Using genomic data to estimate genetic correlations between countries with different levels of connectedness

Bonifazi R, GM Neufeld, T Pook, J Vandenplas, MPL Calus
International evaluations

- Compare animals’ genetic values **across countries**

- **Differences**: trait definition, scale and genetic bases, GxE

- Multi-trait models: *countries as different correlated traits*

- $r_G$ between countries < 1

- Impact on international (G)EBVs → **Crucial** for international evaluations
Estimation of $r_G$ between countries

- Requires **genetic connections** between countries

- “Common Bulls” (CB)

- Beef cattle, small (dairy) populations $\rightarrow$ low use AI/low #CB $\rightarrow$ low connections

- **Maternally influenced traits** $\rightarrow$ “Common Maternal Grand-Sires” (CMGS)

- Estimation methods: **pedigree-based**

- **Low connectedness $\rightarrow$ issues:**
  - Long computational times, large SE
  - Impossible to estimate $r_G$
Genomic data in international evaluations

- Interbeef single-step evaluations (Bonifazi et al. 2022, GSE)
- Genomic data to estimate $r_G$ between countries
- Disconnected populations (according to pedigree) $\rightarrow$ Connected via genomic information (e.g. Wientjes et al. 2015, 2018, GSE)

![Diagram showing the process of estimation of genetic correlations between two countries (Country A and Country B) through international genomic evaluations.](image-url)
Aim of the study

Investigate if genomic data help to estimate $r_G$ between countries more accurately than pedigree at different level of connectedness
Simulated data and (genetic) parameters

- Two beef cattle populations of same breed
- Maternally affected trait: weaning weight
- Genetic parameters (Bonifazi et al. 2020, GSE)

- ~2,000 QTLs
- 30 chr (1 Morgan)

<table>
<thead>
<tr>
<th></th>
<th>POP₁</th>
<th></th>
<th>POP₂</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dir</td>
<td>Mat</td>
<td>Dir</td>
<td>Mat</td>
</tr>
<tr>
<td>POP₁</td>
<td>0.30</td>
<td>-0.2</td>
<td>0.15</td>
<td>0.30</td>
</tr>
<tr>
<td>POP₂</td>
<td>0.80</td>
<td>0</td>
<td>0.30</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.70</td>
<td>-0.2</td>
<td>0.15</td>
</tr>
</tbody>
</table>
• Mimic current beef situation (Bonifazi et al. 2023, GSE)

• Different sires exchange levels
### Scenarios: connectedness between POPs

<table>
<thead>
<tr>
<th>Scenario</th>
<th>n. CB</th>
<th>n. off. from CB</th>
<th>GS</th>
<th>Mean n. from CB</th>
<th>Mean n. grand-off. from CMGS</th>
<th>Mean GS&lt;sub&gt;CMGS&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disconnected</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>10</td>
<td>1,500</td>
<td>0.02</td>
<td>8</td>
<td>2,322</td>
<td>0.04</td>
</tr>
<tr>
<td>Medium</td>
<td>20</td>
<td>3,000</td>
<td>0.05</td>
<td>16</td>
<td>4,544</td>
<td>0.07</td>
</tr>
<tr>
<td>High</td>
<td>80</td>
<td>12,000</td>
<td>0.18</td>
<td>63</td>
<td>15,364</td>
<td>0.23</td>
</tr>
</tbody>
</table>

- Names based on GS thresholds of Bonifazi et al. 2020, GSE
- 10 replicas

GS = Genetic similarity
CB, CMGS = Common Bulls, Common MGS
**Estimation of $r_G$: sources of information**

<table>
<thead>
<tr>
<th>A-matrix</th>
<th>10 gen.</th>
<th>11 gen.</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-matrix</td>
<td>10 gen.</td>
<td>11 gen.</td>
<td>3 gen.</td>
</tr>
<tr>
<td>G-matrix</td>
<td>3 gen.</td>
<td>4 gen.</td>
<td>3 gen.</td>
</tr>
</tbody>
</table>

- Pedigree 14 generations → 11 generations (space limitations)
- H as Legarra et. al. 2009, JDS (scaling G inbreeding to $A_{22}$)
- G as VanRadren 2008, JDS, method 1
- Maternal effects → need prev. gen. relationships (effectively computed as H-mat)
Software

- Simulation: *MoBPS (miraculix & RandomFieldUtils)*
- GRM: *calc_grm* (Calus and Vandenplas, 2013, WUR)
- \( r_G \) estimation: *mtg2* (CORE GREML: maternal effects) ¹
- **Mimic current Interbeef**
- **Bivariate model**: countries as correlated traits (no residual corr.)
- **Raw data**
- Starting values: within-country (co)variances, 0 across-country

¹ Lee and van der Werf, 2016, Bioinformatics; Zhou et al., 2020, Nat. Commun.
Results

- Genetic correlations
- Standard errors
- Computational resources
Direct $r_G$
Simulated = 0.8

Maternal $r_G$
Simulated = 0.7
Standard Errors

Disconnected
Low
Medium
High

SE(dir rG) SE(mat rG)  SE(dir rG) SE(mat rG)  SE(dir rG) SE(mat rG)  SE(dir rG) SE(mat rG)
## Computational requirements

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals in matrix</td>
<td>66,000</td>
<td>24,000</td>
<td>66,000</td>
</tr>
<tr>
<td>Elapsed time (hours)</td>
<td>3.1</td>
<td>7.3</td>
<td>2.9</td>
</tr>
<tr>
<td>RAM peak usage (Gbytes)</td>
<td>106</td>
<td>13</td>
<td>102</td>
</tr>
</tbody>
</table>

Averages across scenarios and replicas
Conclusions

- **Genomic data** more accurate estimates of $r_G$ between countries and smaller SE for:
  - disconnected and low connected: beef cattle, small (dairy) populations (e.g. AYR, GUE, JER)
  - maternal $r_G$

- Medium/high connected: no benefits (e.g. HOL)

- **G** with 3 gen. (pheno & geno) similar to **H**, but less comput. resources

- Real data: unbalanced, <5 off/dam, large datasets (connected sub-set → include genomic info in connectedness measures)
Acknowledgements

Martin Schlather, Universität Mannheim

Hong Lee, University of New England

Pascal Duenk and Noraly van Hemert, Wageningen University & Research

Interbeef, Interbull, ICAR, national organisations
Take-home messages

- Genomic data more accurate estimates of $r_G$ between countries and smaller SE for:
  - disconnected and low connected: beef cattle, small (dairy) populations (e.g. AYR, GUE, JER)
  - maternal $r_G$

- Useful for beef cattle, small (dairy) populations

Thanks for your attention

renzo.bonifazi@wur.nl