The Analysis and Use of Genomic Data in the
Genetic Evaluation of Livestock

Samuel Adam Clark
Bachelor of Rural Science (Hons) University of New England

A thesis submitted for the degree of Doctor of Philosophy of the
University of New England

September 2012

School of Environmental and Rural Science
Faculty of Arts and Science
Abstract

This thesis explores various aspects of genetic evaluation using genomic information. Genomic selection is based on the prediction of the effects of thousands of markers that are associated with quantitative trait loci (QTL), usually through linkage disequilibrium. However, genomic predictions of breeding value are also strongly affected by the degree of relatedness between individuals. A better understanding of genomic selection has possible consequences for evaluating genomic prediction accuracy, the design of reference populations and managing the balance between genetic gain and inbreeding. This thesis aims to gain an understanding into the use of genomic information in the genetic evaluation of livestock.

The first experiment of this thesis compared the performance of different methods based on pedigree relationships (Best Linear Unbiased Prediction, BLUP), genomic relationships (gBLUP), and based on a Bayesian variable selection method (Bayes B) to estimate breeding values under a range of different underlying models of genetic variation, marker densities and varying animal relationships. It found that Bayes B was the most accurate method to predict breeding value when genetic variation was controlled by common QTL and rare variants for marker densities that ranged from 5,000 markers to full SNP sequence. However, the superiority of Bayes B over gBLUP is highly dependent on the presence of large QTL effects.

The second experiment examined the accuracy of estimated breeding values for different groups of selection candidates that had varying degrees of relationship to the reference dataset. We found that an animal’s relationship to the reference data set is an important factor for the accuracy of genomic prediction. Animals that share a close relationship to the reference data set
gain the highest accuracy when using either gBLUP or pedigree based BLUP. However a baseline accuracy that is driven by the reference data set size and the overall effective population size enables gBLUP to estimate a breeding value for unrelated animals within a population, using information previously ignored by pedigree based BLUP methods.

The third experiment compared the use of genomic or pedigree information in optimal selection, where genetic gain is balanced with inbreeding. The amount of within family variation that is explained by genomic breeding values was also observed. Selection based on genomic breeding values increased genetic gain and when genomic measures of co-ancestry were used to restrict inbreeding more genetic gain was achieved but this effect was only significant when the population consisted of many large full sib families. In a half sib structure there was no advantage in using genomic relationships to manage inbreeding and increased merit was obtained via the higher accuracy of genomic breeding values. For a group of dairy sires, we found that between 30 and 40% of the variation in genomic breeding values was because of within family variation due to Mendelian sampling.

The final experiment compared the use of identity by state (IBS) and identity by descent (IBD) information to construct the GRM. This study shows that IBD probabilities and information from the fastIBD module of the Beagle software can be used to predict breeding values in real data. Genomic relationship matrices based on IBD performed similar to those based on IBS when comparing the accuracy of genomic prediction in a real sheep dataset on the trait, eye muscle depth. The variance components estimated from either IBD or IBS information are affected by the scale of the GRM based on the relationships between known relatives.
Acknowledgements

Firstly, I would like to thank my principal supervisor Professor Julius van der Werf for his continued encouragement and support. I thank you for your enthusiastic attitude and knowledge which has enabled me to learn both the theoretical and applied aspects of scientific research, something I will value for my entire life. Secondly, I would also like to thank Dr John Hickey who was also a valuable contributor to my thesis. I’m grateful for your supervision and support. Thanks also to all of the other fellow scientists who have contributed to this thesis. Particularly to Professor Brian Kinghorn, his help and support was greatly appreciated.

This thesis would not have been possible without the support of my family and friends. I’d like to thank my family and my wife’s family for their kind support. Finally, my wife Annie, I would like to thank you for your loving and selfless support, understanding and patience.

I would also like to acknowledge the financial support provided by the Cooperative Research Centre for Sheep Industry Innovation, University of New England.
# Table of Contents

Abstract ........................................................................................................................................ 2  
Declaration .................................................................................................................................... 4  
Acknowledgements.......................................................................................................................... 5  
Table of Contents............................................................................................................................. 6  
List of Figures.................................................................................................................................... 10  
List of Tables.................................................................................................................................... 12  
Chapter 1. Introduction..................................................................................................................... 14  
Chapter 2. Review of Literature....................................................................................................... 20  
  2.1. Genetic markers....................................................................................................................... 21  
  2.2. Genome wide association studies and QTL mapping .............................................................. 23  
  2.3. Genetic evaluation using genomic information ....................................................................... 26  
  2.4. Methods used for genomic evaluation .................................................................................... 28  
    2.4.1. Ridge Regression Best Linear Unbiased Prediction (RR-BLUP) ........................................ 28  
    2.4.2. Bayes A and B .................................................................................................................... 29  
    2.4.3. Genomic Best Linear Unbiased Prediction (gBLUP) .......................................................... 32  
    2.4.1. The genomic relationship matrix (GRM) ......................................................................... 34  
  2.5. Results using genomic evaluation methods in real data......................................................... 36  
Chapter 3. Different models of genetic variation and their effect on genomic evaluation............. 39  
  3.1. Abstract .................................................................................................................................... 40  
  3.2. Introduction ............................................................................................................................... 41  
  3.3. Methods .................................................................................................................................... 44  
    3.3.1. Base genotype simulations ................................................................................................ 44  
    3.3.2. Statistical analyses and breeding value estimation ............................................................ 47  
  3.4. Results .................................................................................................................................... 51
Chapter 4. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes.

4.1. Abstract ............................................................................................................. 67
4.2. Introduction ....................................................................................................... 68
4.3. Methods ............................................................................................................. 70
  4.3.1. Simulated Data ............................................................................................. 71
  4.3.2. Merino sheep phenotypic data ..................................................................... 74
4.4. Results ............................................................................................................. 77
  4.4.1. Simulation .................................................................................................... 77
  4.4.2. Merino sheep data analysis ......................................................................... 80
4.5. Discussion ....................................................................................................... 83
4.6. Conclusions ..................................................................................................... 87
4.7. Competing interests ....................................................................................... 88
4.8. Authors' contributions .................................................................................... 88
4.9. Appendix 1: Accuracy estimated using the PEV of the mixed model equations weighted by genomic relationships. ................................................................. 88
4.10. Acknowledgements ......................................................................................... 89

Chapter 5. The effect of genomic information on optimal selection in livestock breeding programs ................................................................................................................ 92

5.1. Abstract .......................................................................................................... 93
5.2. Introduction ..................................................................................................... 94
5.3. Methods .......................................................................................................... 97
  5.3.1. Simulation study............................................................................................ 97
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.2.</td>
<td><strong>Optimal selection</strong></td>
<td>99</td>
</tr>
<tr>
<td>5.3.3.</td>
<td><strong>Real Data Analysis</strong></td>
<td>100</td>
</tr>
<tr>
<td>5.4.</td>
<td>Results</td>
<td>102</td>
</tr>
<tr>
<td>5.4.1.</td>
<td><strong>Simulation study</strong></td>
<td>102</td>
</tr>
<tr>
<td>5.4.2.</td>
<td><strong>Real Data</strong></td>
<td>106</td>
</tr>
<tr>
<td>5.5.</td>
<td>Discussion</td>
<td>111</td>
</tr>
<tr>
<td>5.6.</td>
<td>Conclusions</td>
<td>115</td>
</tr>
<tr>
<td>5.7.</td>
<td>Competing interests</td>
<td>116</td>
</tr>
<tr>
<td>5.8.</td>
<td>Authors' contributions</td>
<td>116</td>
</tr>
<tr>
<td>5.9.</td>
<td>Acknowledgements</td>
<td>116</td>
</tr>
</tbody>
</table>

Chapter 6. **Comparisons of identical by state and identical by descent relationship matrices derived from SNP Markers in genomic evaluation** ................................................................. 119

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.</td>
<td>Abstract</td>
<td>120</td>
</tr>
<tr>
<td>6.2.</td>
<td>Introduction</td>
<td>121</td>
</tr>
<tr>
<td>6.3.</td>
<td>Methods</td>
<td>124</td>
</tr>
<tr>
<td>6.3.1.</td>
<td><strong>Merino Sheep Data</strong></td>
<td>124</td>
</tr>
<tr>
<td>6.3.2.</td>
<td><strong>Relationship Matrices</strong></td>
<td>126</td>
</tr>
<tr>
<td>6.4.</td>
<td>Results</td>
<td>133</td>
</tr>
<tr>
<td>6.5.</td>
<td>Discussion</td>
<td>135</td>
</tr>
<tr>
<td>6.6.</td>
<td>Conclusions</td>
<td>139</td>
</tr>
<tr>
<td>6.7.</td>
<td>Competing interests</td>
<td>140</td>
</tr>
<tr>
<td>6.8.</td>
<td>Authors' contributions</td>
<td>140</td>
</tr>
<tr>
<td>6.9.</td>
<td>Appendix 1- Identity by descent relationship matrix using genotype probabilities</td>
<td>140</td>
</tr>
<tr>
<td>6.10.</td>
<td>Acknowledgements</td>
<td>146</td>
</tr>
</tbody>
</table>

Chapter 7. **General Discussion** .................................................................................. 149

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.1.</td>
<td><strong>The validation of genomic selection methods using simulated and real data</strong></td>
<td>152</td>
</tr>
<tr>
<td>7.1.2.</td>
<td><strong>The number of markers used in GS and the promise of sequence data</strong></td>
<td>155</td>
</tr>
</tbody>
</table>
7.1.3. The makeup of reference populations for genomic selection................................. 159
7.1.4. Across breed prediction .......................................................................................... 161
7.1.5. The value of genomic information to a structured breeding program ............... 164
7.2. Conclusions .................................................................................................................. 166

Chapter 8. Consolidated Reference List................................................................................. 167
List of Figures

Figure 3.1- The effect of the number of QTL and marker density on the accuracy of estimating breeding values in the test set using Bayes B (reference population 1) ........................................ 56

Figure 4.1- Estimates of accuracy based on the PEV from the coefficient matrix ($r_{(PEV)}$) and based on the correlation between estimated and true breeding values ($r_{(cor)}$) for the close and distantly related individuals using genomic (gBLUP) and pedigree (BLUP-S and BLUP-D) based prediction methods ........................................................................................................... 79

Figure 4.2- Estimates of accuracy predicted using gBLUP and plotted against different measures of relationship between an animal in the test data set with animals in the reference data set. These measures include: a) The mean relationship, b) The average of the top ten relationships, c) The average of the top 100 relationships and d) The maximum relationship to the reference population. .................................................................................................................................................. 82

Figure 5.1- Selection response when selecting on EBV or GEBV for various levels of constrained genomic inbreeding ........................................................................................................ 103

Figure 5.2- Selection response when selecting on GEBV and constraining inbreeding either based on pedigree or genomic information ........................................................................................................ 104

Figure 5.3- Selection response when selecting on GEBV and constraining inbreeding either based on pedigree or genomic information in a full sib population ........................................................................... 105

Figure 5.4- Optimized selection of ADHIS bulls at varied rates of genomic co-ancestry using three different measures of merit for protein yield (comparisons based on the scale of the breeding value used). ......................................................................................................................... 106
Figure 5.5- Optimized selection of ADHIS bulls at varied rates of genomic co-ancestry using three different measures of merit for protein yield (comparisons based on PT BV). .................. 107

Figure 5.6- Optimized selection of LIC Holstein bulls at varied rates of genomic co-ancestry using three different measures of merit for protein yield (comparisons based on the scale of the breeding value used). ................................................................. 108

Figure 5.7- Optimized selection of LIC Holstein bulls at varied rates of genomic co-ancestry using three different measures of merit for protein yield (comparisons based on PT BV). ...... 108

Figure 5.8- Selection response when selecting on GEBV and constraining inbreeding either based on pedigree or genomic information (LIC- Protein yield)......................................................... 109

Figure 5.9- Selection response when selecting on GEBV and constraining inbreeding either based on pedigree or genomic information (ADHIS- Protein yield)..................................................... 110
List of Tables

Table 3.1- The average accuracy of breeding value estimates (±SE) in the test set obtained from three methods of analysis of reference population 1 with 60,000 SNPs and different genetic models........................................................................................................................................ 52

Table 3.2- The average accuracy of breeding value estimates (±SE) in the test set obtained from three methods of analysis of reference population 2 with 60,000 SNPs and different genetic models........................................................................................................................................ 53

Table 3.3- The average accuracy of breeding value estimates (±SE) in the test set obtained from three methods of analysis of reference population 3 with 60,000 SNPs and different genetic models........................................................................................................................................ 54

Table 4.1- Empirical accuracy¹ (±S.E.)² using genomic and pedigree based methods in simulated data........................................................................................................................................ 78

Table 4.2- Empirical¹ (r(cor)) and estimated accuracy² (r(pev)) using genomic and pedigree based methods for the Merino EMD data........................................................................................................ 80

Table 4.3- Empirical¹ (r(cor)) and estimated accuracy² (r(pev)) using genomic and pedigree based methods for the Merino SC_WT data........................................................................................................ 81

Table 5.1- Intra class correlation between breeding values from the half sib and full sib populations........................................................................................................................................ 105

Table 5.2- The proportion of variation in breeding value explained by between family (Sire and Dam) and within family (MS) information........................................................................................................ 111
Table 6.1- The accuracy of the estimated breeding values (correlation between gEBV and ASBV) and the regression of gEBV on ASBV, estimated using different methods to define the genomic relationship matrix ................................................................. 133

Table 6.2- Variance components estimated using various methods to define the genomic relationship matrix. .......................................................................................................................... 134

Table 6.3- Mean and Standard deviation of different relationship classes for alternative GRM’s. ........................................................................................................................................ 137

Table 6.4- Genotype probabilities given IBD status, using the proband locus alone............. 141

Table 6.5- Components of prior probability of IBS at the supporting locus, s...................... 142

Table 6.6- Components of prior probability of IBS at the supporting locus, s...................... 143