** matrices
    - Different **G** matrices were estimated using either observed allele frequencies across breeds or breed allele means
- To estimate breeding values (EBV) and direct genomic values (DGV) using different G matrices
    - Estimated coefficients and their respective DGVs were compared

The population

- The Nordic Red dairy cattle (RDC) is a combined population
    - 3 sub-populations from **DNK**, **SWE** & **FIN**
    - 2nd largest breeding population, with *Ne* larger than Holsteins
    - Most animals in the data (~98%) are composites of breeds
- Absence of pure breed animals remains a major limiting factor for the estimation of breed-specific allele frequencies

Materials and Methods

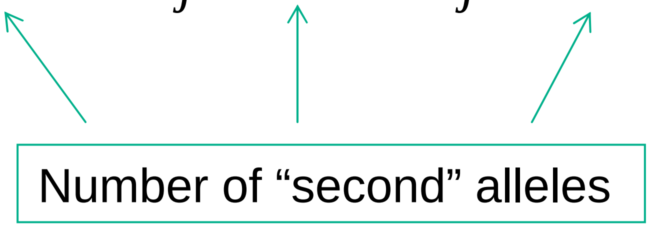
- Data were genotypes of 38194 SNP markers for 4106 bulls
- Breed proportions for bulls were estimated from the full Nordic RDC pedigree (>4m animals)
  - 3 main breeds defined with mean BP>10% were,
    - SRB, FAY & NRF
    - Remaining breeds with mean BP<10% -> breed “OTHER”
- Phenotypes were cow IDDs for milk, protein & fat, from 2010 NAV routine evaluations

Estimation of relationships

- Genomic relationships (**G**) were estimated following methods 1 and 2 by VanRaden (2008)
- **G** estimated using observed allele frequencies (**GOF**)

GOF = ZZ´/k

- $$Z_{i,j} \leftarrow (0-2p_j);(1-2p_j);(2-2p_j)$$



Number of “second” alleles

- p_j is the frequency for the 2nd allele & $k=2\sum_j p_j(1-p_j)$

Estimation of relationships

- **G** matrices estimated using breed allele means (**GBM** and **GBM2**)

GBM = **MM´**/k

- $M_{i,j} \leftarrow (0-2p_{ij});(1-2p_{ij});(2-2p_{ij})$
- p_{ij} is the expected allele frequency of marker *j* for bull *i* given it's base breed proportions

✓ computed by multiple regression of genotypes on BP

Estimation of relationships

- Modification of VanRaden method II

- There,

$$G_{PvRII} = ZDZ'/m = ZD^{0.5}D^{0.5}Z'/m$$

- m is the number of markers

Estimation of relationships

Following the same, we define:

GBM2 = **M*M*´**/m

$$M^*_{i,j} \leftarrow \frac{-2p_{ij}}{\sqrt{2p_{ij}(1-p_{ij})}};\frac{1-2p_{ij}}{\sqrt{2p_{ij}(1-p_{ij})}};\frac{2-2p_{ij}}{\sqrt{2p_{ij}(1-p_{ij})}}$$

- m is the number of markers

Combined **A** and **G** matrices

- Pedigree relationships (**A**) were estimated for genotyped bulls only, using *RelaX2* computer program

- **GOF** and **GBM2** were combined with 20% weight on **A** to yield **GAOF** and **GABM2**
  - **G*** = w**G** + (1-w)**A**

Statistical Analyses

- Variance components, EBVs & DGVs were estimated separately for each matrix, using a GBLUP model

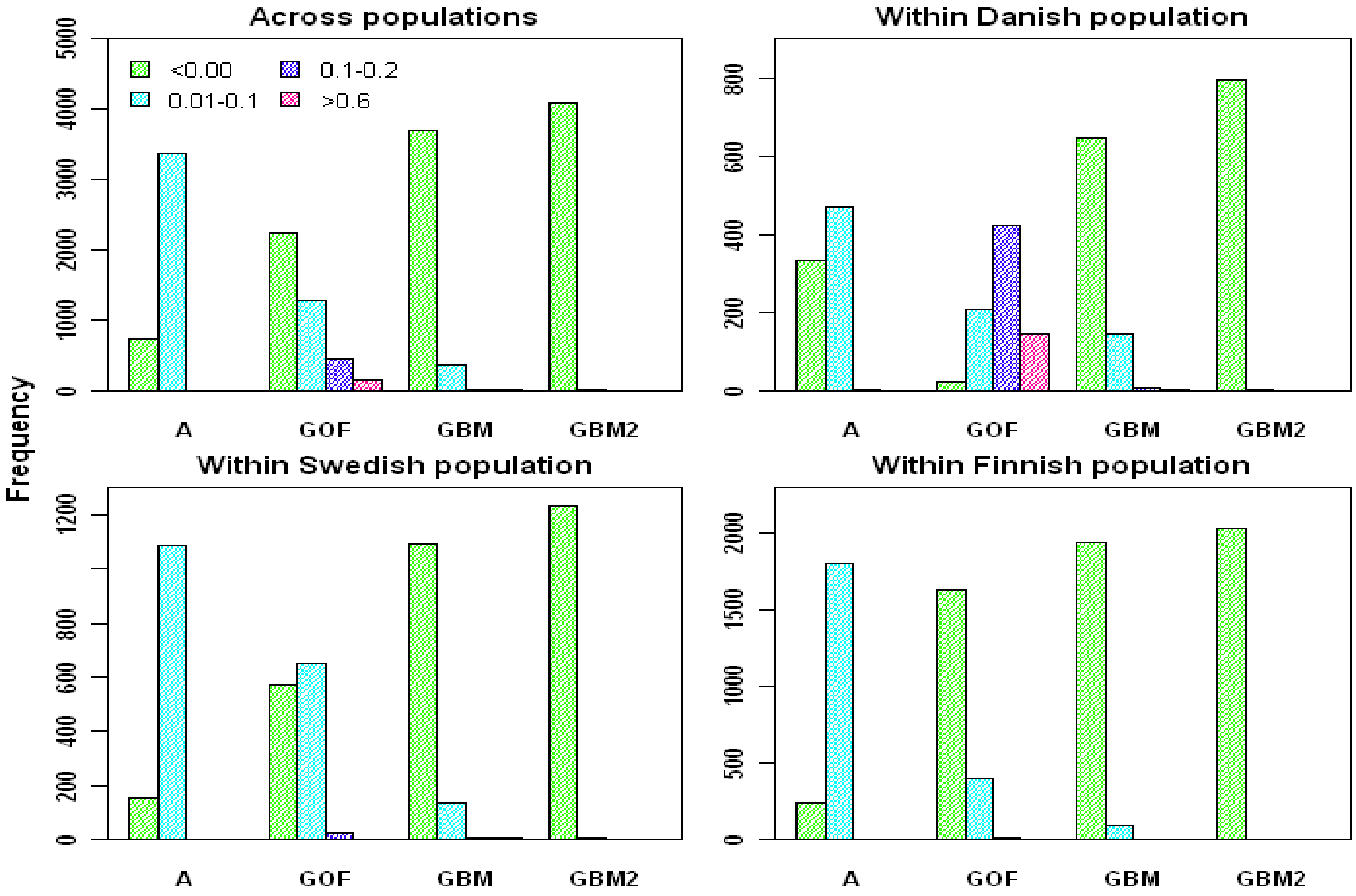
- **y** = **Xb** + **Za** + **e**,
    - **y** is a vector of cow IDD
    - **X** and **Z** are design matrices allocating records to **b** and **a**
    - **b** is a vector of fixed mean and breed regression effects
    - **a** is a vector of breeding values
    - **e** is a vector of residuals

- Breed regression effects were used only for predictions with **GBM** and **GABM2**
    - ✓ Predicted values included fixed regression solutions

Statistics of (diagonals)-1 by estimator

	Mean	Min	Max	Mean	Min	Max
	Across populations			Within Swedish bulls		
A	0.012	0.000	0.135	0.008	0.000	0.081
GOF	0.019	-0.129	0.379	0.006	-0.129	0.184
GBM	-0.051	-0.254	0.310	-0.043	-0.226	0.234
GBM2	-0.242	-0.387	0.093	-0.238	-0.387	0.029
	Within Danish bulls			Within Finnish bulls		
A	0.007	0.000	0.109	0.016	0.000	0.135
GOF	0.136	-0.027	0.328	-0.021	-0.123	0.157
GBM	-0.040	-0.173	0.310	-0.062	-0.217	0.283
GBM2	-0.233	-0.339	0.093	-0.250	-0.377	0.077

Histograms of (diagonals)-1 for A and G matrices



Correlations between A and G matrices for pair-wise relationships

	Across populations		Within Swedish bulls	
A	0.702	0.661	0.789	0.781
GOF		0.537		0.784
GBM				

	Within Danish bulls		Within Finnish bulls	
A	0.644	0.856	0.819	0.759
GOF		0.625		0.876
GBM				

Correlations between EBV & DGV from different estimators for validation bulls

	GOF	GBM	GBM2	GAOF	GABM2
A	0.67	0.67	0.66	0.76	0.76
GOF		1.00	1.00	0.98	0.98
GBM			1.00	0.98	0.98
GBM2				0.98	0.98
GAOF					1.00

EBV i.e. Parent average

Conclusions

- The use of simple observed allele frequencies across breeds over-estimate values in **G** for:
  - Populations with the least number of animals in the combined data and/or,
  - Individuals from distantly related populations
- Estimated breed allele means reduced country differences in coefficients, similarly, but shifted them too much torwards zero or less

Conclusions

- The prediction of DGV converged to similar solutions regardless of allele frequencies used
  - Inclusion of breed regressions for **GBM** & **GBM2** brought breed means back into the DGV
- The validation accuracy slightly increased when **A** and **G** matrices were combined
- A single-step **GBM2** and **A** including non-genotyped animals could increase the prediction of DGV even more

☺ Thank you for your attention
Questions !!!