Genomic predictions for marbling score in Hanwoo cattle using sequence data

Sungbong Jang\textsuperscript{1}
A. Garcia\textsuperscript{1}, S.H. Lee\textsuperscript{2}, S. Tsuruta\textsuperscript{1}, D. Lourenco\textsuperscript{1}

\textsuperscript{1}University of Georgia, \textsuperscript{2}Chungnam National University
Hanwoo cattle

- Korean native breed
- Small size & brown coat color
- High marbling
Why sequence data?

• More information
  • SNPs chip cannot capture all variants in genome
  • Need more SNPs to cover whole genome

• Sequencing price is getting cheaper
  • Imputation
Objectives

• Assess gain in accuracy using
  • Regular SNP chip
  • Sequence data
  • Regular SNP chip + sequence data

• Test different methods for genomic prediction
Animal data

- Hanwoo Research Institute in South Korea
- 1.3M pedigree
- 545K marbling score phenotype
  - 1 to 9 score (9: High-marbling, 1: Low-marbling)
- 1,160 genotyped animals
- Pedigree-based $h^2 = 0.51$
Genomic data

- 50K, DEG, 50k + DEG, Imputed_WGS
  - DEG: Selected SNPs from Imputed_WGS

<table>
<thead>
<tr>
<th>SNP panel</th>
<th>Number of SNPs after QC</th>
</tr>
</thead>
<tbody>
<tr>
<td>50k</td>
<td>39,822</td>
</tr>
<tr>
<td>DEG</td>
<td>321,614</td>
</tr>
<tr>
<td>50k+DEG</td>
<td>360,407</td>
</tr>
<tr>
<td>Imputed_WGS</td>
<td>11,146,536</td>
</tr>
</tbody>
</table>
Imputation

- Eagle 2.4v (Phasing)
- Minimac3 (Imputation, Filtered: $R^2 > 0.6$)
- Two-step

Das et al., (2016)
Howie et al., (2012)

50k → 777k → Sequence

N of Ref = 1,295
N of Ref = 203
**DEG identification**

- Differentially Expressed Genes
- RNA-seq analysis Ⅰ (n = 24, Limma package)
- RNA-seq analysis Ⅱ (n = 40, Linear regression)
- Total 336 DEGs were identified

High marbling vs Low marbling
High fat (%) vs Low fat (%)

Ritchie et al., (2015)
Methods

• BayesR

\[ p(\beta_j \mid \pi, \sigma^2_g) = \pi_1 \times N(0, 0 \times \sigma^2_g) + \pi_2 \times N(0, 10^{-4} \times \sigma^2_g) + \pi_3 \times N(0, 10^{-3} \times \sigma^2_g) + \pi_4 \times N(0, 10^{-2} \times \sigma^2_g) \]

Erbe et al., (2012)
Methods

- GBLUP (BLUPF90 programs)
\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + G^{-1}\lambda
\end{bmatrix}
\begin{bmatrix}
b \\
u
\end{bmatrix}
= 
\begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

- ssGBLUP (BLUPF90 programs)
\[
\begin{bmatrix}
X'X & X'Z \\
Z'X & Z'Z + H^{-1}\lambda
\end{bmatrix}
\begin{bmatrix}
b \\
u
\end{bmatrix}
= 
\begin{bmatrix}
X'y \\
Z'y
\end{bmatrix}
\]

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

- \( A_{22} \) = pedigree relationship matrix for genotyped animals
- \( G \) = genomic relationship matrix
Methods

• SNP weighting (BLUPF90 programs)
  • SNPs are assumed to have different effects on the traits

1. \( D = I, \ G = \frac{\text{ZDZ}'}{2\Sigma p_i(1-p_i)} \)
2. GEBV from ssGBLUP
3. \( \hat{a} = \lambda DZ'G^{-1}\hat{u} \)
4. \( d_i = \frac{|\hat{u}_i|}{CT\text{sd}(\hat{u})^{-2}} \)
5. Iteration from 2

VanRaden (2008)
Wang et al., (2012)
Fragomeni et al., (2019)
Methods

• Linear weights: \( d_i = \hat{u}_i^2 2p_i (1 - p_i) \)
  
  (Falconer & Mackay, 1996)

• Non-linearA weights: \( d_i = \frac{|\hat{u}_i|}{CT \text{sd}(\hat{u})^2} \)
  
  (VanRaden, 2008)
Validation

• Youngest genotyped animals (n = 169)
• Predictive ability: $\text{cor}(\text{GEBV}, y_{adj})$
• Inflation (b1): $y_{adj} = b_0 + b_1 \times \text{GEBV}$
Predictive ability

BayesR
- 50k: 0.19
- DEG: 0.16
- 50k+DEG: 0.18
- WGS: 0.18

GBLUP
- 50k: 0.16
- DEG: 0.18
- 50k+DEG: 0.17
- WGS: 0.17

ssGBLUP
- 50k: 0.27
- DEG: 0.26
- 50k+DEG: 0.27
- WGS: 0.26
Predictive ability

<table>
<thead>
<tr>
<th></th>
<th>ssGBLUP</th>
<th>WssGBLUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>50k</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>DEG</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>50k+DEG</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>WGS</td>
<td>0.26</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Inflation

GBLUP  0.98  0.63
ssGBLUP  0.92  0.77
WssGBLUP  0.94  0.92
BayesR  1.22  1.29  1.35

- GBLUP
- ssGBLUP
- WssGBLUP
- BayesR

- 50k
- DEG
- 50k+DEG
- WGS

UNIVERSITY OF GEORGIA
Conclusion

• No gain using sequence data
• ssGBLUP outperforms GBLUP and BayesR
  • More data is used in ssGBLUP
• No increase in predictive ability using SNP weighting
Limitations

• Small number of genotyped animals with records
• Non-stringent way to select SNPs (DEG)

• Are those selected SNPs (DEG) really causative?
Acknowledgements
Thank you!

jsbng@uga.edu