

USE OF MOLECULAR MARKERS IN
GENETIC EVALUATION OF ANIMALS
WITH APPLICATIONS IN AUSTRALIAN
MERINOS

Bronwyn Elizabeth CLARKE

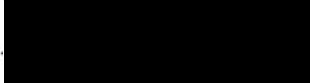
B.Sc. Wool and Pastoral (HONS) UNSW

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

October 1997

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree or qualification.

I certify that any help received in preparing this thesis, and all sources used, have been acknowledged in this thesis.

..... 

Acknowledgments

This thesis is dedicated to my family and to Geoffrey Rasmussen. Thank you for putting up with the ups and downs of PhD research and writing for the last three years and for your immense support and encouragement over this period. Also, thank you to my friends in Armidale, who heard the day to day roller coaster ride of PhDing ... *The roller coaster ride has finally come to its end!*

I am indebted to my three supervisors who have all provided valuable assistance in different areas throughout the course of my study. Thanks must first go to Brian Kinghorn my principle supervisor. Brian provided the initial questions that became chapters in this thesis and has provided support and encouragement throughout my study. Ian Purvis is thanked for keeping me on track with the sheep side of things and for his support. A huge thanks must go to Julius van der Werf. “Thank you Julius for taking over where Brian left off and for putting up with the *hard work* time of the PhD.” Julius provided immense support at the crunch time of finishing chapters and assisting in the whole thesis coming together.

The support of a CSIRO/UNE Postgraduate Research Scholarship provided the funds necessary to carry out this research project. In combination with this, I would especially like to thank Johan van Arendonk for organising a Fellowship from Wageningen Agricultural University, and for the three months I was able to spend working with the animal breeders at Wageningen. This visit was vital to the research that I have done. Further thanks to Agriculture Western Australia who provided support and time for me to finish this thesis while working for them.

A number of people have provided computing assistance and programming code that has assisted my research, so thanks also to John Henshall, Richard Kerr and Andrew Swan. AGBU, Animal Science and CSIRO Postgraduates and Staff have all provided for a great three years at UNE. The seminars, journal club discussions and lunchtime discussion have all been helpful. A final thanks, again, to John Henshall, for competitive coding, and to Susanne Hermesch for reading the thesis along the way and providing help from the beginning to the end.

Abstract

This thesis has examined aspects of detection and utilisation of genetic marker technology in animal breeding systems. Models have been proposed to include genetic marker information into mixed model equations for estimating major gene, or Quantitative Trait Loci (QTL) effects. These involve building a Gametic Relationship Matrix (GRM) with probabilities of identity by descent of QTL alleles between gametes. Algorithms to build the GRM were compared and a further method was proposed to build the GRM based on simulated QTL information. The simulated QTL method was identified as an approximation of the true GRM, as it involved simulating all possible QTL–marker genotype combinations.

Detection of linkage between genetic markers and QTL is usually based on data collected from animals in the field which are assumed to be unselected. The effect of this assumption on parameter estimates is unknown. Results from use of a restricted maximum likelihood detection method and a granddaughter design population, show that in the presence of within family (or sire) selection, large underestimates of variance occur and QTL position and effect estimates are inaccurate. However, selection of grandsires only has little effect on estimation of variance.

A non-iterative method for estimating genotype probabilities using genetic marker information is proposed. Random QTL effects are estimated using best linear unbiased prediction and genotype probabilities are calculated conditional on each possible QTL genotype. Stochastic simulated used to compare the non-iterative method with an iterative method, which incorporates segregation analysis, showed that the non-iterative method does not provide as accurate estimates, however, it is able to handle complex pedigrees, making it feasible over any pedigree structure.

A practical application of the use of genotype probability estimates is in mate selection decisions. Strategies for including genotype information on the recessive gene for black lambs in a sheep breeding program using mate selection were evaluated. Results show that for an observable trait, such as the gene for black lambs, there is little benefit in genotyping all animals. Similar gains can be made by use of genotype probability estimates or by selective genotyping and using mate selection to optimise genetic value of the progeny.

Table of Contents

Acknowledgments.....	iii
Abstract	iv
Table of Contents	v
List of Figures	ix
List of Tables.....	x

Chapter 1 - General Introduction

Chapter 2 - The Use of Genetic Markers in QTL Detection and Evaluation

2.1 Introduction	7
2.2 Genetic Markers	9
2.2.1 Polymerase Chain Reaction	10
2.2.2 Types of Genetic Markers.....	10
2.2.2.1 Restriction Fragment Length Polymorphisms.....	10
2.2.2.2 Variable Number Tandem Repeats.....	11
2.2.2.3 Short Tandem Repeats.....	11
2.2.2.4 Randomly Amplified Polymorphic DNA	12
2.2.2.5 Further Developments	12
2.3 Genetic Maps.....	13
2.4 Detection of Linkage Between Genetic Markers and QTL.....	14
2.4.1 Cross Between Inbred Lines	14
2.4.2 Within Family Segregation	15
2.4.3 Half-Sib Versus Full-Sib Design.....	16
2.4.4 Single Marker Versus Interval Mapping.....	17
2.4.5 Power of Experiments to Detect Linkage	18
2.5 Methods Available for QTL Detection and Evaluation	20
2.5.1 Modelling of Phenotype and Genetic Marker Information.....	20
2.5.2 Segregation Analysis.....	22
2.5.2.1 Maximum Likelihood	23
2.5.2.2 Regression	26
2.5.2.3 Gibbs Sampling	28
2.5.2.4 Application of Segregation Analysis	30
2.5.3 Linkage Analysis.....	31
2.5.3.1 Maximum Likelihood	31
2.5.3.2 Regression	36
2.5.3.3 Restricted Maximum Likelihood	38
2.5.3.4 Gibbs Sampling	42
2.5.3.5 Estimated Breeding Value Only	43
2.5.4 Estimated Parameters	44
2.5.5 Discussion	46

2.6 Genetic Selection Strategies using Molecular Information.....	48
2.6.1 Marker Assisted Selection.....	48
2.6.1.1 MAS Utilising Linkage Disequilibrium	49
2.6.1.2 MAS Using Estimated Breeding Values.....	50
2.6.2 Genetic Gains using Marker Assisted Selection	51
2.6.3 Discussion	55
2.7 Genetic Gains and Molecular Technology in the Merino Wool Industry.....	56
2.7.1 Genetic Variation	56
2.7.2 Conventional Genetic Evaluation.....	58
2.7.3 Molecular and Reproductive Technology	59
2.7.3.1 DNA Fingerprinting.....	59
2.7.3.2 Advanced Breeding Techniques	59
2.7.4 Mapping the Sheep Genome and Marker Assisted Selection	60
2.7.5 Discussion	65
2.8 Conclusions	66

Chapter 3 - Comparison of Algorithms for Incorporating Genetic Marker Data in BLUP

3.1 Introduction	69
3.2 Development of Mixed Model Equations to Include Genetic Marker Information.....	71
3.3 Methods to Build the GRM and its Inverse.....	73
3.3.1 Method of Fernando and Grossman (1989)	74
3.3.2 Method of van Arendonk et al. (1994c)	76
3.3.3 Wang et al. (1995).....	78
3.3.4 Quantitative Trait Loci Simulation Method to build Approximation of True GRM.....	79
3.4 Comparison of Elements in the GRM and Differences in Resultant Estimated Breeding Values.....	82
3.4.1 Simulated Pedigree.....	82
3.4.2 Comparison of Elements in the GRM and its Inverse.....	84
3.4.3 Comparison of Accuracy of Estimated Breeding Values.....	85
3.5 Results	87
3.5.1 Comparison of Elements in the GRM and its Inverse.....	87
3.5.2 Comparison of Estimated Breeding Values	90
3.6 Discussion	92
3.6.1 The Effect of Unknown Parental Origin of Marker Alleles on Elements of the GRM.....	92
3.6.2 The Effect of Inbreeding on Elements of the GRM and Prediction Error Variance.....	93
3.6.3 Simulated Pedigree and Genetic Model	94

Chapter 4 - Analysis of Linkage Between Genetic Markers and Quantitative Trait Loci Using REML: The Effect of Selection on Parameter Estimates

4.1 Introduction	95
4.2 Materials and Methods	97
4.2.1 Simulated Data	97
4.2.2 Selection Scenarios	99
4.2.3 Genetic Model	100
4.2.4 Starting Values and Bounds	102
4.2.5 Likelihood Ratio	103
4.3 Results	103
4.3.1 Selection of Grandsires - increasing selection intensity	103
4.3.2 Further Selection Scenarios	105
4.3.3 Decreased Population Size	107
4.4 Discussion	108
4.4.1 REML Detection Method	108
4.4.2 Changes in Genetic Variance with Selection	109
4.4.3 Simulated Pedigree	110
Appendix A - The Effect of Marker Spacing on the Biasedness of QTL Position Estimates	112
Appendix B - Reduction in Genetic Variance When Only Selected Grandsires are Available for Genotyping	117

Chapter 5 - Testing a Non-Iterative Genotype Probability Estimation Method

5.1 Introduction	120
5.2 Method	122
5.2.1 Non-Iterative Method to Calculate Genotype Probabilities	122
5.2.1.1 Estimation of Breeding Values Using Marker Information	123
5.2.1.2 Genotype Probability Estimation	124
5.2.2 Segregation Analysis Method of Fernando et al. (1993)	128
5.2.2.1 Calculation of Genotype Probabilities	129
5.2.3 Comparison of Methods to Calculate Genotype Probabilities	132
5.2.4 Simulation of Test Data Set	132
5.3 Results	134
5.4 Discussion	138
5.4.1 Effect of Genetic Markers	138
5.4.2 Effect of Polygenes	138
5.4.3 Effect of Selection	139
5.4.4 Effect of Method of Analysis	139
5.5 Conclusions	141

Chapter 6 - Use of Marker Information and Genotype Probabilities in Mate Allocation and Selection Decisions

6.1 Introduction	142
6.2 Materials and Methods	142
6.2.1 Simulation Procedures.....	145
6.2.1.1 Parameters for Simulated Flock.....	145
6.2.1.2 Simulation of Records.....	146
6.2.1.3 Calculation of Selection Criterion.....	148
6.2.2 Selection Alternatives	150
6.2.2.1 Phenotypic Selection Against Black Lambs	150
6.2.2.2 Selecting Mating Pairs using Linear Programming - all animals genotyped	150
6.2.2.3 Selecting Mating Pairs using Linear Programming - only males genotyped.....	152
6.2.2.4 Selecting Mating Pairs using Linear Programming - using genotype probabilities	153
6.2.3 Mate Selection Procedures using Linear Programming	154
6.2.4 Summary Statistics.....	155
6.3 Results	156
6.3.1 Reference Selection Options	156
6.3.2 High Cost Selection Options	159
6.3.3 Medium Cost Selection Options	162
6.3.4 Low Cost Selection Options.....	165
6.4 Discussion	168
6.4.1 Impact of Genotyping all Animals	169
6.4.2 Impact of Genotyping Only Males	169
6.4.3 Impact of No Genotyping.....	170
6.4.4 Simulation Performance.....	170
6.5 Conclusions.....	171

Chapter 7 - General Discussion

7.1 Introduction	172
7.2 Detection of QTL-marker linkage in the Wool Industry.....	174
7.3 General Conclusions	176
References	177

List of Figures

Figure 1.1 Illustration of the three steps from identification of genetic markers to utilisation in breeding programs.....	1
Figure 3.1 Example of chromosome segment simulated for each animal	80
Figure 3.2 Flow chart for QTL simulated Genetic Relationship Matrix.	81
Figure 3.3 Difference between empirical prediction error variance and prediction error variance from the diagonal of the coefficient matrix.	92
Figure 5.1 An illustration of the input and output parameters required for the 2-step genotype probability estimation method.....	122
Figure 5.2 Heights of the distribution of \hat{v}_i (EBV at QTL for animal i) at the expectations conditional on QTL genotype are proportional to genotype probabilities.	125
Figure 6.1 Difference in true breeding value (TBV) between combined selection options (PHS, AS2, AS3, AS4, AS2*,AS3*,AS4*) and selection for FD EBV (AS5) only.	162
Figure 6.2 Difference in true breeding value (TBV) between combined selection options with only males genotyped (MS2, MS3, MS4, MS2*, MS3*, MS4*) and selection for FD EBV (AS5) only.	165
Figure 6.3: Difference in true breeding value (TBV) between combined selection options involving segregation analysis to obtain genotype probabilities (PS2, PS3, PS4, PS2*, PS3*, PS4*) and selection for FD EBV (AS5) only.....	168

List of Tables

Table 2.1 Methods of segregation and linkage analysis and parameters estimated by each method. X indicates method of analysis used, or parameter estimated, ~ indicates that expansion to this task is possible.	45
Table 2.2 Marker assisted selection in different animal breeding populations: Realised gains per generation from simulation studies	54
Table 2.3 Ranges for heritabilities (on diagonal), genetic correlations (below diagonal) and phenotypic correlations (above diagonal) for wool traits in Merino sheep (from Atkins, 1997).....	57
Table 3.1 Pedigree and Marker information for population given simulated QTL alleles.	83
Table 3.2 Simulated quantitative trait locus genotypes and associated values	84
Table 3.3 The gametic relationship matrix built using Fernando and Grossman (1989) and van Arendonk <i>et al.</i> (1994c) [upper triangle] and Wang <i>et al.</i> (1995) [lower triangle] - elements differing between the two GRMs are in italics. (recombination rate = 0.1)	87
Table 3.4 Sum of squares of differences between elements in the four different gametic relationship matrices - elements in the bottom left of the table are the same as those above the diagonal (recombination rate = 0.1).....	88
Table 3.5 The inverse of the gametic relationship matrix built using the rules of Fernando and Grossman (1989) [upper triangle] and by van Arendonk <i>et al.</i> (1994c) [lower triangle] - elements differing between the two matrices are in italics. (recombination rate = 0.1) ..	89
Table 3.6 Sum of squares of differences between elements in the inverse of four different gametic relationship matrices - elements in the bottom left of the table are the same as those above the diagonal (recombination rate = 0.1)	90
Table 3.7 Empirical prediction error variance [$PEV = \text{var}(a - \hat{a})$] for animal number 6 calculated over 100,000 replicates with varying recombination rate (r)	90
Table 3.8 Prediction error variance calculated using the diagonals of the inverted coefficient matrix [$PEV = C^{22}\sigma_e^2$] for animal number 6 for varying recombination rate (r).....	91
Table 4.1 Input parameters for simulation of base population.....	99
Table 4.2 Parameter estimates (empirical standard errors in brackets) for an unselected population and a population with increasing proportions of selected grandsires.	104
Table 4.3 Parameter estimates (empirical standard errors in brackets) for five further selection schemes.	106
Table 4.4 Parameter estimates (empirical standard errors in brackets) for an unselected	

population and a population with increasing proportions of selected grandsires. As in Table 4.2, but for a reduced population size.	107
Table A1 Estimates of QTL position within marker bracket with varying marker bracket interval and simulated QTL position over 250 replicates (standard errors in brackets).	113
Table A2 Percent of 250 replicates in each class according to estimates of QTL position (simulated position = 0.25) within marker brackets - those to the left of the marker bracket (< 0.0), those within the marker bracket (0.0 - 1.0) and those to the right of the marker bracket (> 1.0).	114
Table A3 Percentage of 250 replicates in given classes according to estimates of QTL position - those <i>above</i> and <i>below</i> the simulated QTL position (distance between bracketing marker = 10 cM).	114
Table A4 Different methods of reporting QTL position estimates averaged over 250 replicate populations. The number of replicates included in averages are given in brackets. (Simulated QTL position = 0.25)	115
Table A5 Effects of setting bounds for QTL position estimates and subsequent effects on other parameter estimates for a simulated QTL position of 0.25 relative to a marker bracket length of 10 cM (standard errors in brackets).	116
Table B1 Analysis of variance table for estimates of mean squares.	118
Table 5.1 QTL Genotype classes, frequencies and values (Falconer, 1989).	133
Table 5.2 Average correlations between true and estimated number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.2	134
Table 5.3 Average correlations between true and estimated number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.5.	135
Table 5.4 Average number of favourable QTL alleles captured in top 5% of population when simulated populations were ranked on the number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.2	136
Table 5.5 Average number of favourable QTL alleles captured in top 5% of population when simulated populations were ranked on the number of favourable QTL alleles. Initial Favourable QTL Allele Frequency = 0.5	137
Table 6.1 Parameter values used for flock simulation	146
Table 6.2 Fibre diameter parameters used for phenotypic simulations.	147
Table 6.3 Transmission probabilities of inheriting QTL alleles from sires and dams (assuming Mendelian inheritance).	148
Table 6.4 Paired merit function ($f_{ij}(PM)$) for sire i and dam j , based on measure of estimated breeding value EBV_{ij} and b genotype probability P_{ij} for five mate selection schemes	

(AS1-AS5). [Paired Merit $f_{ij}(PM)^{\wedge}$ is similar to $f_{ij}(PM)$, except bb genotype probabilities are minimised and four mate selection schemes are used (AS1*-AS4*)]	152
Table 6.5 Transmission probabilities of inheriting QTL alleles from genotyped sires and ungenotyped dams (assuming Mendelian inheritance).	153
Table 6.6 Results of selection on EBV for FID and random mating with black animals (bb) not able to be parents (PHS), selection for genotype probabilities, P_{ij} (AS1) and P_{ij}^{\wedge} (AS1*) and estimated breeding value for fibre diameter EBV_{ij} (AS5).....	157
Table 6.7 Results of selection for genotype probabilities with only males genotyped, P_{ij} (MS1) and P_{ij}^{\wedge} (MS1*) and with no animals genotyped (using segregation analysis) P_{ij} (PS1) and P_{ij}^{\wedge} (PS1*)	159
Table 6.8 Results of combined selection on genotype probabilities aimed at minimising frequency of b allele (P_{ij}) and estimated breeding value for fibre diameter (EBV_{ij}) with all animals genotyped.....	160
Table 6.9 Results of combined selection on genotype probabilities aimed at minimising black lambs (P_{ij}^{\wedge}) and estimated breeding value for fibre diameter (EBV_{ij}) with all animals genotyped	161
Table 6.10 Results of combined selection on genotype probabilities aimed at minimising frequency of b allele (P_{ij}) and estimated breeding value for fibre diameter (EBV_{ij}) with only male animals genotyped.....	163
Table 6.11 Results of combined selection on genotype probabilities aimed at minimising black lambs (P_{ij}^{\wedge}) and estimated breeding value for fibre diameter (EBV_{ij}) with only male animals genotyped.....	164
Table 6.12 Results of combined selection on genotype probabilities aimed at minimising frequency of b allele (P_{ij}) and estimated breeding value for fibre diameter (EBV_{ij}) with no animals genotyped.....	166
Table 6.13 Results of combined selection on genotype probabilities aimed at minimising black lambs (P_{ij}^{\wedge}) and estimated breeding value for fibre diameter (EBV_{ij}) with no animals genotyped.....	167