Putting the sheep SNP50 Beadchip to work: case studies in gene mapping and genomic selection

John McEwan, N Pickering, K Dodds, B Auvray, P Johnson, R Tecofsky, T Wilson
AgResearch & University of Otago

PAG Jan 2010
Overview

- Industry
- GWS
- Major Genes
- Industry
# Productivity comparison

<table>
<thead>
<tr>
<th></th>
<th>1990-91</th>
<th>2006-07</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambing Percentage (ewe)</td>
<td>101.6</td>
<td>117.9</td>
<td>+17%</td>
</tr>
<tr>
<td>Hogget lambs as % all lambs</td>
<td>-</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>Average Lamb Wt (kg)</td>
<td>14.35</td>
<td>16.90</td>
<td>+17%</td>
</tr>
<tr>
<td>Lamb sold Kg/Ewe</td>
<td>9.76</td>
<td>16.83</td>
<td>+72%</td>
</tr>
<tr>
<td>Wool kg/head</td>
<td>5.28</td>
<td>5.60</td>
<td>+6%</td>
</tr>
<tr>
<td>Average Steer Wt (kg)</td>
<td>297</td>
<td>320</td>
<td>+8%</td>
</tr>
<tr>
<td>Milksolids per cow (kg)</td>
<td>260</td>
<td>325</td>
<td>+25%</td>
</tr>
</tbody>
</table>

Source: M&WNZ
Genetic progress: period & flock type

<table>
<thead>
<tr>
<th>Year</th>
<th>Terminal</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990-94</td>
<td>0.37</td>
<td>0.25</td>
</tr>
<tr>
<td>1995-98</td>
<td>0.34</td>
<td>0.29</td>
</tr>
<tr>
<td>1999-2003</td>
<td>SIL</td>
<td>0.61</td>
</tr>
<tr>
<td>2003-2006</td>
<td>CPT/ACE</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Nimmo Bell: return on levy
- SIL: 16.3
- CPT ACE: 18.4

Potential existing technology ~$1.40 but we can do better…. ~$2.00
The future: Whole genome selection (WGS)

- Major advance
- Genome sequencing & SNP chips = “genome wide selection”
- As accurate as progeny testing,
  - but can be done at birth
- Suitable for
  - sex limited,
  - difficult to measure traits or
  - traits measured late in life
- Dairy cattle:
  - increase genetic gain 50-70% 
  - while decreasing progeny testing costs
- Application in sheep is still being explored
5 steps to whole genome selection in NZ

- Sequence & ID SNPs (ISGC)
- Create SNP chip (ISGC)
- Collect DNA
  - from measured animals (progeny tested sires)
  - Genotype
- Create “SNP product”
  - Estimate coefficients
  - Validate
- Use to estimate GBVs
  - GBV =a*EBV +b*MBV

- 3X coverage 6 sheep
  - Dec 2007

- Create and validate 60K SNP chip
  - Aug 2008

- Genotype first resources
  - Dec 2008

- Validate
  - Oct 2009

- 1st SNP Product
  - early 2010
Ovita

- Ovita owned by M&WNZ & AgResearch
- DNA genomics in sheep
- Industry good outcomes
- First right of refusal to Pfizer for commercialisation
- First 5 years developed and commercialised 5 DNA tests
- 2009 one new test Polled
- Next year 1st genomic selection tests
Current NZ AgR/SIL Database structure

- GenomNZ
  - 700K (10K)
  - 60K
  - DNA Sequence

- DNAdb
  - ~45Gbp
  - ~array data

- SIL
  - >10M

- AnimalID
  - [arrow]

- SIL GE
  - ~3M

- [checklist]
  - Species
  - IDs
  - Real time
  - Research/commercial
Traits

• Top traits include:
  – A
    • NLB and twinning
    • Lamb survival
    • Weaning weight (maternal + direct)
    • Carcass
    • Wool weight
  – B
    • Adult live weight
    • Parasites
    • Meat yield (B/C)
    • Dags/flystrike/breech score (B/C)
    • FE (B/C)
    • Longevity
    • Fibre diameter
  – C
    • Feed efficiency
    • GHG
    • Meat quality
What NZ animals?

• Experiment 1 (Jan 2009) n=3000
  – 2034 industry sires (Rom, Coop, Peren, Composite)
    • Av 43 FEC measured prog/sire
    • Av 187 weaning records/sire
    • Traits: WWT, LW8, FWT, NLB, FEC, SUR, ULT, dags, FE..
  – 693 FEC validation animals, 259 controls (duplicates, mapping, breed standard)

• Experiment 2 (May 2009) n=2500
  – FE resistance (Rom, Coop, Composite)

• Experiment 3 (May 2009) n=460
  – industry sires used 2008

• Experiment 4 (Mar 2010) n=3500
  – 2000 animals meat yield
  – 1500 additional sires and validation animals
Prelim results: Data QC and method

- Standard genotype checks: final 47,676 SNPs
- Data obtained from SIL BV analysis ~2.5M records
- Converted to adjusted records (Mrode and Swanson 2003)
- R, C, P >50% breed included in training analysis
- Those with phenotype reliability 80% of heritability retained
- Broken into training and validation sets based on birth year
- Used BLUP methodology and adjusted for first 6 breed PCA components
- Validation animals estimates based on own plus progeny records…
PCA plots

Key (for both plots)
- ○ Romney
- ■ Coopworth
- ✤ Perendale
- △ Texel
- × Marshall Romney
- X Other
Cut off years and number of animals in training and validation sets.

<table>
<thead>
<tr>
<th>Trait</th>
<th>1st validation year</th>
<th>nTrain</th>
<th>nValidation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>C</td>
<td>P</td>
</tr>
<tr>
<td>wwt</td>
<td>2006</td>
<td>2007</td>
<td>2004</td>
</tr>
<tr>
<td>cw</td>
<td>2006</td>
<td>2007</td>
<td>2004</td>
</tr>
<tr>
<td>NLB</td>
<td>2004</td>
<td>2002</td>
<td>2003</td>
</tr>
<tr>
<td>Fec1</td>
<td>2005</td>
<td>2007</td>
<td>2004</td>
</tr>
<tr>
<td>Fec2</td>
<td>2005</td>
<td>2007</td>
<td>2004</td>
</tr>
<tr>
<td>AEFec</td>
<td>2005</td>
<td>2002</td>
<td>2004</td>
</tr>
</tbody>
</table>

2917 animals in total
FEC1

The graph shows the accuracy of different models (Train, R, C, P) as a function of the number of SNPs. The accuracy increases with the number of SNPs for all models, with the Train model showing the highest accuracy overall.
What do we expect?

- SNP density
  - \(2 \times Ne \times L\)
  - \(~30,000\)
  - Rom \(~450\)
- Accuracy depends on
  - Ne
  - Effective \(h^2\)
- Across breed predictions
  - Need more SNPs
- Need more animals preferably progeny tested sires
SIL Implementation

- Genetic params
- Pedigree, traits → EBVs, accuracies
- Animal
- DNA sample, Genotypes
- MBVs, accuracies
- SNP Coeffs &
- Blend → GBV

Ovita™

MEAT & WOOL
NEW ZEALAND

agresearch
Future NZ AgR/SIL Database structure

- GenomNZ: 1000K DNA, 100K/yr, 50K/yr, Imp Geno?
- DNAdb: ~1000 Genomes, ~200/yr
- SIL: >10M Genomes
- SIL GE: ~3M
- Meat Co: NAIT
SNP chips: single gene traits

- Very powerful only need 10-15 affected animals for a recessive trait
- Already used in sheep to map
  - SMA
  - “dwarf Texels… Ireland and NZ”
  - Polycystic kidney disease
  - Microphthalmia
  - MyoMax
  - Horns/poll
  - Yellow fat
  - ...........
- Easily added to genomic selection tests
1. Sheep Domestication: Poll
2. Breed and Trait Specific: meat, microphthalmia
3. Single mutation

- Related to a specific recessive mutation
  - Yellow fat in Perendales
  - Northern Europe origin
  - Carcasses downgraded
  - Partially recessive?
Cases Vs Controls
Summary

- Created ovine chip, genotyped animals, analysed initial results
- Looks like can predict across 3 breeds and crosses…..
- 50K sire product
- More progeny tested sires will improve predictions
- Starting with WWT, CWT, NLB and FEC
  - Have other traits
  - FE already genotyped
  - Meat yield and Survival this year?
  - Excellent for finding single gene traits
- Results SAME FORMAT as existing breeding values
Acknowledgements

- Ovita
- Meat & Wool New Zealand
- ISGC
- New Zealand sheep breeders
An example

$$GBV = \frac{(1 - Rel_{\text{pheno}}) MBV + (1 - Rel_{\text{markers}}) BV)}{(1 - Rel_{\text{pheno}} \times Rel_{\text{markers}})}$$

<table>
<thead>
<tr>
<th>Trait</th>
<th>Markers</th>
<th>Phenotype</th>
<th>Blended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MBV</td>
<td>Acc</td>
<td>BV</td>
</tr>
<tr>
<td>animal 1</td>
<td>WWT</td>
<td>1.10</td>
<td>2.00</td>
</tr>
<tr>
<td>animal 2</td>
<td>WWT</td>
<td>1.10</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>animal 1</td>
<td>LW8</td>
<td>2.20</td>
<td>3.00</td>
</tr>
<tr>
<td>animal 2</td>
<td>LW8</td>
<td>2.20</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>animal 1</td>
<td>NLB</td>
<td>0.24</td>
<td>0.05</td>
</tr>
<tr>
<td>animal 2</td>
<td>NLB</td>
<td>0.17</td>
<td>0.20</td>
</tr>
</tbody>
</table>

- BVs are shrunken
- GBVs have higher accuracies
- More accurate BVs more extreme values
- Not just “adding together”
An example

\[
GBV = \frac{(1 - \text{Rel}_{\text{pheno}}) \cdot MBV + (1 - \text{Rel}_{\text{markers}}) \cdot BV}{1 - \text{Rel}_{\text{pheno}} \cdot \text{Rel}_{\text{markers}}}
\]

<table>
<thead>
<tr>
<th>Trait</th>
<th>Markers</th>
<th>Phenotype</th>
<th>Blended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MBV</td>
<td>Acc</td>
<td>BV</td>
</tr>
<tr>
<td>animal 1</td>
<td>WWT</td>
<td>1.10</td>
<td>2.00</td>
</tr>
<tr>
<td>animal 2</td>
<td>WWT</td>
<td>1.10</td>
<td>0.50</td>
</tr>
<tr>
<td>animal 1</td>
<td>NLB</td>
<td>0.24</td>
<td>0.05</td>
</tr>
<tr>
<td>animal 2</td>
<td>NLB</td>
<td>0.17</td>
<td>0.20</td>
</tr>
</tbody>
</table>

- BVs are shrunken
- GBVs have higher accuracies
- More accurate BVs more extreme values
- Not just “adding together”
An example

\[
GBV = \frac{\left(1 - \text{Rel}_\text{pheno}\right) \text{MBV} + \left(1 - \text{Rel}_\text{markers}\right) \text{BV}}{1 - \text{Rel}_\text{pheno} \times \text{Rel}_\text{markers}}
\]

<table>
<thead>
<tr>
<th>Trait</th>
<th>Markers</th>
<th>Phenotype</th>
<th>Blended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MBV</td>
<td>Acc</td>
<td>BV</td>
</tr>
<tr>
<td>animal 1</td>
<td>WWT</td>
<td>1.10</td>
<td>0.45</td>
</tr>
<tr>
<td>animal 2</td>
<td>WWT</td>
<td>1.10</td>
<td>0.45</td>
</tr>
<tr>
<td>animal 2</td>
<td>NLB</td>
<td>0.17</td>
<td>0.40</td>
</tr>
</tbody>
</table>

- BVs are shrunken
- GBVs have higher accuracies
- More accurate BVs more extreme values
- Not just “adding together”