# METHODS AND MODELS FOR THE ACCURATE ESTIMATION OF THE EFFECTS OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNP) IN BEEF CATTLE

By
Kirsty L. Moore
BSc (Animal Science) (Latrobe), BSc (Hons) (UNE)

**A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY OF THE UNIVERSITY OF NEW ENGLAND**

February 2008

# Declaration

---

*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.*

*I certify that to the best of my knowledge any help in preparing this thesis, and all sources used, have been acknowledged in this thesis.*

……  ….

---

# Acknowledgements

I would like to thank Meat and Livestock Australia for their financial support during this thesis. I would also like to thank my supervisors, Dr David Johnston and Professor John Gibson. Thank you for your suggestions, guidance and assistance during the past three years. I would also like to acknowledge and thank Dr Bruce Tier, whose help over the last three years has been invaluable. I would like to acknowledge the Cooperative Research Centre for the Cattle and Beef Industry for including me in their post graduate program, and for providing the dataset that was analysed in Chapter 8.

I am grateful for the opportunity to complete my PhD with the Animal Genetics and Breeding Unit, and would like to thank Dr Hans Graser and all the staff and students at AGBU for their support and friendship. I would especially like to thank Dr Kath Donoghue, Anna Hansson, Dr Gilbert Jeyaruban, Rob Jones and Matt Wolcott for reading and commenting on sections of this thesis.

I have made many friends in Armidale, and I thank them for their friendships and the after work drinks. I would especially like to thank my 'Armidale family', who have made my stay in Armidale memorable. A huge thank you to my family for the support and encouragement I received during this thesis. Finally, I would like to thank Nathan for his company and support during this PhD.

# Abstract

Genetic markers provide the Australian beef industry with the opportunity to increase rates of genetic gains. However, accurate estimates of the gene frequencies and the marker size of effects are first required. Stochastic simulation was used to examine the methods and models required to estimate SNP effects.

Results showed for a single SNP explaining 2% of the phenotypic variation, 1,500 animals were required to estimate the SNP effects when the favourable allele frequency (p) was 0.5. However, increasing the additive SNP effects decreased the number of animals required. SNP effect estimates were inflated when the power to detect genotype effects was low. In addition, when the allele frequency was rare ($p$=0.1), biased dominance effects were estimated. For SNPs that were in linkage disequilibrium with the causative SNP, the SNP effects were accurately estimated when linkage disequilibrium was greater than D′=0.9. This thesis found that when there was linkage disequilibrium between the two direct SNPs (explaining 8% of the phenotypic variation, collectively), and one SNP was ignored in the model, estimates of the SNP effects were biased upwards. Ignoring epistatic effects (1% of the phenotypic variance) also increased the estimates of the SNP effects. To estimate accurate SNP and epistatic effects 3,000 animals were required. If SNPs were excluded in the model, the SNP and epistatic variance was partitioned as polygenic and residual variances, respectively. The inclusion of SNPs was shown to increase the accuracy of the estimated breeding value (EBV); the more phenotypic variation explained by the SNPs the higher the increase in EBV accuracy (or estimated genetic merit when non-additive (i.e. epistasis) effects were included).

Different genotyping strategies resulted in different accuracies for estimating SNP effects. SNP effect estimates were inflated for the truncation selection strategies, but these over-estimates could be correctly adjusted for selection provided the genotype frequencies of the whole population were known. SNP effect estimates were more accurate for truncation selection (correctly adjusted for selection), compared with random genotyping strategies.

An industry dataset for meat tenderness was analysed using the same statistical models as the simulation. In this dataset, the estimated SNP effects were similar to estimates previously reported in literature. However, there is a need for further research on how pooling data across several breeds, potentially with different SNP allele frequencies may affect the ability to estimate SNP effects.

The statistical models were also shown to be important for the estimation of SNP effects. When only one or two SNPs were considered, it was found that fitting the SNPs as fixed effects was preferred to fitting the SNPs as random effects. Different methods of modelling genetic effects had no significant effect on estimates, when animals were both phenotyped and genotyped. However, when phenotyped animals were included that were not genotyped, estimates of the SNP effects and variance components were biased when a full animal model was fitted.

This thesis shows that the population size, allele frequency, statistical power to detect genotype effects, statistical models and the data structure all affect the ability to accurately estimate the size of SNP effects.

# Table of Contents

# List of Figures

# List of Tables