Genomic predictions
New Developments

Esa Mäntysaari

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

Introduction

Genomic evaluation
vs. marker
assisted evaluation

Single-step model
combines genomic and
phenotypic information

Genomic evaluation of small breeds

Holstein bulls in Reference:

- 6670 Nordic bulls
- 18500 EuroGenomics bulls

Red Bulls

- 5670

The reliability 16 traits of

- Reds 0.34
- Holstein 0.45

Validation Reliability of Nordic Genomic evaluations

Ominaisuus	Holstein R^2	Nordic RED R^2
Milk	0.59	0.35
Protein	0.55	0.31
Fat	0.57	0.44
Fertility	0.55	0.30
Mastitis	0.48	0.24
Udder conform.	0.53	0.33
Longevity	0.46	0.43

Genomic evaluation of breed crosses

Experiences with the Illumina high density Bovine BeadChip

B. L. Harris, F. E. Creagh,
A. M. Winkelman and D. L. Johnson
August 2011

Results: Across breed performance

- Bayesian Ridge Regression with a polygenic effect (Campos et al, 2009)

Training	Test	N Training	N Test	Accuracy 50K	Accuracy HD
HF + Jersey	HF-J	3713	498	0.67	0.67
HF	Jersey + HF-J	2290	1423	0.37	0.41
Jersey	HF + HF-J	1423	2290	0.15	0.22

0.45
0.14
0.02

LIC

DairyNZ

LIC

Options to fine tune genomic model:

- More data:
  1. animals – combine reference populations
  2. SNPs – move to HD –chip

- Better genomic models – most likely with 1 & 2 above
  - Bayesian variable selection models, better algorithms
  - Genomic model concept:
    - Add the breed background to the model
    - Address the genes, instead of relationships

“Sustainability” of genomic evaluation?

Currently we use 2-step, or 3-step approach

1. Estimate EBVs using national evaluation model

2. Solve SNP-effects using genotypes and EBVs, (or DGVs with genomic relationships)

3. Combine pedigree information and DGVs

This relies on accuracy of EBVs !!

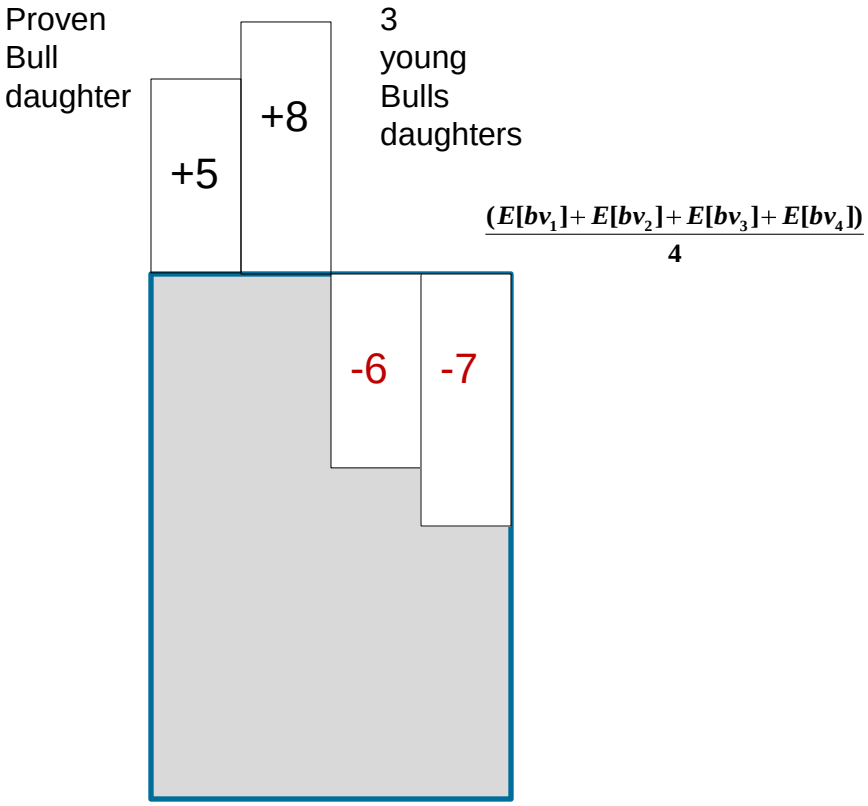
Genomic selection deteriorates the reliability of national EBVs

Henderson 1973, 1975,…

> "BLUP is unbiased by selection if the information used for selection is included in the data"

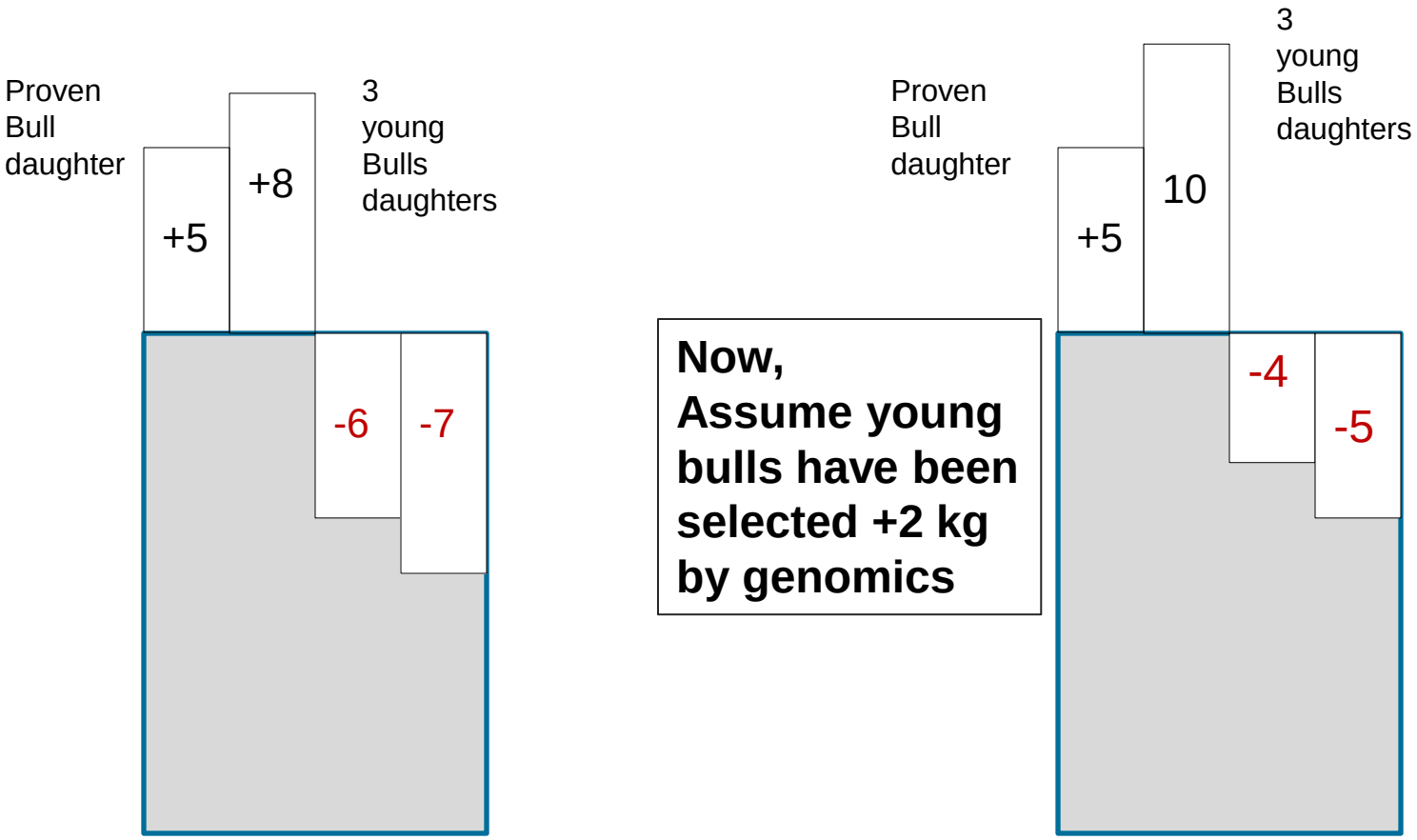
- If the AI bulls entering service are selected group of the bull calves, then their expectation is not PA (Parent average EBV)
  - This leads into wrong partition of environmental and genetic trends
  - young sires will become underestimated
  - Daughters of "GenVik" bulls will get poor indices ?

BLUP models the environment and breeding values simultaneously



- BLUP assumes that the production is a sum of environment and expected genetic value
- Alternatively:
  Environment =
  Actual production – mean(E[BV])

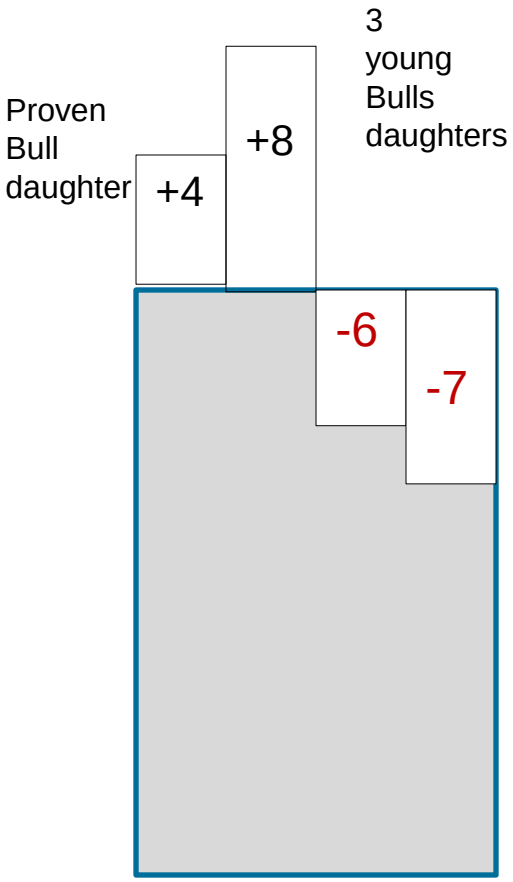
BLUP models the environment and breeding values simultaneously



Now, Assume young bulls have been selected +2 kg by genomics

BLUP models the environment and breeding values simultaneously

- If the **expected value** of genomic young bulls is same as before, BLUP assumes the environment has improved



Genomic selection deteriorates the reliability of national EBVs

Henderson 1973, 1975,…

> "BLUP is unbiased by selection if the information used for selection is included in the data"

- If the AI bulls entering service are selected group of the bull calves, then their expectation is not PA (Parent average EBV)
  - This leads into wrong partition of environmental and genetic trends
  - young sires will become underestimated
  - Daughters of "GenVik" bulls will get poor indices ?
- **WE can not continue using 2-step after daughters of selected bulls start produce**

Options

We need to evaluate EBVs with genomic information included

- So called bi-variate blending helps here
- But we can not use those GEBVs in 2-step, because that would return the genomic information (to genomic information)

We need to evaluate GEBVs and EBVs simultaneously

- Single-step approach
  - Evaluate GEBVs directly in national evaluations

(Christensen and Lund, 2010; Aguilar et al. 2010, Legarra et al 2010; Mistzal et al. 2010)

GENOMIC MODEL IN MARKER ASSISTED SELECTION

Genomic model in Marker Assisted Selection

Genomic selection model (ours)

- Replaces pedigree based relationships of genotyped animals by genomic relationships
  - Elements in G are correlations of genotype “scores”
  - While a_{ij} for paternal half-sibs is 0.25 in g_{ij} ~0.2-0.3
  - Each animal is “related” to all other animals

MAS model in evaluations

- Bases on usual evaluation model, but is **complemented** with QTL-effects
  - If few QTL effects a sum of them fitted
  - If many QTL effects assumed, they are replaced by “marker” effects

French report on the use of genomic evaluation

Vincent Ducrocq[1], *Sébastien Fritz* [2,1], *François Guillaume* [3,1], *Didier Boichard* [1]

- Mixed linear model including a polygenic effect and a regression on identical-by-state haplotypes

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \sum_{i=1}^{n\_QTL} \mathbf{Z}_{\mathbf{h}_i}\mathbf{h}_i + \mathbf{e}$$

data = (weighted) DYD or YD

PROCEEDINGS OF THE INTERBULL

INTERNATIONAL WORKSHOP

Genomic Information in Genetic Evaluations

UPPSALA, SWEDEN

January 26-29, 2009

BULLETIN NO. 39, 2009

French report on the use of genomic evaluation

Vincent Ducrocq[1], *Sébastien Fritz* [2,1], *François Guillaume* [3,1], *Didier Boichard* [1]

- Mixed linear model including a polygenic effect and a regression on identical-by-state haplotypes

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \sum_{i=1}^{n\_QTL} \mathbf{Z}_{\mathbf{h}_i}\mathbf{h}_i + \mathbf{e}$$

data = (weighted) DYD or YD

i=1

- 19 to 32 QTL/trait/breed explaining 50 to 70% of σ_g^2 but artifically bounded at 60 % !

  About 30 haplotype effects to estimate per QTL

PROCEEDINGS OF THE INTERBULL

INTERNATIONAL WORKSHOP

Genomic Information in Genetic Evaluations

UPPSALA, SWEDEN

January 26-29, 2009

BULLETIN NO. 39, 2009

French report on the use of genomic evaluation

Vincent Ducrocq[1], *Sébastien Fritz* [2,1], *François Guillaume* [3,1], *Didier Boichard* [1]

- Mixed linear model including a polygenic effect and a regression on identical-by-state haplotypes

$$y = Xb + Zu + \sum_{i=1}^{n\_QTL} Z_{h_i} h_i + e$$

data = (weighted) DYD or YD

- 19 to 32 QTL/trait/breed explaining 50 to 70% of σ_g^2 but artifically bounded at 60 % !

  About 30 haplotype effects to estimate per QTL

- QTLs have been found using LDLA association analysis, with “stringent” significance levels
- Haploblocks are assembled from blocks based on 5 SNP-markers

PROCEEDINGS OF THE INTERBULL

INTERNATIONAL WORKSHOP

Genomic Information in Genetic Evaluations

UPPSALA, SWEDEN

January 26-29, 2009

BULLETIN NO. 39, 2009

Assets of MAS (Boichard, WCGALP 2010)

- *Haploblock – gene effect* associations are likely to be more consistent than SNP-effects (over generations)
- Haploblocks are more likely to detect *gene effects* than individual SNP-markers

Haploblocks can address origin of the marker-gene association (coalesce)

Haplotype blocks for the Nordic Red work on progress…

- All the (DSF) bulls were phased using Findhap program
  - Recognizes maternal and paternal haplotypes
- Different lengths of physical segments explored to find optimal scheme
  - Compare structures in Reds, FAY and Holsteins
  - E.g. seems that in Holstein 5 haplotypes (50 SNPs) represent 70% of obs in Reds 5 haplotypes represent 25% of obs
- Haploblocks are chosen to model
  1. Using association analysis
  2. Simultaneous estimation of haplotype block variances

SINGLE STEP EVALUATION

Single step evaluation - I

Principle:

- Divide animals to
  non-genotyped (1) and genotyped (2)
- For the (2) we know the relationships: G
- For (1) there are two covariance structures:

var(u_1)=var(u_1 \|u_2) + var(E(u_1\|u_2)), therefore

$$\mathbf{var}(\boldsymbol{u}) = \boldsymbol{H} = \begin{bmatrix} \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{12} + \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{G}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{G} \\ \boldsymbol{G}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{G} \end{bmatrix}$$

Single step evaluation - II

- The beauty of the method:

$$\boldsymbol{H}^{-1} = \boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}^{-1} - \boldsymbol{A}_{22}{}^{-1} \end{bmatrix}$$

- And thereafter the MME:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix}$$

Single step evaluation - III

- What is needed:
  - Genomic relationship matrix and it's inverse
  - Relationship matrix of genotyped animals and its inverse

- Simplified alternative:
  - Use existing EBVs
    → Deregress → fit genomic data via single step
  - \+ Combines DGV and EBV for genotyped animals
  - \+ Introduces genomic information to all relatives
  - – Does not full correct the genomic selection bias problem

Nordic implementations of single step evaluations

Using deregressed bull EBVs (Su et al. 2011; Koivula et al. 2011)

Trait	r_v^2 PIt (%)	r_v^2 DGV (%)	r_v^2 GEBV(o) [b)] (%)
Milk	19.4	35.8	36.7
Fat	25.1	45.4	47.8
Protein	19.9	34.6	37.2
Fertility	16.6	29.7	31.4
Birth index	10.2	19.5	19.8
Calving index	11.1	16.0	15.9
Udder health	17.1	24.4	26.2
Other diseases	32.8	30.1	32.9
Body conform.	30.6	45.0	45.5
Feet & leg	14.8	29.6	26.5
Udder conform.	23.5	32.1	32.0
Milking ability	13.9	29.7	30.1
Temperament	18.9	30.0	29.6
Longevity	24.1	25.9	32.5
Yield	20.6	36.1	38.8
Average	19.9	30.9	32.2

	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

Using deregressed cow EBVs (Mäntysaari et al. 2011)

- 4.6 million animals in data
- 3.4 million cows with records → deregressed EBVs
- Computing time < 1 hour, for milk+protein+fat

\+ Introduces genomic information to **all** relatives

\- Still does not help for genomic selection bias

	Milk			Protein			Fat		
	b_0	b_1	R^2	b_0	b_1	R^2	b_0	b_1	R^2
PA	3.29	0.70	0.22	1.22	0.89	0.25	2.17	0.80	0.28
DGV[§]	3.15	0.76	0.30	4.51	0.77	0.31	2.23	0.85	0.40
Single step sire model[§]	3.67	0.69	0.32	4.70	0.74	0.35	2.69	0.80	0.44
Single step Animal_D	**3.80**	**0.72**	**0.35**	**4.23**	**0.81**	**0.38**	**3.29**	**0.79**	**0.45**
Single step Animal_E	**3.90**	**0.71**	**0.35**	**5.01**	**0.76**	**0.36**	**2.17**	**0.80**	**0.45**

Summary

Project group is working on

- Implementing the single-step
  and perhaps unified single step
- improving the validation accuracy on Red breed
  - Haploblock approach
  - Adding more data (Norwegian genotypes)
  - Tracing the origin of genotypes, genomic structures
  - Multibreed multitrait models
  - Using the HD chip