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 of 16 traits of:
- Reds 0.34
- Holstein 0.45

<table>
<thead>
<tr>
<th>Trait</th>
<th>Holstein R²</th>
<th>Nordic RED R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ominaisuus</td>
<td>0.59</td>
<td>0.35</td>
</tr>
<tr>
<td>Milk</td>
<td>0.59</td>
<td>0.31</td>
</tr>
<tr>
<td>Protein</td>
<td>0.55</td>
<td>0.30</td>
</tr>
<tr>
<td>Fat</td>
<td>0.57</td>
<td>0.44</td>
</tr>
<tr>
<td>Fertility</td>
<td>0.55</td>
<td>0.44</td>
</tr>
<tr>
<td>Mastitis</td>
<td>0.48</td>
<td>0.24</td>
</tr>
<tr>
<td>Udder conform.</td>
<td>0.53</td>
<td>0.33</td>
</tr>
<tr>
<td>Longevity</td>
<td>0.46</td>
<td>0.43</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Breed Combination</th>
<th>Test Breed</th>
<th>N Training</th>
<th>N Test</th>
<th>Accuracy 50K</th>
<th>Accuracy HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF + Jersey</td>
<td>HF-J</td>
<td>3713</td>
<td>498</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>HF</td>
<td>Jersey + HF-J</td>
<td>2290</td>
<td>1423</td>
<td>0.37</td>
<td>0.41</td>
</tr>
<tr>
<td>Jersey</td>
<td>HF + HF-J</td>
<td>1423</td>
<td>2290</td>
<td>0.15</td>
<td>0.22</td>
</tr>
</tbody>
</table>
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 $g_{ij} \sim 0.2-0.3$
  - 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
Mixed linear model including a polygenic effect and a regression on identical-by-state haplotypes:

\[ y = Xb + Zu + \sum_{i=1}^{n_{\text{QTL}}} Z_{hi} h_i + e \]

Data = (weighted) DYD or YD
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 \( \sigma_g^2 \)
  but artificially 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.

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

\[ y = Xb + Zu + \sum_{i=1}^{n_{QTL}} Z_{h_i} h_i + e \]
\[ \text{data} = \text{(weighted) DYD or YD} \]

19 to 32 QTL/trait/breed explaining 50 to 70% of \( \sigma_g^2 \)
but artificially bounded at 60%!

About 30 haplotype effects to estimate per QTL.
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:

\[
\text{var}(u_1) = \text{var}(u_1 | u_2) + \text{var}(E(u_1 | u_2)), \text{ therefore}
\]

\[
\text{var}(u) = H = \begin{bmatrix}
A_{11} - A_{12}A_{22}^{-1}A_{12} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21} & A_{12}A_{22}^{-1}G \\
GA_{22}^{-1}A_{21} & G
\end{bmatrix}
\]
Single step evaluation - II

- The beauty of the method:

\[ H^{-1} = A^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \]

- And thereafter the MME:

\[
\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + H^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}
\]
Single step evaluation - III

- **What is needed:**
  - Genomic relationship matrix and its 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 fully correct the genomic selection bias problem
# Nordic implementations of single step evaluations

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

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

<table>
<thead>
<tr>
<th>Trait</th>
<th>Milk</th>
<th>Protein</th>
<th>Fat</th>
<th>Mastitis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b_1 )</td>
<td>( R^2 )</td>
<td>( b_1 )</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Parent Average</td>
<td>0.73</td>
<td>0.19</td>
<td>0.77</td>
<td>0.20</td>
</tr>
<tr>
<td>SNP-BLUP</td>
<td>0.76</td>
<td>0.30</td>
<td>0.77</td>
<td>0.31</td>
</tr>
<tr>
<td>( G_0 )-BLUP</td>
<td>0.77</td>
<td>0.30</td>
<td>0.78</td>
<td>0.31</td>
</tr>
<tr>
<td>( G_0 )-BLUP</td>
<td>0.76</td>
<td>0.30</td>
<td>0.77</td>
<td>0.31</td>
</tr>
<tr>
<td>H-BLUP</td>
<td>0.80</td>
<td>0.32</td>
<td>0.83</td>
<td>0.34</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Milk</th>
<th></th>
<th>Protein</th>
<th></th>
<th>Fat</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_0$</td>
<td>$b_1$</td>
<td>$R^2$</td>
<td>$b_0$</td>
<td>$b_1$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>PA</td>
<td>3.29</td>
<td>0.70</td>
<td>0.22</td>
<td>1.22</td>
<td>0.89</td>
<td>0.25</td>
</tr>
<tr>
<td>DGV$^S$</td>
<td>3.15</td>
<td>0.76</td>
<td>0.30</td>
<td>4.51</td>
<td>0.77</td>
<td>0.31</td>
</tr>
<tr>
<td>Single step sire model$^S$</td>
<td>3.67</td>
<td>0.69</td>
<td>0.32</td>
<td>4.70</td>
<td>0.74</td>
<td>0.35</td>
</tr>
<tr>
<td>Single step Animal_D</td>
<td>3.80</td>
<td>0.72</td>
<td>0.35</td>
<td>4.23</td>
<td>0.81</td>
<td>0.38</td>
</tr>
<tr>
<td>Single step Animal_E</td>
<td>3.90</td>
<td>0.71</td>
<td>0.35</td>
<td>5.01</td>
<td>0.76</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^S$ Genomic selection bias
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