Exploiting sequence variants for genomic prediction in Australian sheep using Bayesian models

M. Khansefid\textsuperscript{1,2}, S. Bolormaa\textsuperscript{1,2}, A. A. Swan\textsuperscript{1,3}, J. H. J. van der Werf\textsuperscript{1,4}, N. Moghaddar\textsuperscript{1,4}, N. Duijvesteijn\textsuperscript{1,4}, H. D. Daetwyler\textsuperscript{1,2,5} & I. M. MacLeod\textsuperscript{1,2}

1 Sheep CRC (Cooperative Research Centre for Sheep Industry Innovation), Armidale, NSW 2351, Australia
2 Agriculture Victoria, AgriBio Centre for AgriBioscience, Bundoora, VIC 3083, Australia
3 Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia
4 School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia
5 School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3086, Australia
Introduction

Challenges:
In comparison with dairy cattle, the adoption of genomic selection in sheep genetic evaluations needs extra considerations due to:

• The diversity of breeds and composites resulting in small reference sizes within breed.
• High genotyping costs relative to economic returns.

Potential solutions:
• Increase the reference size through the use of multi-breed populations.
• Finding the causal mutations from imputed sequences and designing the inexpensive customized low to medium density SNP chip.
Aim

*Increase the accuracy of genomic prediction using sequence variants!*

- Faster genetic gain and improving the profitability of sheep production.
- The application of genomic selection in the sheep industry looks promising.

*How?*

- Including the SNPs from imputed sequence affecting traits of interest “i.e. Top SNPs” in the genomic prediction model.
  - GBLUP (Moghaddar *et al.*, previous talk)
  - BayesR (Erbe *et al.*, 2012) and BayesRC (MacLeod *et al.*, 2016)
Material and methods (Animals and phenotypes)

- A mixture of breeds and crosses from Sheep CRC dataset and industry evaluations.
- Three Groups: GWAS QTL discovery / Genomic prediction reference / Genomic prediction validation.
- Growth and carcass traits (*6 traits)*:
  - Carcass fat depth at C site (CCFAT)
  - Carcass and post-weaning eye muscle depth (CEMD and PEMD)
  - Intermuscular fat percentage (IMF)
  - Shear force measured at day 5 after slaughter (SF5)
  - Post-weaning weight (PWT)
- Wool traits (*3 traits)*:
  - Yearling greasy and clean fleece weight (YGFW and YCFW)
  - Yearling fibre diameter (YFD)
Material and methods (Genotypes and GWAS)

- 50k panel (real or imputed)
- Genotypes on the X chromosome were excluded
- 50k (≈ 37k SNPs) genotypes were imputed to HD (≈ 500k SNPs) and then to WGS (≈ 31 million variants)
- GWAS:
  - The “Top SNPs” were found in WGS imputed variants using Wombat software (Meyer, 2007)
  - The most significant SNPs below p-value 0.001 within a 100 Kb window
  - Prune one of any pair of SNPs with LD ($r^2$) > 0.95
  - ≈ 4,500 Top SNPs
Material and methods (Genomic prediction)

• The phenotypes were pre-adjusted for data source, and breed proportions.
• Genotypes were centred and standardised to a variance of 1.
• BayesR (Erbe et al., 2012):
  – The SNP effects were modelled as a mixture of four normal distributions with a mean=0 and variance: 
    \[ \sigma^2_1 = 0 \sigma^2_g, \sigma^2_2 = 0.0001 \sigma^2_g, \sigma^2_3 = 0.001 \sigma^2_g \text{ and } \sigma^2_4 = 0.01 \sigma^2_g, \]
    where \( \sigma^2_g \) is the additive genetic variance.
• BayesRC (MacLeod et al., 2016):
  – The top SNPs were allocated to a separate category or class than the remaining 50k SNPs.
• Each model was replicated with 5 MCMC chains, each with 40,000 iterations (20,000 burn-in).
Results (Growth and carcass traits / Merino)

Accuracy (Cor_{\text{Pheno,GEBV}} / \sqrt{h^2})

Bias (b_{\text{Pheno,GEBV}})
Results (Growth and carcass traits / Border Leicester × Merino)

Accuracy ($\text{Cor}_{\text{Pheno,GEBV}} / \sqrt{h^2}$)

<table>
<thead>
<tr>
<th>Trait</th>
<th>CCFAT</th>
<th>CEMD</th>
<th>PEMD</th>
<th>IMF</th>
<th>SF5</th>
<th>PWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.60</td>
<td>0.40</td>
<td>0.20</td>
<td>0.40</td>
<td>0.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Bias ($b_{\text{Pheno,GEBV}}$)

<table>
<thead>
<tr>
<th>Trait</th>
<th>CCFAT</th>
<th>CEMD</th>
<th>PEMD</th>
<th>IMF</th>
<th>SF5</th>
<th>PWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>1.50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Results (Wool traits / Merino) / Comparing GBLUP and BayesR(C)

Accuracy ($\text{Cor}_{\text{Pheno,EBV}} / \sqrt{h^2}$)

Bias ($b_{\text{Pheno,EBV}}$)
Conclusions

- Bayesian models increased the accuracy of genomic prediction by about 5% by adding the top sequence variants to 50k genotypes.

- GWAS top sequence variants account for only a proportion of the expected genetic variance and the average bias of predictions tend to be higher than denser genotypes.

- The accuracy of predictions was highest in BayesRC when the top SNPs were highly predictive (such as PWT).

- Adding top SNPs to low density SNP panels can increase the accuracy of genomic prediction while minimising genotyping costs for industry applications.
Acknowledgments

Funding and data
Meat and Livestock Australia
Sheep CRC
https://www.sheepcrc.org.au/

SheepGenomesDB
https://www.sheepgenomesdb.org/

Agriculture Victoria Sequencing ~390 sheep
Amanda Chamberlain
Brett Mason
Claire Prowse-Wilkins
Tracie Webster
Coralie Reich

SheepCRC colleagues at Uni New England

Co-Authors
Bolormaa Sunduimijid
Andrew Swan
Julius van der Werf
Nasir Moghaddar
Naomi Duijvesteijn
Hans Daetwyler
Iona MacLeod
Questions?
Results (Wool traits / Merino)

Accuracy ($\text{Cor}_{\text{Pheno,GEBV}} / \sqrt{h^2}$)

Bias ($b_{\text{Pheno,GEBV}}$)