Implementing genomic selection in livestock species

Julius van der Werf

CRC for Sheep Industry Innovation
School of Environmental and Rural Science, UNE, Armidale, NSW
Outline

1. Potential benefits of genomic selection in breeding programs

2. Can we predict the accuracy of genomic selection?

3. What information is needed for accurate predictions?

4. Requirements for the reference population
   how large, how related, how long-lasting, multi-breed?

5. Strategies for genotyping
   low density chips, high density chips, sequence data?
Genomic Prediction: basic idea

1) Somebody (else) measures lots of sheep, and their DNA → Reference population

2) A breeder tests DNA on young rams

Prediction from DNA → genomic breeding values - GBV

GBV + Current ASBV → Improved ASBV

Merit depends on trait measurability
Compare: Progeny Testing

- 50% accuracy, 0.5-1 yr old
- 90% accuracy, 2-3 yrs old

Each progeny group only informs one sire

Relationship = 0.5
Genomic Testing

Reference population (too small)

Relationship $= 0.02 \ldots 0.5$

51% accuracy
0.5-1 yrs old

use information on “relatives”
while sire is still young
Genomic Testing

Reference population

Relationship = 0.02……0.5

70% accuracy
0.5-1 yrs old
Summarizing Genomic Prediction
- What information is used?

• Based on very many small – genomic- relationships

• Does not require ‘direct relatives’ to be tested

• Can be based on distant relatives ‘some generations away’

• .....but the number of small relatives needs to be large (thousands)

• Can not predict across breed
Outline: Sheep Genomic Analysis

– What information is used?

– How useful is this information?

– How to use it?
Genomic Selection: Benefit

Overall:
More accurate prediction of genetic merit for breeding objective

Specific:
Traits that are usually difficult to improve
difficult or expensive to measure
can not be measured early
low heritability
Possible Benefits

- Muscle conf (mm)
- Dressing %
- Saleable meat yd %
- Carcass Fat (mm)
- PWT (kg)
- Overall Merit ($Index)

Depends on size of reference population

21% increase

Annual gain in $/head

- no GS
- GS1
- GS2

Annual gain in $/head depends on size of reference population.
Modeling genomic selection in breeding programs

1. Selection index approach: multiple information, multiple traits
   Accuracy component

2. Optimizing selection across age classes
   Generation Interval component

3. For specific breeding objectives
Percent increase in rate of genetic gain when using genomic selection

Selection on a single trait

Predicted accuracy of Molecular EBV = 55% \( (VQTL=30\%) \)

<table>
<thead>
<tr>
<th>Trait Measurability</th>
<th>Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>Measured &lt; 1 year, males and female</td>
<td>37</td>
</tr>
<tr>
<td>Measured &gt; 1 year, males and females</td>
<td>64</td>
</tr>
<tr>
<td>Measured &gt;1 year, females only</td>
<td>109</td>
</tr>
<tr>
<td>Measured on Correlated Trait, Genetic Correlation = 0.9</td>
<td>48</td>
</tr>
<tr>
<td>Measured on Correlated Trait, Genetic Correlation = 0.5</td>
<td>143</td>
</tr>
</tbody>
</table>
1. Potential benefits of genomic selection in breeding programs
2. Can we predict the accuracy of genomic selection?
3. What information is needed for accurate predictions?
4. Requirements for the reference population
   - how large, how related, how long-lasting, multi-breed?
5. Strategies for genotyping
   - low density chips, high density chips, sequence data?
Accuracy of genomic prediction depending on size of reference population

Goddard 2009

Using Goddard 2009

Terminals, Maternals

Merinos
Accuracy, depending on how Me is approximated

\[
\frac{2NeLc}{\ln(2NeL)}
\]

- \(M_e = \frac{2NeLc}{\ln(2NeL)}\), \(h^2 = 0.1\)
- \(M_e = \frac{2NeLc}{\ln(2NeL)}\), \(h^2 = 0.5\)
- \(M_e = 2NeL\), \(h^2 = 0.1\)
- \(M_e = 2NeL\), \(h^2 = 0.5\)

\(2NeL\)
design of reference population

– Relatedness between reference population and selection candidates
– Across breeds or lines?
– Number of sires, nr of progeny per sire, which dams?
Realized accuracy 1

Using Goddard 2009
Realized accuracy 2

Using Goddard 2009
### Accuracy of genomic prediction for Post Weaning Weight from a mixed breed reference population

<table>
<thead>
<tr>
<th>Reference population</th>
<th>GEBV accuracy</th>
<th>Type</th>
<th>Size</th>
<th>BL</th>
<th>Merino</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLxMerino</td>
<td></td>
<td>(1)  = Merino</td>
<td>1000</td>
<td>-0.02&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.53&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>BLxMerino + (1)</td>
<td></td>
<td>(2)  = Merino</td>
<td>2000</td>
<td>-0.04&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.57&lt;sup&gt;bc&lt;/sup&gt;</td>
</tr>
<tr>
<td>BLxMerino + (2)</td>
<td></td>
<td>(3)  = Merino</td>
<td>3000</td>
<td>-0.08&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.59&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>BLxMerino + (3)</td>
<td></td>
<td>BLxMerino</td>
<td>1514</td>
<td>0.49&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.45&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BLxMerino + (1)</td>
<td>2514</td>
<td>0.42&lt;sup&gt;bc&lt;/sup&gt;</td>
<td>0.56&lt;sup&gt;bc&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BLxMerino + (2)</td>
<td>3514</td>
<td>0.37&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.54&lt;sup&gt;bc&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BLxMerino + (3)</td>
<td>4514</td>
<td>0.36&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.56&lt;sup&gt;bc&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Moghaddar, van der Werf and Swan, 2013 AAABG)
Accuracy and Mean Relationship to Ref

→ No Link!

$R^2 = 0.08$
Accuracy and Mean of Top 10 Relationships ➔ Clear Link!

$R^2 = 0.95$
Genomic prediction

\[
\begin{bmatrix}
X'X & X'X & 0 & X'X \\
Z'X & Z'X + G^{11} & G^{12} & 0 \\
0 & G^{21} & G^{22} & 0 \\
\end{bmatrix}
\begin{bmatrix}
b \\
g_1 \\
g_2 \\
\end{bmatrix}
= \begin{bmatrix}
X'y \\
Z'y \\
0 \\
\end{bmatrix}
\]

\[\hat{g}_2 = -(G^{22})^{-1}G^{21}\hat{g}_1\]

Genomic regression

Example:
Data on sire 1, sons 2 and 3, 4 unrelated, want to predict 5

A-matrix (pedigree-based)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.5</th>
<th>0.5</th>
<th>0</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
<td>1</td>
<td>0</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

G-matrix (DNA-based)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.5</th>
<th>0.5</th>
<th>0.02</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.20</td>
<td>1</td>
<td>0.025</td>
<td>0.20</td>
</tr>
<tr>
<td>0.5</td>
<td>0.20</td>
<td>1</td>
<td>0.025</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.015</td>
<td>0.025</td>
<td>1</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
<td>0.30</td>
<td>0.025</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

BLUP
\[\hat{u}_5 = 0.1136.y_1 + 0.0455.y_2 + 0.0455.y_3\]

GBLUP
\[\hat{g}_5 = 0.1135.y_1 + 0.0328.y_2 + 0.0591.y_3 + 0.00519.y_4\]
Genomic prediction

$$\begin{bmatrix}
X'X & X'X \\
Z'X & Z'Z + G^{11}
\end{bmatrix}
\begin{bmatrix}
b \\
g_1 \\
g_2
\end{bmatrix} =
\begin{bmatrix}
X'y \\
Z'y \\
0
\end{bmatrix}$$

$$\hat{g}_2 = -(G^{22})^{-1}G^{21}\hat{g}_1$$

Genomic regression

Example:
Data on sire 1, sons 2 and 3, 4 unrelated,
want to predict 5

A-matrix (pedigree-based)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.5</th>
<th>0.5</th>
<th>0</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

G-matrix (DNA-based)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.5</th>
<th>0.5</th>
<th>0.02</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.20</td>
<td>0.015</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.20</td>
<td>1</td>
<td>0.025</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.015</td>
<td>0.025</td>
<td>1</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.20</td>
<td>0.30</td>
<td>0.025</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

BLUP uses: Family Info
GBLUP uses: Family Info
Segregation within family
Info on ‘unrelated’
Sources of information contributing to GBV accuracy

1. Variation between families
   - BLU P: ++
   - GBLUP: ++
   - Half life: 1 gen

2. Variation within families
   - BLU P: 0
   - GBLUP: +
   - Half life: 1 gen

3. Markers tracking effects of genome segments/LD
   - BLU P: 0
   - GBLUP: ++++
   - Half life: several gen’s

Depending on size of reference population:
- Smaller ref pop
- Larger ref pop
## Results – Simulation

<table>
<thead>
<tr>
<th>Method</th>
<th>Close</th>
<th>Distant</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ped 0 - 0.25 Genom 0.08 – 0.35</td>
<td>0 - 0.125 Genom 0.08 – 0.26</td>
<td>0 - 0.05 Genom 0.08 – 0.16</td>
</tr>
<tr>
<td>BLUP- Shallow pedigree</td>
<td>0.39</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BLUP- Deep Pedigree</td>
<td>0.42</td>
<td>0.21</td>
<td>0.04</td>
</tr>
<tr>
<td>gBLUP</td>
<td>0.57</td>
<td>0.41</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Additional accuracy from family info

'baseline accuracy': graphs predict 0.36 for Ne=100, N=1750, h²=0.3

Sam Clark
### Accuracy Real Data

**Sam Clark**

<table>
<thead>
<tr>
<th>Method</th>
<th>Close related sires</th>
<th>Distantly related sires</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Empirical Acc</td>
<td>Predicted Acc</td>
</tr>
<tr>
<td></td>
<td>actual correlation with ASBV</td>
<td>correlation derived from gBLUP</td>
</tr>
<tr>
<td>BLUP-S</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>BLUP-D</td>
<td>0.62</td>
<td>0.37</td>
</tr>
<tr>
<td>gBLUP</td>
<td>0.65</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Genomic prediction FAQ

• How well can we predict distantly related individuals?
  – Ok if reference population is large enough
  – Can NOT predict across breed  *Daetwyler et al., 2011*

• How quick does the genomic prediction erode?
  – Fast if based on relationships, slower if based on ‘distant relatives’

• Do we need relatives?
  – Relatives give more accuracy, but not everyone can have them

• How large does a reference population need to be?
  – Design based as if prediction is based on ‘unrelated’
Reference Pop: How many are needed?

<table>
<thead>
<tr>
<th>Breed</th>
<th>merino</th>
<th>WS, PD</th>
<th>BL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ne</td>
<td>1000</td>
<td>250</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size of reference pop’n</th>
<th>30,000</th>
<th>10,000</th>
<th>5,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Progeny measured per year</td>
<td>3750</td>
<td>1250</td>
<td>625</td>
</tr>
<tr>
<td>h²=0.1</td>
<td>0.33</td>
<td>0.34</td>
<td>0.35</td>
</tr>
<tr>
<td>h²=0.3</td>
<td>0.51</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>h²=0.5</td>
<td>0.60</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>Predicted benefit in dG</td>
<td>40%</td>
<td>20%</td>
<td>?</td>
</tr>
</tbody>
</table>

assuming the reference population is ‘refreshed’ every 8 years

\%V_A explained by GBV ≅ h²
# Reference Pop: How many are needed?

<table>
<thead>
<tr>
<th>Breed</th>
<th>merino</th>
<th>WS, PD</th>
<th>BL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of reference pop’n</td>
<td>12,000</td>
<td>4,000</td>
<td>2,000</td>
</tr>
<tr>
<td>Progeny measured per year(^1)</td>
<td>1500</td>
<td>500</td>
<td>250</td>
</tr>
<tr>
<td>(h^2=0.1)</td>
<td>0.22</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>(h^2=0.3)</td>
<td>0.36</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>(h^2=0.5)</td>
<td>0.44</td>
<td>0.46</td>
<td>0.47</td>
</tr>
<tr>
<td>Predicted benefit in dG</td>
<td>20%</td>
<td>10%</td>
<td>?</td>
</tr>
</tbody>
</table>

\(\approx \frac{1}{2} h^2\)

*assuming the reference population is ‘refreshed’ every 8 years*
Outline

1. Potential benefits of genomic selection in breeding programs

2. Can we predict the accuracy of genomic selection?

3. What information is needed for accurate predictions?

4. Requirements for the reference population
   how large, how related, how long-lasting, multi-breed?

5. Strategies for genotyping
   low density chips, high density chips, sequence data?
Implication

• To predict a selection candidate
  – It needs to have relatives in reference populations
  – We can afford a lower degree of relationship than with BLUP
    » Can predict several generations away
  – Need large reference population
Optimal Genotyping Strategies

• If genotyping is expensive
  – Genotype males only
  – Genotype only ‘best’ males
  – multi-stage selection
  – But enough to be able to select!
Acknowledgements

DPI Vic:       Hans Daetwyler, Ben Hayes
UNE:          Nasir Moghaddar, Sam Clark, John Hickey,
              Brian Kinghorn, Cedric Gondro
AGBU:         Andrew Swan, Daniel Brown
Sheep CRC:    Ken Geenty, Klint Gore, James Rowe
Sheep Genetics/MLA: Rob Banks, Alex Ball, Sam Gill

Thanks