Single-step genomic predictions for yield traits in US Holsteins with UPG and phenotype-pedigree truncation

Daniela Lourenco
A. Cesarani, Y. Masuda, S. Tsuruta, E. Nicolazzi, P. VanRaden, and I. Misztal
Motivation

• Missing pedigree in the US Holstein data
  • ~ 10% sires
  • ~ 20% dams

• QP-transformation for $A^{-1}$
  • (Quaas & Pollack, 1981; Westell et al., 1988)

\[
A^* = \begin{bmatrix}
A^{-1} & -A^{-1}Q \\
-Q'A^{-1} & Q'A^{-1}Q
\end{bmatrix}
\]

• QP-transformation for $H^{-1}$ (Misztal et al., 2013)

\[
H^* = A^* + \begin{bmatrix}
0 & 0 & 0 \\
0 & G^{-1} - A^{-1}_{22} & -G^{-1} - A^{-1}_{22}Q_2 \\
0 & -Q_2'(G^{-1} - A^{-1}_{22}) & Q_2'(G^{-1} - A^{-1}_{22})Q_2
\end{bmatrix}
\]

• Altered QP-transformation for $H^{-1}$ (Tsuruta et al., 2019)

\[
H^* = A^* + \begin{bmatrix}
0 & 0 & 0 \\
0 & G^{-1} - A^{-1}_{22} & -G^{-1} - A^{-1}_{22}Q_2 \\
0 & -Q_2'(G^{-1} - A^{-1}_{22}) & Q_2'(G^{-1} - A^{-1}_{22})Q_2
\end{bmatrix}
\]

Biased EBV
Motivation

- Matilainen et al. (2016): female fertility traits in Nordic Reds
- Tsuruta et al. (2019): type traits in US Holsteins
- Masuda et al. 2018 (Protein)

UPG poorly estimated
Large number of genotyped females with missing pedigree and no phenotypes

<table>
<thead>
<tr>
<th>Data</th>
<th>UPG</th>
<th>R²</th>
<th>b₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncated 2011</td>
<td>Pedigree</td>
<td>0.52</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Ped. + Genomic</td>
<td>0.32</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>No UPGs</td>
<td>0.50</td>
<td>0.78</td>
</tr>
</tbody>
</table>

- Lourenco et al. (2014): pedigree truncation helped to reduce bias
- due to missingness
Objectives

• Assess bias and reliability of GEBV for bulls and cows in ssGBLUP
  • Milk, Fat, and Protein
  • UPG for $\mathbf{A}^{-1}$ (SS_UPG)
  • UPG for $\mathbf{A}^{-1}$ and $\mathbf{A}_{22}^{-1}$ (SS_UPG2)

• Six phenotype-pedigree truncation scenarios
  • 1980, 1990, or 2000
  • Pedigree depth 2 or 3

• Feasibility of ssGBLUP for dairy evaluations in the US
- US Holstein data up to December 2018

<table>
<thead>
<tr>
<th>Phenotype cut-off scenario</th>
<th>Records</th>
<th>Genotypes</th>
<th>Animals in pedigree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Cows</td>
<td>Depth=3</td>
</tr>
<tr>
<td>Pheno1980</td>
<td>77.8 M</td>
<td>31.5 M</td>
<td>862 K</td>
</tr>
</tbody>
</table>
Analyses

• QP-transformation for $A^{-1}$ in $H^{-1}$

$$H_{UPG}^{*} = A^{*} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & G_{APY}^{-1} - A_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$\rightarrow$ SS_UPG

• Altered QP-transformation for $H^{-1}$ (Tsuruta et al., 2019)

$$H_{UPG2}^{*} = A^{*} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & G_{APY}^{-1} - A_{22}^{-1} & -(-A_{22}^{-1})Q_{2} \\ 0 & -Q_{2}'(-A_{22}^{-1}) & Q_{2}'(-A_{22}^{-1})Q_{2} \end{bmatrix}$$

$\rightarrow$ SS_UPG2

• BLUP90IOD2OMP1 for each phenotype-pedigree truncation scenario
  • APY with 15,000 core animals
• Complete data: 2018
• Reduced data: 2014

• 2,710 bulls
  • Reliability: $[\text{CORR}(\text{DYD}, (G)\text{EBV})]^2$
  • Dispersion: $\text{DYD} = b_0 + b_1(G)\text{EBV}$

• 381,779 Cows
  • Predictive ability: $\text{CORR}(y_{adj}, (G)\text{EBV})$
  • Dispersion: $y_{adj} = b_0 + b_1(G)\text{EBV}$
Reliability for bulls

![Graph showing R^2 values for phenotypes 1980, 1990, and 2000, with methods BLUP, SS_UPG, and SS_UPG2 indicated. The graph includes data points for Depth3 and Fat.]
b1 for bulls
Predictive ability for cows
b1 for cows
Rounds to convergence

Pedigree Depth 3

- Pheno1980: 989 (SS_UPG), 369 (SS_UPG2)
- Pheno1990: 910 (SS_UPG), 391 (SS_UPG2)
- Pheno2000: 734 (SS_UPG), 404 (SS_UPG2)
### ssGBLUP with many more genotypes

<table>
<thead>
<tr>
<th></th>
<th>This study</th>
<th>A. Cesarani (Friday)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genotypes</strong></td>
<td>862k</td>
<td>3.4M + data</td>
</tr>
<tr>
<td>Validation bulls</td>
<td>1,529</td>
<td>1,529</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milk</td>
<td>0.72</td>
<td>0.81</td>
</tr>
<tr>
<td>Fat</td>
<td>0.69</td>
<td>0.80</td>
</tr>
<tr>
<td>Protein</td>
<td>0.67</td>
<td>0.77</td>
</tr>
</tbody>
</table>

**Computing cost**
- Genomic: 6 hours, 2.5 days
- Rounds: 404, 473
Conclusions

- Genomic predictions for yield traits in US Holsteins using single-step
  - UPG for $A^{-1}$ and $A_{22}^{-1}$
  - Reliable and unbiased
- Removing old generations of pedigree and phenotypes do not compromise predictions for young selection candidates
- Large-scale dairy genomic evaluations are feasible
  - Up to 3.4 Million genotyped animals in less than 3 days
  - Indirect predictions could further reduce computing time
Acknowledgements