Multitrait across country genomic evaluations for Eurogenomics countries

Hanni Kärkkäinen\textsuperscript{1}, Vincent Ducrocq\textsuperscript{2}, Sören Borchersen\textsuperscript{3}, Gert Aamand\textsuperscript{4}, Esa Mäntysaari\textsuperscript{1}

Interbull WS, Dubrovnik, August 25, 2018

\textsuperscript{1}Natural Resources Institute Finland,
\textsuperscript{2}French National Institute for Agricultural Research (INRA), France
\textsuperscript{3}Eurogenomics Cooperation, Denmark
\textsuperscript{4}Nordic Cattle Genetic Evaluation (NAV), Denmark
Introduction
Background:

- Interbull has established a SNP MACE project
  - Mike Goddard’s group in Melbourne and Interbull Center
  - Countries can share SNP-solutions and LHS matrices, even if they do not share genotypes
    \[ \Rightarrow \text{MME can be build because pseudo phenotypes are: RHS} = \text{inv(LHS)} \times \text{SNP solutions} \]
- Eurogenomics countries share also genotypes:
  \[ \Rightarrow \text{Possible to build the TRUE multi-trait across country SNP BLUP evaluation using pseudo phenotypes from all countries directly} \]
Project – 2 years in 2018-2020

- Financed jointly by Luke, INRA, Eurogenomics COOP and German Livestock Association
  - Research contract signed on April
- Post-doc Hanni Kärkkäinen started 15. May

Our goal is to demonstrate and validate the performance of Eurogenomics SNP MACE
Genomic evaluations with MACE reference vs. Eurogenomics SNP MACE

Eurogenomics cooperation now

Nat. conventional breeding value estimation
Country A  Country B  Country C

Nat. Genomic breeding value estimation

Interbull MACE

MACE solutions

Deregression

GENOMIC EVALUATIONS

Nat. SNP-Effects with international information

Euro SNP MACE

Nat. conv. breeding value estimation

Country A  Country B  Country C

Genetic Evaluations

DYDs, DRPs or YD

Phenotypes

Genotypes

SNP solutions (or GFBVs or DGVs)

Other

Countries

MACE

GENOMIC EVALUATIONS

National SNP-Effects with Eurogenomics Information (international)
Rough plan

Stage 1 (12 months)

- Run simple MT SNP model across countries
  - Genotypes from 6 countries
  - Phenotypes: deregressed proofs from individual countries
    - Protein, somatic cell score, female fertility trait
  - Genetic parameters from Interbull
  - Validate the model results

- Stage 1 extra developments
  - Phenotypes: DYDs are used in place of DRPs
  - Estimation of correlations across countries
  - Considerations of allele frequencies
Stage 2 (12 months)

- Pinpoint the development priorities using the experiences from stage 1
- Individual bull reliabilities from the model
- Handling of external information from third countries (via MACE proofs)
- Different genomic models in different countries
  
  Residual polygenic effects, different SNPs, haploblock models
Data
Phenotype data

- National genetic evaluation EBVs of AI sampled bulls, that countries\(^1\) send to Interbull
- Kindly provided us by:
  - Jutta Jaitner & Zengting Liu, Germany
  - Ulrik Sander Nielsen & Gert Pedersen Aamand, Nordic countries
  - Julie Promp & Vincent Ducrocq, France
  - Pedro Vessies & Gerben de Jong, The Netherlands
  - Juan Pena, Spain
  - Monika Skarwecka & Andrzej Zarnecki, Poland

\(^1\)Germany (DEU), Nordic countries Denmark, Finland and Sweden (DFS), France (FRA), The Netherlands (NLD), Spain (ESP) and Poland (POL). Order and abbreviations from Interbull practice.
• Reliabilities of the EBV, and EDC also provided
• Deregressed proofs: computed using EBV, EDC and pedigree with MiX99
• Animals with at least 10 EDC in at least 10 herds used in analyses
• Later we will use DYD’s (not yet asked from countries)
Genotype data

- EG genotypes received from NAV, used "as is"
- 46342 SNP genotypes for 62628 bulls
- coded as 0,1,2

Big thanks to

Bernt Guldbrandtsen, Aarhus University, Denmark
Traits considered

1. Protein yield
   - High heritability trait (0.28 – 0.48)
2. Somatic cell score
   - Medium high heritability trait (0.15 – 0.37)
3. Female fertility
   - Lactating cow’s ability to conceive
     - expressed as an interval trait
     - Interbull fertility trait 4, cc2
   - Low heritability trait (0.01 – 0.08)
   - Countries differ on submitted fertility traits:
     - DEU, DFS, FRA and NLD send “interval from first to last insemination cows (days)”
     - ESP and POL send “days open” as trait 4
### Some statistics of the trait records

<table>
<thead>
<tr>
<th>Trait</th>
<th>Country</th>
<th>NofAnim</th>
<th>heritability</th>
<th>meanEDC</th>
<th>medianEDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>pro</td>
<td>DEU</td>
<td>11322</td>
<td>0.48</td>
<td>783</td>
<td>124</td>
</tr>
<tr>
<td>pro</td>
<td>DFS</td>
<td>7079</td>
<td>0.42</td>
<td>496</td>
<td>131</td>
</tr>
<tr>
<td>pro</td>
<td>FRA</td>
<td>7969</td>
<td>0.30</td>
<td>1229</td>
<td>126</td>
</tr>
<tr>
<td>pro</td>
<td>NLD</td>
<td>6853</td>
<td>0.44</td>
<td>925</td>
<td>158</td>
</tr>
<tr>
<td>pro</td>
<td>ESP</td>
<td>4399</td>
<td>0.28</td>
<td>553</td>
<td>172</td>
</tr>
<tr>
<td>pro</td>
<td>POL</td>
<td>4953</td>
<td>0.29</td>
<td>237</td>
<td>98</td>
</tr>
<tr>
<td>scs</td>
<td>DEU</td>
<td>11308</td>
<td>0.23</td>
<td>760</td>
<td>120</td>
</tr>
<tr>
<td>scs</td>
<td>DFS</td>
<td>7096</td>
<td>0.23</td>
<td>445</td>
<td>119</td>
</tr>
<tr>
<td>scs</td>
<td>FRA</td>
<td>7949</td>
<td>0.15</td>
<td>1363</td>
<td>139</td>
</tr>
<tr>
<td>scs</td>
<td>NLD</td>
<td>6914</td>
<td>0.37</td>
<td>789</td>
<td>139</td>
</tr>
<tr>
<td>scs</td>
<td>ESP</td>
<td>4370</td>
<td>0.175</td>
<td>581</td>
<td>179</td>
</tr>
<tr>
<td>scs</td>
<td>POL</td>
<td>4913</td>
<td>0.32</td>
<td>212</td>
<td>88</td>
</tr>
<tr>
<td>cc2</td>
<td>DEU</td>
<td>11100</td>
<td>0.010</td>
<td>1272</td>
<td>283</td>
</tr>
<tr>
<td>cc2</td>
<td>DFS</td>
<td>7019</td>
<td>0.064</td>
<td>432</td>
<td>111</td>
</tr>
<tr>
<td>cc2</td>
<td>FRA</td>
<td>7658</td>
<td>0.041</td>
<td>1484</td>
<td>155</td>
</tr>
<tr>
<td>cc2</td>
<td>NLD</td>
<td>6857</td>
<td>0.073</td>
<td>901</td>
<td>198</td>
</tr>
<tr>
<td>cc2</td>
<td>ESP</td>
<td>3692</td>
<td>0.043</td>
<td>364</td>
<td>147</td>
</tr>
<tr>
<td>cc2</td>
<td>POL</td>
<td>3963</td>
<td>0.080</td>
<td>113</td>
<td>51</td>
</tr>
</tbody>
</table>
### Table 2: Country of origin of the animals with protein yield record

<table>
<thead>
<tr>
<th>Country of origin</th>
<th>DEU</th>
<th>DFS</th>
<th>FRA</th>
<th>NLD</th>
<th>ESP</th>
<th>POL</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>7157</td>
<td>193</td>
<td>160</td>
<td>299</td>
<td>431</td>
<td>436</td>
<td>8676</td>
</tr>
<tr>
<td>DFS</td>
<td>244</td>
<td>5803</td>
<td>19</td>
<td>24</td>
<td>44</td>
<td>129</td>
<td>6263</td>
</tr>
<tr>
<td>FRA</td>
<td>330</td>
<td>52</td>
<td>6286</td>
<td>144</td>
<td>509</td>
<td>595</td>
<td>7916</td>
</tr>
<tr>
<td>NLD</td>
<td>1842</td>
<td>367</td>
<td>403</td>
<td>5044</td>
<td>906</td>
<td>333</td>
<td>8895</td>
</tr>
<tr>
<td>ESP</td>
<td>7</td>
<td>0</td>
<td>25</td>
<td>8</td>
<td>1325</td>
<td>5</td>
<td>1370</td>
</tr>
<tr>
<td>POL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1325</td>
<td>2900</td>
<td>2900</td>
</tr>
<tr>
<td>USA</td>
<td>935</td>
<td>457</td>
<td>605</td>
<td>765</td>
<td>630</td>
<td>399</td>
<td>3791</td>
</tr>
<tr>
<td>CAN</td>
<td>344</td>
<td>134</td>
<td>340</td>
<td>142</td>
<td>85</td>
<td>85</td>
<td>1422</td>
</tr>
<tr>
<td>ITA</td>
<td>209</td>
<td>43</td>
<td>83</td>
<td>62</td>
<td>109</td>
<td>41</td>
<td>547</td>
</tr>
<tr>
<td>other</td>
<td>254</td>
<td>30</td>
<td>48</td>
<td>365</td>
<td>68</td>
<td>30</td>
<td>795</td>
</tr>
</tbody>
</table>
### Records common to countries

#### Table 3: Number of protein records common to countries

<table>
<thead>
<tr>
<th></th>
<th>DEU</th>
<th>DFS</th>
<th>FRA</th>
<th>NLD</th>
<th>ESP</th>
<th>POL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>11322</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFS</td>
<td>960</td>
<td>7079</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRA</td>
<td>870</td>
<td>712</td>
<td>7969</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLD</td>
<td>1557</td>
<td>925</td>
<td>901</td>
<td>6853</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESP</td>
<td>1345</td>
<td>837</td>
<td>1326</td>
<td>1211</td>
<td>4399</td>
<td></td>
</tr>
<tr>
<td>POL</td>
<td>1002</td>
<td>718</td>
<td>690</td>
<td>753</td>
<td>836</td>
<td>4953</td>
</tr>
</tbody>
</table>

- Most of the animals with 5 or 6 records imported from USA
- Other traits show similar patterns

#### Table 4: Number and percentage of protein records common to # countries

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>31643</td>
<td>1784</td>
<td>687</td>
<td>393</td>
<td>355</td>
<td>326</td>
</tr>
<tr>
<td>%</td>
<td>89.93</td>
<td>5.07</td>
<td>1.95</td>
<td>1.12</td>
<td>1.01</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Methods
SNP MACE Model

- Basic SNP MACE model \( y = \mu + Zg + e \)

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_6
\end{bmatrix} = 
\begin{bmatrix}
\mu_1 1^{n_1} \\
\vdots \\
\mu_6 1^{n_6}
\end{bmatrix} + 
\begin{bmatrix}
Z_1 g_1 \\
\vdots \\
Z_6 g_6
\end{bmatrix} + 
\begin{bmatrix}
e_1 \\
\vdots \\
e_6
\end{bmatrix}
\]

- \( y_i \in \mathbb{R}^{n_i} \) is the pseudo phenotype (deregressed national breeding value, later DYD) for country \( i \in [1, \ldots, 6] \) with \( n_i \) observations
- \( \mu_i \) the general mean for country \( i \)
- \( Z_i \in \mathbb{R}^{n_i \times m} \) design matrix for genotypes (\( m \) is the number of markers, all countries have the same set of markers with same 0,1,2 coding)
- \( g_i \in \mathbb{R}^m \) estimated SNP effects for country \( i \)

- \( e_i \in \mathbb{R}^{n_i} \) residual effects for country \( i \) individuals

\[
\text{Var}(g_i) = \sigma_{s_i}^2 \Gamma, \quad \text{where } \Gamma = I^m \times 1/ \sum_{j=1}^{m} 2p_j(1 - p_j)
\]
with \( p_j \) = allele frequency of locus \( j \), \( \sigma_{s_i}^2 = \) sire variance of country \( i \) and \( I^m \in \mathbb{R}^{m \times m} \) identity matrix

- \( \text{Cov}(g_i, g_{i^+}) = \sigma_{ii^+} \Gamma, \quad \text{where } \sigma_{ii^+} = \rho_{ii^+} \times \sigma_{s_i} \sigma_{s_{i^+}} \), \( \text{with } \rho_{ii^+} = \text{genetic correlation between countries } i \text{ and } i^+ \)

\[
\text{Var}(e_i) = \sigma_{e_i}^2 \text{diag}(1/EDC_{ik}) = R_i, \quad \text{where } \sigma_{e_i}^2 = \sigma_{s_i}^2 (4 - h_i^2)/h_i^2 \quad \forall i, \text{for animals } k \in [1, \ldots, n_i]
\]

- \( \text{Cov}(e_i, e_{i^+}) = 0 \quad \forall i \neq i^+ \)
  * \( \sigma_{s_i}^2 \) and \( \rho_{ii^+} \) from Interbull
  * \( EDC_{ik} \) and \( h_i^2 \) from countries
Let the genetic (co)variance \( \text{var}(g) = G \), and its inverse \( G^{-1} \) be

\[
G^{-1} = \begin{bmatrix}
G^{11} & \cdots & G^{16} \\
\vdots & \ddots & \vdots \\
\text{symm.} & \cdots & G^{66}
\end{bmatrix}
\]

then

\[
\begin{bmatrix}
1'R_i^{-1}1' & 1'R_i^{-1}Z_i \\
Z'R_i^{-1}1' & Z'R_i^{-1}Z_i + G^{ii}
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
0 & G^{ii+}
\end{bmatrix}
\begin{bmatrix}
\hat{\mu}_i \\
\hat{g}_i
\end{bmatrix}
= \begin{bmatrix}
1'R_i^{-1}y_i \\
Z'R_i^{-1}y_i
\end{bmatrix}
\]

\[
\begin{bmatrix}
1'R_{i+1}^{-1}1' & 1'R_{i+1}^{-1}Z_{i+} \\
Z'_{i+}R_{i+1}^{-1}1' & Z'_{i+}R_{i+1}^{-1}Z_{i+} + G^{i+ij}
\end{bmatrix}
\begin{bmatrix}
\hat{\mu}_{i+} \\
\hat{g}_{i+}
\end{bmatrix}
= \begin{bmatrix}
1'R_{i+1}^{-1}y_{i+} \\
Z'_{i+}R_{i+1}^{-1}y_{i+}
\end{bmatrix}
\]
Data was split into learning and validation sets by bulls’ birth date
- The youngest 10% from each country → validation set

Animal solutions (DGV) were computed as $\hat{a}_{ik} = z_{ik}\hat{g}_i$ for animal $k$ in country $i$

Validation reliability was defined as $R_v^2 = \frac{(\text{cor}(\text{DRP}_v, \text{DGV}_v))^2}{R_{\text{DRP}_v}^2}$,

(where subscript $v$ refers to validation set records)

The bias $b_1$ was tested with a weighted linear regression of $\text{DRP}_v$ on predicted $\text{DGV}_v$, using $\text{EDC}_v$ as weights

The SNP MACE prediction set solutions were compared to country-wise single trait SNP-BLUP solutions
Results
SNP MACE with MiX99

- Computations were performed with MiX99 release XI/2017 version 17.1107
- Not happy for the convergence properties of the model
  \[\Rightarrow\] Long computation time (around 12h for SNP MACE)
  Possibly result suffers slightly, esp. with low heritability trait cc2
- We computed also equivalent G-BLUP MACE
  - No problems with convergence
  - *Much* faster, around 3h
  - GEBVs practically equal to SNP MACE ones
    - correlation \( \geq 0.98 \) for all traits & countries
  - SNP solutions were solved from G-BLUP â:s,
    - solutions were consistent with SNP MACE
1. Protein yield

**Table 5:** Validation reliability $R_v^2$ of DGV predicted either by single trait SNP-BLUP or SNP MACE, and the gain acquired by using SNP MACE

<table>
<thead>
<tr>
<th>Country</th>
<th>Single trait</th>
<th>SNP MACE</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>0.514</td>
<td>0.570</td>
<td>0.056</td>
</tr>
<tr>
<td>DFS</td>
<td>0.457</td>
<td>0.563</td>
<td>0.106</td>
</tr>
<tr>
<td>FRA</td>
<td>0.505</td>
<td>0.579</td>
<td>0.075</td>
</tr>
<tr>
<td>NLD</td>
<td>0.491</td>
<td>0.607</td>
<td>0.116</td>
</tr>
<tr>
<td>ESP</td>
<td>0.448</td>
<td>0.549</td>
<td>0.101</td>
</tr>
<tr>
<td>POL</td>
<td>0.389</td>
<td>0.541</td>
<td>0.152</td>
</tr>
</tbody>
</table>

- Generally reliabilities from 0.54 to 0.61 from MT
- The gain is considerable, gain percentage 11–39%
- Variances may be slightly inflated ($b_1$ in range 0.80–0.90)
2. Somatic cell score

Table 6: Validation reliability $R_v^2$ of DGV predicted either by single trait SNP-BLUP or SNP MACE, and the gain acquired by using SNP MACE

<table>
<thead>
<tr>
<th>Country</th>
<th>Single trait</th>
<th>SNP MACE</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>0.457</td>
<td>0.529</td>
<td>0.072</td>
</tr>
<tr>
<td>DFS</td>
<td>0.407</td>
<td>0.510</td>
<td>0.103</td>
</tr>
<tr>
<td>FRA</td>
<td>0.386</td>
<td>0.480</td>
<td>0.094</td>
</tr>
<tr>
<td>NLD</td>
<td>0.445</td>
<td>0.560</td>
<td>0.115</td>
</tr>
<tr>
<td>ESP</td>
<td>0.311</td>
<td>0.448</td>
<td>0.138</td>
</tr>
<tr>
<td>POL</td>
<td>0.478</td>
<td>0.611</td>
<td>0.133</td>
</tr>
</tbody>
</table>

- Lower $h^2 \Rightarrow$ slightly lower values than with protein
- Generally reliabilities from 0.45 to 0.61
- Variances may be slightly inflated ($b_1$ in range 0.77 – 0.89)
### Table 7: Validation reliability $R^2_v$ of DGV predicted either by single trait SNP-BLUP or SNP MACE, and the gain acquired by using SNP MACE

<table>
<thead>
<tr>
<th>Country</th>
<th>Single trait</th>
<th>SNP MACE</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEU</td>
<td>0.582</td>
<td>0.687</td>
<td>0.105</td>
</tr>
<tr>
<td>DFS</td>
<td>0.289</td>
<td>0.392</td>
<td>0.103</td>
</tr>
<tr>
<td>FRA</td>
<td>0.255</td>
<td>0.338</td>
<td>0.083</td>
</tr>
<tr>
<td>NLD</td>
<td>0.382</td>
<td>0.440</td>
<td>0.058</td>
</tr>
<tr>
<td>ESP</td>
<td>0.505</td>
<td>0.590</td>
<td>0.085</td>
</tr>
<tr>
<td>POL</td>
<td>0.130</td>
<td>0.212</td>
<td>0.083</td>
</tr>
</tbody>
</table>

- Very low $h^2$, still method seemed to work
- Relative gain from MT approach was bigger when the single trait $R^2_v$ was low
- Some variances inflated ($b_1$ in range 0.62–0.92)
Conclusion

- Fitting SNP MACE with individual animal genotypes is feasible, and countries gain from cooperation
- Next steps:
  1. We quantify whether SNP MACE is better than using (the current practice) single trait SNP BLUPs on MACE DRPs
  2. Consider models that account better the country wise definitions of genomic evaluations
Thank You!