View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Edinburgh Research Archive

QTL mapping technology using variance components in general pedigrees applied to the poultry industry

Suzanne J Rowe

This thesis is presented for the degree of

Doctor of Philosophy

The University of Edinburgh

2008

On First Looking into Chapman's Homer

Much have I travelled in the realms of gold,
And many goodly states and kingdoms seen;
Round many western islands have I been
Which bards in fealty to Apollo hold.
Oft of one wide expanse had I been told
That deep-browed Homer ruled as his demesne;
Yet did I never breathe its pure serene
Till I heard Chapman speak out loud and bold:
Then felt I like some watcher of the skies
When a new planet swims into his ken;
Or like stout Cortez when with eagle eyes
He stared at the Pacific – and all his men
Looked at each other with a wild surmise–
Silent, upon a peak in Darien.

John Keats,1816.

Table of Contents

Declaration 6
Acknowledgements 7
List of Tables and Appendices 8
List of Figures 11
Peer reviewed publications 14
Abstract 15

CHAPTER ONE Introduction 16
1.1 Overview 17
1.2 QTL mapping. 18
1.2.1 QTL mapping in experimental populations 18
1.2.2 QTL mapping within outbred pedigrees 19
1.2.3 QTL mapping in Livestock 20
1.2.3.1 Sibship methods 20
1.2.3.2 Variance Component methods 21
1.2.3.3 QTL mapping in human pedigrees 23
1.3 Development of variance component QTL mapping 23
1.3.1 Humans 23
1.3.3 Linkage disequilibrium fine mapping 27
1.4 Aspects of variance component mapping 28
1.5 IBD Probabilities 28
1.5.1 Recursive algorithms 28
1.5.2 Correlation-based algorithms 29
1.5.3 Simulation based algorithms 29
1.5.4 Model fitting in VC analysis - Imprinting, dominance, and epistasis 30
1.6 Statistical analysis of the linear mixed model 32
1.7 Application of VC to livestock data 33
1.8 Conclusions 35

CHAPTER TWO Theory of variance component mapping 38
2.0 Introduction 39
2.1 IBD Estimation 39
2.1.1 The numerator relationship matrix 39
2.1.2 IBD estimation for the QTL 40
2.1.3 Protocol 43
2.1.4 Genetic covariances between relatives conditional on genetic markers 44
2.2 Calculating the gametic IBD matrix and relationship matrices G and D needed for the mixed model analysis 45
2.3 The Linear model and mixed model equations 50
2.4 Extensions 52
2.4.1 Common environment 52
2.4.2 Dominance 52
2.4.3 Imprinting 53
2.5 Solving the mixed model equations 55

CHAPTER THREE QTL analysis of bodyweight and conformation score in commercial broiler chickens comparing variance component and half-sib methods **56**

	Summary	57
3.1	Introduction	57
3.2	Materials and methods	59
3.2.1	Trait measurements	59
3.2.2	Analysis of phenotypic data	59
3.2.3	Linkage maps	60
3.2.4	Half sib QTL analysis	60
3.2.5	Variance component QTL analysis	62
3.3	Results	63
3.3.1	Analysis of phenotypic data	63
3.3.2	QTL analysis	63
3.3.3	Evidence for QTL	65
3.3.4	QTL positions	65
3.3.5	Significant family analysis	65
3.4	Discussion	67
3.4.1	Body Weight	67
3.4.2	Conformation-score	68
3.4.3	Comparison with previous results	70
3.5	Conclusions	71

CHAPTER FOUR Detecting dominant QTL with variance components analysis in simulated pedigrees **73**

	Summary	74
4.1	Introduction	74
4.2	Materials and methods	77
4.2.1	Statistical Genetic Models for Variance Component Analysis	77
4.2.2	Calculating the relationship matrices A, G and D needed for the mixed model analysis	78
4.2.3	Test statistic	78
4.2.4	Population structure	79
4.2.5	Maternal effect	81
4.2.6	Null distribution	81
4.3.	Results	82
4.3.1	Power to detect dominance effects	82
4.3.2	Over dominant, spurious additive and dominant QTL effects	86
4.3.3	Maternal effects and dominance in pig scenario	88
4.3.4	Null distribution	90
4.4	Discussion	94
4.5	Conclusions	97

CHAPTER FIVE Detection of parent of origin effects using variance component analysis in simulated pedigrees **110**
Summary 111
5.1 Introduction 111
5.1.1 Genomic imprinting 113
5.2 Materials and methods 113
5.2.1 Statistical Genetic Models for Variance Component Analysis 113
5.2.2 Calculating the relationship matrices A, G and D needed for the mixed model analysis 114
5.2.3 Simulated populations 115
5.2.4.1 Polygenic variance 116
5.2.4.2 Dominant and imprinted QTL effects 116
5.2.4.3 Common environment and polygenic dominance 116
5.2.5 Test statistic 119
5.2.6 Null distribution 121
5.3 Results 121
5.3.1 Spurious detection of parent of origin effects 122
5.3.2 Power to detect imprinted QTL 127
5.3.3 False positives under full imprinting 131
5.4 Discussion 132
5.5 Conclusions 135

CHAPTER SIX Detecting Parent of Origin and dominant QTL in a two-generation commercial poultry pedigree using variance component methodology **140**
Summary 141
6.1. Introduction 141
6.2 Materials and methods 143
6.2.1. Data 143
6.2.2 Statistical Genetic Models for Variance Component Analysis 145
6.2.3 IBD Estimation 146
6.2.4 Test statistic 146
6.2.4.1 Detection of dominant QTL effects 147
6.2.4.2 Imprinting 147
6.2.5 Maternal effect 149
6.3. Results 149
6.3.1 Additive and Dominant QTL effects 149
6.3.2 Parent of origin QTL effects 154
6.4. Discussion 157
6.4.1 Chromosome 1 157
6.4.2 Chromosome 4 158
6.4.3 Chromosome 5 158
6.4.5 Testing Strategy 159
6.5 Conclusions 160

CHAPTER SEVEN General Discussion **161**
7.1 Summary of results 162
7.1.1 Commercial poultry data 162
7.1.2 Simulation 163
7.2 Theory 165
7.2.1 Design 165
7.2.2 IBD estimation 166
7.2.3 Models and testing strategy implemented 167
7.4 Expectation of distributional properties of the Likelihood Ratio Test statistic (LRT) 169
7.5 Findings from Simulation studies 170
7.5.1 Type 1 error rate 170
7.5.2 Power 176
7.6 Multiple testing 180
7.7 Permutation analysis 182
7.8 Results from poultry data 183
7.9 Implications for poultry breeding 186

BIBLIOGRAPHY **191**

Declaration

I declare that this thesis is my own composition and is an account of analyses performed by me whilst studying for the degree of Doctor of Philosophy at the University of Edinburgh

Suzanne J. Rowe

Acknowledgements

Heartfelt thanks to DJ for boundless optimism, enthusiasm and inspiration. For a door always open and help never far away, for so much laughter, and so little strife.

Thanks to all at the coffee table for musings on life,
Thanks to Janet, Stuart, Matt, Laurie and Robert for help and support
For laptops and flat screens, dongals and ports
Thanks to Liz for always being fab
Thanks to Ricardo for springing from your seat
To help to decode, unravel and explete
Thanks to Alex for being a star
For all your expertise, in particular, in R
Thanks to Cecile for your scripts of perl
To Carol, Charlie & Gail for support and fun
And finally a special thanks
For everyone!

Thanks to Chris Haley, Sara Knott, & Ricardo Pong-Wong for our lively meetings full of help, encouragement and advice.

Thanks to Cobb for data, advice, financial support and a great trip to Arkansas

Thanks to Jayne - always there beside me for the last mile

Thanks to Peter & Katie, for being the best of distractions and keeping mummy sane and finally to Tim for support, patience, humour and unerring faith.

List of Tables

1.1	Software Packages available for Variance Component Mapping in Human pedigrees.	24
2.1	Henderson's rules for estimating the inverse of the A matrix	40
2.2	Probability of descent of QTL allele given parental inheritance of flanking markers	42
2.3	Probability of IBD at QTL between sibs given IBD state at flanking markers ([a])	46
2.4	Calculating additive, dominant, maternal and paternal relationship coefficients from the Gametic IBD matrix	44
3.1	Variance component estimates for 40 day Bodyweight and Conformation Score	63
3.2	Proportion of within family variance explained by fitting a QTL for half sib families using different analyses	66
3.3	Confidence intervals (CI) obtained by bootstrapping for half-sib analyses for 40-day bodyweight and conformation score using only families showing a significant allele substitution effect	67
4.1	Population Parameters for simulated pedigrees	79
4.2	Summary of scenarios	80
4.3	Estimates of variance due to additive QTL and additive and dominant QTL effects under over-dominance when additive QTL effect of zero is simulated	88
4.4	Variance estimates for dominant QTL effect of 0.8 and maternal effects	88
4.5	Empirical 5% thresholds for LRT test statistic (and corresponding P value under χ^2 distribution). 1000 replicates simulated for single point-wise and multiple chromosome-wise testing under null scenario of no QTL effects	90
4.6	Empirical 5% thresholds for LRT test statistic when testing for dominance and corresponding P value under χ^2_1 distribution. 1000 replicates simulated for chromosome-wise testing under null scenario of no QTL effects	93
5.1	Population Parameters for simulated pedigrees.	115

5.2	Table 5.2 Simulated scenarios, genetic and non genetic parameters	118
5.3	Tests for QTL effects and corresponding empirical thresholds for 5% type 1 error based on 1000 simulated replicates	120
5.4	Thresholds for 5% type 1 error rate for *imp** test statistic derived from 1000 replicates when simulating no QTL and simulating fully dominant QTL.	122
5.5	False positive rates for *imp** statistic from empirical 5% thresholds, based on 100 replicates of each scenario.	125
5.6	Proportion of replicates (n=100) where there is maternal expression (patvfull test) and/or paternal expression (matvfull test) in a range of scenarios simulating additive, dominant, overdominant and imprinted QTL	129
5.7	Power (proportion of significant replicates, based on 100 replicates) to detect fully imprinted QTL under different QTL models using 5% empirical threshold derived under null scenario of no QTL.	130
6.1	Summary statistics and heritabilities for trait data	144
6.2	Tests for QTL effects and corresponding empirical thresholds for 5% type 1 error based on 1000 permutations.	148
6.3	Test statistics and proportion of phenotypic variance explained at most likely QTL position when fitting polygenic, additive QTL and dominance QTL effects for 40 day bodyweight and conformation score on chicken chromosomes 1, 4 and 5	151
6.4	Table 6.4 Test statistics for all models at highest test statistic for a *pat+mat* QTL model fitting a maternal and paternal QTL effect versus no QTL	155
6.5	Table 6.5 Proportion of phenotypic variance explained by polygenic, direct maternal (dam), paternal QTL and maternal QTL effects fitted in a *pat+mat* model at the position of the highest test statistic for *pat+mat* model versus no QTL	156
7.1	Table 7.1 Empirical 5% thresholds for LRT test statistic when testing for dominance and corresponding P value under χ^2_1 distribution. 1000 replicates simulated for chromosome-wise testing under null scenario of no QTL effects	172
7.2	Empirical 5% thresholds for the null distribution of the test for dominance *dom* in human and poultry pedigrees. 1000 replicates were simulated for chromosome-wise testing under	173

null scenario of no dominance and varying additive QTL h^2

7.3 Comparison of 5% Type 1 error thresholds set by permutation analysis and empirical distribution for simulated poultry pedigree using 1000 replicates/permutation tests. 175

7.4 Table 7.4 Comparison of power to detect additive and dominant QTL using empirically derived, tabulated, and using web based calculator approximating deterministic formulae* for prediction of power in human sibships. 178

7.5 Comparison of Power to detect imprinted QTL using the test of a full (pat + mat) model against an additive model for simulated human pedigrees 180

Appendices 72

Appendix 3.1 Marker distances and corresponding position on the consensus map. 72

Appendix 4.1 Power (proportion of replicates) to detect significant additive and dominant QTL for Poultry, Human and Pig Scenarios across range of simulated additive and dominant QTL effects at 5 and 1% thresholds based on tabulated χ2 values for likelihood ratio statistic 99

Appendix 4.2 Power (proportion of 100 replicates) to detect significant additive and dominant QTL for Maternal Scenarios using pig scenario across range of simulated additive and dominant QTL effects at 5 and 1% thresholds based on tabulated χ2 values for likelihood ratio statistic. 104

Appendix 5.1 Proportion of significant replicates (n=100) for all tests based on 5% empirical type 1 error rate derived under null scenario of no QTL (Threshold 1) 136

Appendix 5.2 Proportion of significant replicates (n=100) for dom test statistic based on 5% empirical type 1 error rate derived under null scenario of no QTL (Threshold 1) 137

Appendix 5.3 Estimates of variance components for all models for maternally expressed QTL 138

Appendix 5.4 Estimates of variance components for all models for paternally expressed QTL 139

List of Figures

2.1	An example pedigree for a single sire (3) mated to two dams (1 & 2) each with four offspring, creating a nested full sib/half sib family structure.	45
3.1	Log-transformed nominal P values along chromosomes for sire, dam and variance components analysis. Bodyweight on left, conformation on right, top to bottom chromosomes 1, 4 and 5. dashed line for P=0.05 and dotted line for P=0.01	64
3.3	Estimated power to detect QTL based on three-generation pedigree of 15 grandsires, with 40 daughters, each with 24 offspring (grand HS); two-generation pedigree half sib sire families of 59 (sire HS) and half sib or full sib dam families of 27 (dam HS and dam FS). Using a heritability of 0.15, distance between markers of 16 cM, and QTL heterozygosity of 0.5	70
4.1 a, b and c	Proportion of replicates where test for dominance (2v1) is significant ($P < 0.05$) when comparing full model vs additive model. 100 chromosome-wise replicates in top to bottom (a) poultry, (b) pig, and (c) human pedigrees under partial to complete dominance. Simulated additive effect fixed at 0.8. comparing tabulated 5% χ^2_1, χ^2_{1-0} thresholds and 5% empirical threshold. Mixture threshold is estimated by using tabulated 10% χ^2_1 threshold	84
4.2	Estimates of variance components from simulated poultry data. Box plots showing the range of variance estimates. Full dominance is simulated. Variance estimates for single marker position (for 1000 replicates of each scenario) for additive and dominant QTL effects. The black circles indicate the expected variance components. All replicates were significant for a QTL when testing under the full model (additive and dominance QTL effects vs null)	85
4.3	Percentage of replicates detecting additive QTL effects ($p<0.05$) using full model (add + dom), additive model (add) and testing difference between the two (dom) in simulated pig population. A dominance effect of zero is simulated	86
4.4	Over dominance. Percentage of replicates detecting over dominant QTL effects ($p<0.05$) using full model (add + dom), additive model (add) and testing difference between the two (dom) in simulated poultry population over a range of dominant QTL effects when an additive effect of zero is simulated.	87

4.5 Effects of simulating and/or fitting direct maternal effects on proportion of replicates where test for dominance (2v1) is significant ($P < 0.05$) when comparing full model vs additive model. 100 chromosome-wise replicates in pig population under partial to complete dominance (additive QTL effect fixed at 0.8). No mat effect – no maternal effect simulated or fitted, mat effect – maternal variance of 0.1 simulated but not fitted, mat effect fitted – maternal variance of 0.1 simulated and fitted. 89

4.6 Distribution of empirical point-wise test statistic in pig, poultry and human pedigrees for top to bottom (a) additive and dominance effects, (b) dominance effects and (c) additive effects compared with χ^2_1 and χ^2_{1-0} distributions. Top 300 values of 1000 replicates are displayed. 92

4.7 Comparison of distribution of empirical chromosome-wise test statistic for dominance effects under null hypothesis of no QTL in pedigrees with varying full sib (FS) and half sib (HS) structures. χ^2_{1-0}.is also plotted for comparison. All pedigrees have 1900 total offspring. Top 150 values of 1000 replicates are displayed for clarity. 94

5.1 False positive rate for test for imprinting (*imp*) when simulating additive, dominant and over dominant QTL under an empirical 5% threshold, derived from 1000 replicates of a null scenario with QTL effects of zero. The *Imp* test statistic is a comparison between a model fitting a paternal QTL and a maternal QTL *mat+pat* (H1) with *add* model fitting an additive QTL (H0). p denotes allele frequency, a denotes additive QTL effect and d denotes dominance QTL effect, QTL var is proportion of phenotypic variance explained by the QTL. 121

5.2 Empirical distribution of the test statistic for the *imp** test under a) QTL effect of zero, b) fully dominant QTL with additive and dominant QTL effects of 0.6 and c) fully dominant QTL with additive and dominant effects of 0.8. Allele frequencies for the dominant QTL are 0.3 and 0.7. First 500 of 1000 ranked replicates shown. **Imp* test statistic comprises the comparison of model fitting a paternal QTL and a maternal QTL *pat+mat* (H1) with *add* model (H0). 124

5.3 Proportion of replicates significant for the *imp* test for full imprinting for maternally (left) or paternally (right) expressed QTL based on 100 replicates, under three different empirical thresholds: From top to bottom empirical thresholds derived from a) no QTL effect, b) a fully dominant QTL with additive and dominant QTL effects of 0.6 and c) a fully dominant QTL with additive and dominant effects of 0.8. Favourable Allele frequencies for the dominant QTL are 0.5 for a and 0.3 for b and c. Results are regardless of significance at the QTL. 126

5.4 Estimation of maternal and paternal variance components using the *pat+mat* model fitting a maternal and paternal QTL. Based on 100 replicates that were all significant for the simulated QTL. a) paternally expressed QTL *pat* with a simulated imprinted effect of 0.8 and b) maternally expressed QTL *mat* with a simulated imprinted effect of 0.8. Dashed line denotes the expectation of QTL variance under the correct model 128

5.5 Spurious dominance under paternal and maternal expression 131

6.2 Test Statistic along chicken chromosomes 1, 4 and 5 using additive and dominant QTL models for weight (top) and conformation-score (bottom) 150

6.3 Test Statistic along chicken chromosomes 1, 4 and 5 for maternal and paternal QTL models for Body-weight. 152

6.4 Test Statistic along chicken chromosomes 1, 4 and 5 for maternal and paternal QTL models for conformation score. 153

Peer Reviewed Publications

Rowe, S J., Pong-Wong,R., Haley, C.S., Knott,S. and De Koning, DJ. 2008. Detecting dominant QTL with variance components analysis in simulated pedigrees. *Genetics Research*. **90**, 363-374

Rowe, S. J., D. Windsor, C. S. Haley, D. W. Burt, P. M. Hocking, H. Griffin, J. Vincent, & D. J. De Koning, (2006). QTL analysis of body weight and conformation score in commercial broiler chickens using variance component and half-sib analyses. *Animal Genetics* **37,** 269-272.

Abstract

The subject area for this thesis is detection of chromosomal regions or QTL causing complex variation at the phenotypic level. In particular, the differentiation of sources of additive and non additive variation. Unlike QTL mapping using divergent or inbred lines, this study aims to explore methods within populations, facilitating direct application of techniques such as marker assisted selection. Specifically, objectives were to evaluate a linear model or variance components (VC) approach to explore the existence and magnitude of variation caused by additive, dominant and imprinted QTL segregating in general pedigrees. This has been achieved by combining extensive simulation and analysis of real commercial poultry data. Linear models were constructed to simultaneously estimate fixed, polygenic and QTL effects. Different genetic models were compared by hierarchical extension to incorporate more variance components, and likelihood ratio test statistics derived from the comparison of full with reduced or null models. A range of additive, dominant and imprinted QTL effects were simulated within two-generation poultry, pig and human type pedigrees. Effects of family size and structure on power, accuracy of variance component estimation, and distribution of the test statistic, were evaluated. Empirical thresholds were derived by simulating populations under the null hypotheses for each type of simulated pedigree and permutation analysis in real data. In the commercial poultry data, dominant and imprinted QTL effects were found for bodyweight and conformation score. Under simulation, although power to detect QTL effects was high in two-generation livestock pedigrees, considerable variation was found in power and behaviour of test statistics. Power to detect dominance was greater in pig and poultry than human type pedigrees with theoretical thresholds increasingly conservative as the number of dams per sire decreased, highlighting the need for empirical derivation of the critical test statistic. The detection of variance caused by imprinted genes and in particular estimates of variance components were also heavily dependent upon the number of sire and dam families used to estimate them. Results showed that VC analysis can be used to routinely detect genetic effects including imprinting and dominance in complex pedigrees. The work presented is the most extensive evaluation of the detection of non additive QTL using VC methods to date. Results challenge standard assumptions made about power and null distributions and show that optimal use of methodology is dependent on pedigree structure.

CHAPTER ONE

Introduction

1.1 Overview

Current advances in genomics applied to the livestock sector, have enabled progression away from the black box procedure of phenotypic selection into the realms of DNA based technologies. Studies in livestock species have the advantages of large numbers, controlled breeding programs and variation segregating within and between populations. Furthermore, the results often have direct statistical and biological application in human genetics (De Koning *et al*., 2007). The ability to isolate, and explore genetic variants underlying traits of economic importance is an exciting prospect. There is potential for greater understanding of the underlying biological mechanisms and the ability to apply selection with much greater accuracy.

New selection tools incorporating genomic information such as marker assisted and genomic selection (reviewed by Goddard and Hayes, 2007) must, however, be economically feasible and offer tangible benefits over phenotypic selection. Foreseeable drawbacks are that intense selection on specific variants might impede selection on other traits and that the complex interactions between genetic and environmental background could produce unpredictable results. The greatest barrier is likely to be the continuing success of phenotypic selection, thus marker based analysis is unlikely to be applied to traits, which are easily, cheaply and routinely measured. The ability to produce large families in a closely monitored commercial rearing environment enables the poultry industry to make rapid genetic progress (with conventional selection) utilising short generation times, intensive selection and low environmental variance. Scope for the application of genomic technologies lies in the identification of specific disease variants, or to select for livestock traits that are expensive to measure or only recordable in one sex. The poultry breeding structure offers good protection of any investments in MAS because the elite birds do not enter the open market. The sequencing of the chicken genome (Wong et al., 2004) and with it the discovery of over 2 million SNPs places the industry ideally to take advantage of, and to develop new technologies. This Chapter aims to review the methodology currently available for QTL mapping within general pedigrees.

1.2 QTL mapping.

The general concept of QTL mapping is based upon linkage and the violation of Mendel's principle of independent assortment. The basis of mapping functions is that alleles at loci in close proximity are less likely to recombine independently and the strength of linkage between two loci is a function of the distance between them. This concept of genetic distance is close but not equal to physical distance due to differing rates of recombination. If (co dominant) polymorphic DNA variants or 'markers' can be identified such that their inheritance pattern can be monitored they can be used to test for phenotypic associations with regions of the genome. QTL mapping is based on the likelihood of a marker being associated with a gene variant affecting the trait of interest. Power depends on the strength of linkage disequilibrium between the two and the ability to detect statistically the effect of QTL allelic substitution via the linked marker(s).

Statistical methods to determine the association of phenotype and marker genotype assuming a fixed QTL effect include regression or least squares (Knott et al., 1992; Lander et al., 1989; Weller 1986), maximum likelihood (Heath 1997; Jansen et al., 1998), and Bayesian analyses (Thaller et al., 2000; Uimari et al., 1997). Choice of method is largely dependent on the assumptions made and the parameters estimated. The main parameters of interest are means and variances of QTL genotypes and recombination fractions between markers and QTL. QTL mapping methods are reviewed by (Lynch and Walsh, 1998; Lui, 1997, Zheng *et al.*, 1994; Ott,1999; Weller, 2001) and (Haley et al., 1994; Knott et al., 1996; Knott et al., 1996; Lander et al., 1989; Weller et al., 2002)

1.2.1 QTL mapping in experimental populations

In experimental populations linkage disequilibrium can be created relatively easily. Crossing inbred lines with alternate alleles fixed at marker and QTL (unobserved) creates the maximum amount of linkage disequilibrium with each locus heterozygous and fixed for alternate alleles hence the pattern of inheritance and phase of multiple loci is known. Data can be treated as if from a single family because all individuals share the same parental genotype. The additive and dominant effects of the QTL can

be estimated directly. Designs based upon inbred lines therefore have the greatest power to detect QTL. They can use relatively straightforward statistical techniques such as regression and various designs using multiple markers to estimate the mean effect and location of one or multiple QTL. More complex algorithms such as Composite and multiple interval mapping (Zeng 2005) facilitate the study of multiple QTL and complex interactions between loci.

For line cross experiments in livestock or outbred plants, designs for inbred populations can be extended based on the assumption that the two lines although not inbred will be sufficiently different phenotypically to assume each line is fixed for alternative alleles. The analytical approach was developed by Haley *et al* (1994) and pioneered by Andersson *et al*., (1994)

1.2.2 QTL mapping within outbred pedigrees

In many livestock populations experimental crosses are impractical, unviable or uneconomical, and in many natural or human populations they are unfeasible. It is often more practical to explore QTL segregating within a population, particularly if it is to facilitate selection within that population. Most evolutionary important variation occurs within lines (Erickson, 2004) and despite intense selection there is evidence to suggest that there is still much variation that might be exploited within commercial populations. De Koning *et al*., (2004) found QTL previously identified in extreme crosses, explained considerable variation for production traits in a commercial poultry population. Andersson and Georges, (2004) describe as much variation between lines of white leghorn as present between white leghorn and its progenitor, red jungle fowl.

Mapping QTL within an existing or commercial population structure is challenging. The number of QTL alleles segregating in the population together with marker phase is often unknown, and together with allele frequencies must be estimated. Informativeness and therefore power to detect a QTL effect is dependent upon the amount of missing genotype information, heterozygosity of parents at markers and QTL and the fraction of offspring for which the inheritance of the markers is known.

Linkage equilibrium between markers and QTL may exist on a population level. If marker and QTL alleles are segregating independently, even if the marker and QTL

are linked within families or lines, the average effect of the marker segregation across the population will be zero. More recent linkage disequilibrium within families must therefore be utilized and QTL effects either estimated within parents, or phase known QTL genotypes inferred for each parent. Methods can be divided into two broad categories, those that divide the pedigree into sib-ships and estimate an effect within each parent or those that use all pedigree information simultaneously. The former tends to treat the QTL as a fixed effect and the latter as a random effect with variance to be estimated. These methods generally measure the variance of the QTL effect as a proportion of the within family variance (assumed to be constant across families) either by summing effects across families or by estimating variance components across the entire pedigree.

1.2.3 QTL mapping in Livestock

1.2.3.1 Sibship methods

Methods involve dividing a pedigree into, and estimating fixed effects within, sibships, and summing over all families. This is commonly used in livestock species where large full or, more commonly, half-sib families are available. Siblings are sorted according to the haplotype inherited from a common parent for a given map position and phenotypic means of the resulting groups are compared. Scaled, squared or nested ANOVA can be used to combine results into an estimate of the genetic variance associated with the inheritance of parental alleles (Lynch and Walsh, 1998). The Haley-Knott regression has been adapted from inbred lines for the analysis of full and half sib pedigree structures (Haley et al., 1994; Knott et al., 1992; Knott et al., 1996). It is implemented in the QTL Express software (Seaton et al., 2002) and described more fully in terms of a half-sib model in chapter 2. The method has been used extensively in livestock QTL mapping. Assumptions include unrelated sires and dams, with a single progeny per dam. Despite these assumptions being commonly violated in commercial pedigrees, the method has proven to be robust and computationally straightforward.

In dairy cattle, the daughter design, based on half sib contrasts has been extended to incorporate an extra generation. The resulting granddaughter design (Bovenhuis et al., 1997; Weller et al., 1990; Weller et al., 2002) uses phenotypes from a third generation

to estimate breeding values for the second generation. Using breeding values or average effects lowers the residual variance associated with the phenotype providing greater power to detect QTL effects. These within family methods are approximate as heterogeneity may occur due to families with larger or more QTL segregating, maternally inherited QTL alleles, or the polygenic effects in the residual term

Advantages of sibship methods are the transparency and robustness of design. Drawbacks are that in most livestock pedigrees, there are other relationships and inbreeding loops across pedigrees, which are not taken into account. Half sib designs only analyse segregation in one parent thus exploiting only half of the information available. In natural or human populations where large sib ships are generally uncommon, mapping methods using all genetic relationships within complex pedigrees may be a more powerful option.

1.2.3.2 Variance Component methods

An alternative to parameterized likelihood functions is to construct likelihood functions using the variance components (VC) associated with a QTL (or linked group of QTL's) in a region of interest. This allows for arbitrary and potentially complex pedigrees.

Variance component methodology is based on the assumption that individuals that are similar phenotypically are more likely to share alleles identical by descent or IBD. A linear model is constructed to partition phenotypic variance into components derived from fixed effects, and random effects of additive polygenic, additive QTL and residual variance simultaneously. A variance/covariance matrix is constructed for the average relationship for the polygenic effects and the relationship at the putative QTL position inferred by markers. For any pair of individuals the genetic covariance between them is a function of $2\Theta_{ij}$ where Θ_{ij} is the coefficient of ancestry or the probability that an allele randomly drawn from individual i is IBD with an allele from individual j. This coefficient is obtained purely from pedigree data thus is an average across the genome. In the case of fitting a QTL, marker data are used to infer R_{ij}, the proportion of alleles that the two individuals actually share IBD at a chromosomal location. Variance component methodology and IBD estimation is reviewed further in Chapter 2.

The test for the presence of a QTL involves comparing the full model with a model that does not include a QTL effect using a likelihood ratio statistic. Advantages are that many alleles or allelic effects can be modelled, and relationships between families can be used to provide information, increasing power to detect QTL. The assumption that QTL effects are randomly distributed circumvents the estimation of QTL allele frequencies and is robust to violation (George, 2000). This provides a much less parameterised environment by assuming the QTL has an infinite number of alleles and that both the QTL and the polygenic variance are normally distributed. Statistical analysis of the subsequent linear model can be derived using regression, maximum likelihood, moments or Bayesian procedures;(Almasy et al., 1998; Bink et al., 1998; Hoeschele et al., 1997; Xu et al., 1995).

1.2.3.3 QTL mapping in human pedigrees

There are two broad approaches to QTL mapping in human populations within which either sampled pairs of relatives or large sets of relatives from extended or nuclear families are analysed. The most widely used non-parametric method for linkage analysis of quantitative traits is sib pair analysis based on the regression method of Haseman and Elston (1972). The squared difference between the trait values for a pair of relatives is regressed against the proportion of marker alleles IBD. A negative co-efficient reflects a tendency for individuals to be more similar with respect to the trait as they share a greater proportion of the alleles IBD thus implying linkage between trait and marker.

1.3 Development of variance component QTL mapping

1.3.1 Humans

Sib methods such as the Haseman Elston method were first developed to treat QTL as random effects based on a single marker with genetic variance and strength of linkage were confounded. Goldgar (1990) suggested a multipoint IBD method based on maximum likelihood, which was extended by Schork (1993) for several chromosomal regions and common environment effects. Fulker and Cardon (1994) extended this further to include sib pair interval mapping. Xu and Atchley (1995) showed that although regression was robust, a maximum likelihood procedure was more efficient

with greater power and flexibility due to its ability to take the distributional properties into account. Xu and Atchley furthermore developed extensions to Goldgars methods for interval mapping and used simulation to compare it with the regression methods described by Fulker and Cardon (1995). Xu and Atchley found ML more powerful and suggested that the use of the squared difference used in regression could not take advantage of the properties of the normal distribution therefore lost information. The regression method does have the advantage of speed. Fulker and Cherny (1996) showed variance components was a more powerful approach than regression or maximum likelihood and Almasy and Blangero (1998) extended the sib pair multi point mapping approach of Fulker *et al*., (1995)., to general relative pairs using a regression method to calculate IBD coefficients. Once calculated, these multipoint relative pair IBDs were utilised in VC linkage analysis, which considers the likelihood of the entire pedigree jointly using maximum likelihood. Table 1 shows the main software packages available for variance component mapping in human pedigrees.

Variance component literature for humans is reviewed in a special issue of Behaviour Genetics (vol 34 2 March 2000) and includes papers on association tests, haplotype analysis adjusting for covariates and non-normality. Almasy and Blangero, (2004) discuss incorporating individual SNPs into the model as covariates in positional candidate gene approaches.

Table 1.1 Software Packages available for Variance Component Mapping in Human pedigrees

Software	Method	reference
Solar	SOLAR allows for fairly flexible mean modeling and uses the regression method to calculate multipoint IBD matrices. ML for linear model	Almasy and Blangero (1998)
Merlin	Multipoint Engine for Rapid Likelihood Inference The program uses sparse binary trees to represent patterns of gene flow in general pedigrees useful for analyses using dense SNP maps.	(Abecasis et al., 2002)
QTDT	QTDT can be used for linkage and association analysis. It is an implementation of the Fulker et al (1999) family-based test of association using a variance components framework. QTDT will perform the Fulker et al (1999) test, for an arbitrary number of markers and alleles, in general pedigrees	(Abecasis et al., 2000)
Genehunter	uses HMM to estimate IBD matrices but does not handle the VCA as easily and flexibly as SOLAR and Merlin.	(Kruglyak et al., 1995; Pratt et al., 2000)
QTL Express	Loki to estimate IBD using gibbs sampler and ASREML for variance component estimates	(Seaton et al., 2002)
Mx	Based on Fulker case control	(Neale et al., 1997)

1.3.2 livestock

In livestock QTL mapping, random effects methodology is based around the variance component approach and builds upon the well-established animal model. Fernando and Grossman (1989) proposed an extension to include random QTL effects by inclusion of the covariance structure of the QTL inferred by marker information. Each individual with unknown ancestors is assumed to have two unique QTL alleles sampled from an infinite population. The probability of receiving a specific parental allele for a QTL linked to the marker will be a function of the progeny marker genotype and the recombination fraction between them. Based on these probabilities Fernando and Grossman demonstrated how a variance/covariance matrix constructed for the QTL gametic effects could be included in the animal model to test for linkage to putative positions on the genome. This polygenic and QTL model returns estimates of the additive genetic variance, the variance due to the QTL at the test location and the likelihood value of the solution. It can also be used to provide breeding values for the QTL for all individuals in a population.

Schork (1993) and van Arendonk (1994) showed that ML has no problem incorporating fixed effects under mixed methodology and postulated that for such a model VC could easily be estimated using restricted or residual maximum likelihood (REML) techniques. Patterson and Thompson (1971) developed a derivative free algorithm for REML, which maximizes the likelihood of error contrasts with respect to parameter estimates under assumption of joint multivariate normal distribution.

The Fernando and Grossman model included a single marker effect with no inbreeding and assumed completely informative markers. They discussed expansion to multiple markers but this did not happen until Cantet and Smith (1992), Goddard (1992) and Hoeschele, (1993), Van Arendonk (1994) (proposed methods to reduce the number of equations per animal. Cantet and Smith (1992) proposed the reduced animal model (RAM), which only models individuals who are parents. This method was used by Goddard to extend Fernando and Grossmans method to include many linked markers providing a single QTL was present within a marker bracket. In comparison to the allelic matrix used by Fernando and Grossman requiring the estimation of 2n +1 effects for each individual, Hoeschele (1993) proposed estimating the QTL at the animal level. The genotypic IBD matrix is the proportion of alleles

shared IBD and models the sum of the genotypic effects at the loci. This resulted in 2 linearly equivalent mixed models estimating QTL effects at the gametic or genotypic level. Hoeschele also discusses including QTL effects only for genotyped animals and their tie ancestors, and incorporating missing marker information. Van Arendonk proposed reducing the gametic matrix to an additive effects matrix, thus reducing the number of equations necessary to m+1 and furthermore summing these additive QTL effects to create one equation per animal. Van Arendonk (1994) used matrix partitioning (Tier and Solkner, 1994) to construct the relationship matrix and include multiple unlinked markers each associated with a QTL whilst accounting for inbreeding.

Van Arendonk (1994) showed using a half-sib design that the variance and position of QTL cannot be separated when using a single marker. Grignola *et al.*, (1996a) used REML within a simulated granddaughter design to develop models for estimating position and variance of a single QTL separately. They developed interval mapping initially using flanking markers with known linkage phases in sires and no relationships between sires. This was extended for relationships across families. Using the reduced animal model and accounting for missing data Grignola *et al.*, (1996a; Grignola et al., 1996b) went on to show that these methods were robust to the number of alleles at the QTL, regardless of whether relationships between sires and marker linkage phases of sires and ancestors were known fully. (Grignola et al., 1997) went on to look at linked QTL using variance component mapping.

Other methods such as Bayesian and ML have been used to estimate parameters in livestock pedigrees (Bink et al., 1998; Hoeschele et al., 1997; Uimari et al., 1997; Uimari et al., 1996). Unlike ML and Bayesian methods, the VC method does not require number of alleles to be specified and has been shown to be quite robust to number of alleles at a QTL (Almasy and Blangero, 1998; (Xu et al., 1995); (Hoeschele et al., 1997). A further advantage of variance component mapping over Bayesian methods is that the IBD coefficients only need calculating once and therefore with many analyses on the same pedigree is much less computationally demanding. Ronnegard et al., (2008) increase computational efficiency using score statistics using flexible intercross analysis for the detection of QTL effects between and within lines.

Problems associated with ML and regression methods are accounting for more complex structures associated with several families, relationships across families, unknown linkage phases in parents, no of QTL in the population, and varying amounts of data on different QTL or in different families. REML handles any population structure, incorporates fixed effects easily and is robust to deviations from normality (Patterson and Thompson, 1971, Gilmour, 1995, George, 2000).

George *et al*., (2000) describe a two-step process to map QTL based on estimating IBD coefficients with MCMC, and variance components within ASREML. Loki handles large pedigrees using an alternative sampling strategy developed by Thompson and Heath (1997) to calculate identical by descent (IBD) scores for the pedigree at each position, simultaneously estimating missing marker data and unknown haplotype information. Georges at al., (2000) used simulation to assess performance over unknown marker genotypes, inbred individuals, partially or known marker phases and multigenerational data. The difference between the method put forward by George *et al* and earlier methods calculating IBD probabilities (Fernando and Grossman, 1989; van Arendonk, 1994; and Wang, 1995) is the use of a Gibbs sampler. George *et al*., (2000) showed their two-step approach to be capable of detecting QTL in simulated pig and sheep pedigrees varying in structure and in completeness of genotypic information.

To date, applications of variance component mapping to real data are limited. Most use the approach described by George *et al* (2000) and are reviewed in a further section. Examples of the use of this method in human pedigrees include analysis of neurological disorders, obesity, arthritis, alcoholism and bipolar disorder (Dong et al., 2005; Nicholls 2000; Visscher et al., 1999b; Zhou et al., 2007)

1.3.3 Linkage disequilibrium fine mapping

The VC framework can also be extended to include linkage disequilibrium information in order to fine map QTL. LD fine mapping methods assume that LD is primarily due to the introduction of a variant on an ancestral haplotype via mutation (or migration) which is partially preserved in descendents of the current generation

(Hoeschele, 2001). In livestock populations a high level of haplotype sharing reflects long range LD. If this is converted into an IBD probability between two haplotypes conditional on flanking marker data it would seem quite straightforward to include LD information in the REML/random effects framework to fine map QTL (Lee et al., 2004; 2002; Meuwissen et al., 2001; 2000). Combined linkage disequilibrium and linkage can also be used in a variance components framework to apply genomic selection (Calus et al., 2008; de Roos et al., 2007; Goddard et al., 2007).

1.4 Aspects of variance component mapping

As discussed by George *et al*., (2000) the methodology can be split into a two-stage process. Firstly, IBD probabilities and the subsequent allelic or genotypic matrices can be calculated recursively or estimated using correlation and simulation based techniques, depending on the level of information available. IBD probabilities tend to be estimated at the allelic level in animals and the genotypic level in humans. Furthermore IBD probabilities can be estimated at marker locations only or in the case of multipoint mapping they are estimated at set points along the linkage map. The second step involves using the variance/(co) variance matrix to estimate QTL effects and breeding values based on the premise that individuals sharing more alleles identical by descent will be more alike phenotypically. Variations in the model and the fitting of effects on the gametic or animal level can be made in order to model dominance and epistasis. Finally, the model must be tested typically using a likelihood ratio statistic comparing the likelihoods of the models either fitting a QTL or not.

1.5 IBD Probabilities

1.5.1 Recursive algorithms

IBD probabilities can be computed recursively from a chronologically ordered pedigree as described by van Arendonk, (1994), Wang *et al*., (1995), Pong-Wong *et al*., (2001), Liu *et al*., (2002) and Nagamine *et al*., (2004). A major problem faced by recursive algorithms is the inability to handle large amounts of missing information. These algorithms follow a top down strategy so that missing information in individuals early in the pedigree introduces estimation errors throughout the pedigree

due to inability to utilise information that is not otherwise passed down through the parents.

1.5.2 Correlation-based algorithms

Almasy and Blangero (1998) use the IBD correlation relationships of Amos (1994) between the proportion of alleles shared IBD at the fully genotyped marker and a putative QTL. The regression model was suggested by Goldgar, extended to two marker interval mapping by Fulker and Cardon 1994, for considering any number of markers proposed by Fulker 1995 for sib pairs and extended to include general pedigrees (Almasy and Blangero, 1998). Almasy and Blangero use the averaging method of Fulker *et al*., 1995 to extend this equation to allow the calculation of the gametic IBD matrix to be conditional on all of the marker information. The coefficients become increasingly difficult to estimate in a complex pedigree and with missing data.

1.5.3 Simulation based algorithms

For pedigrees with incomplete marker information direct application of recursive or correlation based IBD algorithms is unfeasible. A solution is to use MCMC approximation to calculate the expectation of the IBD matrix. MCMC methods have the advantage of coping with complex pedigree structure and missing information but are computationally intensive and slow to use. They also require certain expertise to assess convergence diagnostics often not easily apparent in software packages such as LOKI.

Pong Wong *et al* (2001) get around the problem of missing information by using the first phase known flanking marker bracket in a recursive method. With the advent of denser maps utilising SNP technology this could be an effective alternative to MCMC methods and has been shown by Sorenson (2002) and De Koning *et al*., (2003) to perform well in both simulated and commercial data. Pong Wong's method differs in that it is able to estimate IBD coefficients for all relatives. Pong Wong's method builds on the recursive method described by Wang for calculating IBD at single locus combining it with a method to estimate IBD within sibs using multiple markers. Complicated calculation of haplotype probabilities are avoided by using the closest informative marker bracket with absolute certainty. Pong Wong partially reconstructs

haplotype phases and then recursively calculates IBD from oldest to youngest. Although this method fails to use information from marker brackets with incomplete information it is a fast approximation well suited to take advantage of the increasing number of markers and genome coverage available in livestock species.

1.5.4 Model fitting in VC analysis - Imprinting, dominance, and epistasis

Calculation of the gametic and genotypic relationship matrices facilitates the genetic dissection of complex traits. Linear mixed models assuming zero covariance between polygenic and QTL effects and between QTL can be extended to include IBD matrices modelling underlying genotypic effects such as imprinting, dominance and epistasis.

The gametic relationship matrix has been used to construct a relationship matrix due to dominance effects (Schaeffer et al., 1989) and for the analysis of gametic imprinting effects (Gibson *et al*., 1998). Genomic imprinting is defined by Hanson *et al*., (2001) as ‘where the genomic segment inherited from one parent is inactivated such that the expression of an allele in one of these regions is dependent upon the sex of the parent from whom it was inherited’. Examples include Prader-Willi syndrome (Nicholls 2000b), and Beckwith-Wiedemann syndrome (Shete et al., 2007) in humans, Callipyge in sheep (Charlier et al., 2001) and Igf2 in pigs (Nezer et al., 1999). Morison *et al* 2001 have compiled an imprinted gene database that contains more than 200 imprinted genes in humans and other organisms.

Hanson *et al*., (2001) discuss parent of origin effects in linkage analysis of quantitative traits using the gametic relationship matrix where the estimated proportion of marker alleles shared IBD is partitioned into paternal and maternal components. Parent specific estimates of allele sharing can then be used in VC or Haseman Elston methods of linkage analysis so that the effect of the QTL on the maternally derived chromosome is potentially different from that of the paternally derived chromosome. If markers are not informative IBD coefficients need to be estimated and it is a straightforward extension to employ separate recombination

fractions for males and females reported to be on average 1.6 greater for females than males in mammals (Broman et al., 1998)

Under simulation, power to detect an imprinted QTL using VC analysis was significantly increased when modelling separate parental contributions (Hanson *et al.*, 2001). The authors note that the splitting of the IBD matrix is expected to inflate type 1 error rate due to multiple testing unless a correction such as the Bonferroni is applied (Ott, 1991). This was not found with data incorporating variable family sizes and missing data suggesting that these factors do not produce substantial inflation. Hanson *et al* conclude that the VC method was more powerful than sib pair based methods both for imprinting and linkage effects as noted by Pratt, (2000).

Shete and Amos, (2002) proceeded with Hanson's method to formalise the model and provide non-centrality parameters that can be used to determine sample sizes to attain specified power for a given significance level. Shete and Amos decompose the total additive genetic variance into parent specific additive genetic variances and the dominance variance. The variance components approach was developed from sibship analysis to imprinting in extended pedigrees by the incorporation of an extra parameter to benefit from information from relationships such as double first cousins (Shete et al., 2003).

Epistasis is widely reported to play an important role in genetic variation (Carlborg et al., 2004a; Carlborg et al., 2004b; Kerje et al., 2003) and could play a key role in marker assisted selection as shown by introgression studies in maize and mice (Shimomura et al., 2001). Liu (2002) outlines an algorithm for computing the conditional covariance between relatives given genetic markers. Procedures are described for additive, dominance, additive by additive, additive by dominance, dominance by additive and dominance by dominance conditional relationship coefficients. Purcell and Sham (2004) review the inclusion of epistasis within a variance components framework concluding that power to detect these interactions is low as apparent variance components in sub models 'soak up' a large proportion of the variance due to epistatic effects in the main model.

A new software program R'Tools', has been developed by Ricardo Pong Wong. Pong Wong's deterministic method is used to incorporate epistatic and dominance effects in

the manner described by Liu. In this way R'Tools calculates the IBD matrices necessary to estimate dominance and epistatic effects.

1.6 Statistical analysis of the linear mixed model

Once a linear model has been constructed and parameters estimated, a test statistic indicating the presence of a QTL can be obtained, from which size and position can be determined. Hypothesis testing for a ML approach is essentially performed by comparing the likelihood of the data under the alternative hypothesis (H1) postulating the presence of QTL at the examined location with the likelihood of the data under the null hypothesis of no QTL at that position. The latter is computed using a reduced model without the haplotype or genotype effects.

A likelihood ratio statistic can then be obtained by twice the difference of the Log likelihoods from the two models. For a single point test, under the null hypothesis of no QTL the test statistic follows a 50:50 mixture distribution where one component is a point of mass 0 and the other is a χ^2_1 distribution. For testing an entire chromosome the LR can be approximated by a χ^2_1 distribution (Allison et al., 1999; Self et al., 1987; Williams et al., 1999b). The empirical distribution of the test statistic under the null hypothesis has been empirically found to follow a chi-square distributed between 1 and 2 df for chromosome wide tests by others (Hoescele, 1997; Xu and Atcley 1995; Grignola, 1996).

To account for the one sided nature of the test the *P* value for the LR test statistic is typically calculated by dividing the corresponding *P* value by 2. Such *P* values are valid providing that the assumption of multivariate normality is not violated. LR can also be converted to a LOD score by dividing the LR by $2\log_e(10)$ (~4.65). Simulations suggest that for sample sizes typical of linkage studies QTL need to have moderate effects accounting for at least 20-30% of the phenotypic variance in order to have reasonable power to detect a LOD greater than 3 (Hanson, 2001)

Using simulated data, George *et al.*, (2000) found the empirical distribution of H0 for a single test to agree with the theoretical 50:50 mixture and that this distribution was relatively unaffected by population structure. For testing an entire chromosome they found that the empirical distribution was more likely to follow χ^2_1 distribution

although this was more conservative at the 5% threshold than the empirical. They suggest applying the Bonferroni correction or permutation analysis. However, because each analysis took approximately one hour the anticipated computational demands are high.

Proponents of the regression method claim that VC relies heavily on normality assumptions. Severe departures from multivariate normality e.g. kurtosis have been shown to inflate type 1 error rates (Allison 1999). Blangero *et al.*, 2000 suggest overcoming this by adjusting the likelihood ratio by a correction factor based on heritability and kurtosis in order to provide a more robust LOD score. Shete *et al.*, (2004) found that Winsorization of non-normal data increased power but did not greatly diminish type I error. Zeegers *et al.*, (2004) have suggested that power could be increased by the use of covariates and explore various methods of doing so. The use of permutation tests to control type I error are an attractive but relatively unexplored option.

1.7 Application of VC to livestock data

The two step method described by George *et al.*, (2000) has subsequently been used to map a locus influencing bipolar disorder in a complex human pedigree (Visscher *et al.* 1999), growth and carcass traits in pigs (De Koning *et al* 2003, Nagamine *et al.*, 2004) and birth weight in red deer (Slate et al., 2002). Nagamine, de Koning and Slate all use VC methods after identifying candidate regions with a Least Squares approach.

De Koning *et al.*, (2003) used 10 pig lines and 10 chromosomal regions previously analysed by Evans *et al.,* (2004; De Koning *et al.*, 2003; 2003) and Knott *et al.*, (1996) to compare VC and half sib methods. Nagamine *et al.*, (2004) showed that QTL for growth traits and back fat were segregating within several commercial pig populations and Slate *et al.*, (2002) used Georges two step approach of Loki and REML to compare half sib and VC analysis for putative birth weight QTL on three separate linkage groups in a wild population of red deer.

All three studies found that IBD coefficients were similar regardless of method used. De Koning *et al.* found Pong Wongs deterministic method for IBD estimation was 9

times faster than LOKI using 10,000 iterations although the number of iteration was higher than recommended (10 times number of individuals in pedigree). Convergence diagnostics were not analysed. Similarly, Nagamine *et al.* (2002) (compared results using their simple deterministic method (SMD) with LOKI and found that SMD and MCMC had highly correlated test statistics of 0.95 and least squares analysis correlation with SMD and MCMC was 0.70 and 0.71 respectively. Slate *et al.*, (2002) used SOLAR 1.7.3 (Almasy and Blangero, 1998) to perform QTL analysis by VC. Although SOLAR uses a different algorithm and was able to calculate only single IBD coefficients at marker locations rather than multipoint IBD coefficients LOKI and SOLAR provided similar IBD estimates at the marker locations. SOLAR also provided the same maximum-likelihood solutions (yielding a test statistic of zero) as the REML software, even when handling IBD coefficients derived from LOKI. Both LOKI and SOLAR were subsequently used to conduct a VC analysis within the half-sib ships where the linear regression approach had found evidence for segregating QTL. The VC methods found evidence (sometimes highly significant) for segregating QTL within these families, but generally with higher *P* values (i.e., less significant) than those obtained by linear regression.

Nagamine *et al* reported a large range of genotypic values across sires but more importantly in terms of marker assisted selection large differences between allelic values within sires, verified by large t-statistics within the respective sire families. High QTL heritabilities could have been due to selective genotyping and although optimistic in terms of possibility of marker assisted selection authors felt that a more complex model should be fitted such as one including dominance effects.

Discrepancies between QTL detection using least squares and variance component analysis have been reported by multiple studies. (De Koning *et al.*, 2003; Nagamine *et al.*, 2004; Slate *et al.*, 2002; Zhang *et al.*, 1998). Differences range from the QTL being almost significant with one method and significant with another to completely undetected with one and very significant with another. In general least squares detected a greater number of QTL than VC. Slate *et al.*, postulate that reduced power due to missing marker data may mean that the VC method simply failed to detect a genuine QTL i.e. type II error. De Koning *et al*, report that QTL significant at the 5% level tend to be detected by both methods and that the comparison was robust to the

choice of threshold suggesting differences amounted to more than merely type 1 error. Where QTL were segregating in the other parent in half sib analyses VC tended to be more powerful.

To date applications of VC in livestock have been used as a comparison with or to verify results from more traditional methods. As methods have only been compared within structured sibships, perhaps it is unsurprising that sibship methods were more powerful. Intuitively, the VC method might be expected to have greater power than the linear regression approach as more phenotypic records are used. The different assumptions underlying the linear regression and VC methods could be responsible for some inconsistencies. The HS model estimates an allele substitution fixed effect for every sire, with maternally inherited alleles assumed to be randomly distributed and used only to improve sire information. VC is a variance estimate across the entire population accounting for both maternal and paternal alleles assuming segregation in both. The VC method assumes that QTL effects are additive and could be confounded by maternal effects or QTL acting in a non additive fashion (e.g., dominance). For 23 out of 46 QTL detected only by HS in the De Koning analysis only a single sire was heterozygous and for 5 others only 2 sires were heterozygous. Differences in allele frequencies in males and females could be sampling, selection, or imprinting. HS methods missed QTL when none of the sires were segregating, for example, the halothane mutation in Spanish Pietrain was detected by VC as none of the 5 sires tested were heterozygous in comparison to 13 out of 60 dams that were heterozygous. QTL segregating at low frequency could be missed by VC as power of detection depends on variance explained by QTL across the population.

1.8 Conclusions

Large enough QTL effects have been detected within commercial populations to make the implementation of MAS potentially worthwhile. Estimates of gene substitution or allelic effects and or QTL variance are important for MAS. Implementations of MAS for selection of young dairy sires before progeny testing and for selection in nucleus breeding schemes have been shown to potentially produce genetic and economic gains (Mackinnon et al., 1998; Meuwissen et al., 1997; Van Arendonk et al., 1994). The past decade has seen huge advances in the mapping and

identification of quantitative trait loci (QTL) in both experimental crosses and commercial populations of livestock. In order to successfully implement marker assisted selection (MAS) in future breeding programs it is important to identify QTLs segregating within lines as well as understand the interactions between the QTL that are being selected for and other genes. There is a need to model underlying effects and opportunity to dissect complex quantitative variation. It is important to understand why theses alleles are segregating as there could be pleiotropic effects upon life traits.

Methods that fully account for all relationships are expected to provide greater power to detect QTL (Almasy and Blangero 1998). This has been the case in human pedigrees (Visscher et al., 1999; Williams et al., 1999) although is less clear cut within structured livestock populations (Slate *et al*., 2002, George *et al*., 2000, De Koning *et al* 2003). Variance component mapping does provide a less parameterised statistical environment than traditional fixed effects models and can be used where sib structures are not available in human, natural, and commercial livestock populations. VC has been shown to be robust to number of alleles or QTL, and can estimate genotypic and allelic effects across an entire population. The methodology is less computationally intense than extending maximum likelihood and Bayesian methods. The method is based on that currently used within livestock production to identify individuals with greatest genetic merit and select them as parents for the next generation. It is, therefore, ideally suited for current livestock systems as despite high computational demands, much of the technology is already in place.
MCMC iterative methods are available and efficient for estimating IBD coefficients with varying levels of missing genotypic and phase estimation. With the advent of SNP technology and cheaper genotyping, however, deterministic methods of IBD estimation look promising for fast efficient calculation of relationship matrices. REML has been shown to be efficient and robust in the estimation of variance components with software such as ASREML able to handle as many as six user defined matrices there is scope to extend models to incorporate higher order effects such as imprinting, dominance and epistasis.

Questions of how missing marker genotypes, unknown marker phase, pedigree-size, map density and QTL size influence the distribution of the test statistic remain

unanswered. Permutation and bootstrapping methods employed within regression analyses seem at present too time consuming and computationally intensive to provide a feasible option. It remains that there is a clear need for more testing of the methodology within various population structures and whether there is enough power to detect interactions within and between loci in existing commercial population structures is not yet clear.

CHAPTER TWO

Theory of variance component mapping

2.0 Introduction

The following chapter gives an account of the more general methods throughout the thesis underpinning variance component analysis. Relationships between relatives must be estimated conditional on marker information. The linear model is defined to indicate sources of variation. The mixed model equations are then constructed and solved by the partitioning the phenotypic variance proportional to the genetic covariances estimated in the relationship matrices. Likelihoods are estimates using residual maximum likelihood in the ASreml software package (Gilmour *et al*., 1995).

2.1 IBD Estimation

2.1.1 The numerator relationship matrix

The probability of identical genes by descent occurring in two individuals is termed the coefficient of kinship thus ϕ_{ij} is the probability that a random allele from i is IBD with a random allele from j. The additive genetic relationship between two individuals is twice the coefficient of kinship.

The matrix indicating additive genetic relationships amongst individuals is termed the numerator or A matrix. A is symmetric with diagonal element for animal i (a_{ii}) equal to $1 + F_i$, where F_i is the inbreeding coefficient of animal i (Wright, 1922). The diagonal element represents twice the probability that two gametes taken at random from animal i will carry identical alleles by descent. The off diagonal a_{ij} is twice the probability that a random allele from i is IBD with a random allele from j so $a_{ij} = 2\phi_{ij.}$ and $A = 2\Phi$.

Elements of the A matrix are the expected proportion of alleles IBD in the genome between two individuals for example full sibs 0.5, half sibs 0.25, first cousins 0.125. When multiplied by the additive genetic variance in the population (σ_u^2), $A\sigma_u^2$ is the covariance amongst breeding values (Mrode, 2005).

The inverse of the A matrix (A^{-1}) is used to solve the animal model. There are fast algorithms to calculate A and its inverse (Henderson 1975; Quaas 1976).

Table 2.1 gives Hendersons method to obtain the inverse of the A matrix, which can be derived directly from pedigree information, by adding the elements of A^{-1} in sequence using the following rules.

Table 2.1 Henderson's rules for estimating the inverse of the A matrix

Animal	Sire	Dam	Add to element (and its transpose) in A^{-1}					
(a)	(s)	(d)	(a,a)	(a.s)	(a,d)	(s,s)	(d,d)	(s,d)
	Known	Known	2	-1	-1	1/2	1/2	½
	Known	-	4/3	-2/3		1/3		
	-	Known	4/3		-2/3		1/3	
	-	-	1					

(Cameron, 1997)

Best linear unbiased prediction or BLUP is a method developed by Henderson (1949) by which fixed effects and breeding values can be simultaneously estimated. Measurements on relatives of an individual are comparable with repeated measures of the individual and can therefore contribute information to that animals breeding value once account has been taken of the genetic relationship between the individual and its relatives. The covariance between relatives is r σ^2_A where r is the relationship coefficient and σ^2_A is the additive genetic variance of the population.

2.1.2 IBD estimation for the QTL

The QTL effect associated with each individual is considered as a random effect with a covariance structure proportional to the IBD probability at the QTL position. The predominant method used throughout the thesis was that of Pong-Wong *et al*., (2001) implemented in the R'Tools software package. The method is based on a recursive method of Wang *et al*., (1995) where the IBD probability between the gamete of an individual and ancestral gametes is function of the probability of descent and the IBD probability of the ancestor's gamete with the two gametes of the parent. In the previous method (Wang *et al*., 1995), if the true haplotype phases for the nearest

informative marker bracket are uncertain IBD then estimation should be integrated over all possible phases across the whole population. With the R'Tools method the haplotype phase is estimated using the nearest marker bracket that can be used with absolute certainty.

From Pong-Wong et al., (2001) the method assumes markers with n loci at known positions with recombination as expected from the Haldane mapping function. Similar to the numerator matrix described above, the Gametic IBD matrix (**G**) is a matrix containing IBD probabilities between the two gametes and the probability of these alleles being the same gamete originating from a common ancestor in the base population. The IBD probability between the gamete of individual i inherited from parent x (A_i^x) and the gamete of an ancestor j inherited from parent y (A_j^y) conditional on the linked marker genotypes (**M**) is equal to

$$P\,(A_i^x \equiv A_j^y \mid \mathbf{M}) = P\,(A_j^y \equiv A_x^p \mid \mathbf{M}) * PDQ\,(A_i^x \leftarrow A_x^p \mid \mathbf{M}) + P\,(A_j^y \equiv A_x^m \mid \mathbf{M}) * PDQ\,(A_i^x \leftarrow A_x^m \mid \mathbf{M})$$

Where ≡ stands for the identity between alleles, and $P\,(A_j^y \equiv A_x^p \mid \mathbf{M})$ and $P\,(A_j^y \equiv A_x^m \mid \mathbf{M})$ are the IBD probabilities between gamete A_j^y and the paternal (A_x^p) and maternal (A_x^m) gametes of parent x respectively.

$PDQ\,(A_i^x \leftarrow A_x^p \mid \mathbf{M})$ and $PDQ\,(A_i^x \leftarrow A_x^m \mid \mathbf{M})$ are the probabilities of gamete A_i^x of individual *i* being the same as gamete A_x^p or A_x^m of parent x.

The *PDQ* of the gamete is the probability that the gamete of an individual inherited from one of its parents is either the parent's paternal or maternal gamete. When the parent is not inbred the *PDQ* is the same as the IBD between the individual's gamete and the parents gamete. The probability of descent is calculated conditional on the closest marker genotype of the individual and its parents. Probability given inheritance of markers is given in table 2.2.

Table 2.2. Probability of descent of QTL allele given parental inheritance of flanking markers

Marker Descent			
M1	M2	$PDQ\ (A_i^x \leftarrow A_x^p \mid \mathbf{M})$	$PDQ\ (A_i^x \leftarrow A_x^m \mid \mathbf{M})$
P	P	$(1-\theta_1)\ (1-\theta_2)\ /\ (1-\theta)$	$(\theta_1\theta_2\ /\ (1-\theta)$
P	M	$(1-\theta_1)\ \theta_2\ /\ (1-\theta)$	$\theta_1(1-\theta_2)\ /\ \theta$
M	P	$\theta_1\ (1-\theta_2)\ /\ \theta$	$(1-\theta_1)\ \theta_2\ /\ \theta$
M	M	$\theta_1\theta_2/\ (1-\theta)$	$(1-\theta_1)\ (1-\theta_2)\ /\ (1-\theta)$
P	-	$(1-\theta_1)$	θ_1
M	-	θ_1	$(1-\theta_1)$
-	P	$(1-\theta_2)$	θ_2
-	M	θ_2	$(1-\theta_2)$
-	-	0.5	0.5

θ_1, θ_2, θ recombination rate between the first marker and the QTL, the second marker and the QTL and the two markers, respectively (assuming the haldane mapping function.

P the individual inherited the paternal allele from the parent

M the individual inherited the maternal allele from the parent

When a marker genotype is missing for a given individual the method does not attempt to reconstruct the genotype (unless the genotype can be inferred with absolute certainty given its parents' and offspring's genotypes). The missing genotype is said to be uninformative and the next marker locus is used.

To estimate the IBD between sibs whose common parent is a base individual for example in a two-generation pedigree, the method is combined with that proposed by Knott and Haley (1998) given in table 2.3.

Table 2.3. Probability of IBD at QTL between sibs given IBD state at flanking markers [a])

IBD state at flanking markers		IBD[b]
M1	M2	
1[c]	1	$((1\text{-}\theta_1)^2 + \theta_1^2)\,((1\text{-}\theta_2)^2 + \theta_2^2\,)\,/\,((1\text{-}\theta)^2 + \theta^2\,)$
1	0	$((1\text{-}\theta_1)^2 + \theta_1^2)\,((1\text{-}\theta_2)\theta_2\,)\,/\,((1\text{-}\theta)\theta\,)$
0	1	$((1\text{-}\theta_1)\theta_1)\,((1\text{-}\theta_2)^2 + \theta_2^2\,)\,/\,((1\text{-}\theta)\,\theta\,)$
0	0	$4((1\text{-}\theta_1)\theta_1)\,((1\text{-}\theta_2)\,\theta_2\,)\,/\,((1\text{-}\theta)^2 + \theta^2\,)$
1	-[d]	$((1\text{-}\theta_1)^2 + \theta_1^2)$
0	-	$2((1\text{-}\theta_1)\,\theta_1)$

a) from Knott and Haley, (1998)

b) Formula is the IBD probability assuming that the common parent is non inbred

c) Both sibs inherited the same/different marker from the parent

d) No informative marker found (the parent is a homozygote or inheritance in sibs is unknown)

2.1.3 Protocol

The protocol therefore for the estimation of the gametic IBD relationship matrix using the method of Pong-Wong is as follows.

1) Marker haplotype phases for all possible markers are reconstructed given the individual and the individual's parents' marker genotypes.

2) IBD is calculated recursively starting from gametes with the oldest ancestors to the youngest descendents assuming that

a) The diagonal of the IBD matrix is always 1

b) If the individual is from the base population its IBD between its gametes and its ancestors is 0

c) If the individual is not from the base population the probability of descent for each gamete is calculated using the closest informative marker bracket with a known haplotype phase using (1) to calculate the IBD probability between parental gametes and gametes of previous ancestors, and if IBD probability is to be calculated

between two gametes that originated from a common parent (i.e. sibs) using the formulae given by Knott and Haley (Table 2.3) including offspring of the base animals for which *PDQ* cannot be estimated.

2.1.4 Genetic covariances between relatives conditional on genetic markers

Liu *et al*., (2002) define genetic covariances of additive and non-additive effects and show how relationship matrices for additive and non additive effects can be simply derived from the gametic IBD matrix.

The genetic covariance of two individuals i and j at a single locus is

$$\mathrm{Cov}\,(g_i,g_j) = r_{ij}\sigma^2_A + u_{ij}\sigma^2_D$$

Where σ^2_A and σ^2_D are the additive and dominance variance associated with the locus and g is the genotypic value. The relationship coefficients of r and u depend on the conditional probability of QTL allelic identities between individuals i and j.

These allelic identities can be estimated at putative QTL positions based on observed flanking marker genotypes. Q_i^{l1} and Q_i^{l2} denote homologous alleles 1 and 2 of individual i at the *l*th QTL locus given *M* marker information. The symbol ≡ stands for the identity between alleles.

$$r_{ij}^{l} = \tfrac{1}{2}\begin{bmatrix} \Pr(Q_i^{l1} \equiv Q_j^{l1} \mid M) + \\ \Pr(Q_i^{l1} \equiv Q_j^{l2} \mid M) + \\ \Pr(Q_i^{l2} \equiv Q_j^{l1} \mid M) + \\ \Pr(Q_i^{l2} \equiv Q_j^{l2} \mid M) \end{bmatrix} \qquad u_{ij}^{l} = \begin{matrix} \Pr(Q_i^{l1} \equiv Q_j^{l1} \mid M) * \Pr(Q_i^{l2} \equiv Q_j^{l1} \mid M) + \\ \Pr(Q_i^{l1} \equiv Q_j^{l2} \mid M) * \Pr(Q_i^{l2} \equiv Q_j^{l2} \mid M) \end{matrix}$$

Therefore from the gametic IBD matrix

$$G_{ij} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

The additive covariance between i and j is $r_{ij} = 1/2\,(P_{11} + P_{12} + P_{21} + P_{22})$

and the covariance due to dominance i.e. the inheritance of two alleles identical by descent is $u_{ij} = P_{11}P_{22} + P_{12}P_{21}$

2.2 Calculating the gametic IBD matrix and relationship matrices G and D needed for the mixed model analysis

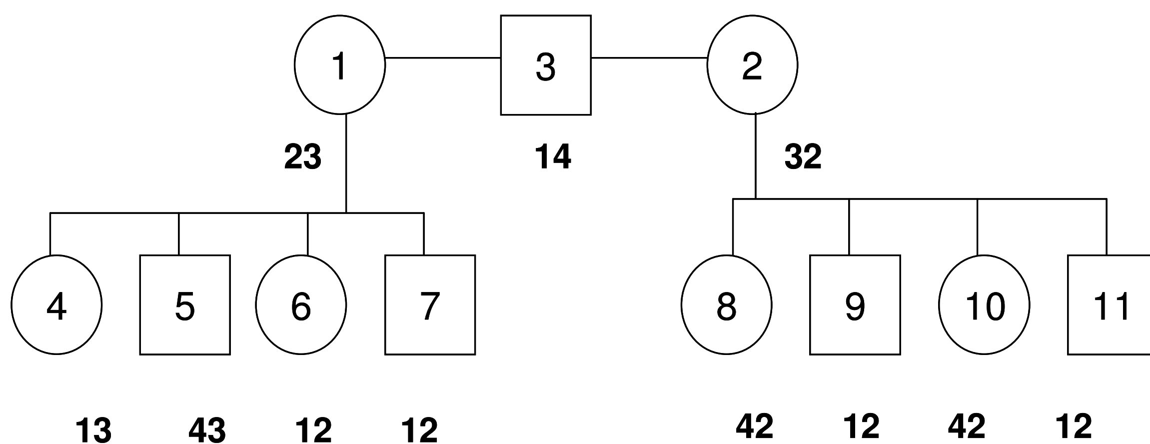


Figure 2.1 An example pedigree for a single sire (3) mated to two dams (1 & 2) each with four offspring, creating a nested full sib/half sib family structure. Figures in bold denote marker genotypes for each individual. The first allele is paternally derived and second is maternally derived i.e. individual 5 inherited marker allele 4 from its father and marker allele 3 from its mother.

A matrix

Individual	1	2	3	4	5	6	7	8	9	10	11
1	1										
2	0	1									
3	0	0	1								
4	0.5	0.5	0	1							
5	0.5	0.5	0	0.5	1						
6	0.5	0.5	0	0.5	0.5	1					
7	0.5	0.5	0	0.5	0.5	0.5	1				
8	0	0.5	0.5	0.25	0.25	0.25	0.25	1			
9	0	0.5	0.5	0.25	0.25	0.25	0.5	0.5	1		
10	0	0.5	0.5	0.25	0.25	0.25	0.5	0.5	0.5	1	
11	0	0.5	0.5	0.25	0.25	0.25	0.5	0.5	0.5	0.5	1

The A matrix is derived purely from the pedigree information i.e. ignoring information from markers and is based on the expectation or average degree or relatedness between relatives over all loci.

Gametic IBD Matrix from marker information

		4		5		6		7		8		9		10		11	
		P	M	P	M	P	M	P	M	P	M	P	M	P	M	P	M
	P	1															
4	M		1														
	P	0.08	0	1													
5	M	0	0.99		1												
	P	0.91	0	0.08	0	1											
6	M	0	0.001	0	0.001		1										
	P	0.91	0	0.08	0	0.91	0	1									
7	M	0	0.001	0	0.001	0	0.99		1								
	P	0.08	0	0.91	0	0.08	0	0.08	0	1							
8	M	0	0	0	0	0	0	0	0		1						
	P	0.91	0	0.08	0	0.91	0	0.91	0	0.08	0	1					
9	M	0	0	0	0	0	0	0	0	0	0.99		1				
	P	0.08	0	0.91	0	0.08	0	0.08	0	0.91	0	0.08	0	1			
10	M	0	0	0	0	0	0	0	0	0	0.99	0	0.99		1		
	P	0.91	0	0.08	0	0.91	0	0.91	0	0.08	0	0.91	0	0.08	0	1	
11	M	0	0	0	0	0	0	0	0	0	0.99	0	0.99	0	0.99	0	1

Where P and M denote paternal and maternal allele respectively.

For the Gametic IBD matrix IBD probabilities are estimates of IBD between individuals at the position of a putative QTL based on information from the linked marker genotypes shown in the pedigree and taking into account the probability of recombination between marker and QTL.

Table 2.4 Calculating additive, dominant, maternal and paternal relationship coefficients from the Gametic IBD matrix

Genetic covariance	**(4,5)**	**(6,8)**	**(8,10)**
A matrix	0.5	0.25	0.5
Additive **G**	0.5*(0.08+0.0+0.0+0.99) **= 0.54**	0.5*(0.08+0+0+0) **= 0.04**	0.5*(0.91+0+0+0.99) **= 0.95**
Dominant **D**	0.08*0.99 + 0*0 **= 0.08**	0.08*0 + 0*0 **= 0**	0.91*0.99 + 0 * 0 **= 0.91**
Paternal $\mathbf{G_P}$	**0.08**	**0.08**	**0.91**
Maternal $\mathbf{G_M}$	**0.99**	**0**	**0.99**

G, $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** are the appropriate relationship matrices used to model the additive, maternal, paternal and dominant QTL effects at each position tested. It can be shown that these relationship matrices are easily estimated from the gametic IBD matrix (Table 2.4). Here it can be seen that although the expectation of IBD between individuals 6 and 8 is 0.25, in fact they do not share any marker alleles in common and therefore the only probability of relatedness is based on the probability of 0.04 of

recombination between QTL and marker genotype. The Dominance coefficient is based on the probability of two individuals sharing both alleles identical by descent i.e. in this example where full sibs inherit both the same paternal and maternal alleles as seen for individuals 8 and 10.

The full additive and dominance matrices are given below.

Additive matrix G

	4	5	6	7	8	9	10	11
4	1							
5	0.54	1						
6	0.45	0.04	1					
7	0.45	0.04	0.95	1				
8	0.04	0.45	0.04	0.04	1			
9	0.45	0.04	0.45	0.45	0.54	1		
10	0.04	0.45	0.04	0.04	0.95	0.54	1	
11	0.45	0.04	0.45	0.45	0.54	0.95	0.54	1

Dominance Matrix D

	4	5	6	7	8	9	10	11
4	1							
5	0.08	1						
6	0.001	0.001	1					
7	0.001	0	0.91	1				
8					1			
9					0.08	1		
10					0.91	0.08	1	
11					0.08	0.91	0.08	1

The R'Tools software uses this method to calculate the genotypic matrix by a linear transformation of the gametic IBD matrix **G** by

$$\mathbf{Q} = \tfrac{1}{2}\,\mathbf{KGK'}$$

where **K** is **I***[1,1], **I** is an identity matrix of equal rank as the number of individuals and * denotes the Kronecker product of the two matrices. Hence the Q matrix is the overall IBD status between individuals i and j and is not the probability of all gametes among two individuals being inherited from a common ancestor but is twice the coefficient of coancestry among them. Elements of Q are not strictly probabilities as with inbreeding they can be greater than 1. The Q matrix is the equivalent of Wright's numerator matrix described above with either no markers or completely informative markers. Paternal and maternal matrices to model imprinting are obtained by

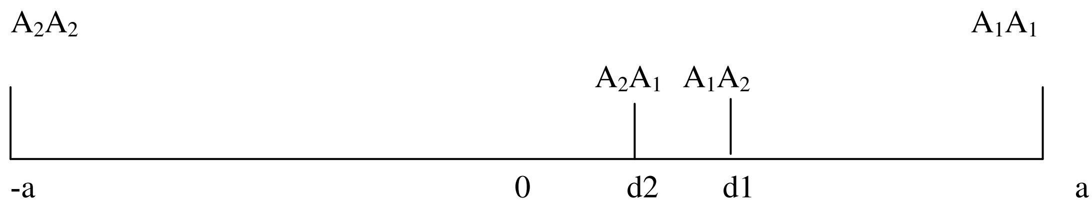
$\mathbf{Q_p} = \mathbf{KGK'}$ $\mathbf{K}$=$\mathbf{I}$*[1,0]

$\mathbf{Q_m} = \mathbf{KGK'}$ $\mathbf{K}$=$\mathbf{I}$*[0,1]

2.3 Genetic variances

Observed phenotypic value is derived from the genetic contribution at all loci and an environmental deviation often described using P = G + E.

Considering a single locus with 2 alleles the genotypic values can be arbitrarily assigned.



(Falconer and Mackay, 1996)

The value d of the heterozygote depends upon the degree of dominance. Dominance is the deviation from the midpoint of the genotypic values and can be defined as an intralocus interaction where the genotypic value of the combination of alleles differs from the value of taking each gene singly. A further source of interaction is imprinting where the heterozygotes may take different values dependent upon parental inheritance. For example A_1 may take a different value depending on whether it is paternally or maternally inherited. The dominance effect (d1+d2/2) and the imprinted effect (i) is equal to (d1-d2)/2.

The relative magnitude of these components determines the genetic properties of a population and in particular the degree of resemblance between relatives. The variances of the genetic effects depend on the frequency of the alleles in the population, if frequency of A_1 is p and A_2 is q then $V(a) = 2pq[a+d(q-p)]^2$, $V(d) = (2pqd)^2$, and $V(i) = 2pq[p^2 + i^2 -2ad(p-q) + p^2d^2 + q^2d^2]$.

The variance of phenotypic values in a population can be divided into genetic and environmental components and furthermore the genetic components attributed to those due to additive genetic variance and those due to dominance or within locus interactions thus $V_P = V_G + V_E$ and $V_G = V_A + V_D + V_i$. In section 2.1 it was shown that the genetic covariance between 2 individuals was derived from additive and dominance components. The coefficients r and u for these genetic covariances can be obtained by using information from relatives see Falconer and Mackay, (1996) for an in depth description.

Relationship		Coefficient	
		r	u
MZ twins		1	1
First degree	Offspring:parent	½	0
	Full sib	½	¼
Second degree	Half sib		
	Offspring:grandparent	¼	0
	Uncle (aunt): nephew(niece)		
	Double first cousin	¼	1/16
Third degree	Offspring:great grandparent	1/8	0
	Single first cousin		

(Falconer and Mackay, 1996)

Furthermore with a nested sib design variation within dams also includes variation due to common environment and maternal genetic effects.

Variance component	additive	dominance	maternal	Common env	Environmental
within sire	¼				
within dam	¼	¼	1	1	
residual	½	¾			1

Thus it can be seen that the degree of resemblance or IBD coefficients can be used to divide phenotypic variance into genetic and non genetic components. In order to estimate dominance monozygotic twins, full sibs or double first cousins are needed. The following section shows how linear models can be used in general pedigrees to apportion phenotypic variance using covariance matrices derived from relationships between all individuals.

2.4 The Linear model and mixed model equations

From Mrode (2005)

For animal i in a population of (n), a linear model can be used to estimate fixed and additive genetic effects (a_i).

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + a_i + e_i \quad (1)$$

where y_i is the phenotypic observation of individual i, $\mathbf{x}_i$ is a known incidence vector, $\boldsymbol{\beta}$ is an unknown vector of fixed effects and e_i is a random error.

Information from relatives contributes to the prediction of a_i through the covariance matrix of a_i values. This covariance matrix is built by the use of the numerator relationship matrix **A** and the variance explained by the genetic effects.

The model can be extended to include a QTL effect (Fernando *et al*., 1989).

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + q_i^p + q_i^m + u_i + e_i \quad (2)$$

where q_i^p and q_i^m are the paternal and maternal additive gametic effects of individual i at the QTL and u_i is the polygenic effect. Again this model uses the information from relatives to contribute to the predictors of the additive effects through the corresponding covariance matrices. For the polygenic term the covariance matrix is built by the use of the usual numerator relationship matrix among individuals as in (1) and the variance explained by this term. The covariance matrix of q_i^p and q_i^m is determined by the gametic identical by descent (IBD) matrix of the alleles at the QTL (**G**) and the variance explained by the QTL. In matrix notation model (2) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{Z}\mathbf{q} + \mathbf{e} \quad (3)$$

where **y** is a vector of phenotypic observations for the trait under consideration (dimension n), $\boldsymbol{\beta}$ is a vector of fixed effects, **u** is a random vector with additive genetic effects of the polygenic terms (dimension n), **q** is a random vector with

gametic effects at the QTL (dimension 2 n), and **e** is a random vector of residuals effects (dimension n).

The matrices **X** and **Z** are incidence matrices relating the phenotypic observation to the corresponding effects of the animals. The expectations of the random variables are

$$E(\mathbf{u}) = 0, E(\mathbf{q}) = 0, \text{and } E(\mathbf{y}) = \mathbf{X\beta} \quad (4)$$

The variance–covariance structure of the random variables is

$$\mathbf{var}\begin{bmatrix}\hat{\mathbf{u}}\\ \hat{\mathbf{q}}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A\sigma_u^2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G\sigma_q^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I\sigma_e^2}\end{bmatrix} \quad (5)$$

where **I** is an identity matrix and $\sigma_u^2, \sigma_q^2, \text{and}\, \sigma_e^2$ are the variances explained by the polygenic term, the QTL and the residual variance, respectively. The total additive genetic variance (σ_a^2) is equal to $\sigma_a^2 = \sigma_u^2 + \sigma_q^2$. Because of (5)

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ'}\,\sigma_u^2 + \mathbf{ZGZ'}\sigma_q^2 + \mathbf{I}\sigma_e^2$$

Let $\sigma_u^2 = \sigma_e^2 / \sigma_u^2 \; and \; \sigma_q^2 = \sigma_e^2 / \sigma_q^2$ then the mixed model equations of (3) are

$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z} & \mathbf{Y'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\sigma_u^2 & \mathbf{Z'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} + \mathbf{G}^{-1}\sigma_q^2\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}}\\ \hat{\mathbf{q}}\end{bmatrix} = \begin{bmatrix}\mathbf{X'y}\\ \mathbf{Z'y}\\ \mathbf{Z'y}\end{bmatrix} \quad (6)$$

The total breeding value for an animal i is the sum of the estimate for the polygenic term u_i and the estimates of q_i. because of (5)

$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_u^2 + \mathbf{G}\sigma_q^2 \quad (7)$$

with expectation $E(\mathbf{a}) = E(\mathbf{u}) + E(\mathbf{q}) = 0$.

2.5 Extensions

This model can be extended further to incorporate breeding values for other genetic models providing the appropriate relationship matrices can be estimated.

2.5.1 Common environment

$$\mathbf{y = X\beta + Zu + Wc + Za + e}$$

$$\mathbf{var(y)} = \mathbf{ZAZ'}\ \sigma_u^2 + \mathbf{ZGZ'}\ \sigma_q^2 + \mathbf{W}\sigma_c^2 + \mathbf{I}\,\sigma_e^2$$

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z + A^{-1}\sigma_u^2} & \mathbf{Z'Z} & \mathbf{Z'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z + G^{-1}\sigma_q^2} & \mathbf{Z,W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'Z} & \mathbf{W'W + I\sigma_c^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{q}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{Z'y} \\ \mathbf{Zc'y} \end{bmatrix}$$

Where W is an incidence matrix relating individuals to dam families and c is a common environment effect. **G** is the genetic covariance matrix for the additive QTL effects.

2.5.2 Dominance

For a model including dominance

$$\mathbf{y = X\beta + Zu + Zq + Zd + e}$$

Where d is a random vector of dominant effects at the QTL and the matrices **X** and **Z** are incidence matrices relating the phenotypic observation to the corresponding effects of the animals. The expectations of the random variables are

$$E(\mathbf{u}) = 0, E(\mathbf{v}) = 0, E(\mathbf{d}) = 0 \text{ and } E(\mathbf{y}) = \mathbf{X\beta}$$

The variance–covariance structure of the random variables is

$$\mathbf{var}\begin{bmatrix}\mathbf{u}\\ \mathbf{q}\\ \mathbf{d}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{A}\sigma_u^2 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}\sigma_q^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{D}\sigma_d^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where **I** is an identity matrix and $\sigma_u^2, \sigma_v^2, \sigma_d^2$, and σ_e^2 are the variances explained by the polygenic term, explained by the additive QTL, dominant QTL and the residual variance, respectively. **G** and **D** are the genetic covariance matrices for the additive and dominance QTL effects respectively.

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ'}\,\sigma_u^2 + \mathbf{ZGZ'}\sigma_q^2 + \mathbf{ZDZ'}\sigma_\delta^2 + \mathbf{I}\sigma_e^2$$

and the mixed model equations of are

$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'Z} & \mathbf{X'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\sigma_u^2 & \mathbf{Z'Z} & \mathbf{Z'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z}+\mathbf{G}^{-1}\sigma_q^2 & \mathbf{Z{,}Z}\\ \mathbf{Z'X} & \mathbf{Z'Z} & \mathbf{Z'Z} & \mathbf{Z'Z}+\mathbf{D}^{-1}\sigma_d^2\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol\beta}\\ \hat{\mathbf{u}}\\ \hat{\mathbf{q}}\\ \hat{\mathbf{d}}\end{bmatrix}=\begin{bmatrix}\mathbf{X'y}\\ \mathbf{Z'y}\\ \mathbf{Z'y}\\ \mathbf{Z'y}\end{bmatrix}$$

$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_u^2 + \mathbf{G}\sigma_q^2 + \mathbf{D}\sigma_d^2$$

Its expectation is $E(\mathbf{a}) = E(\mathbf{u}) + E(\mathbf{v}) = 0$.

2.5.3 Imprinting

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{Z}_p\mathbf{p} + \mathbf{Z}_m\mathbf{m} + \mathbf{e}$$

The matrices **X**, **Z**, $\mathbf{Z_p}$ and $\mathbf{Z_m}$ are incidence matrices relating to fixed, polygenic, paternal QTL and maternal QTL effects respectively.

The variance–covariance structure of the random variables is

$$\mathbf{var}\begin{bmatrix}\mathbf{u}\\\mathbf{p}\\\mathbf{m}\\\mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{A}\sigma_u^2 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_p\sigma_p^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{G}_m\sigma_m^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where **I** is an identity matrix and $\sigma_u^2, \sigma_p^2, \sigma_m^2$, and σ_e^2 are the variances explained by the polygenic term, explained by the paternal QTL, maternal QTL and the residual variance, respectively. $\mathbf{G_p}$ and $\mathbf{G_m}$ are the genetic covariance matrices for the paternal and maternal QTL effects respectively.

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_u^2 + \mathbf{Z}_p\mathbf{G}_p\mathbf{Z}_p'\sigma_p^2 + \mathbf{Z}_m\mathbf{G}_m\mathbf{Z}_m'\sigma_m^2 + \mathbf{I}\sigma_e^2$$

the mixed model equations are

$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'Z}_p & \mathbf{X'Z}_m\\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{A}^{-1}\sigma_u^2 & \mathbf{Z'Z}_p & \mathbf{Z'Z}_m\\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z} & \mathbf{Z}_p'\mathbf{Z}_p+\mathbf{G}_p^{-1}\sigma_p^2 & \mathbf{Z}_p'\mathbf{Z}_m\\ \mathbf{Z}_m'\mathbf{X} & \mathbf{Z}_m'\mathbf{Z} & \mathbf{Z}_m'\mathbf{Z}_p & \mathbf{Z}_m'\mathbf{Z}_m+\mathbf{G}_m^{-1}\sigma_m^2\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol\beta}\\\hat{\mathbf{u}}\\\mathbf{m}\\\hat{\mathbf{p}}\end{bmatrix}=\begin{bmatrix}\mathbf{X'y}\\\mathbf{Z'y}\\\mathbf{Z}_p'\mathbf{y}\\\mathbf{Z}_m'\mathbf{y}\end{bmatrix}$$

$$\text{and var}(\mathbf{a}) = \mathbf{A}\sigma_u^2 + \mathbf{G}_p\sigma_p^2 + \mathbf{G}_m\sigma_m^2$$

Model Summary

(1)	$\mathbf{y = X}\beta\mathbf{ + Zu + Wc + e}$	(null or polygenic)
(2)	$\mathbf{y = X\boldsymbol\beta + Zu + Wc + Za + e}$	(additive)
(3)	$\mathbf{y = X\boldsymbol\beta + Zu + Wc + Za + Zd + e}$	(additive + dominance)
(4)	$\mathbf{y = X\boldsymbol\beta + Zu + Wc + Z_m m + Z_p p + e}$	(maternal + paternal)
(5)	$\mathbf{y = X\boldsymbol\beta + Zu + Wc + Z_p p + e}$	(paternal)
(6)	$\mathbf{y = X\boldsymbol\beta + Zu + Wc + Z_m m + e}$	(maternal)

Thus the polygenic model (1) uses the A matrix and average relationships between sibs to estimate the variance due to all genes. Model 2 also includes the additive

matrix modelling genetic relationships within the pedigree, ignoring dominance. Model 3 divides the genetic covariances into additive and dominant components by also estimating the probability of individuals sharing both alleles at a QTL locus IBD and the deviation of this from variance due to additive effects. Model 4 ignores dominance but divides the genetic variance into maternal and paternal components by attributing covariance coefficients to sibs sharing maternal and paternal alleles and allowing variance of alleles inherited maternally or paternally to differ. Models 5 and 6 only incorporate relationships according to inheritance of the paternal or maternal allele respectively.

2.6 Solving the mixed model equations

Variance components for each model were estimated using REML (Patterson *et al.*, 1971) implemented in the ASReml package (Gilmour *et al.*, 1995). In order to estimate the variance components for the different models, ASReml requires the knowledge of the inverse of the relationship matrices. ASReml calculates the inverse of the **A** matrix directly from pedigree data, but a separate routine was used to invert **G,** $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** before using them in ASReml.

The mixed model equations are solved in ASReml using residual maximum likelihood. The reml procedure maximises the joint likelihood of all error contrasts rather than of all contrasts as in ordinary maximum likelihood. It is based on a log likelihood of the form

$$L\alpha(1/2)\{-(\mathbf{y}-\mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{Xb})-\text{logdet}(\mathbf{V})-\text{logdet}(\mathbf{X'V}^{-1}\mathbf{X})\}$$

Where b is the generalised least square solution (GLS) and satisfies

$$\mathbf{X'V^{-1}Xb=X'V^{-1}y}$$

Comparison of models and tests specific for each analysis are detailed in individual chapters.

CHAPTER 3

QTL analysis of bodyweight and conformation score in commercial broiler chickens comparing variance component and half-sib methods

Summary

The aim of the study was to investigate Quantitative Trait Loci in previously identified regions of chicken chromosomes 1, 4 and 5 relating to 40-day bodyweight and conformation score using a two-generation design. Half-sib analyses and variance component analysis were implemented and compared. The data consisted of 100 full-sib families (46 paternal half-sib families) with trait data for a total of 2,708 offspring. Both generations were genotyped for markers spaced approximately every 16, 14 and 8 cM on chromosomes 1, 4, and 5, respectively. For QTL mapping the trait values were adjusted for fixed and random dam effects using a mixed linear model. Half-sib QTL mapping was performed using the regression method in QTL Express for both paternal and maternal families. Confidence intervals and significance thresholds were estimated using bootstrapping and permutation analysis. Variance component mapping was done testing a novel module in QTL Express using MCMC to estimate IBD coefficients and ASReml to estimate QTL effects. Chromosome 4 showed nominal significance for QTL affecting bodyweight and conformation, and linkage was confirmed for both traits on chromosome 5. Results varied according to method of analysis and common parent in the half-sib method. Variance components analysis (VCA) tended to detect effects segregating from both parents. Analysis of dam families gave the strongest evidence for segregation of QTL. The results suggest that conformation score segregates as a separate trait in sires and dams.

3.1 Introduction

There is growing evidence to suggest that much of the segregating variation identified in line cross experiments can also explain variation within lines (Andersson et al., 2004; De Koning et al., 2003; Evans et al., 2003; Wong et al., 2004). Even after 50 generations of selection, De Koning *et al.*, (2003) detected QTL previously identified by line cross experiments between broilers and layers segregating within a commercial broiler population. An important advantage of within population methods of QTL detection is their immediate potential for use in marker-assisted selection.

Using a three generation half-sib analysis De Koning *et al.*, (2004) measured birds from a commercial broiler dam line (Cobb Breeding Co. Ltd., Chelmsford, UK) for many traits relating to bodyweight, conformation and carcass composition. QTL explaining a large proportion of phenotypic difference for bodyweight ($p<0.001$), and residual feed intake ($p<0.01$) were confirmed, and evidence found for QTL affecting the relative bodyweight of bone and muscle in the thigh $p<0.05$, carcass weight, and conformation score. Previous experimental crosses in broilers have located multiple QTL for bodyweight, growth and carcass traits on chicken chromosomes 1, 4 and 5; (Carlborg et al., 2003; Ikeobi et al., 2002; Sewalem et al., 2002; van Kaam et al., 1999) reviewed by Hocking (2005) and Abasht (Abasht et al., 2006).

Because a three-generation design detects QTL segregating in the grand parental generation, the relevance of the results in terms of selection in the current populations can be reduced. Marker assisted selection is most beneficial where phenotypes are difficult to measure and with the advent of much more cost effective genotyping it could be argued that the results of a two-generation design would provide marker-trait information that was closer to the selection line. To fully exploit variation, there is a need to assess the efficiency of available methods for detecting QTL within general pedigree structures.

Variance component analysis (VCA) is potentially a powerful tool for commercial pedigree structures enabling QTL analyses to take place within a population and without the need to construct sub pedigrees for analysis purposes. Half sib designs (Knott et al., 1996) only examine the segregation of alleles in the common parent assuming a single offspring per mating and that all parents are unrelated, therefore, potentially ignoring important information. George (2000) describes a two-step variance component mapping approach, simulating complex pedigree structures to detect QTL effects. Little work, however, has been done on VCA using real data. Comparisons between half sib and variance component designs have been made by Nagamine (2002), and De Koning (2003) in commercial pig populations and Slate (2002) using a natural deer population. Variance component designs were more powerful where QTL were segregating in the parent not used as the common parent in the half-sib analysis.

The aim of the current study was to use full-sib families from the same broiler dam line as De Koning *et al.,* (2004) (i.e. a mostly independent sample of the same population) to re-examine previously identified candidate regions for QTL affecting bodyweight and conformation score using a two-generation design. Half-sib (HS) and variance component analysis (VCA) methods of analysis were implemented and compared. Comparisons were also drawn with results from the previous three-generation design.

3.2 Materials and methods

3.2.1 Trait measurements

Phenotypes were available for 40-day bodyweight and conformation score. Markers on chromosomes 1, 4 and 5 were used because these linkage groups explained the most variation for these traits in previous studies. The 100 largest dam families were selected from 10,286 population records, resulting in 2,708 offspring with phenotype records from 46 half-sib sire families. There were an average of 59 and 27 offspring per sire and dam, respectively. Progeny were from two flocks across 17 hatch weeks. Birds were genotyped for markers spaced approximately every 16, 14 and 8 cM on chromosomes 1, 4, and 5 respectively. Markers were selected from the consensus linkage map (Schmid et al., 2000) and tested for heterozygosity (see De Koning *et al.*, (2004) for full details). Information content was generally high with 24 sires and 25 dams on average being heterozygous at a given marker.

3.2.2 Analysis of phenotypic data

Variance components and fixed effects were estimated prior to QTL analysis with an animal model (1) using ASReml (Gilmour *et al.*, 2000). Variance components were estimated using 10,286 phenotypic records in the population, representing the birds in the QTL experiment as well as all their contemporaries. Direct maternal effects were estimated by fitting Dam as a random effect.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{e} \qquad (1)$$

where **y** is a vector of phenotypic observations, **β** is a vector of fixed effects of sex, hatch and dam age within flock, **u, c** and **e** are vectors of additive polygenic effects,

random maternal effects, and random residuals, and **X, Z** and **W**are incidence matrices relating to fixed, polygenic, and maternal effects.

3.2.3 Linkage maps

Linkage maps were estimated and evaluated using CriMap (Green *et al.,* 1990). Options build and flips were used to build and test alternatives for the consensus maps for chicken chromosomes 1, 4 and 5. Map distances given in Appendix 3.1 were referenced against, and found to be largely in agreement with, the previous study and existing published maps. The three linkage groups corresponded to the consensus map at approximately 128-205cM, 75 – 182cM, and 57-104cM for chicken chromosomes 1, 4 and 5 respectively. The recently published map from the chicken genome project (Wong et al.,2004) was longer than the estimated linkage map by 12, 23, and 5 cM with published distances of 77, 108 and 47 cM for linkage groups on chromosomes 1, 4 and 5 respectively.

3.2.4 Half sib QTL analysis

Prior to half sib analyses phenotypic values for bodyweight and conformation score were estimated using residual values from a mixed model in GENSTAT (Lawes Agricultural Trust, Harpenden, U.K.)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{d} + \mathbf{e} \qquad (2)$$

where **y** is a vector of observations, **β, d** and **e** are vectors of; fixed effects (sex, hatch and dam age within flock), random dam effects, and random residuals. **X** and **Z** are incidence matrices relating to fixed and random dam effects.

Half sib families were analysed with Sire and Dam fitted independently as common parent. In each case information from the other parent was incorporated to improve phase estimation. Where possible, missing genotypes were inferred from pedigree information.

Genotype, phenotype and map information were used to run single QTL analyses using 3.1 QTL express software at http://QTL.cap.ed.ac.uk/ (Seaton et al., 2002). QTL Express uses a multi marker approach to interval mapping in half sib families

(Knott *et al.,* 1996). The probability of QTL genotype at 1 cM intervals was estimated conditional on marker genotypes and recombination fraction/distance from marker. Phenotype of offspring was then regressed onto QTL genotype using within family least squares analysis to test for a significant effect of allele substitution.

$$y_{ij} = m_i + b_i p_{ij} + e_{ij} \qquad (3)$$

Where y_{ij} is trait score for offspring j originating form parent i, b_i is the substitution effect for a putative QTL, p_{ij} is the conditional probability for individual j of inheriting the first parental haplotype and e_{ij} is the residual effect.

Within families, t-statistics were used to test the significance of the QTL effect at the overall best position of the QTL. Across families a test statistic was calculated as an F ratio for every map position obtained using the ratio of mean squares of a model fitting a QTL to not fitting a QTL. The within family analysis was combined across families to estimate variance of QTL effects. The proportion of within family variance explained by the QTL was estimated by:

$h^2_{QTL} = 4*[1-(MSE_{full}/MSE_{reduced})]$
following Knott *et al.,*(1996)

Empirical significance thresholds were estimated by using permutation tests (Churchill et al., 1994) involving 1,000 randomisations to estimate the 5 and 1% thresholds. The confidence interval of the best position was determined using 1,000 bootstrap replicates, sampling all individuals with replacement (Visscher et al., 1996). Confidence intervals obtained by bootstrapping are often large due to heterogeneity in best QTL positions amongst individual families (de Koning et al., 1998). In order to narrow confidence intervals, further bootstrap analyses were carried out using only families with significant allele substitution effects from (3). Heterogeneity of QTL position was also explored by estimating the putative position of the QTL indicated by each individual family in turn. Families showing a significant allele substitution effect for both traits were investigated as potential indicators of pleiotropic effects.

3.2.5 Variance component QTL analysis

Variance component analysis (VCA) was carried out using a two–step approach following (George et al., 2000) based on Loki and ASREML (Gilmour *et al.*, 2000) software as implemented within QTL express (http://latte.cap.ed.ac.uk/hkcServletLoki.html). Loki, as described by Thompson and Heath (1999), uses a Monte Carlo Markov Chain (MCMC) method to calculate identical by descent (IBD) scores for the pedigree at each position, simultaneously estimating missing marker data and unknown haplotype information. A Q matrix containing the proportion of alleles IBD at the QTL for each relationship in the pedigree, and an A matrix containing the average proportion of alleles identical by descent over all other loci is calculated. In the second step, ASReml uses the IBD proportions to model the phenotypic covariance for a putative QTL. The analysis is based on the premise that individuals sharing more alleles identical by descent will be more alike phenotypically. By fitting QTL and polygenic effects simultaneously, VCA generates the proportion of variance explained by the polygenic component, and by the QTL.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e} \qquad (4)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e} \qquad (5)$$

where $\mathbf{y}$ is a vector of phenotypic observations, $\boldsymbol{\beta}$ is a vector of fixed effects, **u, v** and **e** are vectors of additive polygenic effects, QTL effects and random residuals, and **X** and **Z** are incidence matrices relating to fixed and polygenic effects.

Fixed effects used to estimate variance components in (1) were included in the ASReml step to model a putative QTL at 1 cM intervals. A test statistic for a given location was obtained by running the animal model without a QTL effect (4). Twice the difference between logarithms of the likelihood of (5) vs (4) was used as a log likelihood ratio (LR) test. Following (Self et al., 1987) it is assumed that LR will follow a mixture of a chi-squared distribution with 1 df and a zero-peaked distribution. This was attained by halving the p value for chi-square with one df to account for the one sided nature of the test, as discussed by (Allison et al., 1999).

3.3 Results

3.3.1 Analysis of phenotypic data

Summary statistics and heritabilities for bodyweight and conformation score are given in Table 3.1. Bodyweight was normally distributed within males and females. Conformation score was approximately normally distributed and although slightly negatively skewed was treated as normal. The traits were correlated with a phenotypic correlation of r = 0.37. Results from the animal model (1) indicated a significant direct maternal effect. Possible dominance effects confounded with maternal effects were assumed negligible.

Table 3.1. Variance component estimates for 40 day Bodyweight and Conformation Score

Trait	Mean[a]	SD	Range	Animal model	
				Heritability	Heritability fitting dam effect
Bodyweight (g)	2514	298	820 – 3700	0.19	0.07
Conformation	3.34	0.88	1.0 - 6.0	0.21	0.11

[a] raw phenotypic means.[b] Direct maternal effect from mixed model assuming no dominance.

3.3.2 QTL analysis

Results are given for the half-sib QTL analyses using residual phenotypic values from (2). In general, the fitting of fixed effects had a greater effect on the F ratios and the estimated QTL position than the direct maternal effect. Fixed effects for VCA were estimated using the ASREML step within QTL Express. Fitting a random dam effect did not affect QTL heritability or power to detect an effect. It did affect putative position of the QTL and polygenic heritability reducing the polygenic heritability to almost zero, possibly indicating that the maternal effect was overestimated in the mixed model. Results are shown for QTL analyses accounting for fixed effects only with the random dam effect omitted.

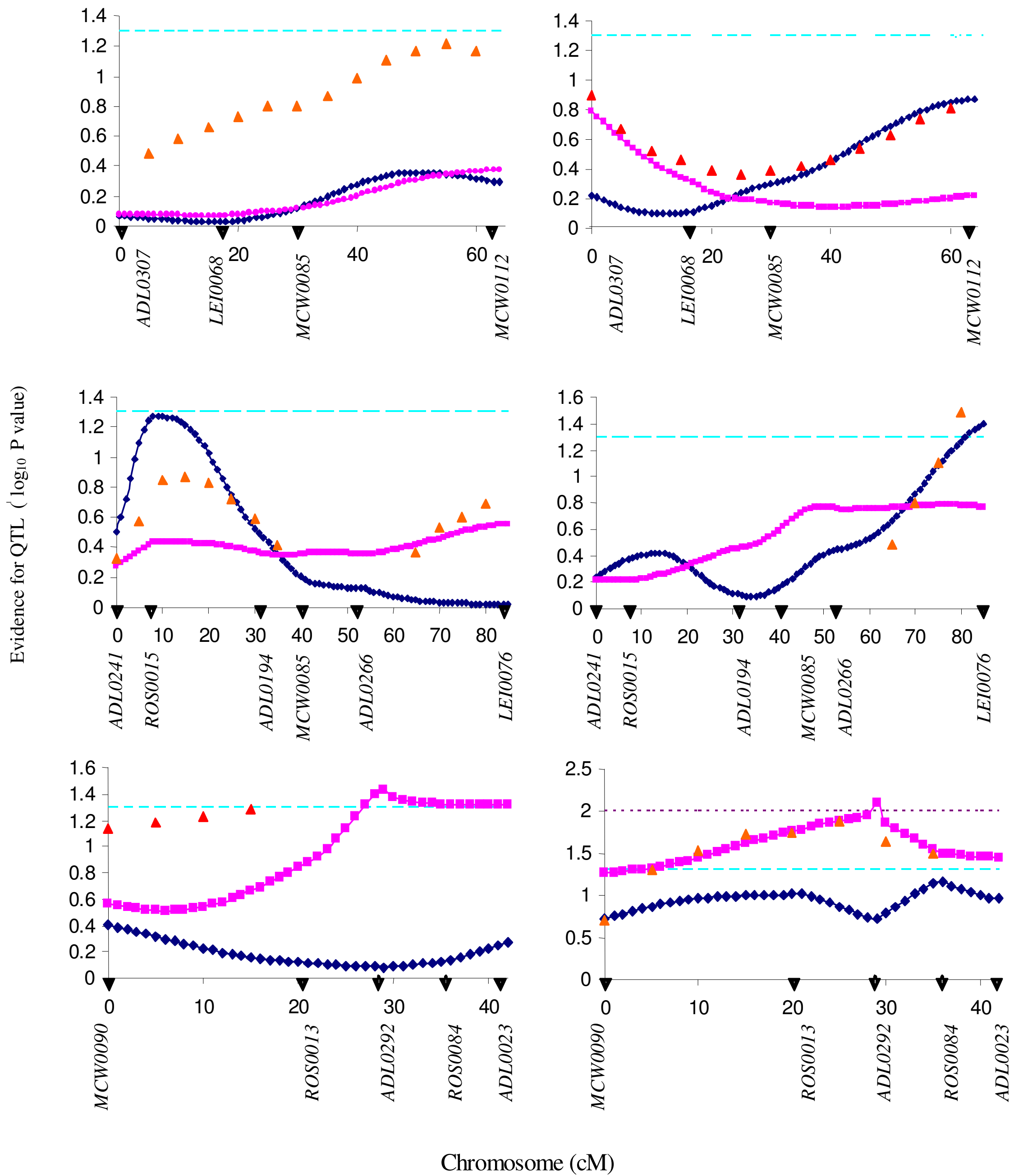


Figure 3.1. Log-transformed nominal P values along chromosomes for sire, dam and variance components analysis. Bodyweight on left, conformation on right, top to bottom chromosomes 1, 4 and 5. dashed line for P=0.05 and dotted line for P=0.01

3.3.3 Evidence for QTL

5% Significance thresholds based on permutation analysis are indicated in figure 3.1. VCA and HS analysis of sire families achieved nominal significance for conformation on chromosome 4. Analysis of dam families exceeded the threshold to confirm linkage on chromosome 5 for QTL affecting bodyweight and conformation score.

Figure 3.1 compares evidence along the chromosomes for the VC and half sib analyses using a logarithmic scale. There was some evidence for a QTL affecting bodyweight on chromosome 1. However this does not reach significance and is not replicated in the half sib analyses. QTL for weight on chromosomes 1 and 5 were close to significance with P values of 0.06. VCA found significant QTL for bodyweight on chromosome 4 at 80 cM and conformation on chromosome 5 at 25 cM in agreement with the half-sib analyses. Conformation was significant at the nominal 1% level to confirm linkage.

3.3.4 QTL positions

Putative positions for QTL varied between analyses. For the half-sib analyses position often appeared to vary according to common parent, for example conformation on chromosome 4. Bootstrapping, however, failed to estimate confidence intervals less than the entire chromosome. The VCA did not alter the putative position of the QTL and where it appears there are separate optima for sire and dam analyses, the VCA showed evidence at both positions. For QTL affecting weight on chromosomes 4 and 5 the VCA was uninformative along much of the chromosome although did agree with the most significant position.

3.3.5 Significant family analysis

For the half-sib analyses using only families segregating with a significant QTL, there were on average 5 and 6 families segregating with 270 and 180 total progeny for sire and dam families, respectively. With the exception of body weight on chromosome 1, all analyses achieved genome wide significance under permutation analysis explaining between 16 and 80% of the within family variance.

Table 3.3 shows that these families were useful for gaining more insight into the position of the QTL. Families segregating for conformation on chromosome 4 map to separate positions with discrete confidence intervals for sire and dam families. On chromosome 5 in the half-sib dam families there is a narrow confidence interval for conformation score compared with a large CI in sire families.

Table 3.2 Proportion of within family variance explained by fitting a QTL for half sib families using different analyses

Trait	Chr	V_w QTL (%)			
		Sire	Dam	VCA	De Koning et al (2004)
Bodyweight	1	0.003	0.003	0.04	0.07**
	4	0.03*	0.03	0.03†	0.24***
	5	0.002	0.06**	0.03	-
Conformation	1	0.02	0.03	0.02	-
Score	4	0.03*	0.02	0.02	-
	5	0.02*	0.09**	0.04††	0.11**

Sire and dam denote common parent, VCA refers to QTL heritability from variance component analysis and previous refers to three-generation analysis by De Koning *et al.*, 2004)

* pointwise significance $P < 0.05$, ** chromosome wide significance, *** genome wide significance

† pointwise 5% significance assuming $\chi^2_{0.5}$, †† pointwise 1% significance assuming $\chi^2_{0.5}$

Table 3.3. Confidence intervals (CI) obtained by bootstrapping for half-sib analyses for 40-day bodyweight and conformation score using only families showing a significant allele substitution effect

Trait		Sire		Dam	
	Chr	Putative QTL position (cM)	Confidence Interval (cM)	Putative QTL position (cM)	Confidence Interval (cM)
Bodyweight	1	51	31-64	64	5.5-63
	4	85	38.5-85	79	59.5-85
	5	0	0-42	29	0-42
Conformation Score	1	64	50-64	0	0.0-0.0
	4	10	5-34.5	85	70-85
	5	36	0-42	29	21-29

Sire and Dam denote common parent

3.4 Discussion

The largest QTL effects for both traits were observed within half sib dam families, achieving chromosome wide significance and confirming previously reported QTL for conformation score on chromosome 5 (Table 3.2).

3.4.1 Body Weight

All three analyses failed to confirm the bodyweight QTL explaining 7% of the variance on chromosome 1 in the three generation analysis (De Koning, 2004). VCA presented the strongest evidence with 4% of the residual variance explained at the linkage peak, ten fold greater than that explained by the HS analyses. De Koning *et al.*, report considerable a QTL on chromosome 1, affecting the direct maternal effect on weight of 6%,

On chromosome 4, the 2-generation analyses showed a putative weight QTL around 80cM with nominal significance attained in sire families and VCA. Only 3% of the within family variance is explained in contrast to the genome wide significant QTL

with a large effect (24%) found by the three generation analysis. The three-generation analysis found most evidence close to marker ADL0194 whereas all analyses presented here suggest the QTL is linked to marker LEI0076 around 50cM away. There was considerable heterogeneity in the putative QTL positions for individual significant families particularly in sire families. This is in line with evidence found by Wong *et al*., (2004) that two QTL affecting weight exist on chromosome 4.

There was a tendency to find weight and conformation QTL in the same marker interval on chromosome 5. The 5 and 1% thresholds for chromosome wide significance set by permutation analysis were reached for weight and conformation respectively in dam families at 29 cM. The three-generation analysis found evidence for QTL affecting weight and conformation around 10cM. Sire and VCA found evidence for a weight QTL in the same marker interval at 0cM. Ruy *et al* (Ruy et al., 2005) also found a significant weight QTL ($P < 0.001$) associated with marker MCW0090.

3.4.2 Conformation-score

A putative QTL for conformation score on chromosome 4 was found only using sire analysis. Sire families achieved nominal significance at 10 cM. When only significant families were analysed dam families indicated a putative QTL at around 80 cM whilst VCA found nominal significance for QTL at both positions. Effects segregating in dam families tended to be significant for both weight and conformation. Analysis of significant sire families, however, resulted in weight and conformation QTL in discrete confidence intervals indicating two QTL rather than pleiotropic effects of a single QTL affecting both traits.

The QTL for conformation on chromosome 5 was in the same interval as the putative QTL for weight in the HS dam and three generation analyses. In the Dam analysis some families were showing a significant QTL effect for both traits. Nominal significance for a conformation QTL was also reached in the HS sire families and VCA. De Koning et al report the QTL for conformation score explained 11% of the within family variance with a direct maternal effect of 19%.

There is evidence to suggest that conformation score is a slightly different trait in sire and dam lines. For the putative conformation QTL on chromosome 5, VCA found evidence at both putative positions indicating the possibility of two QTL. This might also explain the high variance ratio across the entire linkage group. There was a tendency for QTL in dam families to explain more of the within family variance of the trait and have narrower confidence intervals. This could be due to the full sib structure or the possibility of an unaccounted for maternal genetic effect for conformation. Greater power using dam families might also be a result of stronger selection in the sire lines, leading to fixation of favourable alleles. Where alleles are segregating only in the dams, the effect is diluted across the population making a specific dam family analysis more powerful than an analysis across the population. This would make dam families a better resource to detect QTL for traits traditionally selected for in sire lines as there are more likely to be alleles segregating. There is a possibility of imprinting where alleles inherited from the dam have greater expression. De Koning *et al*., (2002, 2003b) present evidence that differences in QTL allele frequencies between sexes cause discrepancies between VCA and HS analyses, postulating that these could reflect effects of selection when the parents originate from specific dam and sire lines, or that differences between paternal and maternal models could be explained by genomic imprinting.

Figure 3.3 compares the empirical power of the present two-generation design with that of the previous three-generation design (De Koning et al., 2004). This shows that the three-generation design has the greatest power to detect an effect and that a smaller number of larger half sib sire families is more powerful than many small half-sib dam families. If, however, the dam families consist of full sibs, power to detect QTL effects increases with effects greater than 0.2 phenotypic standard deviations. Differences are largest at intermediate effects. At 0.1 and 0.3 standard deviations, there is little difference between the full sib analysis and the three-generation design.

Because VCA uses all genetic relationships in the pedigree it could detect effects from both the HS sire and dam analyses. The variance component method is more computationally intensive than the HS methods. However, given current breeding value estimation techniques, it would not require specialist software to integrate VCA QTL methods into existing genetic evaluation programs. The analysis would also

supply QTL breeding values for the population. For the purpose of comparing selection candidates, the half-sib analysis is more time consuming and disjointed with analysis repeated in its entirety for both sexes. Advantages of the HS methods are a greater insight into the difference between QTL segregating in the sire or dam and the influence of a maternal dam effect. This is only the case if the analysis can be carried out using both parents as common parent, which is impractical in many species where size of dam family limits power.

3.4.3 Comparison with previous results

It would appear that the two-generation analysis does not have sufficient power to detect effects of the magnitude previously reported or that the effects of the QTL are inflated in the previous analysis as suggested by the authors who note that the total within family variance explained by QTL and cofactors was unrealistic.

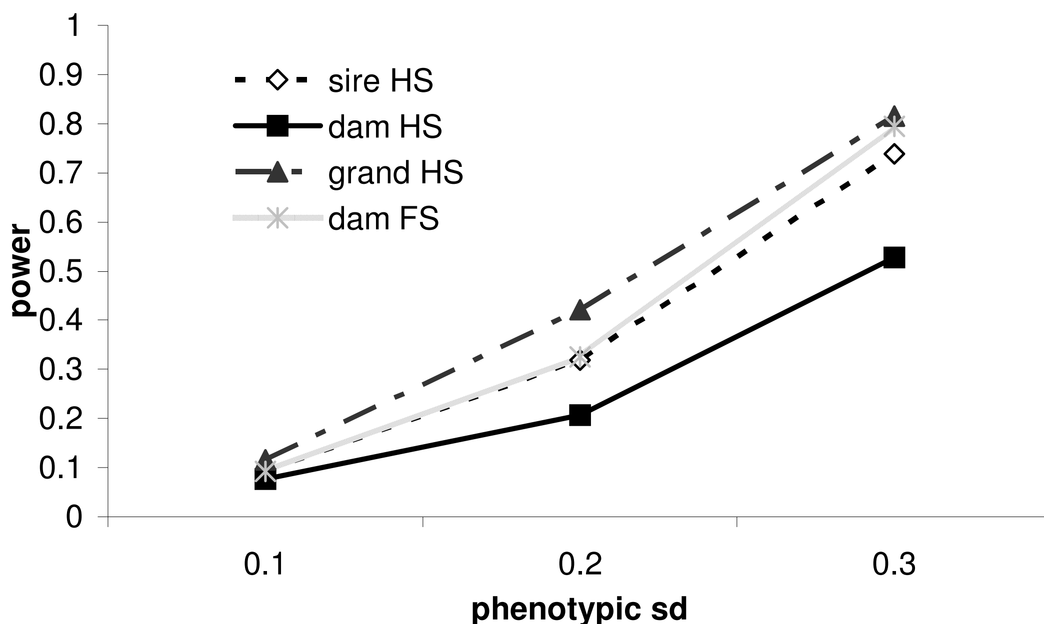


Figure 3.3 Estimated power to detect QTL based on three-generation pedigree of 15 grandsires, with 40 daughters, each with 24 offspring (grand HS); two-generation pedigree half sib sire families of 59 (sire HS) and half sib or full sib dam families of 27 (dam HS and dam FS). Using a heritability of 0.15, distance between markers of 16 cM, and QTL heterozygosity of 0.5

3.5 Conclusions

Half-sib dam families achieved chromosome-wide significance for weight and conformation on chromosome 5 explaining 6 and 9% of the within family variance respectively. Analysis of significant families narrowed the confidence interval for a conformation QTL down to 9 cM but failed to narrow down the confidence interval for a weight QTL to less than the entire chromosome. These results are sufficient to confirm previously published QTL for weight and conformation on chromosome 5. Evidence was also found for QTL affecting weight and conformation-score on chromosome 4.

This study confirms QTL segregating in a commercial population for traits that have been under intensive selection for > 40 years. Linkage under the two-generation design peaked at QTL positions found by the three-generation design but effects were much smaller and in most cases linkage was not confirmed. Further comparisons are needed to predict the consequences of design choice. It is clear that QTL detection is not confined to line crosses but can be studied in populations where they are directly relevant in terms of selection.

Appendix 3.1 Marker distances and corresponding position on the consensus map.

		Size of interval (cM)				
Marker Interval	Chr	Female	Male	Sex Averaged	consensus	Position consensus (cM)
ADL0307-LEI0068	1	19	15.3	17.1	23	128-151
LEI0068-MCW0297	1	2.6	7.7	10.1	11	151-162
MCW0297-MCW0112	1	35.8	38.6	37.4	42	162-205
Total linkage group	1	57.4	61.6	64.6	76	
ADL0241-ROS0015	4	6	8.9	7.6		
ROS0015-ADL0194	4	24.3	27.2	25.5	38	80-118
ADL0194-MCW0085	4	9.1	7.6	8.2	2	118-120
MCW0085-ADL0266	4	14.3	10.7	12.5	17	120-137
ADL0266-LEI0076	4	31.3	32.5	31.9	45	137-182
Total linkage group	4	85	86.9	85.7	102	
MCW0090-ROS0013	5	23.6	19	21.4	21	57-78
ROS0013-ADL0292	5	9.2	5.5	7.4	5	78-83
ADL0292-ROS0084	5	8	5.8	6.8		
ROS0084-ADL0023	5	3.3	9.1	6.5		(77-104)
Total linkage group	5	44.2	39.4	42.1		

CHAPTER 4

Detecting dominant QTL with variance components analysis in simulated pedigrees

Summary

Dominance is an important source of variation in complex traits. Here we have carried out the first thorough investigation of QTL detection using variance component models extended to incorporate both additive and dominant QTL effects. Simulation results showed that the empirical distribution of the test statistic when testing for dominant QTL effects did not behave in accordance with existing theoretical expectations and varied with pedigree structure. Extensive simulations were carried out to assess accuracy of estimates, type 1 error and statistical power in two generation human, poultry and pig type pedigrees each with 1900 progeny in small, medium and large-sized families, respectively. The distribution of the likelihood ratio test statistic was heavily dependent on family structure, with empirical thresholds lowest for human pedigrees. Power to detect QTL was high (0.84-1.0) in pig and poultry scenarios for dominance effects accounting for >7% of phenotypic variance but much lower (0.42) in human type pedigrees. Maternal or common environment effects can be partially confounded with dominance and must be fitted in the QTL model. Including dominance in the QTL model did not affect power to detect additive QTL effects. Also, detection of spurious dominance QTL effects only occurred when maternal effects were not included in the QTL model. When dominance effects were present in the data but were not in the analysis model this resulted in both spurious detection of additive QTL or inflated estimates of additive QTL effects. The study demonstrates that dominance can be routinely included in QTL analysis of general pedigrees, however optimal power is dependent on selection of the appropriate thresholds for pedigree structure.

4.1 Introduction

Historically dominance has often been ignored or treated as a nuisance parameter, for example in genetic evaluations of livestock and quantification of variance components. The importance of the detection and quantification of dominance effects underlying complex traits, however, is underlined by an increasing body of evidence for dominant QTL with major effects on human disease and agricultural traits of economic importance. Duong *et al*. (2006) found eight completely dominant QTL associated with hypertension in congenic rat lines and in agriculture examples of

dominant QTL include fertility and production traits in cattle (Cohen-Zinder *et al.* 2005), chicken (Ikeobi *et al.* 2002; Hocking 2005), tomatoes (Semel *et al.* 2006), and maize (Zhang *et al.* 2006). Liu *et al*., (2007) performed a genome wide scan on an F_2 Duroc Pietrain cross and found 40 additive QTL and 31 QTL showing overdominance effects. Although definitions vary, overdominance is a phenomenon for which there is increasing evidence in plants as the underlying mechanism for heterosis (Xiao *et al.* 1995; Frascaroli *et al.* 2007). Lippman *et al.* (2007) review detection and characterization of heterosis, overdominance and pseudo-overdominance.

To date, the detection of these dominant QTL effects has predominantly involved model species or experimental crosses requiring inbred or genetically divergent populations. Reproductive constraints render these test crosses impractical for many agricultural species while for human and natural populations they are unethical, and/or untenable. In commercial livestock populations it is often more relevant, practical and cost effective to explore QTL segregating within a population, particularly if the objective is to facilitate selection within that population. There is evidence to suggest that much of the variation found between lines is segregating within lines (De Koning *et al.* 2004) and furthermore, most evolutionarily important variation appears to occur within lines (Erickson *et al.* 2004).

There is, therefore, an increasing need for QTL methodology to routinely account for genetic interactions such as dominance within any population structure. Independently developed within human and livestock research, variance component (VC) based linkage methods (Fernando & Grossman 1989; Goldgar 1990; Amos 1994; Grignola *et al.* 1997; Almasy & Blangero 1998; Allison *et al.* 1999; George *et al.* 2000), have the advantage of simultaneously locating and estimating genetic effects within arbitrary pedigrees. Genetic parameters associated with the polygenic effect, and at specific loci using marker and pedigree information can be estimated simultaneously. The incorporation of many alleles or allelic effects and all relationships within a pedigree has been shown to increase power to detect QTL over sib based methods (Williams & Blangero 1999; Sham *et al.* 2000b; Kolbehdari *et al.* 2005; Rowe *et al.* 2006). Furthermore, linkage disequilibrium and haplotype information can be incorporated to provide greater accuracy (Meuwissen *et al.* 2002; Lee & Van der Werf 2006). Most importantly there is the potential for flexibility to incorporate

random effects and their interactions, for example, dominance, epistasis and maternal effects, limited only by the size and structure of the experimental population.

Although undeniably an important source of variation, non-additive effects are notoriously difficult to estimate due to confounding with other sources such as common maternal environment (Gengler *et al.* 1997; Misztal 1997). Computational complexity combined with the more generic problems of setting appropriate thresholds to account for multiple testing and lack of suitable data have inhibited the extension of variance component methodology to incorporate interactions such as dominance and epistasis. Although extensions to VC QTL linkage models to incorporate dominance are widely discussed (Sham *et al.* 2000a; Diao & Lin 2005) they have rarely been implemented, indicating a need for further investigation before the full potential of these methods can be unleashed.

In the present study, extensive simulations have been used to explore the power and potential for partitioning QTL effects into additive and dominant components using VC methods for linkage analysis. Varying full sib and half sib population structures have been used to evaluate accuracy and power to detect additive and dominant genetic effects in pedigrees that are representative of commercial livestock and human scenarios.

4.2 Materials and methods

4.2.1 Statistical Genetic Models for Variance Component Analysis

Population wide linkage equilibrium between QTL and marker alleles was assumed for all analyses. Following the two–step approach described by George *et al*. (2000), for each putative QTL position, marker information was used to estimate IBD coefficients for all relationships in the pedigree. In the second step, different QTL models were fitted for given genome locations using the following models:

(1) $\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{e}$ (null or polygenic)

(2) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Za} + \mathbf{e}$ (null + additive QTL)

(3) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e}$ (null + additive QTL + dominance QTL)

(4) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e}$ (null + maternal + additive QTL + dominance QTL)

where **y** is a vector of phenotypic observations, **β** is a vector of fixed effects, **u**, **a**, **d**, **m** and **e** are vectors of random additive polygenic effects, additive and dominance QTL effects at the putative QTL position, non genetic maternal effects and residuals respectively, and **X, Z** and **W** are incidence matrices relating records to fixed and random genetic and maternal effects respectively.

Variances for polygenic and QTL effects are distributed as follows: Var(**a**) = $\mathbf{G}\sigma^2_q$, Var(**d**) = $\mathbf{D}\sigma^2_d$, var(**e**) = $\mathbf{I}\sigma^2_e$, var(**u**) =$\mathbf{A}\sigma^2_a$. For the non-genetic maternal effect Var(**m**) = $\mathbf{I}\sigma^2_m$.

where **A** is the standard additive relationship matrix based on pedigree data only, **G** is the QTL additive genetic relationship matrix based on marker information and **D** is the QTL dominance genetic relationship matrix representing the probability that two individuals have the same pair of alleles in common based on marker information. Variance components for each model were estimated using REML (Patterson & Thompson 1971) implemented in the ASReml package (Gilmour *et al.* 1995)

4.2.2 Calculating the relationship matrices A, G and D needed for the mixed model analysis

The relationship matrices **G** and **D** for a given QTL position are calculated from the gametic IBD matrix as outlined by Liu *et al* (2002). The gametic IBD matrix is a 2n x 2n matrix containing the probability of identity of descent between either of the two gametes of an individual with the gametes of the remaining individuals in the pedigree. In contrast to George *et al*. (2000) who used a Monte-Carlo method, the gametic IBD matrix was estimated with the recursive method of Pong-Wong *et al*., (2001), which uses the two first available fully informative or phase known flanking markers. The **G** and **D** matrices are conditional on flanking marker information and therefore unique for each position evaluated for a QTL, hence, the calculation of **G** and **D** requires the prior calculation of the gametic IBD matrix conditional on linked marker information at the position of the putative QTL. Here the matrices were calculated every 5 cM. In order to estimate the variance components for the different models, ASReml requires the inverse of the relationship matrices **A, G** and **D**. The version of ASReml used calculates the inverse of the **A** matrix directly from pedigree data, but the inverse for **G** and **D** were calculated from the gametic matrix, inverted using a separate routine then passed to ASReml.

4.2.3 Test statistic

A test statistic for a given location was obtained by comparing the likelihood of the full versus the null model. The log likelihood ratio test statistic (LRT) was calculated as twice the difference between the log likelihood of the full and the reduced model. Power was estimated both empirically using thresholds derived from 1000 chromosome-wise replicates, and using tabulated values assuming that the test statistic is chi squared distributed with degrees of freedom equal to the number of extra parameters estimated in the full model compared with the reduced. This is conservative for a test at a single location in the genome as the test statistic under the null hypothesis is likely to be distributed as a complex mixture of distributions (Self & Liang 1987; Stram & Lee 1994; Allison *et al.* 1999; Visscher 2006). For QTL mapping, it has been suggested that the most straightforward method of achieving the critical null value is to halve the *P* value obtained for χ^2_k where k is the number of

extra variance components in the full model (Visscher, 2006). In practice this mixture of distributions $\chi^2_{0\text{-}1}$ equates to using the 10% critical threshold for a 5% type 1 error rate. This result is only valid for one extra variance component. When k is greater than 1 the appropriate P value can be obtained from an appropriately weighted combination of *P* values corresponding to the LRT statistic. For example where k = 2, and LRT = 4, the appropriate *P* value is a weighted mixture of 0, 1 and 2 degrees of freedom at ¼, ½, and ¼ respectively which corresponds to a P value of 0.08.

Three tests were carried out; (i) additive QTL (2) versus null (1) to test significance of the QTL variance component under a purely additive model (denoted 1v0); (ii) additive QTL + dominance QTL (3) versus null (1) to test significance of QTL variance components under a model including additive and dominance effects (denoted 2v0); and (iii) additive QTL + dominance QTL (3) vs. additive QTL (2) to test the significance of the dominance variance component (denoted 2v1). To estimate the effect of common environment the model including additive and dominance QTL effects was further extended to incorporate a random dam effect (4), representing a maternal or full-sib family effect.

4.2.4 Population structure

The method was implemented in three simulated populations, representative of poultry, pig and human pedigrees (Table 4.1). The parental generation was simulated by random sampling without replacement from an unrelated base population. Under each scenario, random mating of parents was simulated to obtain a second generation of 1900 progeny.

Table 4.1 Population Parameters for simulated pedigrees

	Sires	Dams per sire	no. of HS per sire	no. of FS per dam
Chicken	19	5	100	20
Pig	10	19	190	10
Human	633	1	-	3

A 20 cM chromosome was simulated with 5 markers spaced at 5 cM intervals and a bi-allelic QTL between the second and third marker at 7.5 cM. To simulate polygenic variance 10 unlinked additive effects of 0.2 were simulated each with an allele

Table 4.2 Summary of scenarios

	QTL effect		Heritability (h^2)		
	Additive	Dominant	Additive QTL	Dominant QTL	Total*
Scenario	(a)	(d)	σ^2_q / σ^2_P	σ^2_d / σ^2_P	$\sigma^2_a + \sigma^2_q + \sigma^2_d / \sigma^2_P$
1	0.00	0.00	0.00	0.00	0.11
2	0.10	0.00	0.00	0.00	0.12
3	0.20	0.00	0.01	0.00	0.13
4	0.30	0.00	0.03	0.00	0.14
5	0.00	0.50	0.00	0.04	0.15
6	0.40	0.00	0.04	0.00	0.16
7	0.00	0.60	0.00	0.05	0.16
8	0.00	0.70	0.00	0.07	0.18
9	0.50	0.00	0.07	0.00	0.18
10	0.00	0.80	0.00	0.09	0.19
11	0.50	0.50	0.07	0.03	0.21
12	0.60	0.60	0.09	0.05	0.24
13	0.80	0.40	0.16	0.02	0.27
14	0.70	0.70	0.12	0.06	0.27
15	0.80	0.50	0.15	0.03	0.28
16	0.80	0.60	0.15	0.04	0.29
17	0.80	0.70	0.15	0.06	0.30
18	0.80	0.80	0.15	0.07	0.31

σ^2_P = phenotypic variance. *total heritability includes polygenic variance (σ^2_a) of 0.2, residual variance 1.5, expected additive QTL variance (σ^2_q) estimated by ($a^2/2$), and expected dominant QTL variance (σ^2_d) estimated by ($d^2/4$)

frequency of 0.5 following Alfonso and Haley (1998). The phenotypes generated under a polygenic model were normally distributed indicating that these unlinked QTL were sufficient to provide a reasonably structured polygenic variance.

Dominance effects were simulated ranging from partial to overdominance over a range of additive effects. These are summarized in Table 4.2. The variances of the QTL additive (a) and dominance (d) effects were calculated as $0.5a^2$ and $0.25d^2$, respectively because allele frequencies were set to 0.5.For each individual, a residual effect was sampled from a normal distribution with mean 0 and a variance of 1.5. As the error variance was constant, phenotypic variance and overall heritability increased with genetic effects. In the base scenario with no QTL simulated, polygenic heritability was 0.11. Total heritability (polygenic and QTL) ranged from 0.1 to 0.31 with dominance QTL effects ranging from 0 to 9% of the phenotypic variance. For each scenario where QTL were simulated, 100 replicates were analysed and the test statistics described above calculated at 2,7,12, and 17cM.

4.2.5 Maternal effect

Common environment or direct maternal effects are often, at least partially, confounded with dominance. To explore the effect on detection of dominance QTL, random maternal effects were simulated in the pig population. A maternal effect was simulated by sampling for each full sib family from a normal distribution with variance of 0.1 and assigning this value to each full-sib offspring. A residual effect was sampled from a normal distribution with mean 0 and a variance of 0.75. The implication of potential maternal effects were evaluated in three different ways: a maternal effect was simulated with a range of dominance QTL effects. First, the maternal effect was not fitted in the model to test for spurious detection of dominance. Secondly, a maternal effect was included in the linear model to test whether the model correctly accounts for the maternal variance. Finally, no maternal effect was simulated but a maternal component was included in the linear model to test whether the dominance variance was correctly identified or incorrectly estimated as a maternal effect.

4.2.6 Null distribution

Chromosome-wise type 1 error rates were determined empirically for all three population structures. 1000 replicates were used to explore 1, 5 and 10% thresholds

under the null scenario (both additive and dominance QTL effects set to zero). Point-wise test statistics were determined with 1,000 replicates at the QTL position.

Empirical distributions for point-wise tests were compared to tabulated values for χ^2_k where k is equal to the number of extra variance components estimated using *P* values for 1, 5, and 10%. A 5% threshold was determined for empirical power calculations and comparisons based on the analysis of 1000 replicates.

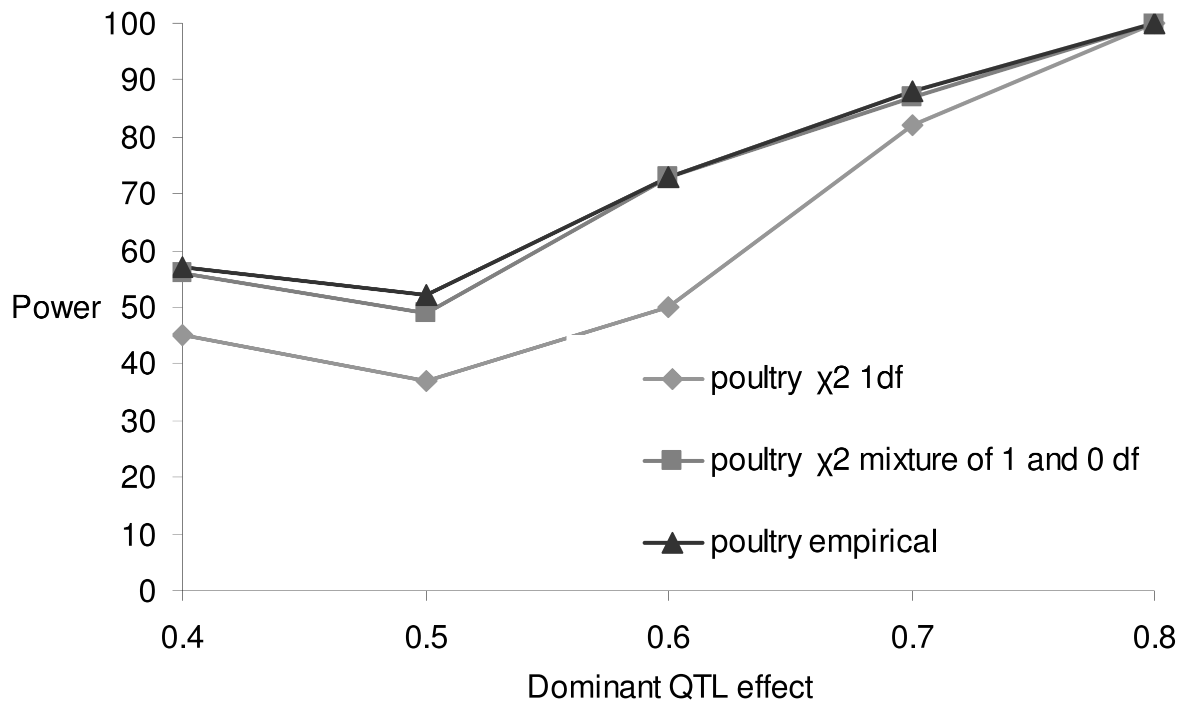
The empirical distribution of the LRT under the null scenario appeared to vary across pedigrees, in particular, differing between human and livestock. This could be due to the difference in number of full sibs per family (three for humans compared with ten or twenty for livestock) or the lack of half sib relationships in the human pedigree. To explore this, chromosome-wise null LRT distributions were determined for five additional pedigrees. The number of offspring remained constant at 1900 but pedigrees with 1 or 2 dams per sire and 3, 5 or 10 offspring per dam were compared with pig and poultry pedigrees to explore the effects of family structure on the distribution of the null test statistic.

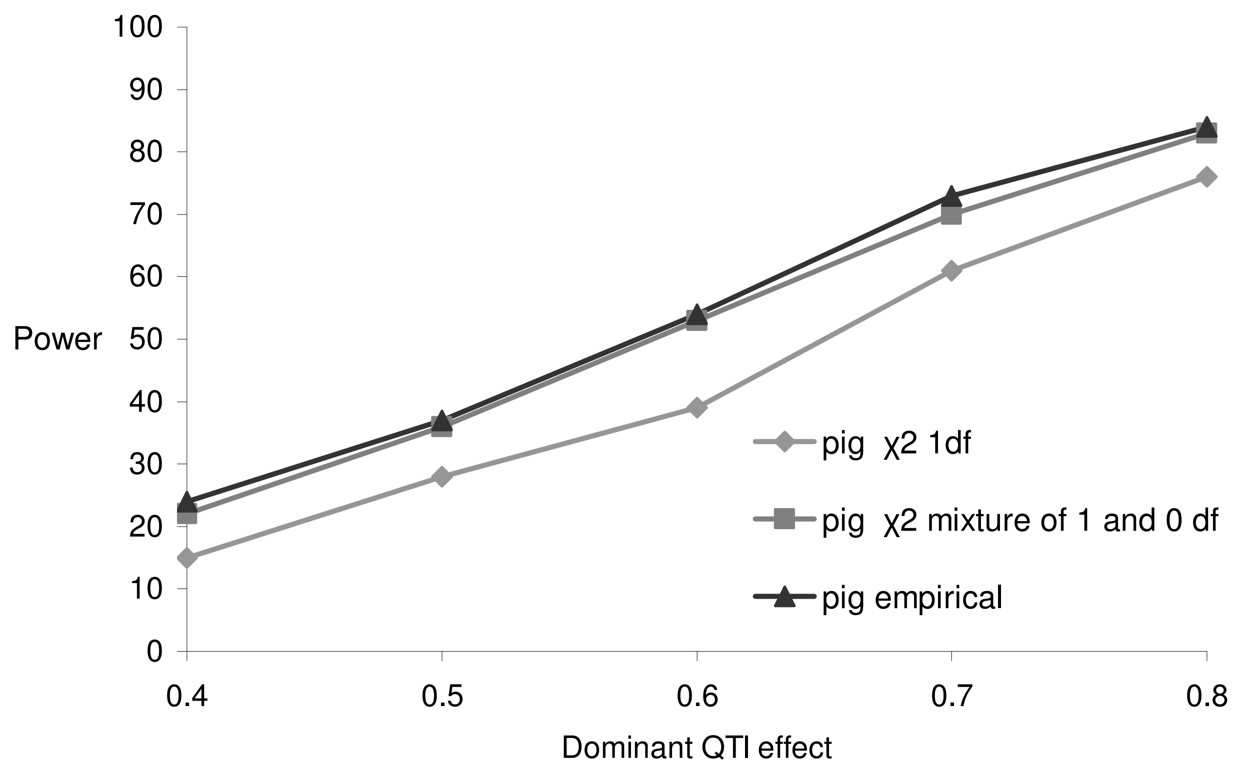
4.3. Results

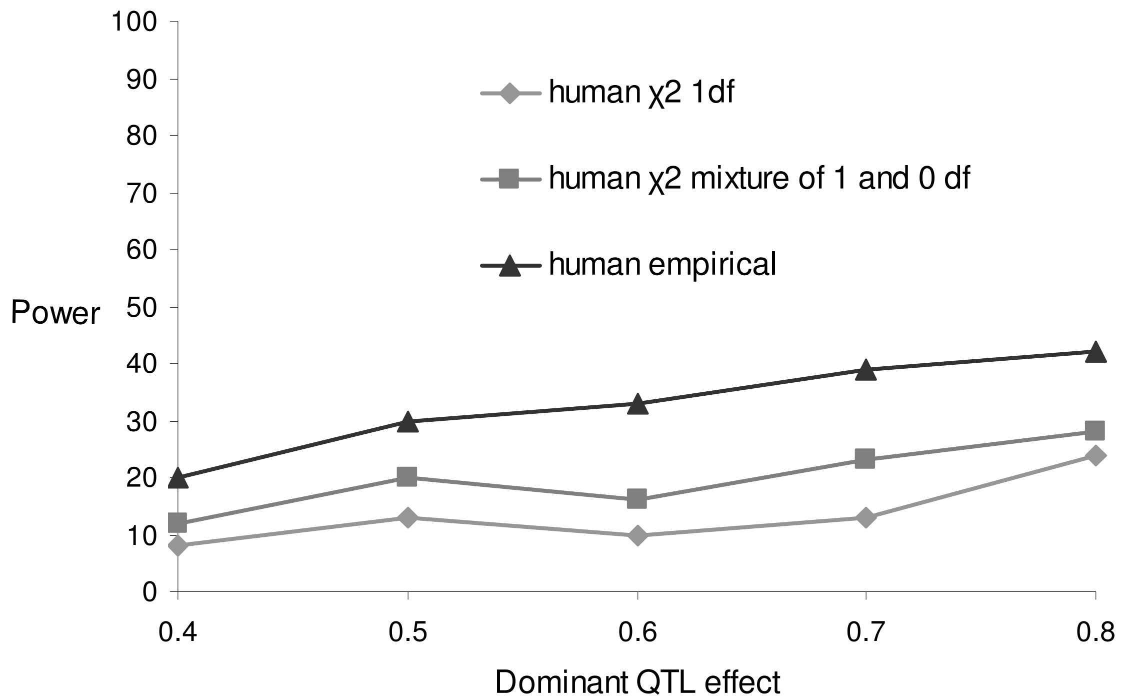
All results shown are based on 5% empirical thresholds from 1000 chromosome-wise replicates. Full results for all populations and scenarios can be found in Appendix 4.1 for power to detect additive and dominant QTL effects, Appendix 4.2 for power in pig populations with maternal or common environment effects and Appendix 4.3 for estimates of variance components.

4.3.1 Power to detect dominance effects

Figures 4.1a, 4.1b and 4.1c give the proportion of replicates detecting dominance using 5% empirical and tabulated thresholds for χ^2_1 and a mixture $\chi^2_{1-0.}$. Testing for dominance involved comparing the additive QTL model with the full model incorporating both an additive QTL and a dominant QTL effect (denoted 2v1). For


100
90
80
70
60
Power 50
40
30
20
10
0
0.4
0.5
0.6
0.7
0.8
Dominant QTL effect
poultry χ2 1df
poultry χ2 mixture of 1 and 0 df
poultry empirical


100
90
80
70
60
Power 50
40
30
20
10
0
0.4
0.5
0.6
0.7
0.8
Dominant QTI effect
pig χ2 1df
pig χ2 mixture of 1 and 0 df
pig empirical



Figures 4.1 a, b and c. Proportion of replicates where test for dominance (2v1) is significant (P < 0.05) when comparing full model vs additive model. 100 chromosome-wise replicates in top to bottom (a) poultry, (b) pig, and (c) human pedigrees under partial to complete dominance. Simulated additive effect fixed at 0.8. comparing tabulated 5% χ^2_1, $\chi^2_{1\text{-}0}$ thresholds and 5% empirical threshold. Mixture threshold is estimated by using tabulated 10% χ^2_1 threshold

the pig and poultry pedigrees power under empirical and the $\chi^2_{1\text{-}0}$ mixture of distributions was in close agreement. For human pedigrees both χ^2_1 and $\chi^2_{1\text{-}0}$.were conservative when compared to empirical thresholds. Power under the empirical 5% threshold was ~100% in the poultry scenario, ~84% for the pig scenario and ~ 42% for humans when the QTL dominance variance was around 7% of the phenotypic variance (dominant effect = 0.8, i.e. complete dominance). Under χ^2_1 thresholds this dropped to ~95, ~75 and ~25. Power to detect dominance was greater for all pedigrees using empirical thresholds. Although the ranking did not change when using tabulated values, power to detect dominance, particularly in humans was much lower and differences between models greater. When comparing the full model with the null model (denoted 2v0) all replicates detected a QTL for the pig and poultry scenarios, and 96% (84% under tabulated thresholds) of replicates detected a QTL for the human scenario (Appendix 4.1).

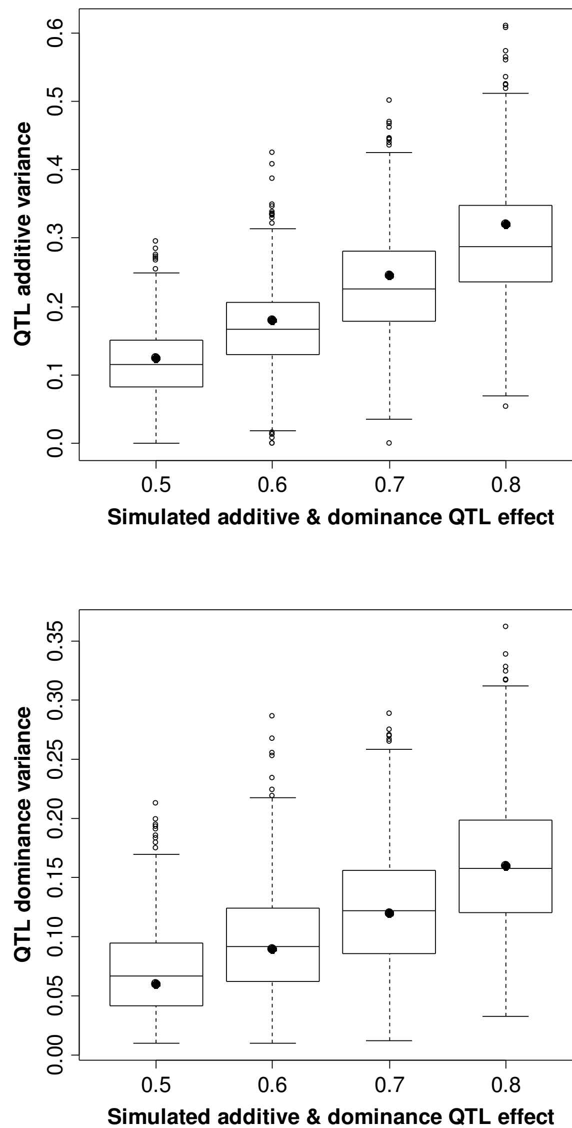


Figure 4.2. Estimates of variance components from simulated poultry data. Box plots showing the range of variance estimates. Full dominance is simulated. Variance estimates for single marker position (for 1000 replicates of each scenario) for additive and dominant QTL effects. The black circles indicate the expected variance components. All replicates were significant for a QTL when testing under the full model (additive and dominance QTL effects vs null)

Figure 4.2 shows the estimates for the additive and dominance QTL variance components from the comparison of the full model with the null model (2v0). In all replicates a QTL was detected at the 5% significance level. Scenarios shown were for complete dominance with effects ranging from 0.5 to 0.8 (also given in Appendix

4.2). These show that although estimates are wide ranging they appear to accurately estimate the mean.

4.3.2 Over dominant, spurious additive and dominant QTL effects

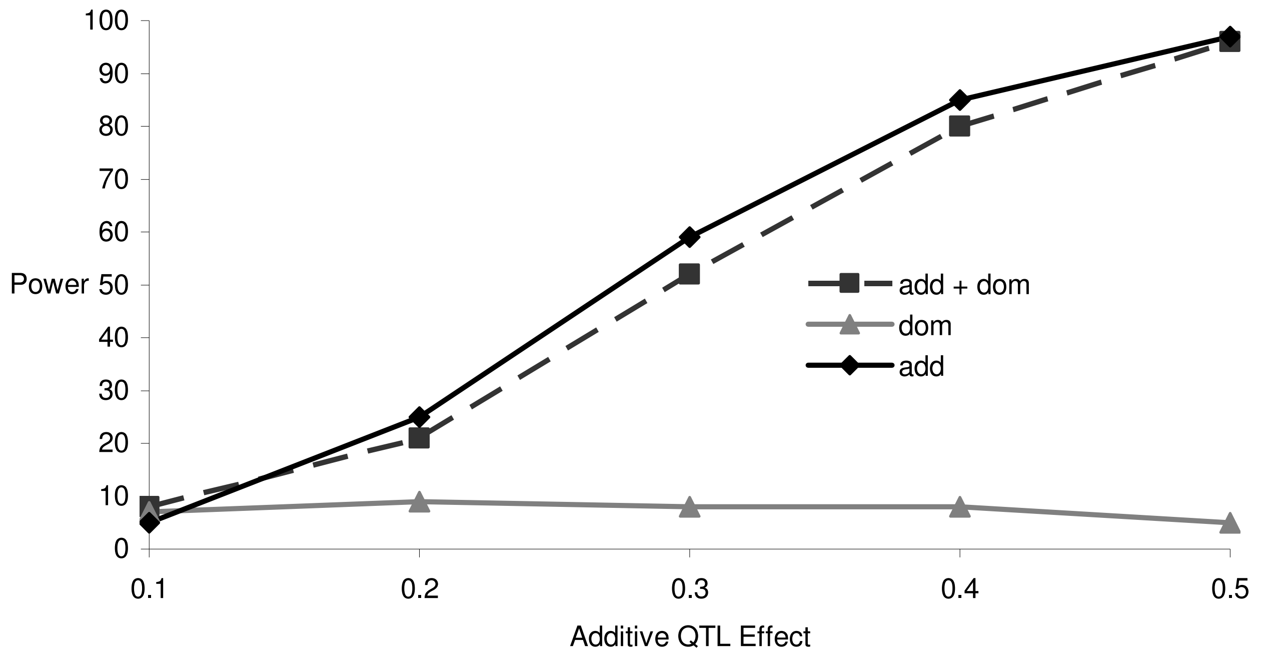


Figure 4.3. Percentage of replicates detecting additive QTL effects ($p<0.05$) using full model (add + dom), additive model (add) and testing difference between the two (dom) in simulated pig population. A dominance effect of zero is simulated

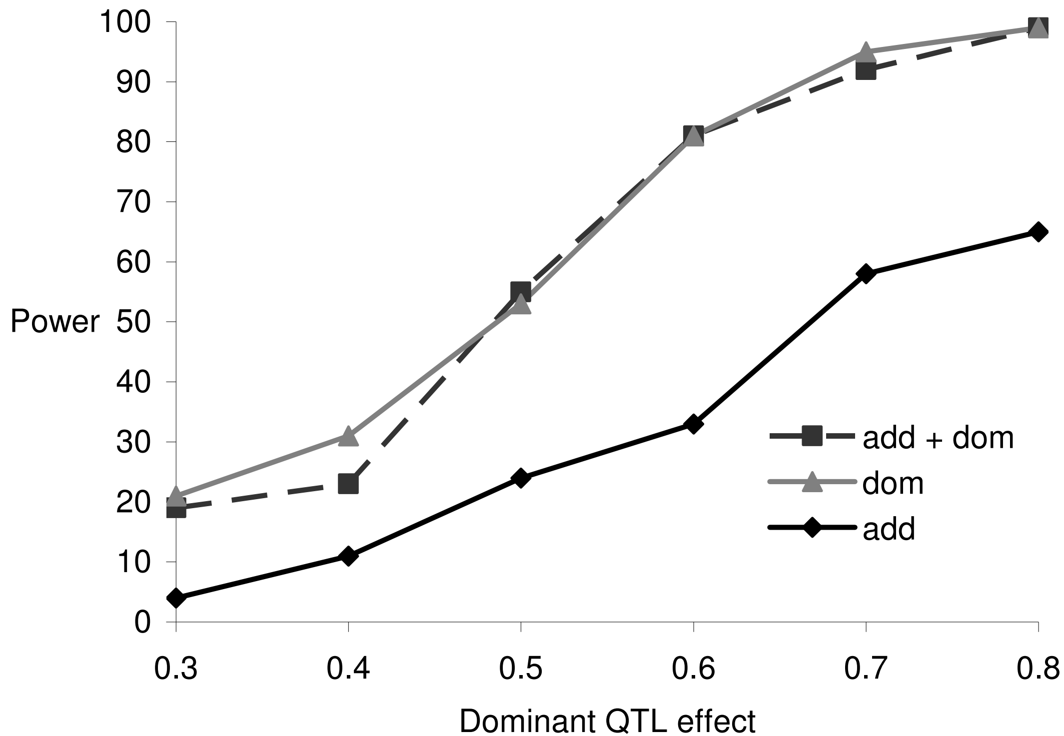


Figure 4.4. Over dominance. Percentage of replicates detecting over dominant QTL effects (p<0.05) using full model (add + dom), additive model (add) and testing difference between the two (dom) in simulated poultry population over a range of dominant QTL effects when an additive effect of zero is simulated.

Figure 4.3 shows power to detect simulated additive effects ranging from 0.1 to 0.5, or 1 – 7% of phenotypic variance. Power reached 90% when the additive variance amounted to >4% of the phenotypic variance. In this case, no dominance effect was simulated and there was little spurious dominance detected. Furthermore, power to detect an additive effect was similar whether or not the extra dominance component was included in the analysis. This shows that a routine scan including dominance would not result in too great a loss of power even in the absence of any dominant effects. Although results are shown only for the pig population, the same pattern was seen for poultry and human scenarios. Figure 4.4, however, shows that spurious additive QTL effects were found when dominant QTL effects were not fitted. With dominant QTL effects ranging from 0.3 to 0.8 and simulated additive effects of zero i.e. overdominance, there is both spurious detection of an additive QTL effect if dominance was not included in the model and inflated estimates of additive variance (Table 4.3).

Table 4.3 Estimates of variance due to additive QTL and additive and dominant QTL effects under over-dominance when additive QTL effect of zero is simulated

Expectation		Additive QTL Vs null model (1v0)	Additive + Dominant QTL Vs null model (2v0)	
Add	Dom	Add	Add	Dom
0	0.06	0.02	0.01	0.07
0	0.09	0.04	0.01	0.10
0	0.12	0.04	0.01	0.12
0	0.16	0.07	0.01	0.16

Variances are mean estimates at highest chromosome-wise test statistic across 100 iterations in simulated pig population

4.3.3 Maternal effects and dominance in pig scenario

Table 4.4. Variance estimates for dominant QTL effect of 0.8 and maternal effects

Maternal effect Simulated	Maternal effect fitted	Dom variance (0.16)	Maternal variance (0.10)
N	N	0.17	-
Y	N	0.25	-
Y	Y	0.20	0.09
N	Y	0.17	0.02

Figures in brackets are simulated or expected variance. Variances are mean estimates at highest chromosome-wise test statistic across 100 iterations in a simulated pig population

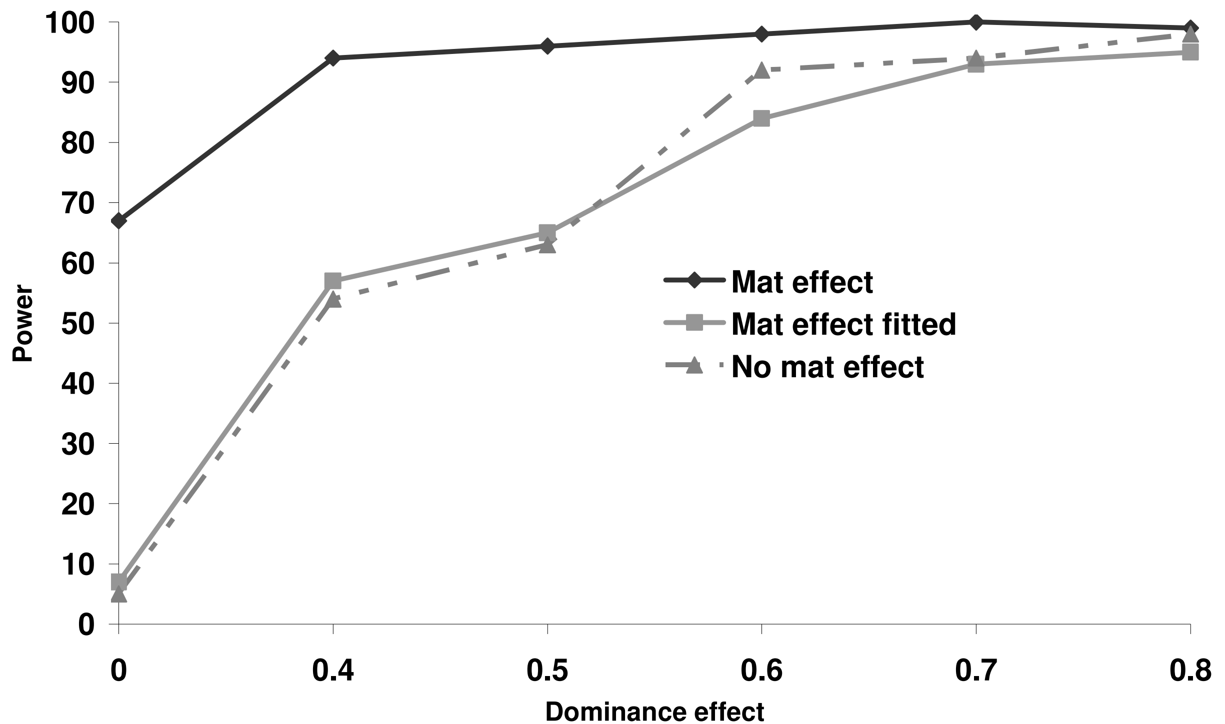


Figure 4.5. Effects of simulating and/or fitting direct maternal effects on proportion of replicates where test for dominance (2v1) is significant (P < 0.05) when comparing full model vs additive model. 100 chromosome-wise replicates in pig population under partial to complete dominance (additive QTL effect fixed at 0.8). No mat effect – no maternal effect simulated or fitted, mat effect – maternal variance of 0.1 simulated but not fitted, mat effect fitted – maternal variance of 0.1 simulated and fitted.

Figure 4.5 clearly shows that a simulated maternal effect can masquerade as a dominant QTL effect. In the most extreme case when a dominance effect of 0 was simulated, ignoring common environment resulted in a type 1 error of 67%. When a maternal effect was fitted in the model, maternal effects and dominance appear to be successfully separated with little or no loss of power (Table 4.5). If a maternal effect was fitted when not present there was little loss of power (for results see Appendices 2 and 3) indicating that a maternal component fitted in the absence of a maternal effect should not prevent detection of variance due to dominance.

4.3.4 Null distribution

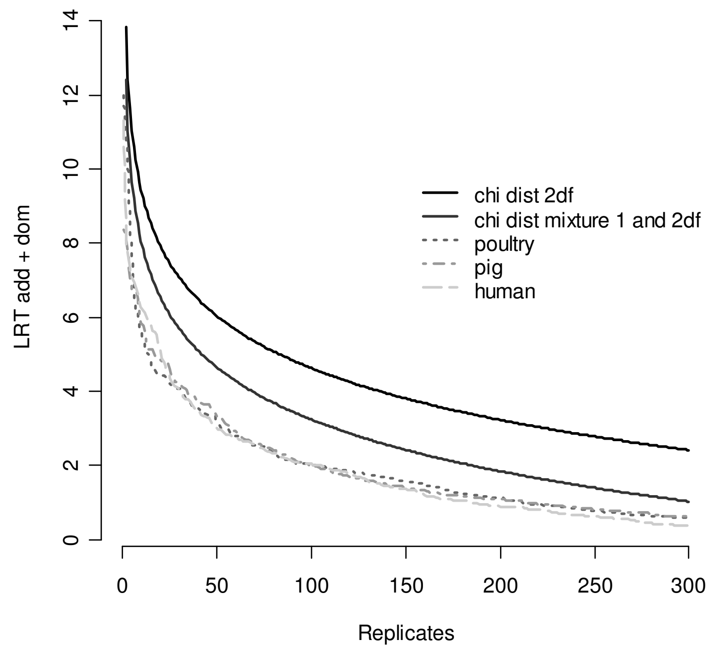
Table 4.5 Empirical 5% thresholds for LRT test statistic (and corresponding P value under χ^2 distribution). 1000 replicates simulated for single point-wise and multiple chromosome-wise testing under null scenario of no QTL effects

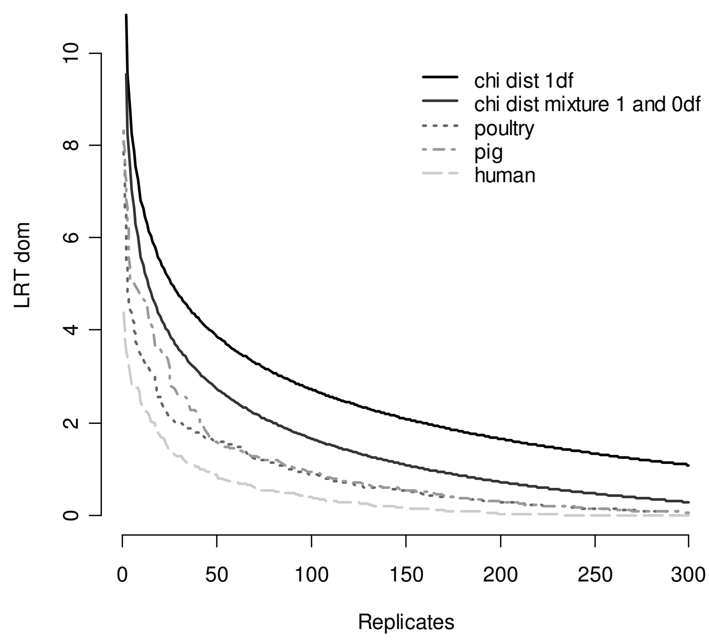
	Point-wise						Chromosome-wise					
	Poultry		**Pig**		**Human**		**Poultry**		**Pig**		**Human**	
2v0	**3.18**	(0.20)	**3.30**	(0.19)	**3.00**	(0.22)	**4.86**	(0.09)	**4.94**	(0.08)	**4.58**	(0.10)
2v1	**1.60**	(0.21)	**1.58**	(0.21)	**0.86**	(0.35)	**2.62**	(0.11)	**2.70**	(0.10)	**1.48**	(0.22)
1v0	**2.20**	(0.14)	**2.08**	(0.15)	**2.62**	(0.11)	**3.78**	(0.05)	**3.46**	(0.06)	**3.78**	(0.05)

2v0 testing model including an additive QTL and a dominant QTL effect versus null model with P value in brackets assuming χ^2_2 ; 2v1 testing model including an additive and a dominant QTL effect versus an additive QTL model with P value in brackets assuming χ^2_1 and 1v0 testing a model including an additive QTL versus a null model with P value in brackets assuming χ^2_1.

Table 4.5 shows that the point-wise test statistic was conservative for all models when compared to tabulated χ^2 values. This was also true if the mixture of distributions was taken into account by assuming a *P* value of 0.1 to derive a 5% critical threshold. Table 4.5 shows that for the test for dominance the equivalent P value under the χ^2_1 distribution to a 5% empirical threshold was actually 0.21 for pig and poultry and 0.35 for human pedigrees. The test for additive QTL effects although still conservative was much closer to tabulated values, particularly 10% thresholds with equivalent P value under the χ^2_1 distributions of 0.14, 0.15 and 0.11. for poultry, pig and human pedigrees respectively.

Chromosome-wise type 1 error rates were close to tabulated thresholds for the additive model, for all three simulated pedigrees. Type 1 error rates for the full model and for dominance, however, remained conservative when compared to tabulated values. None of these type 1 error rates were corrected for multiple testing.


LRT add + dom
Replicates
chi dist 2df
chi dist mixture 1 and 2df
poultry
pig
human
0
2
4
6
8
10
12
14
0
50
100
150
200
250
300


LRT dom
Replicates
chi dist 1df
chi dist mixture 1 and 0df
poultry
pig
human
0
2
4
6
8
10
0
50
100
150
200
250
300

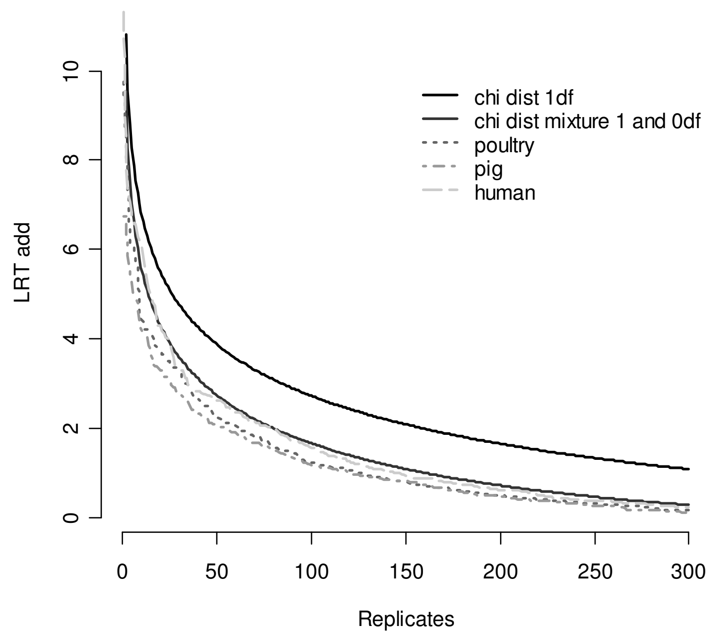


Figure 4.6. Distribution of empirical point-wise test statistic in pig, poultry and human pedigrees for top to bottom (a) additive and dominance effects, (b) dominance effects and (c) additive effects compared with χ^2_1 and $\chi^2_{1\text{-}0}$ distributions. Top 300 values of 1000 replicates are displayed.

Figure 4.6 compares distributions of the empirical test statistic with χ^2 distributions. In particular for the test for dominance the empirically derived null statistic appears to vary according to pedigree structure i.e. is lower in humans. This is apparent to a lesser extent in the model testing for both additive and dominance and somewhat reversed in the additive test.

Table 4.6 Empirical 5% thresholds for LRT test statistic when testing for dominance and corresponding P value under χ^2_1 distribution. 1000 replicates simulated for chromosome-wise testing under null scenario of no QTL effects

Pedigree	Sires	Dams per sire	Progeny per dam	LRT 5% empirical threshold
1 (human)	633	1	3	1.46
2	380	1	5	1.14
3	190	1	10	1.38
4	317	2	3	2.1
5	190	2	5	2.02
6	195	2	10	2.06
7 (chick)	19	5	20	2.62
8 (pig)	10	19	10	2.70

Figure 4.7/Table 4.6 compares distributions of the test statistic under the null hypothesis of no QTL for 8 pedigree structures. These vary from human families with 3 full sibs to more complex structures such as poultry with 20 full-sibs and 200 half sibs. It is apparent that the three human pedigrees i.e. single dam families have very similar distributions seemingly regardless of the number of full sibs. Similarly the three pedigrees with two dam families have similar distributions to each other but clearly different from those of the human or larger livestock families. The pig and poultry distributions are similar to each other although the pig distribution appears slightly more conservative. The $\chi^2_{1\text{-}0}$ distribution appears comparable with the pig and poultry although as these were chromosome-wise tests which are uncorrected for multiple testing they cannot be directly compared.

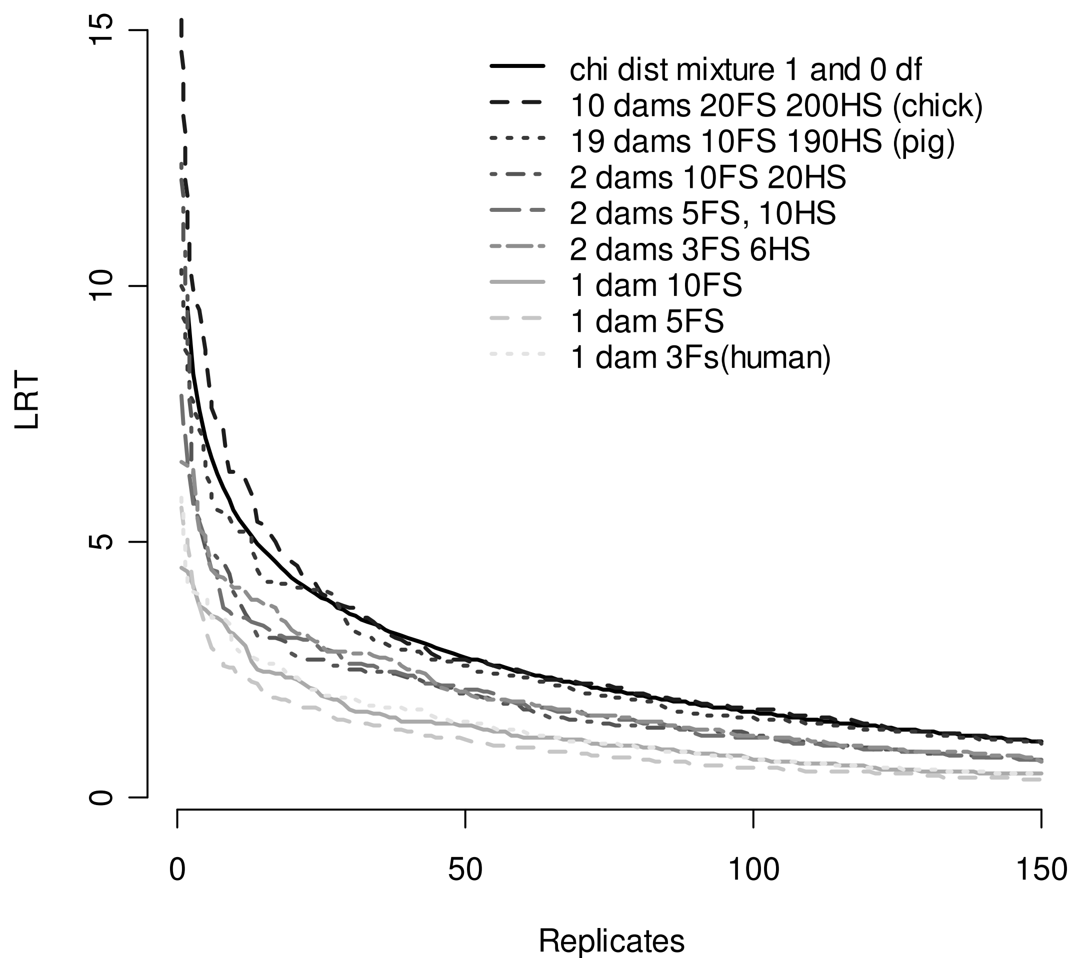


Figure 4.7. Comparison of distribution of empirical chromosome-wise test statistic for dominance effects under null hypothesis of no QTL in pedigrees with varying full sib (FS) and half sib (HS) structures. χ^2_{1-0}.is also plotted for comparison. All pedigrees have 1900 total offspring. Top 150 values of 1000 replicates are displayed for clarity.

4.4 Discussion

This study provides a comprehensive evaluation of the performance of variance component analysis over a range of dominant QTL effects. The method was successfully used to estimate and apportion variances due to polygenic, additive, dominant and non-genetic effects. Power > 95% was achieved when testing for dominant QTL effects accounting for 7% of the total variance for a simulated poultry population. Power to detect additive QTL was also high (~ 97% for an additive effect

explaining 7% of phenotypic variance. Although the upper range of the simulated QTL effects is high, these values are plausible in terms of published literature.

Simulation results showed that, unlike the test for additive QTL effects, the empirical distribution of the test statistic when testing for dominant QTL effects did not behave in accordance with existing theoretical expectations. Theoretically, the asymptotic distribution of the LRT is a mixture of chi squared with different degrees of freedom when testing variance components under the null hypothesis that they are zero (Self & Liang 1987; Stram & Lee 1994). For example with one extra variance component the null distribution is a mixture of ½ χ_0 (i.e.variance is zero) and ½ χ_1 (i.e. variance is non-zero. With a model including two variance components, such as additive and dominant QTL effects the expectation of the distribution would be a mixture of ¼ χ_0 (both variance components are zero, ½ χ_1 (one is non zero) and ¼ χ_2 (both are non zero). Visscher (2006) provides a thorough review.
Problems with incorrect assumptions about the distribution include inflated type II errors leading to reduced power. Extending the linear model to include a dominance component resulted in a conservative test when imposing a χ^2_1 distribution for the likelihood ratio test statistic. The test remained conservative even if thresholds were halved under the assumption of a mixture of distributions. One explanation might be that additive and dominant QTL effects are not entirely independent. Furthermore, the null distribution for the dominance test varies with family structure, in particular, with the number of dam families per sire. Distributions of the test statistic in 4. 7/Table 4.6 appear to group by number of dam families, with human type pedigrees with a single dam per sire most conservative, regardless of family size. The number of full sibs within dam families did not appear to affect the distribution. It is possible that the lack of half sib families, might result in confounding of additive and dominance effects at the QTL. Theoretically if both components need to be estimated within dam, lack of information might lead to a higher probability of variance components being zero.

Results showed that power is also affected by population structure. Power to detect dominance at the QTL was similar for pig and poultry populations but much lower for humans.

This was unsurprising as the human population consisted of many small families with low numbers of full sibs making it difficult to detect dominance. Increased power might be achieved in human studies from a pedigree with more generations providing information from relationships such as grandparents and cousins but this needs to be explored further.

It is anticipated that further correction for multiple testing for large linkage groups or genome wide testing would be necessary. The distribution, however, of H_0 when testing for multiple linked positions is unresolved and authors have used different approximations (Xu & Atchley 1995; Pratt *et al.* 2000; Piepho 2001; Nagamine *et al.* 2004), Procedures such as permutation and bootstrapping enable the setting of empirical thresholds and circumvent problems associated with failure of distributional assumptions and independence of multiple tests (Churchill & Doerge 1994; Visscher *et al.* 1996), although computational complexity can restrict their use within the variance component framework.

The method described by Piepho relies on the gradient of change in likelihood. However, this method still assumes that the test statistic for a single test follows a standard Chi-square distribution under the null hypothesis and therefore does not address the issue of mixture distributions that is apparent for these types of analyses. It is difficult to ascertain whether the method is appropriate here, when the test statistic follows a mixture of distributions and likelihoods under the null scenario are very flat.

There is strong evidence to suggest that a common environment effect should be routinely evaluated in all variance component QTL models as, if unaccounted for, most variation due to common environment masquerades as dominance. I have shown that presence of common environmental effects has little effect on false negative rates but a potentially huge impact on false positive rates.

I have demonstrated that incorporating a dominance effect in a genome scan has very limited detrimental effect on the power to detect purely additive QTL. Detection of spurious dominance was also rare suggesting that dominance could be routinely included in genome scans. I have also shown by simulation that, if not fitted in the

analysis model, dominance may be detected as spurious additive effects or inflated estimates of additive genetic variance. This suggests that dominant QTL effects can be detected as additive QTL when additive-only models are used; see also Misztal *et al*. (1998) and Pante *et al*. (2002) for similar effects with polygenic dominance. This has important implications for predicting response to selection as the success of any selection programme is dependent on correctly identifying the mode of inheritance and proportion of variance explained by the QTL. For example, Hayes and Miller (2000) show that including dominance effects in mate selection can be a powerful tool for exploiting previously untapped genetic variation while Dekkers and Chakraborty (2004) discuss maximization of crossbred performance by incorporating information from overdominant QTL.

A further confounding factor not studied here might be polygenic dominance. It is, however, unlikely to have affected the results as most of the information for polygenic dominance would have come from the covariance of full sibs and should have been accounted for by the common environment effect. This might not be the case within other relationships in deeper, more complex, pedigrees suggesting that the inclusion of a polygenic dominance effect may be valuable when examining such data structures.

4.5 Conclusions

Variance component methods were implemented to detect dominant QTL. Type 1 error rates and power were explored using extensive simulation. Results indicate that if the mixture of distributions is taken into account nominal chi square thresholds were appropriate for when testing for additive QTL but conservative when testing for dominant QTL in all pedigrees and particularly in the case of the populations with sires mated to only one or two dams. Ascertaining the correct null distribution is a difficult issue but one that merits revisiting. Here I have shown that although theoretically the tabulated chi square values are fairly robust, the expected probability of non zero variances varies with population structure, thus there are instances when greater power is achieved by empirically deriving the correct distribution of the test statistic. Power to detect dominant QTL effects was high in livestock pedigrees with

little spurious dominance and could be successfully routinely employed under the proviso that common environment or direct maternal effects are accounted for.

Effects of extra generations or extended pedigrees are yet to be explored but may provide greater power for structures with few dam families.

Appendix 4.1 Power (proportion of replicates) to detect significant additive and dominant QTL for Poultry, Human and Pig Scenarios across range of simulated additive and dominant QTL effects at 5 and 1% thresholds based on tabulated χ^2 values for likelihood ratio statistic

				Power Additive + Dominant QTL Vs null test (2df)							
Genetic effect		QTL	QTL	Poultry		Human		Pig		Poultry point-wise	
Additive	Dominant	variance	h^2	$p<0.01$	$p<0.05$	$p<0.01$	$p<0.05$	$p<0.01$	$p<0.05$	$p<0.01$	$p<0.05$
0.0	0.0	0.00	0.00	1	4	0	1	0	4	0.1	1
0.1	0.0	0.01	0.00	1	4	1	3	2	4		
0.2	0.0	0.02	0.01	3	14	1	3	11	22		
0.0	0.3	0.02	0.01	2	11						
0.0	0.4	0.04	0.02	9	18						
0.3	0.0	0.05	0.03	16	34	1	7	25	45		
0.0	0.5	0.06	0.04	19	41						
0.1	0.5	0.07	0.04	25	48					16.7	37
0.4	0.0	0.08	0.04	56	75	1	4	56	75		
0.2	0.5	0.08	0.05	39	61					30	53
0.0	0.6	0.09	0.05	45	73						
0.3	0.5	0.11	0.06	60	81					52.3	71
0.0	0.7	0.12	0.07	82	89						
0.5	0.0	0.13	0.07	80	90	5	12	87	93		92
0.4	0.5	0.14	0.08	82	92					78.3	90
0.0	0.8	0.16	0.09	95	98						
0.5	0.5	0.19	0.10	98	99			93	99	95.5	98
0.6	0.6	0.27	0.14	99	100					99.3	100
0.8	0.4	0.36	0.17	100	100	43	68	100	100		
0.7	0.7	0.37	0.18	100	100						
0.8	0.5	0.38	0.18	100	100	55	70	100	100		
0.8	0.6	0.41	0.19	100	100	52	76	100	100		
0.8	0.7	0.44	0.21	100	100	59	77	100	100		
0.8	0.8	0.48	0.22	100	100	64	81	100	100		

Power Additive + Dom QTL Vs Additive QTL test (1df)								Power Additive QTL Vs null test (1df)							
Poultry		Human		Pig		Poultry point-wise		Poultry		Human		Pig		Poultry point-wise	
p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05
6	4	0	1	0	1	0.1	1			0	6	1	8	0.1	3
5	1	0	1	1	5				4	2	3	2	2		
10	5	0	1	4	6				23	2	5	11	28		
20	15							14	4						
30	21							17	11						
8	3	0	2	0	3					2	12	36	57		
52	38							35	24						
60	48					14.8	32	15	37					8.1	19
9	5	0	1	3	5				84	3	9	69	82		
60	55					16.3	36	14	30					19.7	37
80	66							45	31						
55	44					16.5	34	49	77					45.3	67
95	91							68	57						
3	2	0	1	15	2		4			8	21	76	95		95
58	48					14.4	35	76	90					77.4	89
99	98							78	64						
56	45			17	28	10	23	98	100			93	97	95.3	98
62	55					33.9	57	99	100					99.6	100
30	21	3	7		14			100	100	58	82	100	100		
92	88							100	100						
49	37	2	12	11	27			100	100	64	81	100	100		
73	50	2	9	21	38			100	100	62	83	100	100		
88	84	4	12	42	60			100	100	72	88	100	100		
99	95	8	23	60	75			100	100	78	93	100	100		

Poultry, Pig and Human scenarios are chromosome-wise involving 100 replicates. Poultry point-wise involves a single test

Empirical Power (proportion of replicates) to detect significant additive and dominant QTL for Poultry, Human and Pig Scenarios across range of simulated additive and dominant QTL effects at 5 and 1% thresholds based on critical null threshold from 1000 replicates for likelihood ratio statistic

Genetic effect		QTL		Power (*P*<0.01) Additive + dominant QTL vs null test (2v0)			Power (P<0.01) Additive + dominant QTL vs additive test (2v1)			Power (P<0.01) Additive QTL vs null test (1v0)		
Additive	Dominant	variance	QTL h^2	Poultry	Pig	Human	Poultry	Pig	Human	Poultry	Pig	Human
0	0	0.000	0.00		0	0		0	2		0	0
0.1	0	0.005	0.00	0	2	1	0	2	1	1	2	1
0.2	0	0.020	0.01	6	13	2	1	4	1	8	9	2
0.3	0	0.045	0.03	25	32	3	2	1	3	33	28	2
0.4	0	0.080	0.04	56	61	2	2	4	2	63	62	2
0.5	0	0.125	0.07	87	88	6	0	16	5	92	73	6
0.1	0.1	0.008	0.00		5			3			3	
0.2	0.2	0.030	0.02		12			3			9	
0.3	0.3	0.068	0.04		39			6			36	
0.4	0.4	0.120	0.07		72			3			67	
0.5	0.5	0.188	0.10	98	95		84	21		99	90	
0.6	0.6	0.270	0.14	99			94			99		
0.7	0.7	0.368	0.18	100			99			100		
0.8	0.8	0.480	0.22	100	100	70	100	66	27	100	100	72
0.4	0.8	0.240	0.12	100	99	50	35	8	11	100	99	58
0.5	0.8	0.285	0.14	100	100	59	18	15	19	100	98	60
0.6	0.8	0.340	0.17	100	100	59	33	29	13	100	100	59
0.7	0.8	0.405	0.19	100	100	64	68	46	23	100	100	69
0.8	0.8	0.480	0.22	100	100	70	100	66	27	100	100	72
0	0.3	0.023	0.01	3			5			1		
0	0.4	0.040	0.02	10			10			2		
0	0.5	0.063	0.04	23	9		27	14		5	0	
0	0.6	0.090	0.05	54	24		52	40		19	2	
0	0.7	0.123	0.07	83	46		74	57		31	2	
0	0.8	0.160	0.09	96	77		92	85		47	9	

Genetic effect		QTL		Power (P<0.05) Additive + dominant QTL vs null test (2v0)			Power (P<0.05) Additive + dominant QTL vs additive test (2v1)			Power (P<0.05) Additive QTL vs null test (1v0)		
Additive	Dominant	variance	QTL h^2	Poultry	Pig	Human	Poultry	Pig	Human	Poultry	Pig	Human
0	0	0.000	0.00	9	6	4	9	8	9	5	2	4
0.1	0	0.005	0.00	21	23	4	10	8	7	24	30	5
0.2	0	0.020	0.01	52	51	11	9	8	8	55	57	13
0.3	0	0.045	0.03	82	79	9	8	9	12	84	83	11
0.4	0	0.080	0.04	96	93	19	5	20	9	97	79	21
0.5	0	0.125	0.07		13			10			12	
0.1	0.1	0.008	0.00		23			11			24	
0.2	0.2	0.030	0.02		61			14			62	
0.3	0.3	0.068	0.04		91			15			92	
0.4	0.4	0.120	0.07	100	99		99	37		100	97	
0.5	0.5	0.188	0.10	100			97			100		
0.6	0.6	0.270	0.14	100			100			100		
0.7	0.7	0.368	0.18	100	100		100	83		100	100	
0.8	0.8	0.480	0.22	100	100	86	100	83	41	100	100	93
0.4	0.8	0.240	0.12	100	99	78	57	24	19	100	99	82
0.5	0.8	0.285	0.14	100	100	80	52	36	29	100	98	82
0.6	0.8	0.340	0.17	100	100	82	73	53	31	100	100	85
0.7	0.8	0.405	0.19	100	100	84	88	70	38	100	100	89
0.8	0.8	0.480	0.22	100	100	86	100	83	41	100	100	93
0	0.3	0.023	0.01	19			21			4		
0	0.4	0.040	0.02	23			31			11		
0	0.5	0.063	0.04	55	21		53	39		24	4	
0	0.6	0.090	0.05	81	52		81	62		33	15	
0	0.7	0.123	0.07	92	73		95	84		58	16	
0	0.8	0.160	0.09	99	91		99	93		65	30	

Genetic effect		QTL		Power (P<0.10) Additive + dominant QTL vs null test (2v0)			Power (P<0.10) Additive + dominant QTL vs additive test (2v1)			Power (P<0.10) Additive QTL vs null test (1v0)		
Additive	Dominant	variance	QTL h^2	Poultry	Pig	Human	Poultry	Pig	Human	Poultry	Pig	Human
0	0	0.000	0.00	13	11	14	13	14	18	17	11	13
0.1	0	0.005	0.00	33	37	13	16	16	17	39	43	11
0.2	0	0.020	0.01	68	61	22	16	13	16	74	70	22
0.3	0	0.045	0.03	90	86	22	17	12	21	91	91	21
0.4	0	0.080	0.04	97	97	31	16	25	17	97	81	36
0.5	0	0.125	0.07		19			15			16	
0.1	0.1	0.008	0.00		32			24			32	
0.2	0.2	0.030	0.02		73			28			74	
0.3	0.3	0.068	0.04		95			30			92	
0.4	0.4	0.120	0.07	100	99		99	46		100	98	
0.5	0.5	0.188	0.10	100			100			100		
0.6	0.6	0.270	0.14	100			100			100		
0.7	0.7	0.368	0.18	100	100		100	88		100	100	
0.8	0.8	0.480	0.22	100	100	90	100	88	58	100	100	96
0.4	0.8	0.240	0.12	100	99	84	68	36	28	100	99	89
0.5	0.8	0.285	0.14	100	100	89	65	47	45	100	98	92
0.6	0.8	0.340	0.17	100	100	89	84	63	42	100	100	88
0.7	0.8	0.405	0.19	100	100	91	93	82	50	100	100	91
0.8	0.8	0.480	0.22	100	100	90	100	88	58	100	100	96
0	0.3	0.023	0.01	28			34			17		
0	0.4	0.040	0.02	38			50			22		
0	0.5	0.063	0.04	70	42		63	56		40	12	
0	0.6	0.090	0.05	89	68		89	74		53	20	
0	0.7	0.123	0.07	94	87		97	91		74	21	
0	0.8	0.160	0.09	99	97		99	98		81	43	

Poultry, Pig and Human scenarios are chromosome-wise involving 100 replicates

Appendix 4.2 Power (proportion of 100 replicates) to detect significant additive and dominant QTL for Maternal Scenarios using pig scenario across range of simulated additive and dominant QTL effects at 5 and 1% thresholds based on tabulated χ^2 values for likelihood ratio statistic.

				Power Additive + Dominant QTL Vs null test (2df)							
Genetic effect		QTL	QTL	Maternal effect simulated not fitted		Maternal effect simulated and fitted		Maternal effect fitted		No Maternal effect	
Additive	Dominant	variance	h^2	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05
0.0	0.0	0.00	0.00	32	47	1	4	0	2	0	4
0.1	0.0	0.01	0.00	47	66	0	12	2	9	2	4
0.2	0.0	0.02	0.02	63	86	26	41	18	32	11	22
0.3	0.0	0.05	0.04	89	94	65	80	63	74	25	45
0.4	0.0	0.08	0.07	100	100	93	97	86	95	56	75
0.5	0.0	0.13	0.11	100	100	99	100	98	99	87	93
0.8	0.4	0.36	0.26	100	100	52	52	100	100	100	100
0.8	0.5	0.38	0.27	100	100	100	100	100	100	100	100
0.8	0.6	0.41	0.28	100	100	100	100	100	100	100	100
0.8	0.7	0.44	0.30	100	100	100	100	100	100	100	100
0.8	0.8	0.48	0.31	100	100	100	100	100	100	100	100

Power Additive + Dom QTL Vs Additive QTL test (1df)								Power Additive QTL Vs null test (1df)							
Maternal effect simulated not fitted		Maternal effect simulated and fitted		Maternal effect fitted		No Maternal effect		Maternal effect simulated not fitted		Maternal effect simulated and fitted		Maternal effect fitted		No Maternal effect	
p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05	p<0.01	p<0.05
41	63	1	4	0	0	0	1	0	4	1	8	0	5	1	6
58	79	1	6	0	4	1	5	4	11	2	2	1	4	3	12
52	73	3	8	1	4	4	6	24	47	11	28	31	49	34	53
41	67	3	3	0	2	0	3	71	80	36	57	73	86	71	85
46	72	1	9	1	4	3	5	91	96	69	82	91	98	96	99
46	67	3	7	0	5	15	2	98	100	76	95	98	100	99	100
82	94	36	57	25	44	7	14	100	100	100	100	100	100	100	100
91	96	43	65	46	68	11	27	100	100	100	100	100	100	100	100
98	98	70	84	63	77	21	38	100	100	100	100	100	100	100	100
100	100	88	93	78	84	42	60	100	100	100	100	100	100	100	100
99	99	95	95	90	95	60	75	100	100	100	100	99	99	100	100

Maternal effect simulated to have a variance of 0.1

Empirical Power (proportion of 100 replicates) to detect significant additive and dominant QTL for Maternal Scenarios using pig scenario across range of simulated additive and dominant QTL effects at 1, 5 and 10% thresholds based on critical null statistic for likelihood ratio statistic from 1000 replicates.

		Power (*P*<0.01) Additive + dominant QTL vs null test (2v0)			Power (P<0.01) Additive + dominant QTL vs additive test (2v1)			Power (P<0.01) Additive QTL vs null test (1v0)		
QTL variance	QTL h^2	F	MF	M	F	MF	M	F	MF	M
0.000	0.00	0	2	37	0	3	36	2	1	1
0.005	0.00	3	4	57	1	1	52	1	5	1
0.020	0.01	25	29	77	0	1	46	31	36	32
0.045	0.03	69	74	90	1	2	36	73	73	79
0.080	0.04	89	95	100	0	1	41	92	96	94
0.125	0.07	98	100	100	4	1	42	98	99	98
0.360	0.17	100	52	100	19	29	77	100	100	100
0.383	0.18	100	100	100	42	40	91	100	100	100
0.410	0.19	100	100	100	59	66	98	100	100	100
0.443	0.21	100	100	100	76	84	100	100	100	100
0.480	0.22	100	100	100	87	95	98	100	100	99

Maternal effect simulated to have a variance of 0.1
F maternal effect fitted but not simulated
MF maternal effect simulated and fitted
M maternal effect simulated but not fitted.

QTL variance	QTL h^2	Power (P<0.05) Additive + dominant QTL vs null test (2v0)			Power (P<0.05) Additive + dominant QTL vs additive test (2v1)			Power (P<0.05) Additive QTL vs null test (1v0)		
		F	MF	M	F	MF	M	F	MF	M
0.000	0.00	5	11	68	0	7	63	5	7	6
0.005	0.00	10	20	82	4	4	79	5	16	7
0.020	0.01	52	55	95	2	6	73	52	58	57
0.045	0.03	83	88	97	4	8	67	82	89	86
0.080	0.04	96	100	100	5	3	72	96	99	99
0.125	0.07	100	100	100	17	4	67	100	100	100
0.360	0.17	100	52	100	44	57	94	100	100	100
0.383	0.18	100	100	100	68	65	96	100	100	100
0.410	0.19	100	100	100	77	84	98	100	100	100
0.443	0.21	100	100	100	84	93	100	100	100	100
0.480	0.22	100	100	100	95	95	99	100	100	99

Maternal effect simulated to have a variance of 0.1
F maternal effect fitted but not simulated
MF maternal effect simulated and fitted
M maternal effect simulated but not fitted.

QTL variance	QTL h^2	Power (*P*<0.10) Additive + dominant QTL vs null test (2v0)			Power (P<0.10) Additive + dominant QTL vs additive test (2v1)			Power (P<0.10) Additive QTL vs null test (1v0)		
		F	MF	M	F	MF	M	F	MF	M
0.000	0.00	12	24	82	4	9	75	8	12	11
0.005	0.00	20	30	91	10	14	83	9	22	17
0.020	0.01	60	71	98	8	4	77	61	71	72
0.045	0.03	90	97	98	4	12	78	88	95	88
0.080	0.04	97	100	100	4	8	81	97	99	99
0.125	0.07	100	100	100	28	11	88	100	100	100
0.360	0.17	100	52	100	60	66	94	100	100	100
0.383	0.18	100	100	100	74	72	97	100	100	100
0.410	0.19	100	100	100	84	88	98	100	100	100
0.443	0.21	100	100	100	86	94	100	100	100	100
0.480	0.22	100	100	100	96	96	99	100	100	99

Maternal effect simulated to have a variance of 0.1
F maternal effect fitted but not simulated
MF maternal effect simulated and fitted
M maternal effect simulated but not fitted.

Appendix 4.3 Mean estimates of variance components at highest test statistic under models fitting additive and dominant QTL, and maternal effects for Poultry, Human and Pig Scenarios across range of simulated dominant QTL effects. Additive effect is fixed at 0.8 with dominant QTL effects ranging from partial to full Dominance

	Dominant QTL effect	Expectation Variance			Additive QTL model				Additive + Dominant QTL model				
		Add QTL	Dom QTL	Resid	Add QTL	Poly	Mat	Resid	Add QTL	Dom QTL	Poly	Mat	Resid
Poultry	0.4	0.32	0.04	1.50	0.32	0.22	-	1.53	missing	0.04	0.23	-	1.51
	0.5	0.32	0.06	1.50	0.34	0.20	-	1.55	0.30	0.07	0.22	-	1.50
	0.6	0.32	0.09	1.50	0.35	0.20	-	1.56	0.30	0.09	0.22	-	1.50
	0.7	0.32	0.12	1.50	0.38	0.22	-	1.58	0.29	0.13	0.25	-	1.49
	0.8	0.32	0.16	1.50	0.38	0.20	-	1.60	missing	0.17	0.23	-	1.50
Human	0.4	0.32	0.04	1.50	0.37	0.16	-	1.52	0.28	0.08	0.21	-	1.40
	0.5	0.32	0.06	1.50	0.39	0.14	-	1.57	0.26	0.12	0.23	-	1.41
	0.6	0.32	0.09	1.50	0.40	0.15	-	1.58	0.28	0.11	0.22	-	1.40
	0.7	0.32	0.12	1.50	0.43	0.11	-	1.61	0.28	0.14	0.21	-	1.41
	0.8	0.32	0.16	1.50	0.46	0.09	-	1.64	0.26	0.17	0.21	-	1.35
Pig	0.4	0.32	0.04	1.50	0.28	0.26	-	1.52	0.26	0.05	0.27	-	1.48
	0.5	0.32	0.06	1.50	0.31	0.24	-	1.54	0.28	0.07	0.25	-	1.49
	0.6	0.32	0.09	1.50	0.31	0.23	-	1.57	0.27	0.09	0.24	-	1.50
	0.7	0.32	0.12	1.50	0.34	0.24	-	1.58	0.27	0.13	0.26	-	1.48
	0.8	0.32	0.16	1.50	0.36	0.23	-	1.61	0.28	0.16	0.25	-	1.49
Pig2 Maternal effect simulated not fitted	0.4	0.32	0.04	0.75	0.27	0.34	-	0.79	0.22	0.13	0.36	-	0.68
	0.5	0.32	0.06	0.75	0.27	0.34	-	0.80	0.21	0.15	0.37	-	0.69
	0.6	0.32	0.09	0.75	0.27	0.34	-	0.80	0.21	0.18	0.38	-	0.68
	0.7	0.32	0.12	0.75	0.30	0.33	-	0.84	0.22	0.23	0.37	-	0.68
	0.8	0.32	0.16	0.75	0.29	0.35	-	0.84	0.21	0.25	0.39	-	0.66
Pig2 Maternal effect fitted and simulated	0.4	0.32	0.04	0.75	0.25	0.29	0.11	0.75	0.23	0.07	0.31	0.09	0.72
	0.5	0.32	0.06	0.75	0.24	0.26	0.12	0.75	0.24	0.08	0.30	0.10	0.73
	0.6	0.32	0.09	0.75	0.27	0.28	0.12	0.79	0.23	0.12	0.32	0.09	0.70
	0.7	0.32	0.12	0.75	0.27	0.26	0.13	0.81	0.22	0.15	0.31	0.09	0.71
	0.8	0.32	0.16	0.75	0.29	0.26	0.14	0.83	0.23	0.20	0.32	0.09	0.71
Pig2 maternal effect fitted	0.4	0.32	0.04	0.75	0.24	0.27	0.02	0.77	0.22	0.06	0.28	0.01	0.73
	0.5	0.32	0.06	0.75	0.25	0.25	0.02	0.76	0.23	0.08	0.27	0.01	0.73
	0.6	0.32	0.09	0.75	0.25	0.25	0.03	0.77	0.22	0.10	0.28	0.01	0.71
	0.7	0.32	0.12	0.75	0.25	0.24	0.03	0.79	0.22	0.13	0.28	0.01	0.72
	0.8	0.32	0.16	0.75	0.31	0.22	0.05	0.87	0.24	0.17	0.26	0.02	0.74
Pig2 no maternal effect	0.4	0.32	0.04	0.75	0.24	0.29	-	0.77	0.22	0.06	0.29	-	0.72
	0.5	0.32	0.06	0.75	0.24	0.28	-	0.79	0.21	0.08	0.29	-	0.73
	0.6	0.32	0.09	0.75	0.26	0.28	-	0.80	0.23	0.11	0.28	-	0.72
	0.7	0.32	0.12	0.75	0.26	0.29	-	0.82	0.22	0.13	0.30	-	0.72
	0.8	0.32	0.16	0.75	0.29	0.27	-	0.86	0.23	0.17	0.29	-	0.73

Polygenic variance is fixed at 0.2 for all scenarios. Maternal variance is 0.1 when simulated. Human. Pig, and Poultry scenarios have residual variance sampled from distribution mean 0, variance 1.5. All maternal scenarios are based on pig scenario with residual variance sampled from normal distribution with mean 0 and variance 0.75 denoted pig2

CHAPTER 5

Detection of parent of origin effects using variance component analysis in simulated pedigrees

Summary

A range of additive, dominant and imprinted QTL effects were simulated. Testing strategies for imprinted QTL were evaluated in human, pig and poultry populations for power to detect fully imprinted QTL and for false positive rates under Mendelian inheritance.

Three different empirical thresholds for type 1 error were derived using varying additive and dominant QTL effects, and frequencies of the favourable allele.

The detection of variance caused by imprinted genes and in particular estimates of variance components were also heavily dependent upon the number of sire and dam families used to estimate them.

Type 1 error rates were high for the test of the separate maternal and paternal components against the additive model in the presence of large additive and dominant QTL effects. Type 1 error rates also differed markedly between human and livestock populations. For the detection of imprinting, power was greatest under a model incorporating separate parental components and could be used for an initial QTL search with little loss of power when compared to an additive model. Subsequent comparisons with individual parental effects were an effective test for parental inequalities and more robust than the test of the maternal and paternal model against the additive model.

5.1 Introduction

5.1.1 Genomic imprinting

Genomic imprinting is the preferential expression of genes depending upon the sex of the parent from which they were inherited (Barlow 1995). It is brought about by epigenetic instructions – imprints – that are laid down in the parental germ cells (Reik & Walter, 2001). Examples include Prader-Willi syndrome in humans (Nicholls 2000), Callipyge in sheep (Charlier *et al*. 2001) and Igf2 in pigs (Nezer *et al*. 1999). The underlying mechanisms and evolutionary basis of imprinting is complex and not yet fully understood but genes tend to cluster in imprinted regions that are partially conserved across species.

An increasing number of genome scans incorporating parent of origin effects have further highlighted their importance; see (De Koning *et al*. 2000;Dong *et al*. 2005;Liu *et al*. 2007; McElroy *et al*. 2006a;Wolf *et al*. 2008) for scans revealing imprinted effects in pigs, humans, dogs, chicken and mice respectively. Luedi *et al* (2005) predict that there are over 600 imprinted genes in the mouse implying that the role of imprinting (in other species) is potentially underestimated. Morison *et al* (2005) have compiled an imprinted gene database that contains more than 200 entries for mammals, marsupials and birds.

Within livestock the effectiveness of selection procedures utilising genomic information relies on correctly identifying the mode of inheritance of desired variants. It has been hypothesised that imprinting has a role in heterosis and is potentially a source of variation that could be exploited in cross breeding and reciprocal crosses practised in many livestock species to combine fecundity and production traits (De Koning, Bovenhuis, & Van Arendonk 2002;Tuiskula-Haavisto & Vilkki 2007).

Statistical methods to detect genomic imprinting involve the detection of allelic effects dependent on the parent of origin by genetic mapping. A maternally imprinted gene involves preferential expression of the paternal allele, i.e. shows an allelic effect when inherited from a sire and none when inherited from a dam and vice versa. A partially expressed QTL is one that shows an allelic effect when inherited from both sire and dam but the effect differs according to the parent from which it is inherited.

In outbred populations parent of origin effects can be calculated as the difference between the heterozygous genotypes in a crossbred generation (Knott *et al*. 1998). This was modified to include direct tests for the contribution of maternal and paternal alleles by de Koning *et al* (2000; 2002) and implemented to find imprinted QTL explaining 2-10% of the phenotypic variance for body composition in pigs. Thomsen *et al*., (2004) treated the same model as a special case of the Mendelian model to compile a decision tree involving direct comparison of imprinted and mendelian models leading to the discovery of 33 parent of origin effects in pigs.

To date QTL studies in livestock populations for parent of origin effects have involved line crosses or divergent populations (Charlier *et al*. 2001;De Koning, Bovenhuis, & Van Arendonk 2002;Heuven *et al*. 2005;Lee *et al*. 2003;McElroy *et al*. 2006b;Nezer *et al*. 2002;Thomsen *et al*. 2004b;Tuiskula-Haavisto *et al*. 2004).

Detection of imprinted and dominant effects, however, within model or experimental populations is costly and potentially of limited relevance to populations under selection. It is often more practical to explore QTL segregating within a population, particularly if it is to facilitate selection within that population. Variance component theory has been developed to incorporate imprinted effects in complex pedigrees (Hanson *et al*. 2001;Pratt, Daly, & Kruglyak 2000;Shete & Amos 2002;Shete, Zhou, & Amos 2003). Heuven *et al*., (2005) use simulation of a variance component approach incorporating linkage disequilibrium following Lee *et al*., (2004) to evaluate optimum population structure in pigs to detect parent of origin effects. The detection of imprinted QTL using variance component approaches has been used in human data (Atwood *et al*. 2002;Shete & Yu 2005;Zhou *et al*. 2007) but very little application has been seen in livestock.

The following chapter uses simulation to compare the power of the variance component approach to detect parent of origin effects in human, pig and poultry type pedigrees. The effect of population structure on the partitioning and estimation of variance components and the distributional properties of test statistics are evaluated by the comparison of different genetic models.

5.2 Materials and methods

5.2.1 Statistical Genetic Models for Variance Component Analysis

As described in **chapter 3** IBD coefficients were estimated for all relationships in the pedigree to calculate the covariance matrices for the QTL effects. Variance

components for each model were estimated using REML (Patterson & Thompson 1971) implemented in the ASReml package (Gilmour, 1995). Models were solved using mixed model equations as described in **chapter 2**.

A range of statistical models were evaluated: models used were

(1) $\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{Wc} + \mathbf{e}$ (null or polygenic)

(2) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Za} + \mathbf{e}$ (additive)

(3) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e}$ (additive + dominance)

(4) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Z_m m} + \mathbf{Z_p p} + \mathbf{e}$ (maternal + paternal)

(5) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Z_p p} + \mathbf{e}$ (paternal)

(6) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Z_m m} + \mathbf{e}$ (maternal)

where **y** is a vector of phenotypic observations, **β** is a vector of fixed effects, **u, a, d, m**, **p**, **c** and **e** are vectors of random additive polygenic effects, additive and dominance QTL effects, maternal and paternal QTL effects, non genetic maternal effects and residuals, respectively. **X, Z, W**, $\mathbf{Z_m}$**,** and $\mathbf{Z_p}$ are incidence matrices relating to fixed and random genetic, direct maternal, maternally expressed and paternally expressed QTL effects, respectively.

Variances for polygenic and QTL effects are distributed as follows: var(**u**) =$\mathbf{A}\sigma^2_a$ Var(**a**) = $\mathbf{G}\sigma^2_q$, Var(**d**) = $\mathbf{D}\sigma^2_d$, Var(**m**) = $\mathbf{G_M}\ \sigma^2_m$, Var(**p**) = $\mathbf{G_P}\ \sigma^2_p$ var(**e**) = $\mathbf{I}\sigma^2_e$,. For the non-genetic maternal effect Var(**m**) = $\mathbf{I}\sigma^2_m$. Where **A** is the standard additive relationship matrix based on pedigree data only and the relationship matrices **G,** $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** for a given QTL position are calculated from the gametic IBD matrix as outlined by Liu *et al.,*(2002) described further in **chapter 2**.

5.2.2 Calculating the relationship matrices A, G and D needed for the mixed model analysis

The **G,** $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** are the appropriate relationship matrices used to model the additive, maternal, paternal and dominant QTL effects at each position tested. They

are conditional on flanking marker information and therefore unique for each position evaluated for a QTL. Here the matrices were calculated every 5 cM.
It can be shown that these relationship matrices are easily estimated from the gametic IBD matrix (**chapter 2**), a 2n x 2n matrix containing the probability of identity of descent between any of the two gametes of an individual with the gametes of the remaining individuals in the pedigree.

The gametic IBD matrix was estimated with the recursive method of Pong-Wong *et al.*, (2001) described in **chapter 2**. Variance components for each model were estimated using REML (Patterson & Thompson 1971) implemented in the ASReml package (Gilmour, Thompson, & Cullis 1995). In order to estimate the variance components for the different models, ASReml requires the knowledge of the inverse of the relationship matrices. ASReml calculates the inverse of the **A** matrix directly from pedigree data, but $\mathbf{G}$, $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** were inverted in using the R package for statistical computing (www.r-project.org) before using them in ASReml.

5.2.3 Simulated populations

The method was implemented in three simulated populations, representative of poultry, pig and human pedigrees (Table 5.1). The parental generation was simulated by random sampling without replacement from an unrelated base population. Under each scenario, random mating of parents was simulated to obtain a second generation of 1900 progeny.

Table 5.1 Population Parameters for simulated pedigrees.

	Sires	Dams per sire	no. of HS per sire	no. of FS per dam
Chicken	19	5	100	20
Pig	10	19	190	10
Human	633	1	-	3

2.4 Simulated polygenic and QTL effects

A 20 cM chromosome was simulated with 5 markers spaced at 5 cM intervals and a bi-allelic QTL between the second and third marker at 7.5 cM.

5.2.4.1 Polygenic variance

To simulate polygenic variance, 10 unlinked additive effects of 0.2 were simulated each with an allele frequency of 0.5 following Alfonso and Haley (1998). The phenotypes generated under a polygenic model were normally distributed indicating that these unlinked QTL were sufficient to provide a reasonably structured polygenic variance. A residual effect was sampled from a normal distribution with mean 0 and a variance of 0.75.

5.2.4.2 Dominant and imprinted QTL effects

A range of maternally and paternally expressed QTL effects were simulated, together with various null scenarios. Null scenarios involved a range of additive, and partial to over dominant effects to explore spurious detection of parent of origin effects. The frequency of the favourable allele (p) was also varied. All scenarios were replicated 100 times and are summarized in Table 5.2.

5.2.4.3 Common environment and polygenic dominance

The phenotypic resemblance between full sibs consists of one half of the additive genetic variance, one quarter of the variance due to dominance and any common environment effects. Non genetic sources of resemblance amongst full sibs such as common environment or direct maternal effects are often, at least partially, confounded with dominance as the covariance of phenotypic values is the sum of the covariances arising from genotypic and environmental causes. To ensure that maternal effects were accounted for in all scenarios a common environment or maternal effect was simulated by sampling from a normal distribution with variance

of 0.1 for each full sib family and assigning this value to each full-sib offspring. A direct maternal or dam effect was then fitted in all models.

For one scenario the maternal effect was simulated but not fitted in the model to examine the effect on test statistics denoted mat^{env}.

Polygenic dominance with a variance of 0.1 was simulated by simulating full dominance at the 10 unlinked QTL denoted $poly^{dom}$. For these scenarios no maternal effect was simulated. Test statistics were examined with a direct maternal effect both fitted ($poly^{dom}$ F) and not fitted in the model ($poly^{dom}$ NF).

The variances of the additive (a), dominance (d) and imprinted (a_i) QTL effects were calculated as follows:

$$V(d) = (2pqd)^2$$

$$V(a) = 2pq[a+d(q-p)]^2$$

$$V(a_i) = 2pq[p^2 + a_i^2 -2ad(p-q) + p^2d^2 + q^2d^2]$$

reducing to $V(a_i) = 4pqa^2$ in the case of full imprinting.

where $d = (d1+d2)/2$ and $a_i = (d1-d2)/2$, for a bi-allelic locus with two alleles A and b, where A is the favourable allele and AA – bb = 2a, Ab = d_1 and bA = d_2

Table 5.2 Simulated scenarios, genetic and non genetic parameters

	Allele freq.	Genetic effect at QTL				Genetic variance at QTL						
		a	d1	d2	a-					Phen	QTL	Gen
	p	AA	Ab	bA	bb	V(a)	V(d)	V(a_i)	QTL	Var.	h^2	h^2
Mat^{env}	0.5	0.8	0.0	0.0	-0.8	0.32	0	0.	0.32	1.37	0.23	0.53
Poly^{dom}*	0.5	0.8	0.0	0.0	-0.8	0.32	0	0	0.32	1.37	0.23	0.43
Null1	0.5	0.8	0.0	0.0	-0.8	0.32	0	0	0.32	1.37	0.23	0.43
Null2	0.5	0.8	0.8	0.8	-0.8	0.32	0.16	0	0.48	1.53	0.31	0.51
Null3	0.5	0.8	0.4	0.4	-0.8	0.32	0.04	0	0.36	1.41	0.26	0.46
Null4	0.5	0.8	0.8	0.8	0.0	0.32	0.16	0	0.48	1.53	0.31	0.51
Null5	0.5	0.0	0.8	0.8	0.0	0	0.16	0	0.16	1.21	0.13	0.33
Null6	0.7	0.8	0.0	0.0	-0.8	0.27	0	0	0.27	1.32	0.2	0.4
Null7	0.7	0.8	0.8	0.8	-0.8	0.1	0.11	0	0.21	1.26	0.17	0.37
Null8	0.3	0.8	0.0	0.0	-0.8	0.27	0	0	0.27	1.32	0.2	0.4
Null9	0.3	0.8	0.8	0.8	-0.8	0.53	0.11	0	0.64	1.69	0.38	0.58
Null10	0.5	0.4	0.0	0.0	-0.4	0.08	0	0	0.08	1.13	0.07	0.27
Null11	0.5	0.4	0.4	0.4	-0.4	0.08	0.04	0	0.12	1.17	0.1	0.3
Null12	0.5	0.6	0.0	0.0	-0.6	0.18	0	0	0.18	1.23	0.15	0.35
Null13	0.5	0.6	0.3	0.3	-0.3	0.18	0.02	0	0.2	1.25	0.16	0.36
Null14	0.5	0.6	0.6	0.6	-0.6	0.18	0.09	0	0.27	1.32	0.2	0.4
Mat1	0.5	0.2	-0.2	0.2	-0.2	0.02	0	0.04	0.04	1.09	0.04	0.24
Mat2	0.5	0.4	-0.4	0.4	-0.4	0.08	0	0.16	0.16	1.21	0.13	0.33
Mat3	0.5	0.6	-0.6	0.6	-0.6	0.18	0	0.36	0.36	1.41	0.26	0.46
Mat4	0.5	0.8	-0.8	0.8	-0.8	0.32	0	0.64	0.64	1.69	0.38	0.58
pat1	0.5	0.2	0.2	-0.2	-0.2	0.02	0	0.04	0.04	1.09	0.04	0.24
pat2	0.5	0.4	0.4	-0.4	-0.4	0.08	0	0.16	0.16	1.21	0.13	0.33
pat3	0.5	0.6	0.6	-0.6	-0.6	0.18	0	0.36	0.36	1.41	0.26	0.46
pat4	0.5	0.8	0.8	-0.8	-0.8	0.32	0	0.64	0.64	1.69	0.38	0.58

Based on an environmental variance of 0.75, polygenic additive variance of 0.2, and common environmental effect of 0.1

* poly^{dom} scenario has no maternal effect, polygenic additive variance of 0.2 and polygenic dominance variance of 0.1.

a = additive effect, d = dominance effect = (d1+d2)/2, V(d) = $(2pqd)^2$, and V(a) = $2pq[a+d(q-p)]^2$ a_i is imprinted effect = (d1-d2)/2

For each individual, a residual effect was sampled from a normal distribution with a mean of 0 and a variance of 0.75. Because the error variance was constant, phenotypic variance and overall heritability increased with genetic effects. In the base scenario with no QTL simulated, polygenic heritability was 0.11. Total heritability (polygenic and QTL) ranged from 0.1 to 0.58 with maternally/paternally expressed QTL effects ranging from 4 to 39% of the phenotypic variance.

5.2.5 Test statistic

For each scenario, 100 replicates were analysed and the test statistics calculated at 2, 7, 12, and 17cM. Tests involved comparisons of linear models (1)-(6) and are described in Table 5.3. A log likelihood ratio test statistic (LRT) for a given location was calculated as twice the difference between the log likelihood of the full and the reduced model. Power was estimated using empirical thresholds derived from 1000 chromosome-wise replicates.

Table 5.3 gives the type of QTL effects (i.e. mode of inheritance) estimated under each comparison and the 5% empirical threshold for type 1 error rate. Initially, each model was tested against the null hypothesis of no QTL. Subsequent tests between models were for

i. Dominance by comparing the model incorporating additive and dominance *addom* (3) with a purely additive model *add* (2)
ii. Parent of origin effects by comparing the *pat+mat* model incorporating maternal and paternal QTL (4) with the additive model *add* (2).
iii. Paternal expression by comparing the *pat+mat* model (4) with a paternal model *pat* (5).
iv. Maternal expression by comparing the *pat+mat* model (4) with a maternal model *mat* (6).

Table 5.3 Tests for QTL effects and corresponding empirical thresholds for 5% type 1 error based on 1000 simulated replicates

Test	QTL in Model		QTL effect tested for	Empirical Threshold		
	alternative (H1)	null (H0)		Poultry	Pig	Human
add	add (2)	null (1)	additive	3.78	3.44	3.78
addom	add + dom (3)	null (1)	additive + dominant	5.14	4.86	4.54
pat+mat	pat + mat (4)	null (1)	paternal + maternal	5.40	4.78	4.68
pat	pat (5)	null (1)	paternal	3.52	3.76	3.0
mat	mat (6)	null (1)	maternal	3.67	3.16	3.66
dom	add + dom (3)	add (2)	dominant	2.60	2.40	1.4
imp	pat + mat (4)	add (2)	imprinted	2.80	2.54	2.76
**patvfull	pat + mat (4)	pat (5)	maternally expressed	3.40	2.90	3.72
**matvfull	pat + mat (4)	mat (6)	paternally expressed	3.34	3.12	3.04

* LRT is chromosome–wise empirical threshold for 5% type 1 error rate for test statistic (twice the difference between log likelihoods for the alternative and null model), estimated by 1000 iterations

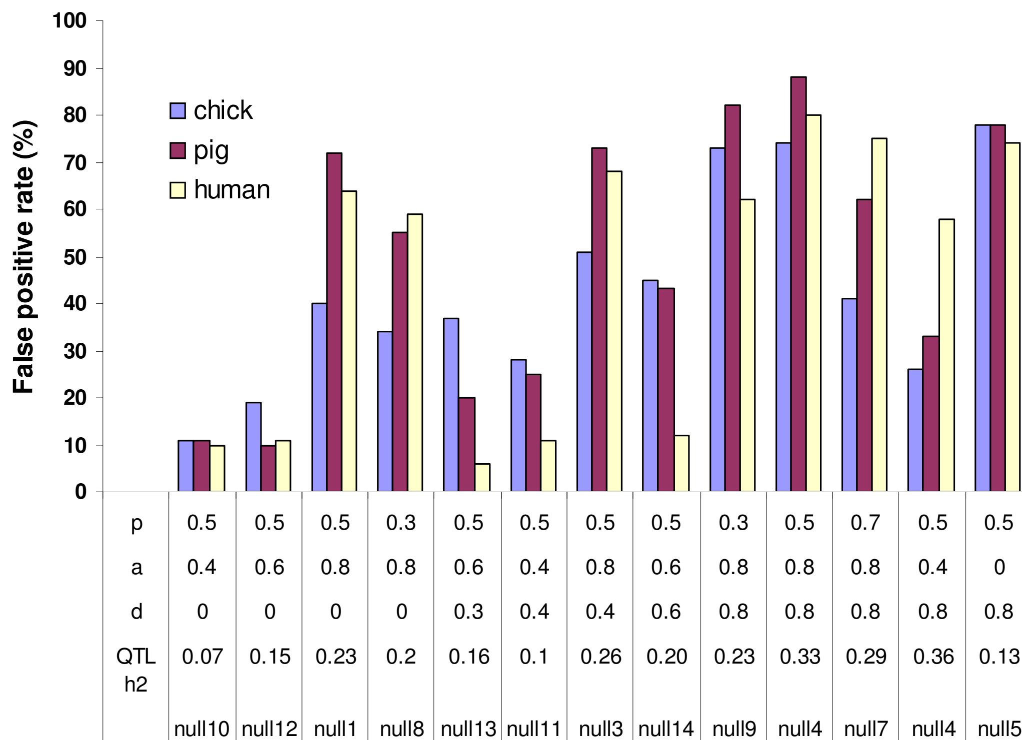
** For example if the test of *patvfull* is significant the model incorporating paternal and maternal QTL is explaining more variation than the paternal QTL indicating some level of maternal expression. If there is no significant difference between the *pat+mat* model and *mat* model the maternal QTl is explaining all of the variation.

5.2.6 Null distribution

Empirical chromosome-wise type 1 error rates were determined separately for all three population structures. One thousand replicates of the null scenario (both additive and dominance QTL effects set to zero), were analysed to determine 5% thresholds for the likelihood ratio statistic from all models and tests defined in Table 5.1.

5.3 Results

Power and variance estimates for all models and all scenarios are given in Appendices 5.1 to 5.4



p	0.5	0.5	0.5	0.3	0.5	0.5	0.5	0.5	0.3	0.5	0.7	0.5	0.5
a	0.4	0.6	0.8	0.8	0.6	0.4	0.8	0.6	0.8	0.8	0.8	0.4	0
d	0	0	0	0	0.3	0.4	0.4	0.6	0.8	0.8	0.8	0.8	0.8
QTL h2	0.07	0.15	0.23	0.2	0.16	0.1	0.26	0.20	0.23	0.33	0.29	0.36	0.13
	null10	null12	null1	null8	null13	null11	null3	null14	null9	null4	null7	null4	null5

Figure 5.1. False positive rate for test for imprinting (*imp*) when simulating additive, dominant and over dominant QTL under an empirical 5% threshold, derived from 1000 replicates of a null scenario with QTL effects of zero. The *Imp* test statistic is a comparison between a model fitting a paternal QTL and a maternal QTL *mat+pat* (H1) with *add* model fitting an additive QTL (H0). p denotes allele frequency, a denotes additive QTL effect and d denotes dominance QTL effect, QTL var is proportion of phenotypic variance explained by the QTL.

5.3.1 Spurious detection of parent of origin effects

Figure 5.1 gives the proportion of replicates significant ($P<0.05$) for the *imp* test under varying additive and dominant QTL effects when there is no imprinting. All of the scenarios exceed the 5% type 1 error rate for the test between the *pat+mat* and the *add* model despite the absence of imprinting. Even with moderate additive effects there is spurious imprinting of 10-20%, and when large additive QTL effects are simulated explaining >20% of the phenotypic variance the false positive rate is 40-70%. The test statistic is inflated by dominance effects, in particular over-dominance where the false positive rate is 70-80% and also affected by changes in allele frequency. False positive rate is also affected by population structure to a varying degree depending on mode of inheritance with the highest false positive rate observed for the largest QTL effects in pigs.

Table 5.4 Thresholds for 5% type 1 error rate for *imp test statistic derived from 1000 replicates when simulating no QTL and simulating fully dominant QTL.**

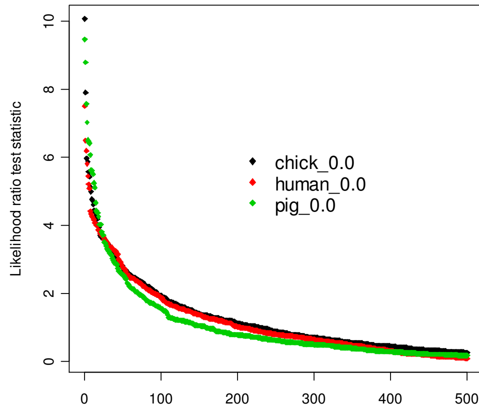
Empirical Threshold	Allele freq. p/q	Additive QTL effect a	Dominant QTL effect d	5% type 1 error rate		
				Poultry	Pig	Human
1	0.5/0.5	0	0	2.8	2.54	2.76
2	0.3/0.7	0.6	0.6	12.6	22.7	10.3
3	0.3/0.7	0.8	0.8	30.6	53.7	11.6

**Imp* test statistic comprises the comparison of model fitting a paternal QTL and a maternal QTL *pat+mat*(H1) with *add* model (H0)

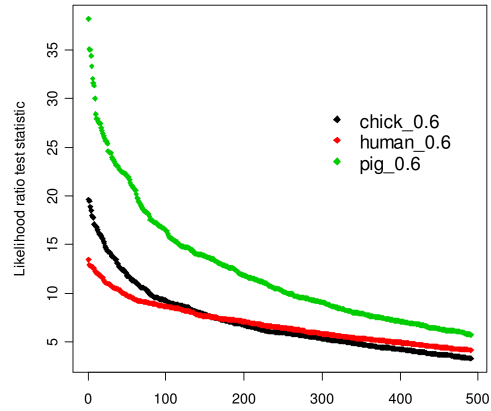
Based on the high false positive rates under the null hypothesis of no QTL effects seen in Figure 5.1, more stringent thresholds were derived for the test for imprinting to try to account for the effects of additive and dominant effects at the QTL and

changes in allele frequency. Two further empirical thresholds for the *imp* test statistic were derived using fully dominant QTL with additive and dominant effects of 0.6 and 0.8 and favourable allele frequency (p) of 0.3. Plots of the three distributions are given in figure 5.2 and thresholds for 5% type 1 error rate in table 5.4.

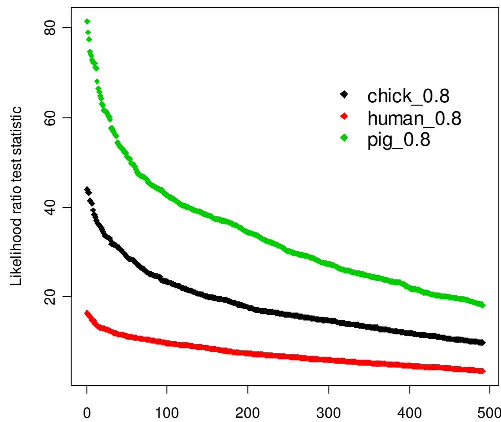
Figure 5.2 shows that under threshold 1 the distributions are similar in all three populations. As additive and dominant effects at the QTL increase the distribution of test statistics become more variable confirming that the test for imprinting is affected both by the size of additive and dominance effects at the QTL and dependent on population structure.



a.



b.



c.

Figure 5.2. Empirical distribution of the test statistic for the *imp test under a) QTL effect of zero, b) fully dominant QTL with additive and dominant QTL effects of 0.6 and c) fully dominant QTL with additive and dominant effects of 0.8. Allele frequencies for the dominant QTL are 0.3 and 0.7. First 500 of 1000 ranked replicates shown. **Imp* test statistic comprises the comparison of model fitting a paternal QTL and a maternal QTL *pat+mat* (H1) with *add* model (H0).**

Table 5.5 False positive rates for *imp statistic from empirical 5% thresholds, based on 100 replicates of each scenario.**

Scenario	Allele freq. p	Genetic effect simulated			QTL var	Threshold 1			Threshold 2			Threshold 3		
		AA	AB BA	BB		Chick	Pig	Hum	Chick	Pig	Hum	Chick	Pig	Hum
null1	0.5	0.8	0	-0.8	0.23	39	72	64	3	3	5	0	1	0
null2	0.5	0.8	0.8	-0.8	0.33	66	88	80	20	15	11	4	1	8
null3	0.5	0.8	0.4	-0.8	0.25	51	73	68	2	7	4	0	2	1
null 4	0.5	0.8	0.8	0	0.36	26	33	58	0	0	1	0	0	0
null 5	0.5	0	0.8	0	0.13	78	78	74	18	2	5	0	0	2
null 7	0.7	0.8	0.8	-0.8	0.2	41	62	75	5	4	1	0	3	1
null 8	0.3	0.8	0	-0.8	0.29	34	55	59	0	1	1	0	1	1
null 9	0.3	0.8	0.8	-0.8	0.2	73	82	62	27	38	6	3	6	4
null 10	0.5	0.4	0	-0.4	0.07	11	11	10	0	0	0	0	0	0
null 11	0.5	0.4	0.4	-0.4	0.1	28	25	11	0	0	0	0	0	0
null 12	0.5	0.6	0	-0.6	0.15	19	10	11	2	1	0	1	0	0
null 13	0.5	0.6	0.3	-0.3	0.16	37	20	6	1	1	0	0	0	0
null 14	0.5	0.6	0.6	-0.6	0.2	45	43	12	2	1	0	0	0	0
Mat^{env} NF	0.5	0.8	0	-0.8	0.23	100	95	93	59	45	45	3	0	38
$Poly^{dom}$ F	0.5	0.8	0	-0.8	0.23	10	37	12	1	0	0	0	0	0
$Poly^{dom}$ NF	0.5	0.8	0	-0.8	0.23	39	66	25	3	0	0	2	0	0

**Imp* test statistic comprises the comparison of model fitting a maternal QTL and a paternal QTL (H1) with a Mendelian model (H0).

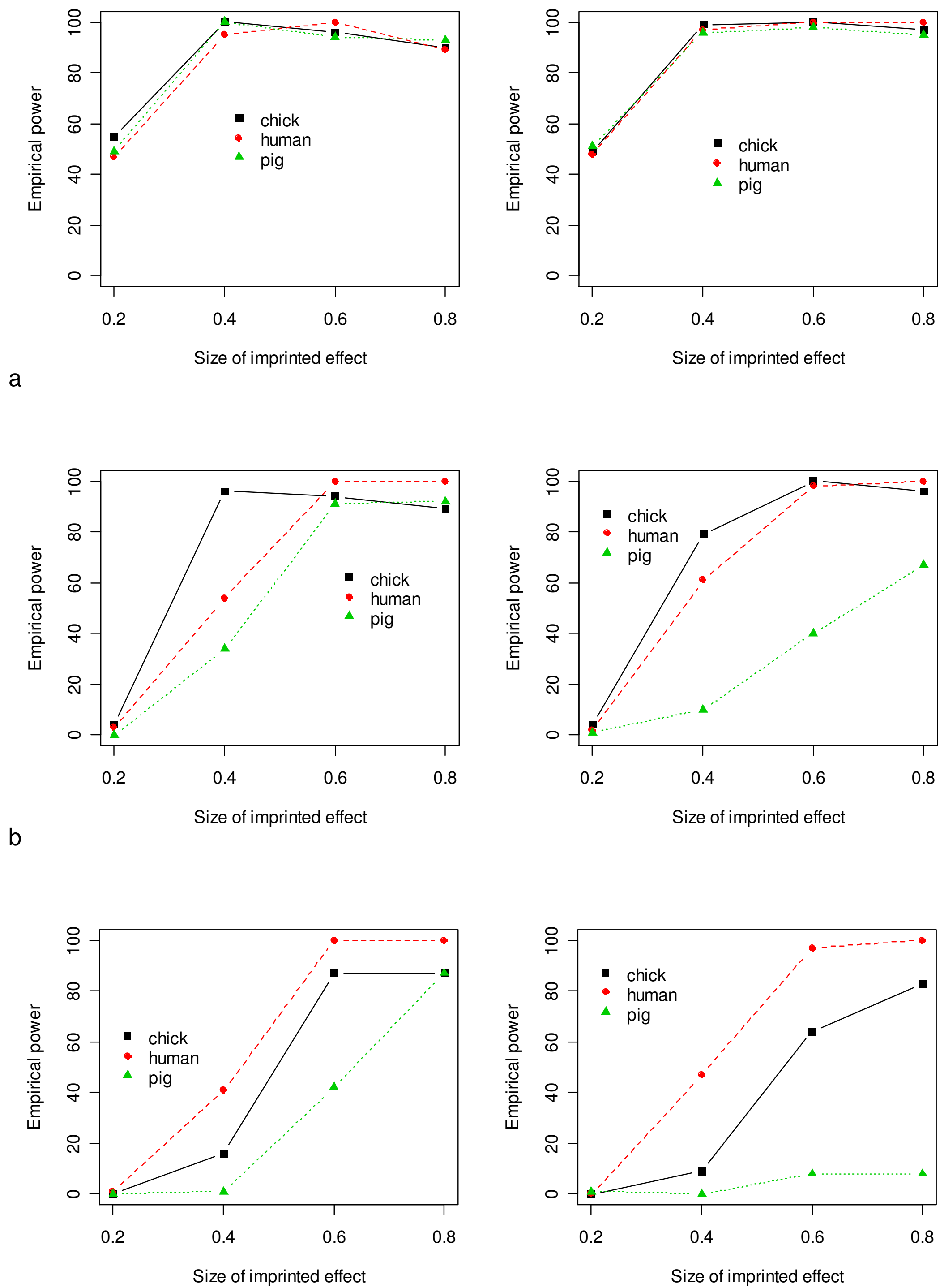


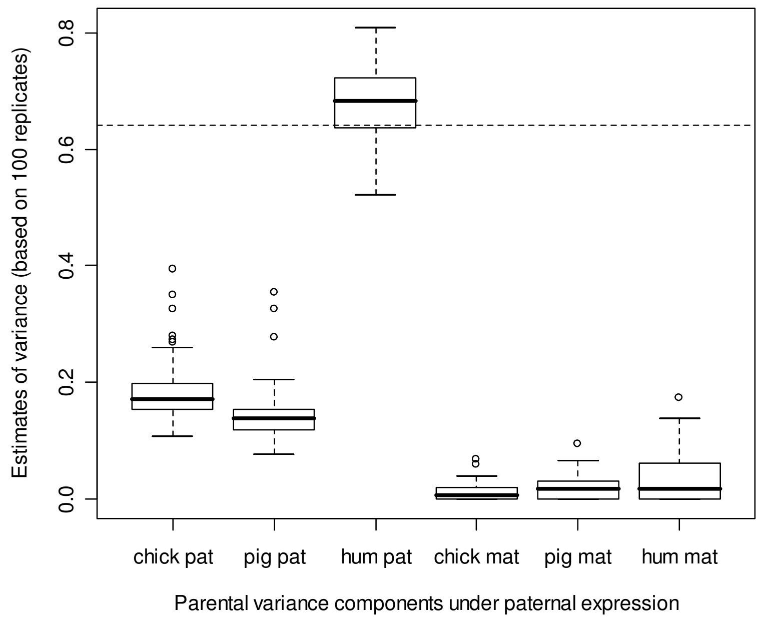
Figure 5.3 Proportion of replicates significant for the *imp* test for full imprinting for maternally (left) or paternally (right) expressed QTL based on 100 replicates, under three different empirical thresholds: From top to bottom empirical thresholds derived from a) no QTL effect, b) a fully dominant QTL with additive and dominant QTL effects of 0.6 and c) a fully dominant QTL with additive and dominant effects of 0.8. Favourable Allele frequencies for the dominant QTL are 0.5 for a and 0.3 for b and c. Results are regardless of significance at the QTL.

Table 5.5 compares the false positive rates for the *imp* statistic under the three thresholds. The more stringent thresholds appear to successfully correct for 5% type 1 error rates in the three different populations when no imprinting is simulated. Not fitting common environment in the model resulted in inflated tests for imprinting; these were controlled by the more stringent thresholds in the chicken and the pig populations but not in humans. Polygenic dominance variance of the same magnitude inflated spurious QTL dominance in chicken and pig populations but did not appear to inflate the imp statistic. Polygenic dominance was controlled by fitting a direct maternal effect.

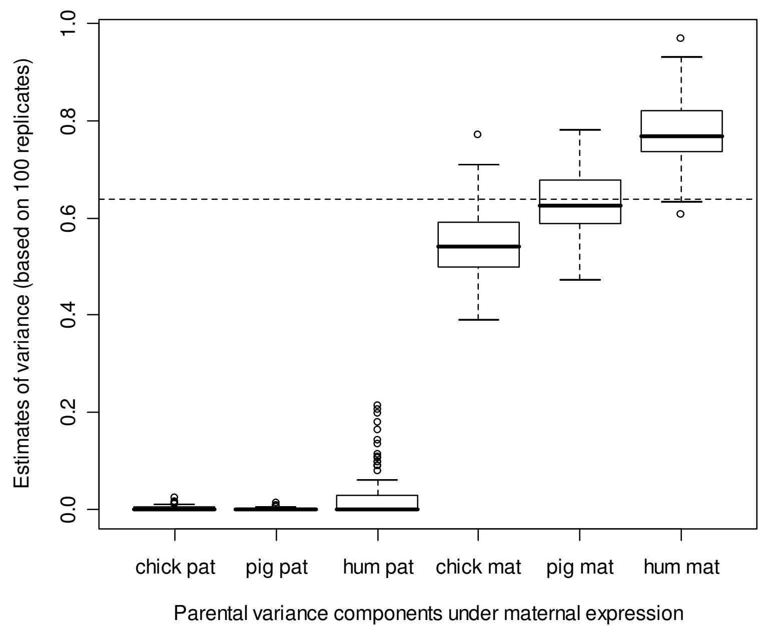
5.3.2 Power to detect imprinted QTL

Figure 5.3 compares the proportion of replicates significant for the *imp* test for paternally/maternally expressed QTL under the three empirical thresholds (table 5.4). Under threshold 1 there was no difference in power amongst populations to detect maternally or paternally expressed QTL. For all populations power was high; > 95% to detect parent of origin effects of 0.2 explaining more than 4% of the phenotypic variance. Under the more stringent thresholds detection of QTL effects of 0.2 and 0.4 explaining 4-13% of the phenotypic variance was most affected. In chicken and pig populations power was lower in the paternal analyses, with the pig population most affected.

Figure 5.4 shows estimates of QTL variance from the *pat+mat* model (4) from scenarios mat4 and pat4 (Table 5.3). Maternally and paternally expressed QTL were simulated with a large fully imprinted effect (a_i) of 0.8 where $E(Va_i)=0.64$. There was little difference between the estimates of the paternal and maternal variance components together in the *pat+mat* model or individually (Appendix 5.1). Paternally expressed QTL variance was underestimated in the pig and the chicken populations. Maternally expressed variance was estimated more accurately, although slightly underestimated in the chicken population and over – estimated in the human population. There was little spurious estimation of variance from the parent from which the QTL is not expressed.



a.



b.

Figure 5.4 Estimation of maternal and paternal variance components using the *pat+mat* model fitting a maternal and paternal QTL. Based on 100 replicates that were all significant for the simulated QTL. a) paternally expressed QTL *pat* with a simulated imprinted effect of 0.8 and b) maternally expressed QTL *mat* with a simulated imprinted effect of 0.8. Dashed line denotes the expectation of QTL variance under the correct model

Table 5.6. Proportion of replicates (n=100) where there is maternal expression (patvfull test) and/or paternal expression (matvfull test) in a range of scenarios simulating additive, dominant, overdominant and imprinted QTL

	Chicken		Pig		Human	
	patvfull	matvfull	patvfull	matvfull	patvfull	matvfull
null1	100	100	99	100	91	94
null2	99	100	95	100	98	99
null3	99	99	96	100	95	93
null4**	91	92	69	91	53	61
null5 **	96	45	83	24	55	70
null7	83	83	84	92	77	81
null8	99	100	98	100	75	88
null9	100	100	97	98	100	100
mat1	69	3	44	2	67	89
mat2	100	0	100	3	4	91
mat3	100	4	100	3	0	87
mat4	100	4	100	0	0	89
pat1	4	99	3	100	91	52
pat2	5	100	3	100	86	0
pat3	13	100	6	100	95	0
pat4	38	100	0	100	97	0

*Regardless of significance of the QTL

**Overdominance is simulated in these scenarios.Scenarios are summarised in Table 5.3

Details of all scenarios are given in table 5.2. mat1-4 are maternally expressed QTL, and pat1-4 are paternally expressed QTL

Table 5.6 shows the test of individual parental QTL against the *pat+mat* model denoted *matvfull* and *patvfull* based on the thresholds given in table 5.3. For an imprinted QTL only one parent is expected to show expression. For example for a maternally expressed QTL the expectation is that the *patvfull* test is significant and the *matvfull* test is not significant. For non imprinted QTL the expectation is that both tests are significant as there is expression from both parents. For the imprinted models, provided there is power to detect the size of effect, a high proportion of replicates indicate parental expression from the correct parent although for the largest

paternally expressed effect in chicken the paternal QTL fails to explain as much variance as the *pat+mat* model in 38% of scenarios.

Table 5.7 Power (proportion of significant replicates, based on 100 replicates) to detect fully imprinted QTL under different QTL models using 5% empirical threshold derived under null scenario of no QTL.

			QTL Model						
			add†	addom††	pat+mat#	pat‡	mat‡‡	Imp*	% of full sig for imp **
Maternal expression	0.2	chicken	31	44	63	5	74	67	93
		pig	12	26	42	8	52	49	83
		human	18	29	37	14	31	47	92
	0.4	chicken	100	100	100	2	100	100	100
		pig	95	99	100	5	100	100	100
		human	55	75	95	3	96	95	100
Paternal expression	0.2	chicken	97	94	96	100	5	49	100
		pig	98	97	96	100	4	51	51
		human	9	18	39	43	4	48	92
	0.4	chicken	100	100	100	100	4	99	100
		pig	100	100	100	100	5	96	100
		human	69	80	99	100	0	97	100

Thresholds are based on 1000 simulated replicates and are given in table 5.3.

† additive test statistic comparing model fitting additive QTL (H1) and no QTL (H0)

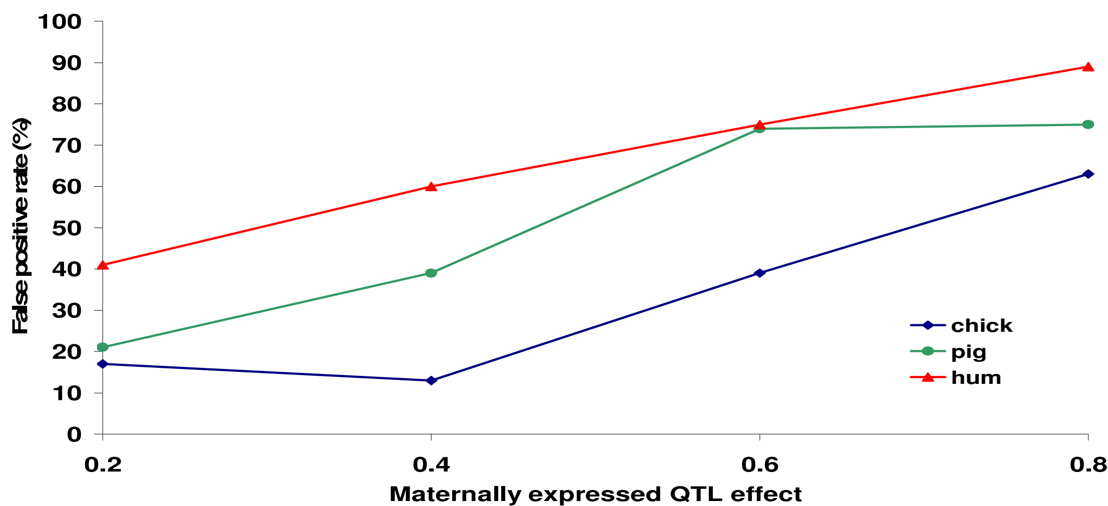
†† *addom* test statistic comparing model fitting additive and dominant QTL (H1) and no QTL (H0)

pat+mat test statistic comparing model fitting maternal and paternal QTL (H1) and no QTL (H0)

**Imp* test statistic comprises the comparison of model fitting a maternal QTL and a paternal QTL (H1) with *add* model (H0).

** % of the significant replicates under the *pat+mat* model also significant for the *imp* test

Table 5.7 compares power to detect imprinted QTL under various QTL models. Power is greatest under the appropriate model incorporating separate paternal and maternal QTL effects, although for small effects only around half of subsequent tests between the *pat+mat* and Mendelian model are significant. When the appropriate model is not used a model including dominance is slightly more powerful than an additive model. For a non imprinted QTL the *pat+mat* model can be used with little loss of power (Appendix 5.1).



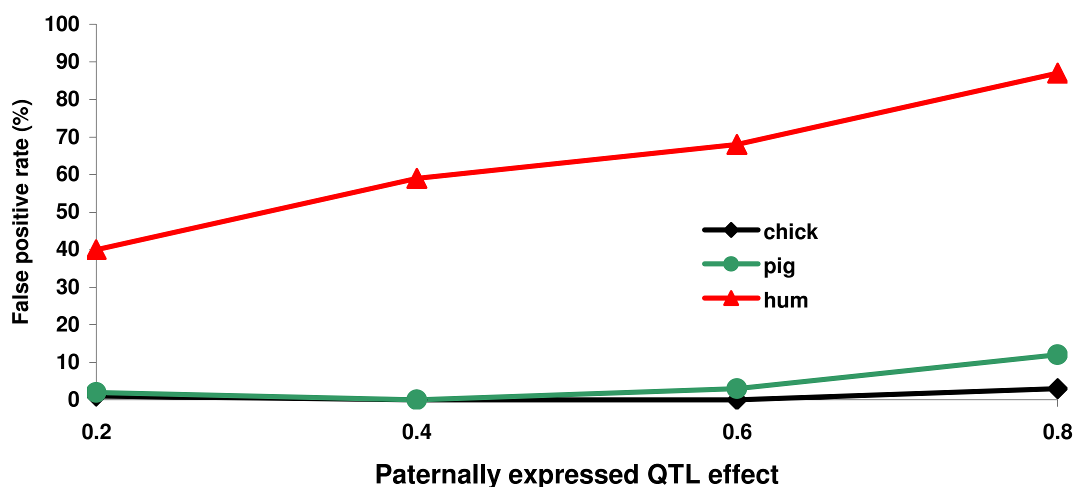


Figure 5.5 Spurious dominance under paternal and maternal expression number of replicates (out of 100) where *dom* test (additive + dominant QTL *addom* (H1) versus additive QTL *add* (H0) is significant when full imprinting is simulated and no dominance.

5.3.3 False positives under full imprinting

Figure 5.5 shows the proportion of replicates significant for the test for dominance when a fully imprinted QTL is simulated. For chicken and pig populations spurious dominance is found under maternal expression with the pig population most affected. For the human population detection of spurious dominance is high and independent of the parent from whom the QTL effect is expressed (Appendix 5.1).

5.4 Discussion

Testing strategies for imprinted QTL were evaluated in human, pig and poultry populations for power ($P < 0.05$) to detect fully imprinted QTL and for false positive rates under Mendelian inheritance using three different empirical thresholds. Empirical thresholds for type 1 error were derived using varying additive and dominant QTL effects, and favourable allele frequencies.

False positive rates for non imprinted QTL from the initial empirical distribution were surprisingly high and imply that the use of the *imp* test particularly using tabulated thresholds should be approached with caution. In particular dominant QTL effects appeared to inflate the test statistic for imprinting although this was also seen for large additive effects. This indicates that using no QTL to derive empirical thresholds is inappropriate and not a true representation of the null hypothesis for this test. It must be noted that at more modest, and potentially more biologically plausible, additive effects, false positive rates for imprinting were much lower. There was also a high rate of spurious dominance when only additive effects were simulated if additive effects were large (appendix 5.2), this was not seen in the chapter 4 because the simulated additive QTL effects were lower.

Thresholds for the imprinting test derived from simulating additive and dominant QTL were extremely high, in particular reflecting the high rate of false positives in the pig population. The 5% type 1 error thresholds for the likelihood ratio test statistic were ~12, 31, and 54 for human, chicken and pig type pedigrees, respectively. When power calculations were made using these thresholds power to detect imprinted QTL was lowest in the pig population, in particular for the detection of paternally expressed QTL. Applying thresholds 1 and 2 across non imprinted scenarios did successfully control the type 1 error rate; however the selection of these thresholds was extreme and arbitrary, given that for real data the magnitude of additive and dominant QTL effects are unknown. Despite this there was still some power to detect imprinted QTL effects in all populations. It is difficult to hypothesise how an appropriate threshold should be selected for the analysis of real data.

One method might be to try to incorporate more variance components in the null hypothesis, for example test a model including maternal, paternal and dominant QTL against a model with additive and dominant QTL. This however would not account for the spurious imprinting found at large additive effects. It is also more difficult to achieve convergence for parameter estimates particularly those close to zero as components in the model increase.

For the initial search power was greatest to detect imprinted QTL when the *pat+mat* model incorporating separate parental QTL when compared to searching under the *add* model. This also resulted in little loss of power when QTL effects were not imprinted. The subsequent testing of individual parental models with the *pat+mat* model using the *patvfull* and *matvfull* tests appears to be a reliable indicator of parental expression although these tests would also indicate expression from a single parent if there was low power to detect an effect segregating from the other parent, for example when favourable allele frequencies are low and/or there are a low number of parents of a given sex contributing to the population under study.

Human pedigrees were in general less powerful to detect QTL under the *add* and *pat+mat* models and more susceptible to spurious dominance and inflated estimates of direct maternal variance when an imprinted QTL was analysed under an incorrect model. Conversely, using the *imp* test, false positive rates were lowest in human pedigrees. When using stringent empirical thresholds, the power to detect imprinting was highest in human pedigrees. The balanced design of the human pedigree also yielded more accurate estimates of the variance components under the correct model and equal power to detect paternally or maternally expressed QTL.

The empirical 5% threshold, derived with no QTL effects simulated, corresponded to a nominal P value around 2.7% for all populations. The expectation is that under the null hypothesis the test statistic for the *pat+mat* versus null model is distributed asymptotically as a mixture of ¼ χ^2_0, ½ χ^2_1 and ¼ χ^2_2 (Hanson *et al*. 2001b;Self & Liang 1987;Shete, Zhou, & Amos 2003). Hanson *et al*., (2001) show, using simulation, that the *imp* statistic, where using the additive model as the null is a

special case of constraining the parental components to be equal, can be controlled using a χ^2_1 distribution. Here, under the null hypothesis of no QTL, the 5% type 1 error rate of 2.7 despite being estimated chromosome-wise, is conservative when compared to the 3.84 tabulated value under χ^2_1 and closer to a mixture of χ^2_1 and χ^2_2 distributions. It is clear, however, that this test statistic cannot be accurately derived under a null QTL model in the presence of moderate to large QTL effects.

Although computationally expensive, permutation analysis is an attractive option. It is crucial to ascertain on what level the permutation should take place, for example in livestock it could be hypothesized that permuting phenotypes and genotypes within large dam families might be effective whilst circumventing re estimation of IBD matrices and polygenic or common environment effects.

Accuracy of variance component estimates for imprinted QTL was dependent upon population structure and increased with the number of parents segregating for the effect. Pig and poultry populations failed to accurately estimate variance from paternally expressed QTL. In the pig population, the average estimate of 100 significant replicates for a paternally expressed QTL with a variance of 0.64, was only 0.14. The pig population has only 10 sires and with even small sampling effects at intermediate allele frequencies there may be too few sires segregating to accurately estimate the paternal variance component. All pedigrees had higher estimates for maternally expressed QTL, possibly due to greater numbers of dam families than sire families in the populations. The lowest estimates were from the chicken population, which had the lowest number of dam families. In the human population, the maternal variance components were overestimated, possibly due to confounding. With no half sib structure, both estimates of the maternal and paternal variance components have to be derived from a single source of information i.e. the covariance of full sibs.

Polygenic dominance resulted in some spurious dominance (Appendix 5.2) but had very little effect on the test for imprinting. A common or maternal environment effect did cause spurious inflation of the *imp* and *dom* test statistics, in particular for

the human populations. Although it is surprising that the maternal effect should have effects of this magnitude at an individual locus, the results reflect findings from the chapter 4. Hager *et al*., (2008) formalize this in an F3 mouse population using simulation to show how maternal genetic and direct effects can result in patterns that mimic those expected under imprinting, bipolar and polar dominance. The net result of gametic imprinting is a reduction of the expected phenotypic covariance between parents and offspring relative to that between siblings, therefore anything that inflates covariance between siblings could result in spurious inheritance patterns.

5.5 Conclusions

Substantially more replicates were used than in previous studies to examine type 1 error rates. Whereas others use deterministic formulae or the assumption of chi-square distributions this study derives empirical distributions and shows that for tests between different linear models, the derivation of the correct null distribution is a difficult issue. It is clear that in many circumstances tabulated thresholds would fail and furthermore type 1 error rates differ markedly between human and livestock populations. Therefore, tests between models although a useful source of information should be regarded with caution. For the detection of imprinting, the *pat+mat* model could be used for an initial QTL search with little loss of power for purely additive QTL. Subsequent comparisons with individual parental effects can be used to test for parental inequalities however this would still require correction for multiple testing. Given the effects of QTL size, environment, and differences in allele frequency care should be taken in interpretation of results.

Appendix 5.1 Proportion of significant replicates (n=100) for all tests based on 5% empirical type 1 error rate derived under null scenario of no QTL (Threshold 1)

		$Poly^{env}$ NF	$Poly^{env}$ F	Mat^{env}	null1	null2	null3	null4	null5	null7	null8	null9	null10	null11	null12	null13	null14
		0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.7	0.3	0.3	0.5	0.5	0.5	0.5	0.5
		0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.3	0.7	0.7	0.5	0.5	0.5	0.5	0.5
		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0	0.8	0.8	0.8	0.4	0.4	0.6	0.6	0.6
		0	0	0	0	0.8	0.4	0.8	0.8	0.8	0	0.8	0	0.4	0	0.3	0.6
		0	0	0	0	0.8	0.4	0.8	0.8	0.8	0	0.8	0	0.4	0	0.3	0.6
		-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	0	0	-0.8	-0.8	-0.8	-0.4	-0.4	-0.6	-0.3	-0.6
	chick	100	100	100	100	100	99	100	95	85	100	100	99	99	100	100	100
	pig	100	99	98	100	100	100	98	60	99	100	99	97	99	100	100	100
add	hum	97	95	97	99	97	100	98	54	55	85	100	29	54	73	78	74
	chick	100	100	100	100	100	99	100	100	85	100	100	100	100	100	100	100
	pig	100	100	98	100	100	100	98	100	98	100	100	95	97	100	100	100
addom	hum	97	94	99	100	94	100	99	68	87	92	100	28	52	72	77	76
	chick	100	100	100	100	100	99	100	97	85	100	100	100	99	100	100	100
	pig	100	100	98	100	100	100	99	85	99	100	100	98	99	100	100	100
pat+mat	hum	95	94	99	100	98	100	99	77	83	95	100	23	43	66	66	72
	chick	100	100	100	100	100	99	94	55	85	100	100	97	92	100	100	100
	pig	100	100	98	100	100	100	92	23	92	100	98	95	94	100	100	100
pat	hum	77	71	95	91	82	96	90	48	62	62	99	15	22	36	52	50
	chick	100	100	100	100	100	99	95	95	83	100	100	70	91	99	99	100
	pig	100	100	98	99	99	100	78	85	90	100	100	53	71	98	92	100
mat	hum	68	68	93	81	69	95	83	42	42	68	99	15	24	43	37	44
	chick	18	9	60	10	98	56	71	100	82	8	92	9	57	5	28	96
	pig	48	21	87	33	93	66	48	100	88	20	87	5	34	10	24	82
dom	hum	28	20	89	41	81	53	53	93	77	36	72	7	13	10	17	19
	chick	39	10	100	40	66	51	26	78	41	34	73	11	28	19	37	45
	pig	66	37	95	72	88	73	33	78	62	55	82	11	25	10	20	43
imp	hum	25	12	93	64	80	68	58	74	75	59	62	10	11	11	6	12

Appendix 5.2 Proportion of significant replicates (n=100) for *dom* test statistic based on 5% empirical type 1 error rate derived under null scenario of no QTL (Threshold 1)

Scenario	Allele Freq.	Genetic effect				QTL	Population		
	p	AA	AB	bA	BB	variance	Chick	Pig	Human
Mat NF	0.5	0.8	0	0	-0.8	0.23	60	87	89
PD NF	0.5	0.8	0	0	-0.8	0.23	18	48	28
PD F	0.5	0.8	0	0	-0.8	0.23	9	21	20
10	0.5	0.4	0	0	-0.4	0.07	9	5	7
12	0.5	0.6	0	0	-0.6	0.15	5	10	10
1	0.5	0.8	0	0	-0.8	0.23	15	33	41
8	0.3	0.8	0	0	-0.8	0.29	8	20	36
11	0.5	0.4	0.4	0.4	-0.4	0.1	57	34	13
13	0.5	0.6	0.3	0.3	-0.3	0.16	28	24	17
14	0.5	0.6	0.6	0.6	-0.6	0.2	96	82	19
2	0.5	0.8	0.8	0.8	-0.8	0.33	98	93	81
3	0.5	0.8	0.4	0.4	-0.8	0.25	56	66	53
4	0.5	0.8	0.8	0.8	0	0.36	71	48	53
7	0.7	0.8	0.8	0.8	-0.8	0.2	82	88	77
5	0.5	0	0.8	0.8	0	0.13	100	100	93
9	0.3	0.8	0.8	0.8	-0.8	0.2	92	87	72

Appendix 5.3 Estimates of variance components for all models for maternally expressed QTL

Model		human				chicken				pig			
		mat1	mat2	mat3	mat4	mat1	mat2	mat3	mat4	mat1	mat2	mat3	mat4
	Simulated variance	0.04	0.16	0.36	0.64	0.04	0.16	0.36	0.64	0.04	0.16	0.36	0.64
Null (1)	residual	0.25	0.32	0.42	0.68	0.27	0.36	0.50	0.74	0.25	0.36	0.53	0.78
	dam	0.08	0.11	0.15	0.73	0.10	0.13	0.19	0.27	0.10	0.14	0.19	0.25
add (2)	add	0.08	0.19	0.39	0.79	0.03	0.20	0.38	0.49	0.02	0.18	0.31	0.38
	poly	0.18	0.13	0.04	0.00	0.24	0.15	0.11	0.19	0.23	0.17	0.20	0.30
	dam	0.08	0.11	0.15	0.73	0.10	0.13	0.19	0.25	0.10	0.14	0.19	0.24
	residual	0.80	0.84	0.91	0.79	0.75	0.80	0.84	0.83	0.76	0.80	0.83	0.86
addom (3)	add	0.02	0.06	0.18	0.78	0.02	0.16	0.35	0.47	0.01	0.10	0.28	0.36
	poly	0.23	0.25	0.23	0.00	0.25	0.19	0.13	0.23	0.24	0.25	0.23	0.37
	dam	0.07	0.09	0.11	0.73	0.10	0.13	0.18	0.24	0.09	0.12	0.16	0.20
	residual	0.73	0.73	0.74	0.78	0.73	0.77	0.80	0.78	0.73	0.73	0.75	0.69
	dom	0.08	0.13	0.21	0.01	0.02	0.03	0.05	0.09	0.03	0.08	0.11	0.20
pat + mat (4)	poly	0.19	0.19	0.18	0.03	0.23	0.23	0.23	0.30	0.21	0.21	0.20	0.23
	dam	0.06	0.07	0.07	0.56	0.09	0.09	0.10	0.13	0.09	0.10	0.09	0.10
	residual	0.76	0.76	0.77	0.82	0.74	0.75	0.75	0.73	0.76	0.75	0.76	0.75
	Pat	0.04	0.04	0.04	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
	Mat	0.07	0.19	0.40	0.77	0.05	0.17	0.36	0.55	0.05	0.17	0.38	0.63
pat (5)	Poly	0.23	0.31	0.42	0.68	0.27	0.36	0.50	0.74	0.24	0.35	0.52	0.77
	Dam	0.08	0.11	0.15	0.73	0.10	0.13	0.19	0.27	0.10	0.14	0.19	0.25
	residual	0.79	0.84	0.91	0.88	0.74	0.77	0.81	0.81	0.76	0.77	0.79	0.81
	Pat2	0.04	0.02	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.01
Mat (6)	Poly	0.22	0.21	0.20	0.04	0.24	0.23	0.23	0.30	0.21	0.22	0.20	0.23
	Dam	0.07	0.08	0.08	0.56	0.09	0.09	0.10	0.13	0.09	0.10	0.09	0.10
	residual	0.78	0.78	0.78	0.84	0.74	0.75	0.75	0.73	0.76	0.75	0.76	0.75
	Mat2	0.06	0.19	0.39	0.77	0.05	0.17	0.37	0.55	0.05	0.17	0.38	0.63

Appendix 5.4 Estimates of variance components for all models for paternally expressed QTL

Model		Human				Chicken				Pig			
		pat1	pat2	pat3	pat4	pat1	pat2	pat3	pat4	pat1	pat2	pat3	pat4
	Simulated variance	0.04	0.16	0.36	0.64	0.04	0.16	0.36	0.64	0.04	0.16	0.36	0.64
Null (1)	residual	0.25	0.31	0.42	0.59	0.27	0.33	0.43	0.57	0.23	0.29	0.36	0.45
	dam	0.09	0.11	0.15	0.21	0.09	0.10	0.09	0.13	0.09	0.09	0.09	0.09
add (2)	add	0.07	0.20	0.37	0.62	0.06	0.14	0.25	0.36	0.06	0.13	0.21	0.26
	poly	0.18	0.11	0.05	0.01	0.21	0.20	0.20	0.25	0.18	0.16	0.16	0.19
	dam	0.09	0.11	0.15	0.21	0.09	0.10	0.10	0.14	0.09	0.09	0.09	0.09
	residual	0.80	0.84	0.91	0.97	0.75	0.77	0.78	0.79	0.78	0.79	0.81	0.85
addom (3)	add	0.02	0.07	0.18	0.39	0.06	0.14	0.25	0.36	0.06	0.13	0.20	0.26
	poly	0.23	0.23	0.23	0.18	0.21	0.20	0.20	0.24	0.18	0.16	0.16	0.19
	dam	0.14	0.09	0.11	0.16	0.09	0.10	0.09	0.14	0.09	0.09	0.09	0.10
	residual	0.74	0.73	0.76	0.77	0.74	0.77	0.78	0.79	0.77	0.79	0.81	0.83
	dom	0.08	0.14	0.19	0.28	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01
pat + mat (4)	poly	0.19	0.17	0.18	0.18	0.22	0.25	0.32	0.39	0.20	0.23	0.27	0.34
	dam	0.06	0.07	0.07	0.07	0.09	0.09	0.08	0.09	0.09	0.09	0.08	0.08
	residual	0.76	0.76	0.78	0.76	0.73	0.75	0.73	0.74	0.76	0.76	0.78	0.80
	Pat	0.08	0.20	0.39	0.68	0.05	0.13	0.19	0.21	0.05	0.10	0.13	0.14
	Mat	0.03	0.04	0.03	0.03	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02
pat (5)	Poly	0.21	0.20	0.20	0.21	0.23	0.26	0.32	0.41	0.21	0.24	0.27	0.35
	Dam	0.07	0.08	0.07	0.08	0.09	0.10	0.08	0.10	0.09	0.09	0.09	0.08
	residual	0.78	0.78	0.79	0.77	0.73	0.75	0.74	0.75	0.76	0.76	0.79	0.81
	Pat2	0.07	0.20	0.39	0.68	0.05	0.13	0.19	0.21	0.05	0.10	0.12	0.14
Mat (6)	Poly	0.23	0.30	0.41	0.59	0.26	0.33	0.42	0.56	0.22	0.28	0.35	0.44
	Dam	0.08	0.11	0.15	0.21	0.09	0.10	0.09	0.12	0.09	0.09	0.09	0.08
	residual	0.79	0.84	0.92	1.00	0.74	0.78	0.84	0.93	0.78	0.81	0.89	1.01
	Mat2	0.03	0.02	0.00	0.00	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.01

Based on residual variance of 0.75, polygenic variance of 0.2, dam variance of 0.1

CHAPTER 6 Detecting Parent of Origin and dominant QTL in a two-generation commercial poultry pedigree using variance component methodology

Summary

Despite increasing evidence for parent of origin effects in crosses between divergent lines of poultry, imprinting in poultry remains a contentious issue. Variance component QTL methodology was used to analyse three candidate regions on chicken chromosomes 1, 4, and 5 for dominant and parent-of-origin effects. Data was available for bodyweight and conformation score measured at 40 days for a two-generation commercial broiler dam line. 100 dams were nested in 46 sires with phenotypes and genotypes on 2708 offspring. Linear models were constructed to simultaneously estimate fixed, polygenic and QTL effects. Different genetic models were compared by hierarchical extension to incorporate more variance components, and likelihood ratio test statistics derived from the comparison of full with reduced or null models. Empirical thresholds were derived by permutation analysis. Dominant QTL were found for bodyweight on chicken chromosome 4 and for bodyweight and conformation score on chicken chromosome 5. A maternally expressed QTL for bodyweight and conformation score was found on chromosome 1 in a region corresponding to orthologous imprinted regions in the human and mouse.

6.1. Introduction

The effectiveness of selection procedures utilising genomic information relies on correctly identifying the mode of inheritance of desired variants. Despite intense selection there is evidence to suggest that there is still much variation that might be exploited within commercial populations (Andersson *et al*., 2004; De Koning *et al*., 2004). This continued segregation together with reciprocal effects and heterosis from line crosses utilized in commercial poultry (Fairfull *et al*., 1983; Liu *et al*., 1995; Marks 1995; Nestor *et al*., 2005) suggests that at least part of the genetic variance may be non-additive.

The underlying genetic architecture of heterosis and reciprocal effects is still not clear. It appears that both maternal effects and dominant or over-dominant genes play a role (Fairfull 1990). Tuiskula-Haavisto & Vilkki. (2007) suggest that there is also recent evidence for the role of parentally imprinted mechanisms in poultry to explain the underlying mechanism for reciprocal effects like those reported by Park *et al* (2006) for physiological traits associated with bodyweight in chicken.

Genomic imprinting affects many mammalian genes (Morison *et al*., 2005) and is brought about by epigenetic instructions or imprints that are laid down in the parental germ cells (Reik *et al*., 2001). Imprinted genes are characteristically found in a clustered organization with 80% physically linked with other imprinted genes. These clusters are conserved in mammals, marsupials and flowering plants. (Reik *et al*., 2001). Imprinting is most prevalent in foetal development and until recently was considered best described by the parental conflict hypothesis (Moore *et al*., 1991). In viviparous animals this occurs where the male exerts selection pressure for offspring to maximise use of maternal resources whereas the female limits this allocation of resources to preserve herself and future offspring. As there is no apparent parental conflict, the presence of imprinting was not thought to occur in oviparous species. Furthermore, IGF2 has been shown to be imprinted and expressed from paternal allele in man rabbits mice and sheep (Nezer *et al*., 2002), but not in the chicken (Yokomine *et al*., 2001). There is, however, recent evidence for imprinted genes in birds and lower vertebrates and for shared orthologues with mammalian imprinted genes (Dunzinger *et al*., 2005; Dunzinger *et al*., 2007). Different species may also have species specific imprinted genes (Okamura *et al*., 2006). Current theory suggests that the evolution of imprinted genes is a dynamic step-wise process with orthologues present on separate chromosomes before imprinting arose. These conserved orthologues were selected during vertebrate evolution becoming imprinted only as the need arose (Dunzinger *et al*., 2005; Edwards *et al*., 2007). Lawton *et al*., (2008) show that transcriptional silencing at imprinted loci has evolved along independent trajectories in mammals and marsupials. In the chicken imprinted effects tend to cluster on a few machrochromosomes with 78% of imprinted gene orthologues

residing on chicken chromosomes 1, 3, and 5 (Dunzinger *et al*., 2005; Morison *et al*., 2005; Tuiskula-Haavisto *et al*., 2007).

Dominant and imprinted QTL effects have been identified in poultry for economically important production and disease resistance traits. Ikeobi *et al* (2002) found that 1/3 of QTL found for fat related traits in a broiler-layer cross showed dominance effects; Yonash *et al* (1999) found both partial and overdominance QTL effects for resistance to Mareks disease, while Kerje (2003) and Tuiskula-Haavisto (2002) report dominant effects for egg production traits. Parent of origin effects in poultry are reviewed by Tuiskula-Haavisto *et al*., (2007) and have been found for bodyweight, carcass and egg production traits (McElroy *et al*., 2006; Sharman *et al*., 2007; Tuiskula-Haavisto *et al*., 2004).

To date, QTL studies in poultry have mostly involved crosses between lines or divergent populations reviewed by Hocking (2005) and Abasht *et al*.,(2006). Detection of QTL effects, however, within model organisms or experimental populations is costly and potentially of limited relevance to populations under selection. It is often more practical to explore QTL segregating within a population, particularly if it is to facilitate selection within that population.

This chapter uses a variance component approach to look for dominant and imprinted QTL associated with bodyweight and conformation score measured at 40 days in a two-generation commercial broiler population.

6.2 Materials and methods

6.2.1. Data

Phenotypes on conformation score and bodyweight, both measured at 40 days, were available for a commercial broiler dam line from Cobb-Vantress Breeding Company

Ltd. Conformation score is a subjective measure of fleshiness scored from 1-10 and was treated as normally distributed. A two-generation pedigree was available with a total of 2708 offspring with phenotypes and genotypes for markers in candidate QTL regions on chicken chromosomes 1, 4 and 5. One hundred dam families were nested within 46 sire families to give an average of 27 full sibs and 59 half sibs. Progeny of Sire and dam family sizes ranged from 9 to 149 and 14 to 44 respectively. Birds were genotyped for markers spaced approximately every 16, 14 and 8 cM on chromosomes 1, 4, and 5, respectively. Markers were selected from the consensus linkage map (Schmid *et al.*, 2000). Linkage groups corresponded to the consensus map at approximately 128-205cM, 75 – 182cM, and 57-104cM for chicken chromosomes 1, 4 and 5 respectively. Linkage maps were estimated using CriMap (Green *et al.*, 1990). Marker distances and consensus map positions are given in Appendix 2.1, and further details can be found in **chapter 2** Rowe *et al.* (2006). Progeny were from two flocks across 17 hatch weeks. Fixed effects of sex, age of dam, and hatch within flock were fitted. Summary statistics and heritabilities can be found in Table 6.1. The phenotypic correlation between the two traits was 0.34 (0.03). These statistics were estimated only on the 2708 records used in the QTL analysis in contrast to chapter 1 where all 10,000 records for the commercial population were used.

Table 6.1 Summary statistics and heritabilities for trait data

	Mean (min,max)	sd	h^2 (s.e.)	c^2 (s.e.)
Bodyweight (g)	2510 (820, 3560)	300.4	0.08 (0.06)	0.045 (0.03)
Conformation Score	3.35 (1, 5)	0.83	0.08 (0.06)	0.03 (0.03)

h^2 polygenic heritability based on animal model, c^2 random common environmental or maternal effect

6.2.2 Statistical Genetic Models for Variance Component Analysis

Following the two–step approach described by George *et al.* (2000), IBD coefficients were estimated for all relationships in the pedigree to calculate the covariance matrices for the QTL effects, which were subsequently used in a linear mixed model.

The statistical models used were:

(1) $\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{Wc} + \mathbf{e}$ (null or polygenic)

(2) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Za} + \mathbf{e}$ (additive)

(3) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e}$ (additive + dominance)

(4) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Z_m m} + \mathbf{Z_p p} + \mathbf{e}$ (maternal + paternal)

(5) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Z_p p} + \mathbf{e}$ (paternal)

(6) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{Wc} + \mathbf{Z_m m} + \mathbf{e}$ (maternal)

where **y** is a vector of phenotypic observations, $\boldsymbol{\beta}$ is a vector of fixed effects, **u, a, d, m**, **p**, **c** and **e** are vectors of random additive polygenic effects, additive and dominance QTL effects, maternal and paternal QTL effects, non genetic maternal effects and residuals, respectively. **X, Z, W**, $\mathbf{Z_m}$**,** and $\mathbf{Z_p}$ are incidence matrices relating to fixed and random genetic, direct maternal, maternally expressed and paternally expressed QTL effects, respectively.

Variances for polygenic and QTL effects are distributed as follows: var(**u**) =$\mathbf{A}\sigma^2_a$ Var(**a**) = $\mathbf{G}\sigma^2_q$, Var(**d**) = $\mathbf{D}\sigma^2_d$, Var(**m**) = $\mathbf{G_M}\ \sigma^2_m$, Var(**p**) = $\mathbf{G_P}\ \sigma^2_p$
var(**e**) = $\mathbf{I}\sigma^2_{e}$,. For the non-genetic maternal effect Var(**m**) = $\mathbf{I}\sigma^2_m$. Where **A** is the standard additive relationship matrix based on pedigree data only and the relationship matrices **G,** $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** for a given QTL position are calculated from the gametic IBD matrix as outlined by Liu *et al.,*(2002) described further in **chapter 2**.

6.2.3 IBD Estimation

The **G**, $\mathbf{G_M}$, $\mathbf{G_P}$ and **D** matrices are conditional on flanking marker information and therefore unique for each position evaluated for a QTL. Here IBD was estimated using the recursive method of Pong-Wong *et al*. (2001) and the matrices calculated every 5 cM, as described in **chapter 2**.

6.2.4 Test statistic

A test statistic for a given location was obtained by comparing the likelihood of the full versus the reduced model. Twice the difference between the log likelihood of the full versus the reduced model was used as a log likelihood ratio test (LRT). For linkage group-wise test statistics genotypes were permuted within dam families to remove associations with IBD status and phenotype. Estimators for polygenic variances remained un-permuted. After each permutation, analyses were repeated for every test position along the chromosome and the highest test statistic at the best position recorded. After 1000 permutations the test statistics were ranked and the 95th percentile used for a linkage group-wise 5% type 1 error rate. Separate permutation analyses were carried out for each trait. Permutation analysis for all three chromosomes was similar so thresholds were set using the results from chromosome 4 as this is the linkage group with the most tests. In each case the highest test statistic for each model was recorded regardless of position. Thresholds and corresponding *P* values for χ^2_k distribution with k df equal to the number of extra QTL components in the full versus the reduced model for each test are given in Table 6.2.

6.2.4.1 Detection of dominant QTL effects

To detect dominance three tests were carried out;

(i) *add*, comparing the additive QTL model (2) versus the null model (1) to test significance of the QTL variance component under a purely additive model

(ii) *addom*, comparing the additive QTL + dominance QTL model (3) versus the null model (1) to test significance of QTL variance components under a model including additive and dominance effects.

(iii) *dom*, comparing the additive QTL + dominance QTL model (3) vs. the additive QTL model (2) to test the significance of the dominance variance component.

Tests (i) and (ii) are used in the initial search for the QTL whereas test 3 is applied subsequently to test specifically for the dominance component.

6.2.4.2 Imprinting

Initially QTL can be searched for using additive, *pat+mat* or single parental models. To test for imprinting four tests were carried out at each position.

(i) The *pat+mat* model (4) was tested against the null model (1)

(ii) The *pat+mat* model (4) was tested against the add model (2)

(iii) and (iv) The paternal and maternal models *pat* (5) and *mat* (6) were tested separately against the *pat+mat* model (denoted *patvfull* and *matvfull* respectively).

Table 6.2 Tests for QTL effects and corresponding empirical thresholds for 5% type 1 error based on 1000 permutations.

Test	QTL in Model		QTL effect tested for	Test		Conformation score	
	alternative (H1)	null (H0)		*LRT (5%)	*P*	*LRT 5%	*P*
add	add (2)	null (1)	additive	5.74	0.02	4.53	0.03
addom	add + dom (3)	null (1)	additive + dominant	6.98	0.03	5.84	0.05
pat+mat	pat + mat (4)	null (1)	paternal + maternal	3.05	0.09	2.94	0.08
pat	pat (5)	null (1)	paternal	7.16	0.03	6.6	0.04
mat	mat (6)	null (1)	maternal	5.38	0.02	4.54	0.03
dom	add + dom (3)	add (2)	dominant	4.80	0.03	5.12	0.02
impvmend	pat + mat (4)	add (2)	Parent of origin	3.18	0.07	3.43	0.06
**patvfull	pat + mat (4)	pat (5)	maternally expressed	4.14	0.04	4.32	0.04
**matvfull	pat + mat (4)	mat (6)	paternally expressed	4.5	0.03	3.58	0.06

* LRT is chromosome–wise empirical threshold for 5% type 1 error rate for test statistic (twice the difference between log likelihoods for the alternative and null model), estimated by 1000 iterations

** For example if the test of *patvfull* is significant the model incorporating paternal and maternal QTL is explaining more variation than the paternal QTL indicating some level of maternal expression. If there is no significant difference between the *pat+mat* model and *mat* model the maternal QTl is explaining all of the variation.

Following Hanson *et al*.,(2001) under an additive model both parents contribute equally whereas for an imprinted QTL only one parent is expected to show expression. For example, for a maternally expressed QTL the expectation is that the *patvfull* test is significant and the *matvfull* test is not significant. For non imprinted QTL the expectation is that both tests are significant as there is expression from both parents.

6.2.5 Maternal effect

Common environment effects are often, at least partially, confounded with dominance and imprinting as seen in chapters 4 and 5 and,Rowe *et al*., 2008). To explore the effect on detection of dominant and imprinted QTL, random dam effects were included in the linear model to account for common environment and evaluate its effect upon the partitioning of QTL variance.

6.3. Results

6.3.1 Additive and Dominant QTL effects

Figure 6.2 shows QTL effects under additive and dominant QTL models for bodyweight and conformation score. There were chromosome-wide significant dominant QTL effects for conformation score on chromosomes 4 and 5. These effects were considerable, explaining 6.2 and 4.5% of the phenotypic variance, respectively. Table 6.3 shows that the dominant QTL explains all of the QTL variance. It also appears that some of the maternal effect is apportioned to the dominant QTL. For bodyweight,the linkage peaks for dominant effects on chromosome 5, and additive effects on chromosomes 1 and 4, failed to reach linkage group-wise significance, regardless of whether a direct maternal effect was fitted.

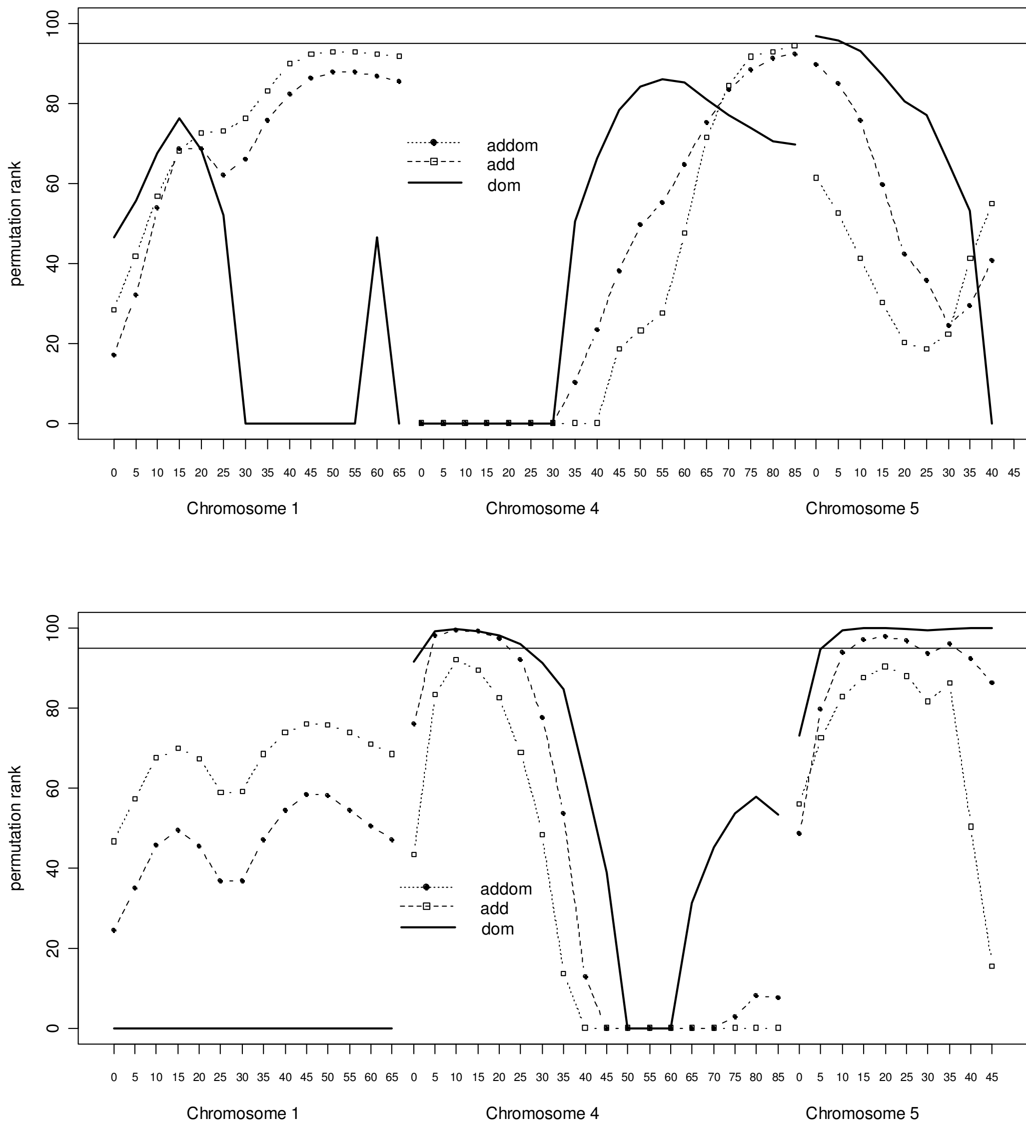


Figure 6.2. Test Statistic along chicken chromosomes 1, 4 and 5 using additive and dominant QTL models for weight (top) and conformation-score (bottom) The Y-axis shows the scaled rank of the test statistic obtained when compared to 1000 permutations of genotype within dam for 18 positions on chromosome 4 for weight and conformation-score. add is rank of test statistic obtained for model testing for additive QTL, addom is test statistic obtained from testing for both additive and dominant QTL effects and dom is test between two models for dominance only. Direct maternal effect was fitted. Solid line at top is 5% empirical linkage group-wise significance

Table 6.3 Test statistics and proportion of phenotypic variance explained at most likely QTL position when fitting polygenic, additive QTL and dominance QTL effects for 40 day bodyweight and conformation score on chicken chromosomes 1, 4 and 5

	Model fitting additive QTL				Model fitting additive and dominant QTL						
	LRT	†variance component			LRT		†variance component				
Chr (pos)	add	add	poly	dam	addom	dom	add	poly	dam	residual	dom
					Bodyweight						
1 (55)	5	0.07	0.09	0.02	5	0	0.07	0.011	0.02	0.89	0.003
4 (85)	5	0.04	0.04	0.03	5.7	0.6	0.03	0.051	0.02	0.88	0.02
5 (1)	1.4	0.02	0.06	0.02	5.3	3.9*	0.00	0.083	0.01	0.86	0.05
					Conformation score						
1 (50)	2.3	0.04	0.04	0.04	2.3	0	0.04	0.04	0.04	0.88	0.00
4 (15)	4.1	0.04	0.04	0.05	10.4*	6.3*	0.00	0.06	0.03	0.84	0.06
5 (25)	3.8	0.03	0.04	0.04	7.9*	8.1*	0.00	0.07	0.04	0.85	0.04

†Proportion of phenotypic variance explained at highest test statistic (LRT). LRT is test statistic obtained from best position (pos) for *pat+mat* versus reduced model, add is additive QTL versus null model, addom is additive and dominant QTL versus null model, dom is additive and dominant QTl versus additive QTL model

* 5 % linkage group-wise significance calculated from 1000 permutations of genotype within dam for 18 positions on chromosome 4 for weight and conformation-score

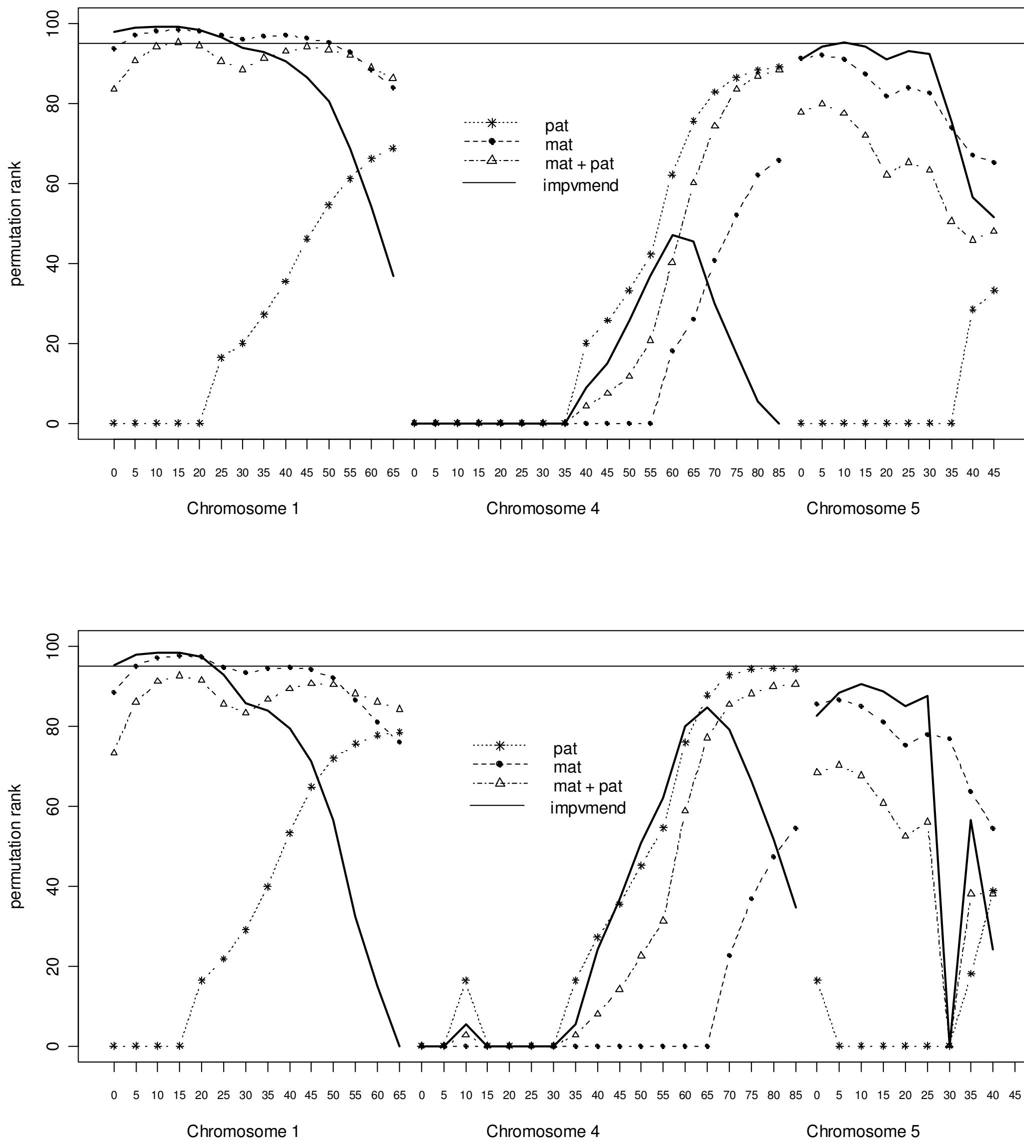


Figure 6.3 Test Statistic along chicken chromosomes 1, 4 and 5 for maternal and paternal QTL models for Body-weight. The Y-axis shows the scaled rank of the test statistic obtained when compared to 1000 permutations of genotype within dam for 18 positions on chromosome 4 for conformation score. *Mat* and *pat* are testing for maternally or paternally expressed QTL respectively. Mat + pat is fitting both maternal and paternal expression and *impvmend* is testing difference between *add* model versus *mat + pat* model. Top is without direct maternal effect fitted, bottom is with direct maternal effect fitted. Solid line at top is 5% empirical linkage group-wise significance

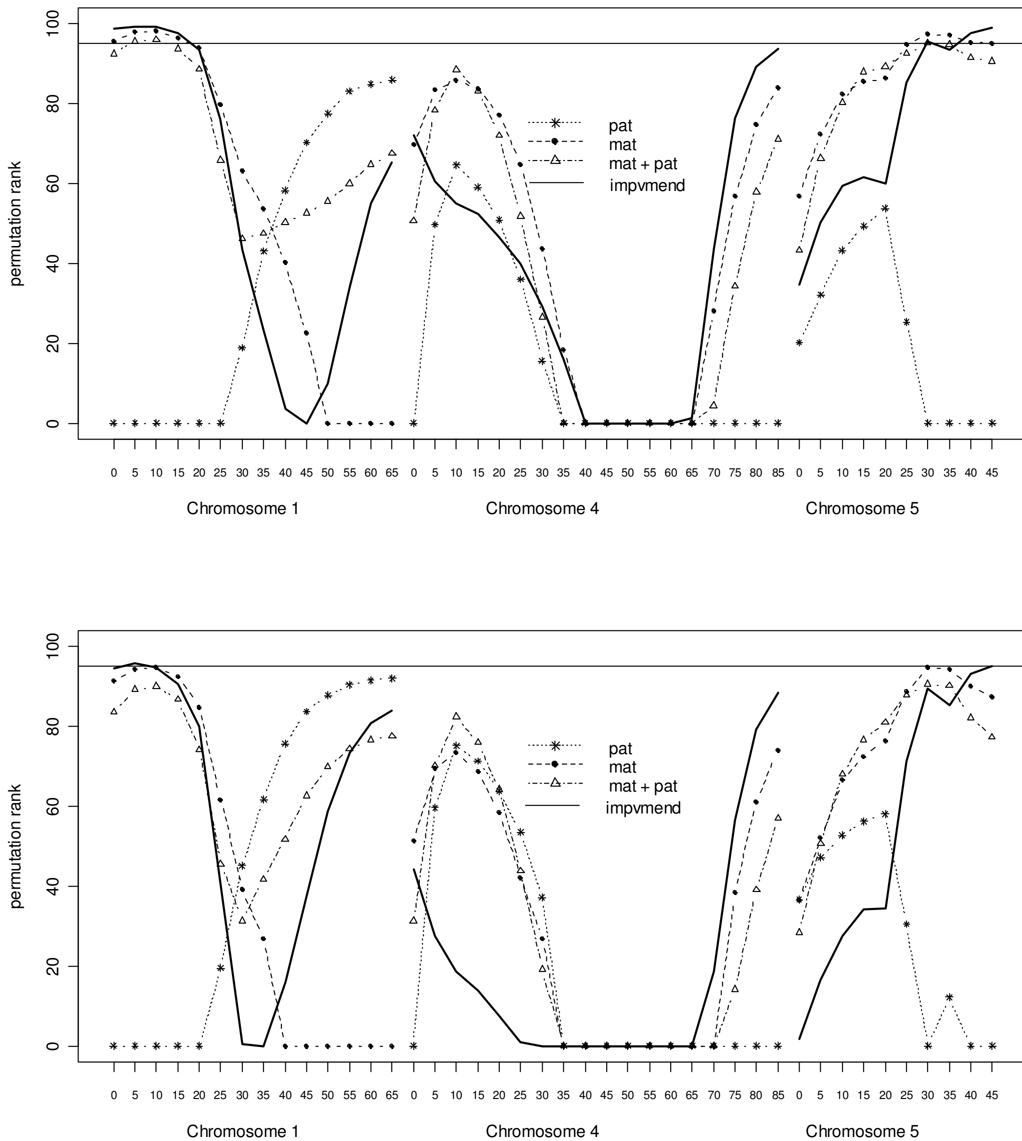


Figure 6.4 Test Statistic along chicken chromosomes 1, 4 and 5 for maternal and paternal QTL models for conformation score. The Y-axis shows the scaled rank of the test statistic obtained when compared to 1000 permutations of genotype within dam for 18 positions on chromosome 4 for conformation score. *Mat* and *pat* are testing for maternally or paternally expressed QTL respectively. Mat + pat is fitting both maternal and paternal expression and impvmend is testing difference between *add* model versus *mat + pat* model. Top is without direct maternal effect fitted, bottom is with direct maternal effect fitted. Solid line at top is 5% empirical linkage group-wise significance

6.3.2 Parent of origin QTL effects

Figure 6.3 shows rank of test statistics when compared to permutation analysis for bodyweight on chromosomes 1, 4 and 5. Figure 6.2 shows that there is little evidence for a purely additive QTL at the beginning of the chromosome 1. Figure 6.3, however, shows that the *pat+mat* model is significantly better than the *add* model and there is evidence for a maternally expressed QTL on chromosome 1 at around 10 cM. Table 6.4 also shows that the *patvfull* test is significant whereas the *matvfull* test is not indicating maternal expression. Furthermore, all of the QTL variance is explained by the maternal QTL (Table 6.5). When a dam effect is fitted, however, although the maternal QTL model and imprinting test (*impvmend*) remain significant there is no longer sufficient evidence to declare a QTL under the *pat+mat* model despite the estimate of the dam variance being zero (Table 6.5).

On chromosome 4 there is some evidence for an, at least partially, paternally expressed QTL under the separate paternal and maternal models when a direct maternal effect is fitted, however the *impvmend* statistic does not reach significance. There is some evidence for imprinting on chromosome 5 but again no linkage peaks reach 5% chromosome-wide significance.

Figure 6.4 shows test statistics for conformation score on chromosomes 1, 4 and 5. For chromosomes 1 and 5 a similar pattern is seen to bodyweight with a maternally expressed QTL when a dam effect is not fitted but insufficient evidence when dam is fitted. Chromosome 4 also has two linkage peaks, however neither reach significance.

Table 6.4 Test statistics for all models at highest test statistic for a pat+mat QTL model fitting a maternal and paternal QTL effect versus no QTL

Chr	Pos (cM)	add	addom	pat+mat	pat	mat	impvmend	patvfull	matvfull	dom
				Bodyweight, no dam effect fitted						
1	10	1.8	3.2	7.3*	0.0	7.3**	5.5*	7.3*	0.0	1.4
4	85	5.3	6.3	5.3	4.1	1.4	0.0	1.3	4.0	0.9
5	5	1.0	5.2	4.1	0.0	4.1	3.0	4.1	0.0	4.2*
				Weight dam, effect fitted						
1	10	1.7	2.6	6.3	0.0	6.3**	4.6	6.3*	0.0	0.8
4	85	5.0	5.7	5.6	5.3	0.6	0.6	0.3	5.0*	0.6
5	5	1.0	4.4	3.2	0.0	3.2	2.2	3.2	0.0	3.4*
				Conformation score, no dam effect fitted						
1	10	1.8	1.8	7.3*	0.0	7.3**	5.4*	7.3*	0.0	0.0
1	65	2.0	2.0	3.2	3.2	0.0	1.2	0.0	3.2	0.0
4	10	4.1	11.5*	5.0	1.5	3.7	0.9	3.5	1.3	7.4*
4	85	0.1	0.8	3.5	0.0	3.5	3.4	3.5	0.0	0.7
5	30	3.1	6.7	6.9*	0.0	6.8**	3.8	6.9*	0.1	3.6*
				Conformation score, dam effect fitted						
1	10	1.8	1.8	5.4	0.0	5.4*	3.6*	5.4*	0.0	0.0
1	65	1.9	1.9	4.0	4.0	0.0	2.1	0.0	4.0*	0.0
4	10	4.1	10.4*	4.4	2.1	2.6	0.2	2.3	1.8	6.3*
4	85	0.1	0.5	2.6	0.0	2.6	2.5	2.6	0.0	0.4
5	30	2.8	5.7	5.5	0.1	5.4*	2.7	5.5*	0.2	5.7*

(Model/Test spans columns add through dom.)

* and ** indicate 5 and 2.5% chromosome wise significance under permutation analysis. top table is model fitting a direct maternal effect, bottom is excluding a direct maternal effect.

Table 6.5 Proportion of phenotypic variance explained by polygenic, direct maternal (dam), paternal QTL and maternal QTL effects fitted in a *pat+mat* model at the position of the highest test statistic for *pat+mat* model versus no QTL

Chr	Position (cM)	Variance component			
		polygenic	dam	pat QTL	mat QTL
		Bodyweight			
1	10	0.09	0.00	0.00	0.06*
4	85	0.03	0.04	0.03	0.01
5	5	0.09	0.01	0.00	0.04
		Conformation score			
1	10	0.08	0.03	0.00	0.05
1	65	0.05	0.06	0.02	0.00
4	10	0.05	0.05	0.02	0.03
4	85	0.08	0.04	0.00	0.03
5	30	0.08	0.04	0.00	0.04

Table shows proportion of phenotypic variance explained by variance components. In null model with no QTL fitted polygenic heritability is 0.08, and dam component (C2) =0.05 for con and 0.03 for weight.

6.4. Discussion

6.4.1 Chromosome 1

There appears to be a maternally expressed QTL on chromosome 1 for both weight and conformation score associated with marker interval ADL0307-LEI0068, a region orthologous with imprinted regions in the mouse and human associated with Prader-willi/Angelman syndrome (Nicholls *et al.*, 2001). This region of chromosome 1, corresponding to approximately 128 to 151 cM on the consensus map, is within marker interval associated with many fat and carcass traits (Abasht *et al.*, 2006a; Ikeobi *et al.*, 2002b; Ikeobi *et al.*, 2004; Kerje *et al.*, 2003; Sewalem *et al.*, 2002). Furthermore, McElroy (2006) and Tuiskula-Haavisto *et al.*, (2004a) both find maternally expressed QTL within the same marker bracket associated with egg production. Sharman *et al* (2007) find imprinted effects for skeletal traits at 135 cM chromosome 1. Tuiskula-Haavisto *et al.*, (2004b) also find a paternally expressed QTL associated with age at first egg in the same marker interval as the putative paternally expressed effect seen here for conformation score.

Fitting a direct maternal effect appeared to reduce evidence for maternally expressed QTL. The *impvmend* test does not reach linkage group-wise significance when a direct maternal effect is fitted. It is difficult to know whether this is due to confounding of effects or that common environment can give spurious variance at the QTL. De Koning *et al* (2004) found significant additive effects for bodyweight and conformation and a strong direct maternal effect associated with this region using a three generation design from the same population. This could indicate that a strong component of the effect upon bodyweight comes from maternally influenced egg traits. Kerje *et al* (2003) report a strong correlation between egg weight and adult bodyweight (r=0.62, 0.0001) and a QTL for growth at the beginning of chromosome 1 at 68cM explaining half the phenotypic variation seen in egg weight.

6.4.2 Chromosome 4

There appear to be two separate QTL segregating for bodyweight and conformation score on chromosome 4. For bodyweight there is an additive QTL in the region of ADL0266 – LEI0076 as found by Kerje *et al*., (2003) and Jacobsson *et al*.,(2005) . There is greater evidence for this from the paternal analysis. Although the paternal QTL appears to explain most of the additive variance there is insufficient evidence for imprinting i.e. the test of the *pat+mat* model versus an additive model is not significant. For conformation, a dominant and potentially over-dominant QTL explaining all of the QTL variance maps to around 80-118cM on the consensus map. Yonash *et al* (1999b) find partial and overdominance for QTL affecting resistance to Mareks disease in this marker bracket. Although Ikeobi *et al* (2004) find many dominant effects for carcass trait QTL, they find the QTL on chromosome 4 tends to behave additively as a single locus affecting many traits. Sharman *et al* (2007) report QTL for many traits associated with skeletal traits on chromosome 4 including a dominant QTL associated with tibial marrow diameter at ADL0266-ROS0024.

6.4.3 Chromosome 5

On chromosome 5 there appear to be dominant effects for bodyweight and conformation traits. Although the test for dominance is significant for bodyweight the actual QTL does not reach linkage group-wise significance. Ikeobi *et al*., (2004) also found modest dominance effects for growth traits in this region. For conformation score, there is significant evidence for most of chromosome 5 for a dominant QTL and maternal expression at the end of the linkage group. Abasht *et al*., (2006b) also find a maternal sex interaction with fat traits in this marker bracket. Chromosome 5 has been associated with many paternally expressed traits (McElroy *et al*., 2006; Sharman *et al*., 2007; Tuiskula-Haavisto *et al*., 2007) and although the linkage group does not span the region, the first marker interval is close to a conserved gene cluster of twelve imprinted gene orthologues shown to replicate asynchronously. Here we see no evidence for paternal

imprinting on chromosome 5. Ikeobi *et al* find many QTL for traits associated with weight and carcass composition in this region although little dominance and no imprinting.

Given that we are only using a two-generation pedigree we cannot confirm that these are truly imprinted effects, only that statistically there is evidence for uniparental expression. It is possible that effects are due to QTL allele frequency differences between the sexes. There is a requirement for enough sires and dams to ensure segregation together with enough offspring to detect QTL. Because sire families were larger it might be anticipated that power to detect paternally expressed traits was greater although this is not reflected in the results. Using simulation, Tuiskula-Haavisto *et al*., (2004a) concluded that segregation is an unlikely source of spurious parent-of-origin effects. Furthermore, imprinted effects on chromosome 1 were found in regions previously identified as parentally expressed in poultry and orthologous with genomically imprinted regions in humans and mice.

6.4.5 Testing Strategy

One problem with the strategy used is that the contrast may not be greatest at the highest test statistic for the *pat+mat* model but at the highest test statistic for the parental QTL. For example, on chromosome 5 the greatest evidence for a maternal QTL and for the *impvmend* test is not at the same position as the highest test statistic for a search under the *pat+mat* model versus null. The *pat+mat* model versus null is perhaps diluted by the non expression from the imprinted parent as it is explaining the same amount of variation with an extra degree of freed. Here we find that a bodyweight QTL on chromosome 4 could be declared as paternally expressed based upon separate parental QTL models but there is insufficient evidence when comparing a Mendelian versus a *pat+mat* or imprinted model. It is difficult to know whether this is due to information source, or perhaps too stringent a threshold on the *impvmend* test and what the true effect is.

6.5 Conclusions

A large dominant and potentially over-dominant QTL for conformation score is segregating on chicken chromosome 4. This QTL is also detected under an additive model. However, the additive variance becomes zero in a model that also fits a dominance component. There is also evidence for dominant QTL affecting bodyweight and conformation on chromosome 5. There is evidence for a paternally imprinted or maternally expressed QTL affecting bodyweight and conformation score on chromosome 1 in a region orthologous with human and mouse imprinted regions and close to previously reported imprinted QTL affecting bodyweight and maternal traits in chicken and Quail. Initial results suggest that variance component analysis can be applied within commercial populations for the direct detection of segregating dominant and parent of origin effects. Further exploratory analysis might be useful to evaluate to what extent QTL effects might confounded with other sources of information or affected by segregating QTL allele frequency.

CHAPTER 7

General Discussion

In the following chapter the main results and conclusions from each of the chapters, and the underlying theory of methods used are summarised briefly. Detailed discussion follows on findings, implications and inferences from simulation and real data in turn. Implications for the livestock industry and specifically poultry are considered. Finally conclusions, lessons learned and recommendations for future research are outlined.

7.1 Summary of results

7.1.1 Commercial poultry data

Data was available for bodyweight and conformation score measured at 40 days for a two generation commercial broiler dam line. 100 dams were nested in 46 sires (i.e. Each sire was mated to on average 2 dams) with phenotypes and genotypes on 2708 offspring. Average family sizes were 59 and 27 for sires and dams respectively. Both generations were genotyped for markers spaced approximately every 16, 14 and 8 cM on chromosomes 1, 4, and 5, respectively.

Chapter 3 Comparison of Half-sib and Variance component analysis in a two generation poultry pedigree

The aim of the study was to investigate QTL in previously identified regions of chicken chromosomes 1, 4 and 5 relating to 40-day bodyweight and conformation score using a two-generation design. Half-sib and variance component analyses were implemented and compared. Half-sib QTL mapping was performed using the regression method in QTL Express for both paternal and maternal families. Confidence intervals and significance thresholds were estimated using bootstrapping and permutation analysis. Variance component mapping was done testing a novel module in QTL Express using MCMC to estimate IBD coefficients and ASReml to estimate QTL effects. Chromosome 4 showed nominal significance for QTL affecting bodyweight and conformation, and linkage was confirmed for both traits on chromosome 5. Results

varied according to method of analysis and common parent in the half-sib method. VCA tended to detect effects segregating from both parents. Analysis of dam families gave the strongest evidence for segregation of QTL. The results suggest that conformation score segregates as a separate trait in sires and dams.

Chapter 6 Detecting dominant and Parent of Origin QTL effects using variance component analysis in poultry

Variance component QTL methodology was used to analyse three candidate regions on chicken chromosomes 1, 4, and 5 for, additive, dominant and parent-of-origin effects. Linear models were constructed to simultaneously estimate fixed, polygenic and QTL effects. Different genetic models were compared by hierarchical extension to incorporate more variance components, and likelihood ratio test statistics derived from the comparison of full with reduced or null models. Empirical thresholds were derived by permutation analysis. Dominant QTL were found for conformation score on chicken chromosome 4 and for bodyweight and conformation score on chicken chromosome 5. A maternally expressed QTL for bodyweight and conformation score was found on chromosome 1 in a region corresponding to orthologous imprinted regions in the human and mouse.

7.1.2 Simulation

Simulation studies involved an investigation of QTL detection using variance component models extended to incorporate additive, dominant and separate parental QTL effects. Extensive simulations were carried out to assess accuracy of estimates, type 1 error and statistical power in two generation human, pig and poultry type pedigrees each with 1900 progeny in small, medium and large-sized families, respectively. 19, 10 and 633 sires were mated to 20, 19 and 1 dam(s) with 20, 10 and 3 progeny for poultry, pig and human pedigrees respectively. A 20cM chromosome was

simulated with 5 markers spaced 5cM apart and a QTL situated between markers 2 and 3 at 7.5cM. Test statistics for each model were computed at mid marker intervals.

Chapter 4 Detecting Dominant QTL with variance component analysis in general pedigrees

Simulation results showed that the empirical distribution of the test statistic when testing for dominant QTL effects did not behave in accordance with existing theoretical expectations and varied with pedigree structure. The distribution of the likelihood ratio test statistic was heavily dependent on family structure, with empirical thresholds lowest for human pedigrees. Power to detect QTL was high (0.84-1.0) in pig and poultry scenarios for dominance effects accounting for >7% of phenotypic variance but much lower (0.42) in human type pedigrees. Maternal or common environment effects were confounded with dominance. Including dominance in the QTL model did not affect power to detect additive QTL effects. Also, detection of spurious dominance QTL effects only occurred when maternal effects were not included in the QTL model. When dominance effects were present in the data but were not in the analysis model this resulted in both spurious detection of additive QTL or inflated estimates of additive QTL effects. Optimal power was dependent on selection of the appropriate thresholds for pedigree structure.

Chapter 5 Detecting Parent of Origin QTL with variance component analysis in general pedigrees

A range of additive, dominant and imprinted QTL effects were simulated. Testing strategies for imprinted QTL were evaluated in human, pig and poultry populations for power to detect fully imprinted QTL and for false positive rates under Mendelian inheritance.

Tests involved a model with separate maternal and paternal QTL effects tested against

No QTL

Additive QTL

Maternal QTL

Paternal QTL

Three different empirical thresholds for type 1 error were derived using varying additive and dominant QTL effects, and frequencies of the favourable allele.

The detection of variance caused by imprinted genes and in particular estimates of variance components were also heavily dependent upon the number of sire and dam families used to estimate them.

Type 1 error rates were high for the test of the separate maternal and paternal components against the additive model in the presence of large additive and dominant QTL effects. Type 1 error rates also differed markedly between human and livestock populations. For the detection of imprinting, power was greatest under a model incorporating separate parental components and could be used for an initial QTL search with little loss of power when compared to an additive model. Subsequent comparisons with individual parental effects were an effective test for parental inequalities and more robust than the test of the maternal and paternal model against the additive model.

7.2 Theory

7.2.1 Method

Using sire as a common parent a half sib design is based on the assumption that sires are unrelated and that each sire is mated to multiple dams with a single progeny. The half sib design fails to exploit information from the parent not used as the common parent or relationships between sires and between dams. Other relationships in the pedigree provide information in their contribution towards the genetic covariance. The phenotypic variance of the trait within the population is the sum of the genetic and environmental

covariances. This means that two related individuals are expected to have similar phenotypes proportional to the amount of genetic information they share measured by the number of genes or proportion of their genes descended from the same ancestral gamete or IBD. In a two generation pedigree this reduces to the alleles inherited from either parent. Variance component analysis can be applied directly within the mapping population, circumventing the need for model populations, divergent lines or crosses or very large families. In humans there is greater power over sibship methods and for livestock, breeding values can be estimated based on marker information with directly application for selection (Bennewitz, 2004).

7.2.2 IBD estimation

The inclusion of random genetic effects requires separate relationship matrices for each genetic effect. These covariances are then used to partition the phenotypic variance into genetic and environmental components. Estimating these relationships is the cornerstone of the methodology and potentially computationally expensive. The approximate method of IBD estimation used by Pong-Wong et al., (2001) has been shown to be an order of magnitude faster than MCMC whilst achieving similar results (Besnier et al., 2007; Sorensen et al., 2002). Comparing the results of chapters 3 and 6 for the detection of additive QTL it can be seen that the methodology implemented throughout the thesis performed better than that implemented in QTL Express. VCA was uninformative for much of the linkage group using LOKI, whereas this was never the case using R'Tools. The R'Tools software has two steps in the method protocol described in **chapter 2**. The first calculates allelic descent in a recursive fashion, however this is irrelevant in the case of a two generation design as the grandparental genotypes are not known. For a two generation design only the second step applies where information from sibs is used to estimate marker phase in offspring and parent. Following Liu (2002) the method uses the gametic IBD matrix to estimate the IBD coefficients for dominance i.e. the probability that two individuals will share two alleles at the QTL identical by descent

given marker information and computes a dominance covariance matrix. The diagonals from the gametic IBD matrix were also used to compute relationship matrices for separate maternal and paternal components as described in **chapter 2**. Computation of these matrices facilitated the extension of the linear model to incorporate dominance; and separate maternal and paternal QTL variances to test for inequalities of expression dependent on parental origin.

7.2.3 Models and testing strategy implemented

Five models were evaluated for the initial detection of QTL, models contained

An additive QTL

An additive and a dominance QTL

A maternal and paternal QTL

A maternal QTL

A paternal QTL

The subsequent test for dominance involved comparing the first and second model to test specifically for dominant QTL effects. An alternative would be to compare a model containing only a dominance effect with one containing additive and dominance effects, i.e. the null hypothesis is that only dominance effects are segregating however, this would seem less plausible biologically.

There are various strategies which can be employed for the detection of imprinting. The initial search can be for an additive QTL, maternal and paternal QTL in the same model i.e. allowing them to vary, or to search for maternal and paternal QTL separately and subsequently comparing each of them in turn to a full model. Here the full model including both maternal and paternal QTL was used for the initial search. The subsequent test for imprinting involved comparison of the full model with an additive model (i.e. where both parents are constrained to be equal); at the highest test statistic from the search with the full model versus no QTL. This is a similar approach to that

recommended by Thomsen et al., (2004) who use a decision tree for line cross models that initially searches under the full/additive model and only subsequently tests for individual parental components if there is evidence for a QTL. They also report that fewer QTL are found using a Mendelian versus imprinting test than tests based on significance of parental alleles.

One problem with this approach is that under imprinting the contrast between models may not be greatest at the highest tests statistic for the full model but at the highest test statistic for the non-imprinted parental QTL. This is seen in **chapter 6.** For example, on chromosome 5 the greatest evidence for a maternal QTL and for the test for imprinting is not at the same position as the highest test statistic for a search under the full model versus null. The full model versus null is diluted by the non expression from the imprinted parent as it explains the same amount of variation as the maternal QTL but the test involves an extra degree of freedom. Using simulation Hanson *et al*., (2001c) show that power to detect an imprinted QTL using VC analysis was significantly increased when modelling separate parental contributions. This is also noted by McElroy *et al* (2006) who test separately for maternal QTL and paternal QTL effects even when there is no evidence for additive effects under an additive QTL or a model including paternal and maternal QTL together.

Shete et al., (2003; 2002) argue that testing the difference between the maternal and paternal components is a valid test for imprinting as the null hypothesis involves maternal QTL=paternal QTL=0. This is implicit as the maternal and the paternal covariances sum to the additive covariance, therefore if there is any degree of uniparental expression they cannot be equal. Testing the difference between the QTL as a search strategy would still require the initial ascertainment of a significant QTL. The test of the full model including a maternal and a paternal QTL component with an additive model (*imp* test) is based on using the additive model as the null as a special case of constraining the parental components to be equal as described by Hanson et al., (2001).

De Koning *et al* (2002) investigate search strategies in an F2 population using simulation and find that although power is increased when parental contributions are modelled separately the probability of spurious detection of QTL is also increased. Here we find that a bodyweight QTL on chromosome 4 could be declared as paternally expressed based upon separate parental QTL models but there is insufficient evidence when comparing an additive versus a full or imprinted model. It is difficult to know whether this is due to information source, or perhaps too stringent a threshold on the *imp* test and what the true effect is.

Common environment was found to be confounded with both the test for dominance denoted *dom* and the test for imprinting denoted *imp*, therefore the effect of fitting a non-genetic maternal effect was also evaluated under QTL models used for the real data. These effects may arise, for example, from the rearing ability of the dam, or maternal nutrition. In poultry juvenile bodyweight has been shown to be correlated with egg size in turn affected by the age of the dam. Another potential source of confounding was polygenic dominance although is unclear to what level a polygenic effect might affect a single locus. It is hypothesised that it might be accounted for when fitting a common environment effect as both are estimated from the inflated covariance of full sibs.

7.4 Expectation of distributional properties of the Likelihood Ratio Test statistic (LRT)

Theoretically, the asymptotic distribution of the LRT is a mixture of chi square distributions with different degrees of freedom according to the number of parameters constrained (Self and Liang., 1987; Stram and Lee., 1994). This is because variance components are constrained to be non negative to make biological sense and therefore it becomes a one sided test. For example with one extra variance component the null distribution is a mixture of ½ χ_0 (i.e.variance is zero) and ½ χ_1 (i.e. variance is non-zero. With a model including two variance components, such as additive and dominant QTL effects the expectation of the distribution would be a mixture of ¼ χ_0 (both variance

components are zero, ½ χ_1 (one is non zero) and ¼ χ_2 (both are non zero). Studies which assume this mixture of distributions include (Hanson et al., 2001; Self et al., 1987; Shete et al., 2003). These assumptions hold providing the trait does not violate multivariate normality. The alternative hypothesis follows a non central chi squared distribution. Visscher (2006) provides a thorough review. Note these assumptions only hold for a point estimate corresponding to a single test (Self and Liang, 1987, Lander and Botstein, 1989). It is therefore, inappropriate to compare theoretical mixture distributions with global empirical estimates for which distributional properties remain unresolved. Here although the mixture of distributions is used for comparison no distributional assumptions about the test statistic were made. The empirical derivation of the null distribution of the test statistic allowed the comparison across populations and genetic effects.

7.5 Findings from Simulation studies

7.5.1 Type 1 error rate

When selecting empirical thresholds either from simulation studies or permutation tests one issue is at what position to select the most appropriate threshold for each test. For example, alternatives might be i) to use the highest test statistic at the position of the QTL, ii) to compare all models at the highest test statistic for the initial search, iii) to select the highest test statistic for each comparison regardless of position. Here the third option was used to ensure a conservative approach.

Dominance

Simulation showed that the empirical distribution of the test statistic for dominance *dom* did not behave in accordance with existing theoretical expectations and varied with pedigree structure. This was apparent in the much lower empirically derived threshold for the human pedigree. In total, eight chromosome-wise distributions each with 1000

replicates were simulated (Table 6.6) reproduced here in Table 7.1. Type 1 error rates were lower than the expectation for a nominal test, under a mixture of chi square distributions of 1 and 0 degrees of freedom, in all pedigrees even when estimated chromosome-wise. As there were only 4 tests this perhaps might not be expected to be very different from the nominal test statistic. When the distributions were compared the most conservative were those with single dams mated to each sire i.e. analogous to human populations. The distributions appear to group according to number of dams regardless of the number of full sibs within dams. It is possible that this is due to all variance components being estimated within dam. The covariance between full sibs consists of one half of the additive genetic variance, one quarter of the variance due to dominance and any common environment effects thus information about dominance comes from the comparison of information from full sibs. When there are half sib families the additive variance can be estimated by the covariance of half sibs which is on average ¼ of the additive genetic variance. If there are no half sib families it could be hypothesised that confounding of the two variances would mean they were more likely to be both zero. Another hypothesis is that the estimation of the relationship between siblings based on markers would be less informative and possibly less accurate. This was not the case with the additive QTL, not only were distributional properties close to the expectation but the IBD matrices estimated by MCMC agreed with those estimated using the deterministic method in R'Tools.

Table 7.1 Empirical 5% thresholds for LRT test statistic when testing for dominance and corresponding P value under χ^2_1 distribution. 1000 replicates simulated for chromosome-wise testing under null scenario of no QTL effects

Pedigree	Sires	Dams per sire	Progeny per dam	LRT 5% empirical threshold
1 (human)	633	1	3	1.46
2	380	1	5	1.14
3	190	1	10	1.38
4	317	2	3	2.1
5	190	2	5	2.02
6	195	2	10	2.06
7 (chick)	19	5	20	2.62
8 (pig)	10	19	10	2.70

Further tests for dominance were carried out under the scenarios simulated to test for false positives in imprinting. Here the residual variance was scaled differently such that the QTL effects explained a greater proportion of the variance. With much greater additive effects type 1 error rates for dominance rose sharply. Regardless of the ratio to residual variance with an additive QTL explaining 15% of the phenotypic variance type 1 error for a test for dominance was 0.05. When the additive QTL increased to 23% of the phenotypic variance however the type 1 error rate increased to 0.15 in poultry, 0.33 in pigs and 0.41 in human pedigrees. These type 1 error rates were based on 100

replicates. This is contrary to the initial result that type 1 error rates are low in human populations. Although the probability of additive effects segregating of this magnitude are low (Hayes et al., 2001) it is clear that there is a threshold over which the distributional properties of the test are altered. Derivation of type 1 error using more replicates may be appropriate but it seems unlikely that sampling variation could account for the variation of the test statistic. When 1000 replicates were used to derive 5% type 1 error thresholds again there is marked increase with larger additive effects as shown in Table 7.2 derived for the chicken and human pedigrees.

Table 7.2 Empirical 5% thresholds for the null distribution of the test for dominance *dom* in human and poultry pedigrees. 1000 replicates were simulated for chromosome-wise testing under null scenario of no dominance and varying additive QTL h^2

Additive QTL (h^2)	Human	Poultry
0.0	1.48	2.78
0.01	1.76	2.48
0.04	1.6	2.16
0.09	1.66	2.62
0.15	1.88	3.72

Although QTL have been reported of this magnitude these are large effects and it might be hypothesised that for the range of QTL effects one might expect that the spurious rates found for the *imp* test in **chapter 6** are unlikely.

Imprinting

The empirical 5% threshold for the *imp* test derived with no QTL effects simulated, corresponded to a nominal LRT value of around 2.7% for all populations in line with expectations under a mixture of $\chi_{1\text{-}0}$. Furthermore type 1 error rates of a similar order were found under permutation analysis for the real poultry data shown in Table 3.
False positive rates for non imprinted QTL from the initial empirical distribution, however, were surprisingly high and imply that the use of tabulated thresholds for this type of test should be approached with caution. In particular dominant QTL effects appeared to inflate the *imp* test statistic with rates of ~0.7 for all pedigrees for an over-dominant QTL explaining 13% of the phenotypic variance. Spurious imprinting was also seen for large additive effects. As seen for spurious dominance at more modest, and potentially more biologically plausible, additive effects, false positive rates for imprinting were much lower dropping to around 10% for an additive QTL explaining 7% of the variance. Despite this there appears to be an over-dominant QTL effect on chromosome 4 in the real data analysis explaining 6.2% of the phenotypic variance with little evidence for imprinting and an imprinted QTL on chromosome 1 which shows no evidence for dominance.

Table 7.3 Comparison of 5% Type 1 error thresholds set by permutation analysis and empirical distribution for simulated poultry pedigree using 1000 replicates/permutation tests.

test	Empirical 5% threshold from simulation of no QTL effects (4 tests)	Permutation linkage-wise for bodyweight (18 tests)	Permutation linkage-wise for Conformation-score (18 tests)
Addom	5.14	6.98	5.84
Add	3.78	5.74	4.53
dom	2.6	3.05	2.94
Full	5.4	7.16	6.6
Pat	3.52	5.38	4.54
Mat	3.67	4.80	5.12
Imp	2.6	3.18	3.43
Patvfull	3.4	4.14	4.32
Matvfull	3.34	4.5	3.58

Alternative thresholds for the imprinting test were derived by simulating fully dominant QTL to account for the false positive rates. The selection of these statistics was fairly arbitrary and applied retrospectively so of little application to real data analysis. It was interesting to note, however, that despite the extremely high thresholds set, 12, 31 and 54 for chicken, human and pig populations respectively there was still power to detect imprinted QTL. Again empirical distributions of the null statistic diverged with the human pedigree much more conservative than the poultry.

One alternative might be to try to incorporate more variance components in the null hypothesis, for example test a model including maternal, paternal and dominant QTL against a model with additive and dominant QTL. This however would not account for the spurious imprinting found at large additive effects. It is also more difficult to achieve convergence for parameter estimates particularly those close to zero as components in the model increase.

Contrary to these results Hanson et al., (2001) show, using simulation in human pedigrees that type 1 error rate for the *imp* test, can be controlled using a χ^2_1 distribution for a range of additive effects. They assess type 1 error differently at a locus unlinked to the QTL. They do report inflated type 1 error rates using the Haseman-Elston sibship method although not of the magnitude described here. They postulate that these might be due to its failure to account for non independence of sibships.

In summary, for comparisons between model such as the *imp* and *dom* tests, failure to select the correct null hypothesis gives spurious inflation of the test statistic. When derived under the null distribution of no QTL thresholds for the *imp* statistic are similar across pedigrees and in line with expectations under a mixture of chi square distributions. In contrast for the dominance test there are population differences even under the null hypothesis with tabulated thresholds becoming more conservative as the number of dam families per sire decreases. It is possible that the lack of independence between the variance components is affecting the distribution of the test statistic or that the null hypothesis selected is inappropriate. It is extremely difficult to hypothesise how an appropriate general threshold could be derived for the analysis of real data.

7.5.2 Power

In the light of the type 1 error rates it is difficult to discuss power of tests between models with any real degree of certainty in particular for the *imp* test.

Power to detect Dominance

Power to detect moderately large dominant QTL effects was high in livestock pedigrees reaching 95% for a QTL explaining 7% of the variance of a similar magnitude to the QTL found in the real data. Power to detect dominance at the QTL was much lower for the human pedigree at 42%. This was unsurprising as the human population structure of many small families with low numbers of full sibs would be expected to be less

informative for the detection of dominance. This is further compounded under the use of tabulated thresholds where power drops to 25%. The low empirical 5% type 1 error threshold means human pedigrees are more susceptible to type 2 errors as shown in chapter 4. Power under the 5% empirical threshold was similar to deterministic expectations incorporating non centrality parameters accounting for size of sibship as shown in Table 4 comparing results from the web based power calculator based on formulae derived by Sham and Purcell, (2000)

Increased power might be achieved in human studies from a pedigree with more generations providing information from relationships such as grandparents and cousins but this requires further exploration. Shete and Amos (2003) develop a method which also incorporates information from an extra IBD coefficient (fm_{ij}, mf_{ij}) derived from tracing descent of grandparental alleles and resulting in higher power. They simulate 500 replicates using parameters a=3, i=2, residual=1, and d=0, with a pedigree of 40 individuals where grandparents have three sons and three daughters and parents have 4-5 progeny. They see a moderate increase in power with 0.65 under an additive model increasing to 0.83 without the extra IBD information and 0.93 when it is included.

Table 7.4 Comparison of power to detect additive and dominant QTL using empirically derived, tabulated, and using web based calculator approximating deterministic formulae* for prediction of power in human sibships.

			Model			
Genetic effect		Calculator	Add + dom QTL		Add QTL	
a	d	predicted	X^2_1	Emp	X^2_{1-0}	emp
0.10	0.0	24	21	20	22	22
0.20	0.0	19	9	9	10	11
0.30	0.0	15	12	11	12	12
0.40	0.0	12	5	4	4	5
0.50	0.0	10	3	4	3	4
0.80	0.4	75	62	72		
0.80	0.5	79	79	83		
0.80	0.6	83	78	85		
0.80	0.7	88	65	73		
0.80	0.8	92	84	96		

*(Sham and Purcell, 2000)

Power to detect imprinting

As seen by others (De Koning et al., 2002; Hanson et al., 2001a; McElroy et al., 2006; Shete et al., 2002; Thomsen et al., 2004) power was greatest to detect imprinted QTL when the full model incorporating separate parental QTL was used. This also resulted in little loss of power when QTL effects were not imprinted. The subsequent comparison of individual parental models with the full model appears a more reliable indicator than the *imp* test although this is still prone to segregation issues when allele frequencies are very low and/or there are few families from which the QTL allele can segregate.

Human pedigrees were in general less powerful to detect QTL under the Mendelian and full model and more susceptible to spurious inflation of dominance and common maternal estimates when an imprinted QTL was examined under an incorrect model. Conversely, using the *imp* test, false positive rates were lowest and under stringent empirical thresholds power to detect imprinting was highest in human pedigrees. The balanced design of the human pedigree also yielded more accurate estimates of the variance components under the correct model and equal power to detect paternally or maternally expressed QTL.

Hanson et al., (2001) also use this test with 956 sibs across 263 nuclear families with median size of 3 and range from 2-11, a comparison of results is given in Table 7.5. For the *imp* test results are largely in agreement despite the clear indication from the simulation results that the empirical threshold does not perform well and fails to control type1 error in many scenarios. With a QTL explaining 30% of the variance Hanson et al (2001) find LOD scores with imprinting fitted of LOD 4.5 versus LOD 3.1 when not fitted. Here, using the threshold derived under the null hypothesis of no QTL, for a QTL explaining 26% of the variance the pat + mat model results in an LRT with an equivalent LOD score of 11.5 versus a LOD score of 4.9 under the additive model.

Table 7.5 Comparison of Power to detect imprinted QTL using the test of a full (pat + mat) model against an additive model for simulated human pedigrees

Hanson* method using *imp* test with tabulated thresholds of X^2_1		Human pedigree *imp* test statistic using 5% empirical threshold derived under no QTL	
heritability	power	heritability	power
0.1	0.20	0.04	0.45
0.2	0.62	0.13	0.90
0.3	0.90	0.26	0.99
0.4	0.99	0.38	1.0
0.5	1.0		

Hanson et al., 2001

7.6 Multiple testing

There are two general sources of multiple testing with these approaches. The first is common to all analyses involving genome scans and involves the testing at multiple positions. Secondly the use of multiple tests at each position must also be resolved.

The distribution, however, of H_0 when testing for multiple linked positions is unresolved and authors have used different approximations (Nagamine et al., 2004; Piepho 2001; Pratt et al., 2000; Xu et al., 1995), Procedures such as permutation and bootstrapping enable the setting of empirical thresholds and circumvent problems associated with failure of distributional assumptions and independence of multiple tests (Churchill et al., 1994; Visscher et al., 1996).

Here only the first issue is addressed either by deriving empirical threshold under the null hypothesis using a chromosome-wise approach for the simulation work and by permutation and bootstrapping analyses using the real data. In both cases linkage groups are small and do not cover whole chromosomes therefore further correction for multiple testing for large linkage groups or genome wide testing would be necessary. Computational complexity is likely to be an issue using this framework, particularly at the genome-wide level.

The approximation method described by Piepho (2001) relies on the gradient of change in likelihood across the linkage group. However, this method still assumes that the test statistic for a single test follows a standard Chi-square distribution under the null hypothesis and therefore does not address the issue of mixture distributions. The application of Piepho's method to either the simulated or the real data results in very small changes in test statistic, analogous to rounding up from three to 2 significant figures. This may be due to few tests in the simulation study where the likelihood profiles are very flat.

Heuven et al (2005) take the approach that if the nominal test statistic is a mixture of distributions, a more conservative approach and therefore way of dealing with multiple testing is to use df equal to the number of parameters. This seems an arbitrary way of dealing with a complex issue, the mixture of distributions are not accounted for, and the resolution is in no way proportional to the amount of multiple testing being done. Furthermore, this still doesn't account for multiple models or differences attributable to population structure.

Testing using multiple models is a more difficult issue and there is little evidence for its application within the variance component framework. For the test of maternal and paternal QTL Hanson et al., (2001) find that type 1 error rates approximately double and therefore suggest a bonferroni correction is appropriate when scanning for a QTL by selecting the highest test statistic from either parent.

7.7 Permutation analysis

The evidence presented does not point towards a single approach in the selection of thresholds but rather shows that the distributional properties of the test statistic are affected by many things and that the selection of a null hypothesis is a difficult issue. Although computationally expensive, permutation analysis is an attractive option circumventing many distributional assumptions and accounting for the genetic background of each specific data set. It could also potentially be adapted to account for multiple models as well as multiple positions.

The crucial issue is to ascertain on what level the permutation should take place. The permutation of phenotypes and genotypes within large dam families circumvented the re-estimation of IBD matrices and kept polygenic effects constant. It is possible that this method would be less appropriate for small full sib families, for example within 3 full sibs in the human pedigree where there would be little room for sampling. In particular, it is difficult to see how the use of permutation testing could be implemented in unstructured populations where there is no obvious sub class within which to permute genotypic and phenotypic information. Using regression techniques, McElroy et al., (2006) suggest the permutation of the maternal and paternal coefficients for a permutation test for imprinting. A novel and potentially exciting test might be the development of a similar approach with variance component analysis by permuting maternal and paternal coefficients at the IBD matrix level, for example with a 0.5 probability of the coefficient being selected from either the maternal or the paternal relationship matrix.

Permutation thresholds were less stringent for the *dom* test and similar to what one expect under a mixture of distributions although cannot be directly compared to the null due to testing at multiple positions. Under simulation test statistics for *addom*, *add* and *dom* empirically derived for a poultry pedigree with 4 tests were similar for the real data. They were higher in the real data given that 18 tests were carried out but ordered

similarly in terms of magnitude. The add and *addom* were more stringent than the nominal cut off for a chi square under a mixture or distributions as expected under multiple testing but the dom statistic much less so.

7.8 Results from poultry data

Additive analyses

There was nominal significance for an additive QTL affecting bodyweight on chromosome 4 and linkage was confirmed for bodyweight and conformation score on chromosome 5. The half sib analyses gave the first indication that QTL effects were dependent upon parent of origin. Test statistic curves differed in significance and shape according to whether the sire or dam was used as a common parent. Given the simplicity and ease of application could be considered a very useful initial approach although it is only applicable to pedigrees with large enough sibships from both parents. The half sib analyses also provided test statistics for individual families useful for narrowing confidence intervals and investigating pleiotropy. On chromosome 1, 3 out of 4 sire families were segregating for both weight and conformation score. This could be due to pleiotropy, or correlation between the trait measurements which was moderate at 0.34 (0.03), or potentially linkage disequilibrium as a result of intense selection. The population has been under intense selection for more than 40 generations.

Dominance

For the detection of the dominant QTL for conformation score the fitting of a direct maternal effect did not have any effect on the test statistic in contrast to the simulation findings, possibly because the dominance effect on chromosome 4 appeared to be segregating in sire families as also apparent in the half sib analysis. Low estimates of polygenic variance were associated with this region. For bodyweight on chromosome 4 the linkage curves for the *dom* and the *imp* test statistics are very similar and are reduced when a maternal effect is fitted.

Again for chromosome 5 the *dom* and *imp* linkage peaks look remarkably similar for bodyweight although these are unchanged by the fitting of a maternal effect despite most of the information coming from maternal expression.

Imprinting

The most striking result is conformation score on chromosome 1. There is no evidence from the half sib analyses for a QTL and little evidence from the variance component approach. However when the maternal and paternal QTL effects are allowed to differ there is much greater evidence for a maternally expressed QTL explaining ~6% of the variance in both traits. There is no evidence for dominance and little evidence for an additive QTL so it is not anticipated that the effects could be due to false positives. The fitting of a dam effect does reduce the power of the full model to detect a QTL although the imp and the maternal QTL remain significant for both bodyweight and conformation score. Furthermore as with the half sib analyses there is another linkage peak for a paternal QTL at the opposite end of the linkage group. This region is orthologous with the Prader-willi/Angelman syndrome region on human and mouse chromosomes (Nicholls et al., 2001). Most of this region is paternally expressed but there is a region maternally expressed and shown to affect Angelman syndrome which can cause obesity. Moreover many others have found imprinted effects here one of which is associated with egg weight which in turn is highly correlated with bodyweight.

Potential sources of type 1 error

It has been hypothesised that truncation selection can lead to spurious QTL detection and differences between sexes. Also for imprinting the assumption under the null of equal parental contributions is only valid if recombination rates are either equal amongst sexes or sex specific recombination is incorporated. Hanson et al., use simulation to show that in humans the method is robust to modest differences in genetic distances less than a ratio of 10. Although higher recombination rates in one sex can lead to the spurious detection of imprinting, imprinting itself also leads to a greater recombination rate in one sex therefore it is difficult to partition cause and effect. Genetic distances

between markers estimated in CRIMAP (Green et al., 1990) were similar for both sexes and for the physical and the consensus map. Kerje et al., (2002) also investigate sex specific recombination rates across the chicken genome and do not find evidence for high ratios of female:male recombination rates in these regions. Mackinnon and Georges (1992) suggest that spurious QTL detection may arise from intense truncation selection with different rates in sexes and that commercial populations would be most susceptible to this. The authors use an extreme scenario and do not differentiate between mutation rates between QTL and marker alleles. Here mapping was done within a commercial broiler dam line so segregation differences would not be expected between sexes although selection intensity and selection criteria for male parents mat differ to female parents. The low trait heritabilities compared to the three generation pedigree might be due to effects of selection, for example the Bulmer effect or potentially masked variation from dominance effects.

In an F2 population power is dependent on allelic differences between lines and enough heterozygote parents for power to detect segregation. For the variance component approach in livestock pedigrees the pig pedigree had the least number of sires and was least powerful at detecting paternally expressed QTL. If sires were fixed for a QTL allele, no variation would be seen in offspring. If dams were segregating then it would appear as though the trait was maternally expressed as there would only be variation in progeny that had inherited marker alleles from the dams regardless of marker segregation. Both poultry and pig pedigrees failed to give accurate estimates of the paternally expressed variance components. This suggests that for a paternally expressed QTL, for example IGF2 explaining on average 2% of the phenotypic variance power to detect the QTL or estimate the variance would be low for the pedigree structure given here. Heuven et al., (2005) also conclude that in an F2 design a balance is needed between number of sires and offspring.
Hager et al., (2008) give a derivation in the case of an F3 pedigree where a direct or maternal genetic effect could be detected as polar bipolar-dominance or imprinting

depending on other genetic parameters, and allele frequencies. Although their derivation is specific to this population it could generalise.

Despite overwhelming evidence from chapter 4 that large spurious type 1 error rates abound, for the real data analysis on chromosome 4 an over-dominant QTL explaining in the region of 6% of the variance is detected yet we see thresholds for the imp statistic under permutation of a similar magnitude to those seen when derived under the null threshold of no QTL. What is more we don't see spurious imprinting where there is a dominant QTL, we find it in a region where there is no evidence for dominance. This implies as suggested by the simulation results that spurious test statistics are controlled for moderate background QTL effects. It might also explain why these anomalies have not been previously reported in other studies as this is the most extensive simulation study of its kind to date evaluating the variance component method across species, models and genetic effects.

7.9 Implications for poultry breeding

The variance estimate for the maternal bodyweight QTL on chromosome 1 with a dam effect fitted is $2500g^2$. If we assume that this QTL is bi-allelic and fully imprinted effect the variance reduces to $4pqa^2$ at intermediate allele frequencies. This could result from additive QTL effects of approximately 50g rising to around 85g if frequency of the QTL favourable allele drops is 0.1. The standard deviation of the trait is 300g so a relatively modest imprinted effect could potentially cause considerable variation at the phenotypic level.

Desired benefits of genomic analyses for commercial populations are the ability to select against specific variants for example the halothane gene in pigs, and to incorporate genomic information to estimate breeding values more accurately to maximise selection. The greatest advantage for many breeding schemes is the reduction of generation

intervals and accurate breeding values for traits difficult to measure or measurable in only one sex.

Potential methods for commercial poultry breeders are marker assisted selection tools such as marker assisted BLUP where markers are fitted into the traditional selection criteria. This is useful for removing specific variants and maximising performance using a few marker traits. Genomic selection is also an increasingly popular tool incorporating marker genotypes or haplotypes to produce individual breeding values. Advantages of genomic selection are that greater resolution is achieved using haplotypes and further fine mapping is not essential. Disadvantages are that it requires a significant amount of genomic information and a huge investment. With the short generation times in poultry it doesn't carry the same advantages as one might see in cattle where progeny testing is an extremely lengthy and expensive process. It also fails to account for non additive information, therefore performance over many generations is still to be evaluated.
At present there is less evidence for the financial advantages of genomic selection over traditional selection or marker assisted BLUP given the considerable investment involved. Advantages will depend on levels and patterns of linkage disequilibrium in the population and the amount of genotyping necessary for a resolution providing population wide linkage. For initial approaches family based analyses are advantageous when compared to association type analyses as there is greater LD and therefore fewer markers needed and spurious effects due to population structure such as migration and admixture are accounted for. Disadvantages are the low resolution renders the initial marker QTL association less likely to be applicable population wide. Variance component analyses can be extended to incorporate linkage disequilibrium, however advantages are relatively small (Heuven et al., 2005; Lee et al., 2005; Lee et al., 2006)

Potentially of greatest benefit is to take advantage of dominance and imprinting in terms of mate selection and reciprocal crossing for important traits. Dekkers and Chakroborty (2004) show that the use of information from dominant QTL within lines for mate

selection can result in substantial benefits over index selection, in particular with over-dominance where reported benefits are as high as 30%.

8.0 Conclusions and Recommendations

Variance component analysis was used to detect additive, dominant and imprinted QTL in a two generation poultry pedigree. Findings agreed with the results of other studies using a variety of methods. The use of an approximate algorithm for the estimation of IBD matrices performed well when compared to the MCMC based method within QTL Express.

When variance component methods were evaluated with simulation they performed well within livestock pedigrees. Type 1 error rates under the null distribution of no QTL were similar for simulated and real poultry populations. Although power was lower to detect additive and dominant effects in human pedigrees simulation results were comparable with previous findings from other studies.

Despite this the derivation of empirical type 1 errors revealed that the underlying assumptions of the test statistic are fundamentally flawed. In particular tests between models such as the *imp* and *dom* test are not robust to changes in population structure, common environment effects or the segregation of large QTL effects. Some of these effects can be systematically included into the model but when using reml fitting too many effects simultaneously results in non convergence of parameters. It could be debated that the *pat+mat* and the *add* models are not truly nested and that this might account for the high type 1 error rates. In this type of analysis it is also difficult to think of any extension to include an extra genetic parameter that could guarantee true independence. Careful thought should be given to how many models and traits to test. Parsimony is important as although accuracy of the estimation of variance components was good when the model was extended there were some issues with non convergence. Results such as parent of origin effects should be viewed as a starting point for further

investigation and could be underpinned by a variety of reasons such as non segregation or selection. It is important that a body of evidence is collated.

The most challenging issue is the appropriate selection of a null hypothesis given that in real data there is no prior knowledge of potentially confounding effects. Permutation analysis is an attractive option as this provides a way to account for systematic effects. It has not been evaluated, however, under the extreme QTL effects which resulted in the deviation from expectation of the likelihood ratio statistic. The similarity in results from permutation testing and theoretical expectation could be interpreted as failure to correct the test statistic appropriately. There are other methods to compare likelihoods such as the Bayesian information criteria BIC, however this also compares likelihoods based on parameters constrained and would not be expected to provide greater accuracy.

Future work

Deriving empirical thresholds is useful and tells us how different family structures are affected but selection is arbitrary therefore an unreliable method of accounting for underlying variation. As computational performance becomes less of an issue with the advent of tools such as grid computing, tests such as permutation will become much faster. Computational restraints might be further removed by faster IBD approximations (Besnier et al., 2007), and faster algorithms such as the score statistic (Ronnegard et al., 2008). Further analysis might involve the use of simulation to evaluate permutation testing strategies within different population structures and subclasses. An interesting extension of this for the detection of imprinting could include the permutation of maternal and paternal relationship matrices with an equal probability of receiving either coefficient/parental allele. The effects of selection and allele frequencies would also be of interest. Finally for human type pedigrees analysis could be extended to incorporate further generations and further information from more complex pedigrees which is where the greatest benefit of the method can be derived.

This thesis presents the most extensive evaluation of variance component methodology to date incorporating different modes of inheritance, genetic effects and population structures. Here we have shown that although theoretically the tabulated chi square values are fairly robust, the expected probability of non zero variances can vary with population structure, thus there are instances when greater power is achieved by empirically deriving the correct distribution of the test statistic. Conversely there are other scenarios when type 1 error rates are high and the likelihood of spurious detection of effects is increased. Recommendations would be that variance component analyses are a useful exploratory tool but should be used with caution. Although the method was robust for large sibships with moderate QTL effects, it is incorrect to assume distributional properties for the likelihood ratio test statistic under the null hypothesis. Thresholds should be derived using permutation or empirically as each population structure and set of genetic parameters presents a unique platform and should be treated discretely.

BIBLIOGRAPHY

Abasht, B., Dekkers, J. C., & Lamont, S. J. 2006, "Review of quantitative trait loci identified in the chicken", *Poult.Sci.*, vol. 85, no. 12, pp. 2079-2096.

Abasht, B., Pitel, F., Lagarrigue, S., Le Bihan-Duval, E., Le, R. P., Demeure, O., Vignoles, F., Simon, J., Cogburn, L., Aggrey, S., Vignal, A., & Douaire, M. 2006, "Fatness QTL on chicken chromosome 5 and interaction with sex", *Genet.Sel Evol.*, vol. 38, no. 3, pp. 297-311.

Abecasis, G. R., Cardon, L. R., & Cookson, W. O. 2000, "A general test of association for quantitative traits in nuclear families", *Am.J.Hum.Genet.*, vol. 66, no. 1, pp. 279-292.

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. 2002, "Merlin--rapid analysis of dense genetic maps using sparse gene flow trees", *Nat.Genet.*, vol. 30, no. 1, pp. 97-101.

Alfonso, L. & Haley, C. S. 1998, "Power of different F-2 schemes for QTL detection in livestock", *Animal Science*, vol. 66, pp. 1-8.

Allison, D. B., Neale, M. C., Zannolli, R., Schork, N. J., Amos, C. I., & Blangero, J. 1999, "Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure", *American Journal of Human Genetics*, vol. 65, no. 2, pp. 531-544.

Almasy, L. & Blangero, J. 1998, "Multipoint quantitative-trait linkage analysis in general pedigrees", *American Journal of Human Genetics*, vol. 62, no. 5, pp. 1198-1211.

Almasy, L. & Blangero, J. 2004, "Exploring positional candidate genes: linkage conditional on measured genotype", *Behav.Genet.*, vol. 34, no. 2, pp. 173-177.

Amos, C. I. 1994, "Robust variance-components approach for assessing genetic linkage in pedigrees", *American Journal of Human Genetics*, vol. 54, no. 3, pp. 535-543.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., ndersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., & . 1994, "Genetic mapping of quantitative trait loci for growth and fatness in pigs", *Science*, vol. 263, no. 5154, pp. 1771-1774.

Andersson, L. & Georges, M. 2004, "Domestic-animal genomics: deciphering the genetics of complex traits
1", *Nat.Rev.Genet.*, vol. 5, no. 3, pp. 202-212.

Atwood, L. D., Heard-Costa, N. L., Cupples, L. A., Jaquish, C. E., Wilson, P. W., & D'Agostino, R. B. 2002, "Genomewide linkage analysis of body mass index across 28 years of the Framingham Heart Study", *Am.J.Hum.Genet.*, vol. 71, no. 5, pp. 1044-1050.

Barlow, D. P. 1995, "Gametic imprinting in mammals", *Science*, vol. 270, no. 5242, pp. 1610-1613.

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L., & Weber, J. L. 1998, "Comprehensive human genetic maps: individual and sex-specific variation in recombination", *Am.J.Hum.Genet.*, vol. 63, no. 3, pp. 861-869.

Calus, M. P., Meuwissen, T. H., de Roos, A. P., & Veerkamp, R. F. 2008, "Accuracy of genomic selection using different methods to define haplotypes", *Genetics*, vol. 178, no. 1, pp. 553-561.

Cantet, R. J., Schaeffer, L. R., & Smith, C. 1992, "Reduced animal model with differential genetic grouping for direct and maternal effects", *J.Anim Sci.*, vol. 70, no. 6, pp. 1730-1741.

Carlborg, O., Kerje, S., Schutz, K., Jacobsson, L., Jensen, P., & Andersson, L. 2003, "A global search reveals epistatic interaction between QTL for early growth in the chicken", *Genome Res.*, vol. 13, no. 3, pp. 413-421.

Carlborg, O., Hocking, P. M., Burt, D. W., & Haley, C. S. 2004, "Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth
3", *Genet.Res.*, vol. 83, no. 3, pp. 197-209.

Carlborg, O. & Haley, C. S. 2004, "Epistasis: too often neglected in complex trait studies?
4", *Nat.Rev.Genet.*, vol. 5, no. 8, pp. 618-625.

Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N., & Georges, M. 2001, "The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status", *Nat.Genet.*, vol. 27, no. 4, pp. 367-369.

Churchill, G. A. & Doerge, R. W. 1994, "Empirical Threshold Values for Quantitative Triat Mapping", *Genetics*, vol. 138, no. 3, pp. 963-971.

Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loor, J. J., Everts-Van Der Wind, A., Lee, J. H., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., Lewin, H. A., Weller, J. I., & Ron, M. 2005, "Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle", *Genome Research*, vol. 15, no. 7, pp. 936-944.

David, M. E. & David, L. D. 2004, "A Simulation Study Concerning the Effect of Varying the Residual Phenotypic Correlation on the Power of Bivariate Quantitative Trait Loci Linkage Analysis", *Behavior Genetics*, vol. V34, no. 2, pp. 135-141.

De Koning, D. J., Visscher, P. M., Knott, S. A., & Haley, C. S. 1998, "A strategy for QTL detection in half-sib populations", *Animal Science*, vol. 67, pp. 257-268.

De Koning, D. J., Haley, C. S., Windsor, D., Hocking, P. M., Griffin, H., Morris, A., Vincent, J., & Burt, D. W. 2004, "Segregation of QTL for production traits in commercial meat- type chickens", *Genetical Research*, vol. 83, no. 3, pp. 211-220.

De Koning, D. J., Archibald, A., & Haley, C. S. 2007, "Livestock genomics: bridging the gap between mice and men", *Trends Biotechnol.*, vol. 25, no. 11, pp. 483-489.

de Roos, A. P., Schrooten, C., Mullaart, E., Calus, M. P., & Veerkamp, R. F. 2007, "Breeding value estimation for fat percentage using dense markers on Bos taurus autosome 14", *Journal of Dairy Science*, vol. 90, no. 10, pp. 4821-4829.

Dekkers, J. C. M. & Chakraborty, R. 2004, "Optimizing purebred selection for crossbred performance using QTL with different degrees of dominance", *Genetics Selection Evolution*, vol. 36, no. 3, pp. 297-324.

Diao, G. & Lin, D. Y. 2005, "A powerful and robust method for mapping quantitative trait loci in general pedigrees", *American Journal of Human Genetics*, vol. 77, no. 1, pp. 97-111.

Dong, C., Li, W. D., Geller, F., Lei, L., Li, D., Gorlova, O. Y., Hebebrand, J., Amos, C. I., Nicholls, R. D., & Price, R. A. 2005, "Possible genomic imprinting of three human obesity-related genetic loci", *Am.J.Hum.Genet.*, vol. 76, no. 3, pp. 427-437.

Dunzinger, U., Haaf, T., & Zechner, U. 2007, "Conserved synteny of mammalian imprinted genes in chicken, frog, and fish genomes", *Cytogenet.Genome Res.*, vol. 117, no. 1-4, pp. 78-85.

Duong, C., Charron, S., Xiao, C., Hamet, P., Menard, A., Roy, J., & Deng, A. Y. 2006, "Distinct quantitative trait loci for kidney, cardiac, and aortic mass dissociated from and associated with blood pressure in Dahl congenic rats", *Mamm.Genome*, vol. 17, no. 12, pp. 1147-1161.

Edwards, C. A., Rens, W., Clarke, O., Mungall, A. J., Hore, T., Graves, J. A., Dunham, I., Ferguson-Smith, A. C., & Ferguson-Smith, M. A. 2007, "The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals", *BMC.Evol.Biol.*, vol. 7, p. 157.

Erickson, D. L., Fenster, C. B., Stenoien, H. K., & Price, D. 2004, "Quantitative trait locus analyses and the study of evolutionary process", *Molecular Ecology*, vol. 13, no. 9, pp. 2505-2522.

Evans, G. J., Giuffra, E., Sanchez, A., Kerje, S., Davalos, G., Vidal, O., Illan, S., Noguera, J. L., Varona, L., Velander, I., Southwood, O. I., De Koning, D. J., Haley, C.

S., Plastow, G. S., & Andersson, L. 2003, "Identification of quantitative trait loci for production traits in commercial pig populations", *Genetics*, vol. 164, no. 2, pp. 621-627.

Falconer, D. S., & Mackay,T.F.C. 1996 "*Introduction to Quantitative Genetics*", Fourth Edition. 3. Longman Scientific and Technical, Essex, UK.

Fairfull, R. W., Gowe, R. S., & Emsley, J. A. 1983, "Diallel cross of six long-term selected leghorn strains with emphasis on heterosis and reciprocal effects", *Br.Poult.Sci.*, vol. 24, no. 2, pp. 133-158.

Fairfull, R. W. 1990, "Heterosis. Elsevier Science.", *In Poultry Breeding and Genetics (ed.R.D.Crawford),* pp. 913-933.

Fernando, R. L. & Grossman, M. 1989, "Marker assisted selection using best linear unbiased prediction", *Genetics Selection Evolution*, vol. 21, pp. 467-477.

Frascaroli, E., Cane, M. A., Landi, P., Pea, G., Gianfranceschi, L., Villa, M., Morgante, M., & Pe, M. E. 2007, "Classical genetic and quantitative trait Loci analyses of heterosis in a maize hybrid between two elite inbred lines", *Genetics*, vol. 176, no. 1, pp. 625-644.

Fulker, D. W. & Cardon, L. R. 1994, "A sib-pair approach to interval mapping of quantitative trait loci", *Am.J.Hum.Genet.*, vol. 54, no. 6, pp. 1092-1103.

Gengler, N., Van Vleck, L. D., MacNeil, M. D., Misztal, I., & Pariacote, F. A. 1997, "Influence of dominance relationships on the estimation of dominance variance with sire-dam subclass effects", *J Anim Sci.*, vol. 75, no. 11, pp. 2885-2891.

George, A. W., Visscher, P. M., & Haley, C. S. 2000, "Mapping quantitative trait loci in complex pedigrees: A two- step variance component approach", *Genetics*, vol. 156, no. 4, pp. 2081-2092.

Gilmour, A. R., Thompson, R., & Cullis, B. R. 1995, "Average Information REML, an efficient algorithm for variance parameter estimation in linear mixed models.", *Biometrics*, vol. 51, pp. 1440-1450.

Goddard, M. E. 1992, "Optimal effective population size for the global population of black and white dairy cattle", *Journal of Dairy Science*, vol. 75, no. 10, pp. 2902-2911.

Goldgar, D. E. 1990, "Multipoint analysis of human quantitative genetic variation", *Am.J.Hum.Genet.*, vol. 47, no. 6, pp. 957-967.

Goldgar, D. E. 1990, "Multipoint analysis of human quantitative genetic variation", *Am.J Hum.Genet.*, vol. 47, no. 6, pp. 957-967.

Green, P., Falls, K., & Crooks, S. 1990, "Documentation for CRI-MAP", *Version 2.4.*

Grignola, F. E., Hoeschele, I., Zhang, Q., & Thaller, G. 1996, "Mapping quantitative trait loci in outcross populations via residual maximum likelihood .2. A simulation study", *Genetics Selection Evolution*, vol. 28, no. 6, pp. 491-504.

Grignola, F. E., Zhang, Q., & Hoeschele, I. 1997, "Mapping linked quantitative trait loci via residual maximum likelihood", *Genetics Selection Evolution*, vol. 29, no. 6, pp. 529-544.

Hager, R., Cheverud, J. M., & Wolf, J. B. 2008, "Maternal effects as the cause of parent-of-origin effects that mimic genomic imprinting", *Genetics*, vol. 178, no. 3, pp. 1755-1762.

Haley, C.S., Knott, S.A., Elsen, J.M. 1994. “Mapping Quantitative trait loci in crosses between outbred lines using least squares”. *Genetics*, vol. 136, no. 3, pp. 1195-1207

Hanson, R. L., Kobes, S., Lindsay, R. S., & Knowler, W. C. 2001, "Assessment of parent-of-origin effects in linkage analysis of quantitative traits", *Am.J.Hum.Genet.*, vol. 68, no. 4, pp. 951-962.

Hayes, B. & Goddard, M. E. 2001, "The distribution of the effects of genes affecting quantitative traits in livestock", *Genet.Sel Evol.*, vol. 33, no. 3, pp. 209-229.

Henderson, C. R. 1975, "Rapid Method for Computing the Inverse of a Relationship Matrix", *Journal of Dairy Science*, vol. 58, no. 11, pp. 1727-1730.

Hocking, P. M. 2005, "Review of QTL mapping results in chickens", *Worlds Poultry Science Journal*, vol. 61, no. 2, pp. 215-226.

Hocking, P. M. 2005, "Review of QTL mapping results in chickens", *World's Poultry Science Journal*, vol. 61, no. 2, pp. 215-226.

Ikeobi, C. O., Woolliams, J. A., Morrice, D. R., Law, A., Windsor, D., Burt, D. W., & Hocking, P. M. 2002, "Quantitative trait loci affecting fatness in the chicken", *Anim Genet.*, vol. 33, no. 6, pp. 428-435.

Ikeobi, C. O. N., Woolliams, J. A., Morrice, D. R., Law, A., Windsor, D., Burt, D. W., & Hocking, P. M. 2002, "Quantitative trait loci affecting fatness in the chicken", *Animal Genetics*, vol. 33, no. 6, pp. 428-435.

Jacobsson, L., Park, H. B., Wahlberg, P., Fredriksson, R., Perez-Enciso, M., Siegel, P. B., & Andersson, L. 2005, "Many QTLs with minor additive effects are associated with a large difference in growth between two selection lines in chickens", *Genetical Research.*, vol. 86, no. 2, pp. 115-125.

Jansen, R. C., Johnson, D. L., & Van Arendonk, J. A. 1998, "A mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families", *Genetics*, vol. 148, no. 1, pp. 391-399.

Kerje, S., Carlborg, O., Jacobsson, L., Schutz, K., Hartmann, C., Jensen, P., & Andersson, L. 2003, "The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs 7", *Anim Genet.*, vol. 34, no. 4, pp. 264-274.

Knott, S. A., & Haley, C.S. “Maximum likelihood mapping of quantitative trait loci using full-sib families.” *Genetics*, vol. 132, no. 4, pp. 1211-1222

Knott, S.A, Elsen, J.M. & Haley , C.S. 1996. “Methods for multiple-marker mapping of quantitative trait loci in half-sib populations”. *Theor Appl Genet* vol. 93, pp. 71-80.

Knott, S. A., Marklund, L., Haley, C. S., Andersson, K., Davies, W., Ellegren, H., Fredholm, M., Hansson, I., Hoyheim, B., Lundstrom, K., Moller, M., & Andersson, L. 1998, "Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs", *Genetics*, vol. 149, no. 2, pp. 1069-1080.

Kolbehdari, D., Jansen, G. B., Schaeffer, L. R., & Allen, B. O. 2005, "Power of QTL detection by either fixed or random models in half-sib designs", *Genetics Selection Evolution*, vol. 37, no. 6, pp. 601-614.

Kruglyak, L. & Lander, E. S. 1995, "Complete multipoint sib-pair analysis of qualitative and quantitative traits", *Am.J.Hum.Genet.*, vol. 57, no. 2, pp. 439-454.

Lawton, B. R., Carone, B. R., Obergfell, C. J., Ferreri, G. C., Gondolphi, C. M., Vandeberg, J. L., Imumorin, I., O'Neill, R. J., & O'Neill, M. J. 2008, "Genomic imprinting of IGF2 in marsupials is methylation dependent", *BMC.Genomics*, vol. 9, p. 205.

Lee, H. K., Lee, S. S., Kim, T. H., Jeon, G. J., Jung, H. W., Shin, Y. S., Han, J. Y., Choi, B. H., & Cheong, I. C. 2003, "Detection of imprinted quantitative trait loci (QTL) for growth traits in pigs", *Asian-Australasian Journal of Animal Sciences*, vol. 16, no. 8, pp. 1087-1092.

Lee, S. H. & van der Werf, J. H. J. 2004, "The efficiency of designs for fine-mapping of quantitative trait loci using combined linkage disequilibrium and linkage", *Genetics Selection Evolution*, vol. 36, no. 2, pp. 145-161.

Lee, S. H. & Van der Werf, J. H. 2005, "The role of pedigree information in combined linkage disequilibrium and linkage mapping of quantitative trait loci in a general complex pedigree", *Genetics*, vol. 169, no. 1, pp. 455-466.

Lee, S. H. & Van der Werf, J. H. 2006, "Using dominance relationship coefficients based on linkage disequilibrium and linkage with a general complex pedigree to increase mapping resolution", *Genetics*, vol. 174, no. 2, pp. 1009-1016.

Lippman, Z. B. & Zamir, D. 2007, "Heterosis: revisiting the magic", *Trends in Genetics*, vol. 23, no. 2, pp. 60-66.

Liu, G., Dunnington, E. A., & Siegel, P. B. 1995, "Growth related traits in body weight selected lines and their crosses reared under different nutritional regimens", *Br.Poult.Sci.*, vol. 36, no. 2, pp. 209-219.

Liu, G., Jennen, D. G., Tholen, E., Juengst, H., Kleinwachter, T., Holker, M., Tesfaye, D., Un, G., Schreinemachers, H. J., Murani, E., Ponsuksili, S., Kim, J. J., Schellander, K., & Wimmers, K. 2007, "A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population", *Animal Genetics*.

Liu, T., Todhunter, R. J., Wu, S., Hou, W., Mateescu, R., Zhang, Z., Burton-Wurster, N. I., Acland, G. M., Lust, G., & Wu, R. 2007, "A random model for mapping imprinted quantitative trait loci in a structured pedigree: an implication for mapping canine hip dysplasia", *Genomics*, vol. 90, no. 2, pp. 276-284.

Liu, Y., Jansen, G. B., & Lin, C. Y. 2002, "The covariance between relatives conditional on genetic markers", *Genetics Selection Evolution*, vol. 34, no. 6, pp. 657-678.

Luedi, P. P., Hartemink, A. J., & Jirtle, R. L. 2005, "Genome-wide prediction of imprinted murine genes", *Genome Res.*, vol. 15, no. 6, pp. 875-884.

Lynch M, & Walsh B. 1998. "*Genetics and analysis of quantitative traits"*. Sunderland, Mass: Sinauer Associates Inc.

Mackinnon, M. J. & Georges, M. A. 1992, "The effects of selection on linkage analysis for quantitative traits", *Genetics*, vol. 132, no. 4, pp. 1177-1185.

Mackinnon, M. J. & Georges, M. A. J. 1998, "Marker-assisted preselection of young dairy sires prior to progeny-testing", *Livestock Production Science*, vol. 54, no. 3, pp. 229-250.

Marks, H. L. 1995, "Heterosis and overdominance following long-term selection for body weight in Japanese quail", *Poult.Sci.* , vol. 74, no. 11, pp. 1730-1744.

Meuwissen, T. H. E. & Goddard, M. E. 1997, "Estimation of Effects of Quantitative Trait Loci in Large Complex Pedigrees", *Genetics*, vol. 146, no. 1, pp. 409-416.

Meuwissen, T. H. E. & Goddard, M. E. 2000, "Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria With Closely Linked Marker Loci", *Genetics*, vol. 155, no. 1, pp. 421-430.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. 2001, "Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps", *Genetics*, vol. 157, no. 4, pp. 1819-1829.

Meuwissen, T. H. E., Karlsen, A., Lien, S., Olsaker, I., & Goddard, M. E. 2002, "Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping", *Genetics*, vol. 161, no. 1, pp. 373-379.

Misztal, I. 1997, "Estimation of Variance Components with Large-Scale Dominance Models", *Journal of Dairy Science*, vol. 80, no. 5, pp. 965-974.

Misztal, I., Varona, L., Culbertson, M. S., Bertrand, J. K., Mabry, J. W., Lawlor, T. J., Van Tassell, C. P., & Gengler, N. 1998, "Studies on the value of incorporating the effect of dominance in genetic evaluations of dairy cattle", *Biotechnologie, Agronomie, Société et Environnement*, vol. 2, no. 4, pp. 227-233.

Moore, T. & Haig, D. 1991, "Genomic imprinting in mammalian development: a parental tug-of-war", *Trends Genet.*, vol. 7, no. 2, pp. 45-49.

Morison, I. M., Ramsay, J. P., & Spencer, H. G. 2005, "A census of mammalian imprinting", *Trends in Genetics*, vol. 21, no. 8, pp. 457-465.

Nagamine, Y., Knott, S. A., Visscher, P. M., & Haley, C. S. 2002, "Simple deterministic identity-by-descent coefficients and estimation of QTL allelic effects in full and half sibs 12", *Genetical Research*, vol. 80, no. 3, pp. 237-243.

Nagamine, Y., Visscher, P. M., & Haley, C. S. 2004, "QTL detection and allelic effects for growth and fat traits in outbred pig populations", *Genet.Sel Evol.*, vol. 36, no. 1, pp. 83-96.

Neale, M. C. & Miller, M. B. 1997, "The use of likelihood-based confidence intervals in genetic models", *Behav.Genet.*, vol. 27, no. 2, pp. 113-120.

Nestor, K. E., Anderson, J. W., & Velleman, S. G. 2005, "Genetic variation in pure lines and crosses of large-bodied turkey lines. 3. Growth-related measurements on live birds", *Poult.Sci.*, vol. 84, no. 9, pp. 1341-1346.

Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J., Hanset, R., Karim, L., Kvasz, A., Leroy, P., & Georges, M. 1999, "An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs", *Nat.Genet.*, vol. 21, no. 2, pp. 155-156.

Nezer, C., Moreau, L., Wagenaar, D., & Georges, M. 2002, "Results of a whole genome scan targeting QTL for growth and carcass traits in a Pie?train x Large White intercross", *Genetics Selection Evolution*, vol. 34, no. 3, pp. 371-387.

Nicholls, R. D. 2000, "The impact of genomic imprinting for neurobehavioral and developmental disorders", *J.Clin.Invest*, vol. 105, no. 4, pp. 413-418.

Nicholls, R. D. & Knepper, J. L. 2001, "Genome organization, function and imprinting in Prader-Willi and Angelman syndromes", *Annual Review of Genomics and Human Genetics*, vol. 2, no. 1, pp. 153-175.

Okamura, K. & Ito, T. 2006, "Lessons from comparative analysis of species-specific imprinted genes", *Cytogenet.Genome Res.*, vol. 113, no. 1-4, pp. 159-164.

Pante, M. J. R., Gjerde, B., McMillan, I., & Misztal, I. 2002, "Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, Oncorhynchus mykiss", *Aquaculture*, vol. 204, no. 3-4, pp. 383-392.

Park, H. B., Jacobsson, L., Wahlberg, P., Siegel, P. B., & Andersson, L. 2006, "QTL analysis of body composition and metabolic traits in an intercross between chicken lines divergently selected for growth", *Physiol Genomics*, vol. 25, no. 2, pp. 216-223.

Patterson, H. D. & Thompson, R. 1971, "Recovery of inter-block information when block sizes are unequal", *Biometrika*, vol. 58, pp. 545-554.

Piepho, H. P. 2001, "A Quick Method for Computing Approximate Thresholds for Quantitative Trait Loci Detection", *Genetics*, vol. 157, no. 1, pp. 425-432.

Pong-Wong, R., George, A. W., Woolliams, J. A., & Haley, C. S. 2001, "A simple and rapid method for calculating identity-by-descent matrices using multiple markers", *Genetics Selection Evolution*, vol. 33, no. 5, pp. 453-471.

Pratt, S. C., Daly, M. J., & Kruglyak, L. 2000, "Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components", *Am.J.Hum.Genet.*, vol. 66, no. 3, pp. 1153-1157.

Pratt, S. C., Daly, M. J., & Kruglyak, L. 2000, "Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components", *Am.J Hum.Genet.*, vol. 66, no. 3, pp. 1153-1157.

Purcell, S. & Sham, P. C. 2004, "Epistasis in quantitative trait locus linkage analysis: interaction or main effect?", *Behav.Genet.*, vol. 34, no. 2, pp. 143-152.

Reik, W. & Walter, J. 2001, "Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote", *Nat.Genet.*, vol. 27, no. 3, pp. 255-256.

Ronnegard, L., Besnier, F., & Carlborg, O. 2008, "An improved method for quantitative trait loci detection and identification of within-line segregation in F2 intercross designs", *Genetics*, vol. 178, no. 4, pp. 2315-2326.

Rowe, S. J., Windsor, D., Haley, C. S., Burt, D. W., Hocking, P. M., Griffin, H., Vincent, J., & De Koning, D. J. 2006, "QTL analysis of body weight and conformation score in commercial broiler chickens using variance component and half-sib analyses", *Animal Genetics*, vol. 37, no. 3, pp. 269-272.

Ruy, D. C., Nones, K., Baron, E. E., Ledur, M. C., de Melo, C. M. R., Ambo, M., Campos, R. D. L., & Coutinho, L. L. 2005, "Strategic marker selection to detect quantitative trait loci in chicken", *Scientia Agricola*, vol. 62, no. 2, pp. 111-116.

Sanjay, S., Mark, B., Carol, J. E., Jos+®, R. F., Jianfang, C., David, B. A., & Christopher, I. A. 2004, "Effect of Winsorization on Power and Type 1 Error of Variance Components and Related Methods of QTL Detection", *Behavior Genetics*, vol. V34, no. 2, pp. 153-159.

Schaeffer, L. R. & Kennedy, B. W. 1989, "Effects of embryo transfer in beef cattle on genetic evaluation methodology", *J.Anim Sci.*, vol. 67, no. 10, pp. 2536-2543.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J. M., Wimmers, K., Burt, D. W., Smith, J., A'Hara, S., Law, A., Griffin, D. K., Bumstead, N., Kaufman, J., Thomson, P. A., Burke, T., Groenen, M. A., Crooijmans, R. P., Vignal, A., Fillon, V., Morisson, M., Pitel, F., Tixier-Boichard, M., Ladjali-Mohammedi, K., Hillel, J., Maki-Tanila, A., Cheng, H. H., Delany, M. E., Burnside, J., & Mizuno, S. 2000, "First report on chicken genes and chromosomes 2000", *Cytogenet.Cell Genet.*, vol. 90, no. 3-4, pp. 169-218.

Schork, N. J. 1993, "Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations", *Am.J.Hum.Genet.*, vol. 53, no. 6, pp. 1306-1319.

Seaton, G., Haley, C. S., Knott, S. A., Kearsey, M., & Visscher, P. M. 2002, "QTL Express: mapping quantitative trait loci in simple and complex pedigrees", *Bioinformatics.*, vol. 18, no. 2, pp. 339-340.

Self, S. G. & Liang, K. 1987, "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions", *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 605-610.

Semel, Y., Nissenbaum, J., Menda, N., Zinder, M., Krieger, U., Issman, N., Pleban, T., Lippman, Z., Gur, A., & Zamir, D. 2006, "Overdominant quantitative trait loci for yield and fitness in tomato", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 35, pp. 12981-12986.

Sewalem, A., Morrice, D. M., Law, A., Windsor, D., Haley, C. S., Ikeobi, C. O., Burt, D. W., & Hocking, P. M. 2002, "Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross", *Poult.Sci.*, vol. 81, no. 12, pp. 1775-1781.

Sewalem, A., Morrice, D. M., Law, A., Windsor, D., Haley, C. S., Ikeobi, C. O., Burt, D. W., & Hocking, P. M. 2002, "Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross
6", *Poultry Science*, vol. 81, no. 12, pp. 1775-1781.

Sham, P. C., Cherny, S. S., Purcell, S., & Hewitt, J. K. 2000, "Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data", *American Journal of Human Genetics*, vol. 66, no. 5, pp. 1616-1630.

Shete, S. & Amos, C. I. 2002, "Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting", *American Journal of Human Genetics*, vol. 70, no. 3, pp. 751-757.

Shete, S., Zhou, X. J., & Amos, C. I. 2003, "Genomic imprinting and linkage test for quantitative-trait loci in extended pedigrees", *American Journal of Human Genetics*, vol. 73, no. 4, pp. 933-938.

Shete, S. & Yu, R. 2005, "Genetic imprinting analysis for alcoholism genes using variance components approach", *BMC.Genet.*, vol. 6 Suppl 1, p. S161.

Shete, S., Elston, R. C., & Lu, Y. 2007, "A novel approach to detect parent-of-origin effects from pedigree data with application to Beckwith-Wiedemann syndrome", *Ann.Hum.Genet.*, vol. 71, no. Pt 6, pp. 804-814.

Shimomura, K., Low-Zeddies, S. S., King, D. P., Steeves, T. D., Whiteley, A., Kushla, J., Zemenides, P. D., Lin, A., Vitaterna, M. H., Churchill, G. A., & Takahashi, J. S. 2001, "Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice", *Genome Res.*, vol. 11, no. 6, pp. 959-980.

Slate, J., Visscher, P. M., MacGregor, S., Stevens, D., Tate, M. L., & Pemberton, J. M. 2002, "A Genome Scan for Quantitative Trait Loci in a Wild Population of Red Deer (Cervus elaphus)", *Genetics*, vol. 162, no. 4, pp. 1863-1873.

Sorensen, A. C., Pong-Wong, R., Windig, J. J., & Woolliams, J. A. 2002, "Precision of methods for calculating identity-by-descent matrices using multiple markers", *Genet.Sel Evol.*, vol. 34, no. 5, pp. 557-579.

Stram, D. O. & Lee, J. W. 1994, "Variance components testing in the longitudinal mixed effects model", *Biometrics*, vol. 50, no. 4, pp. 1171-1177.

Thaller, G. & Hoeschele, I. 2000, "Fine-mapping of quantitative trait loci in half-sib families using current recombinations", *Genet.Res.*, vol. 76, no. 1, pp. 87-104.

Patterson, H.D. & Thompson, R. 1971. “Recovery of interblock information when block sizes are unequal”. *Biometrika* vol. 58, pp. 545–554.

Thomsen, H., Lee, H. K., Rothschild, M. F., Malek, M., & Dekkers, J. C. 2004, "Characterization of quantitative trait loci for growth and meat quality in a cross between commercial breeds of swine", *J.Anim Sci.*, vol. 82, no. 8, pp. 2213-2228.

Tuiskula-Haavisto, M., Honkatukia, M., Vilkki, J., De Koning, D. J., Schulman, N. F., & Maki-Tanila, A. 2002, "Mapping of quantitative trait loci affecting quality and production traits in egg layers", *Poultry Science*, vol. 81, no. 7, pp. 919-927.

Tuiskula-Haavisto, M., De Koning, D. J., Honkatukia, M., Schulman, N. F., Maki-Tanila, A., & Vilkki, J. 2004, "Quantitative trait loci with parent-of-origin effects in chicken", *Genetical Research*, vol. 84, no. 1, pp. 57-66.

Tuiskula-Haavisto, M. & Vilkki, J. 2007, "Parent-of-origin specific QTL--a possibility towards understanding reciprocal effects in chicken and the origin of imprinting", *Cytogenet.Genome Res.*, vol. 117, no. 1-4, pp. 305-312.

Uimari, P., Thaller, G., & Hoeschele, I. 1996, "The Use of Multiple Markers in a Bayesian Method for Mapping Quantitative Trait Loci", *Genetics*, vol. 143, no. 4, pp. 1831-1842.

Van Arendonk, J.A.; Tier, B.; Kinghorn, B.P. 1994. “Use of multiple genetic markers in prediction of breeding values”. *Genetics*, vol. 137, no. 1, pp. 319-329.

van Kaam, J. B., Groenen, M. A., Bovenhuis, H., Veenendaal, A., Vereijken, A. L., & Van Arendonk, J. A. 1999, "Whole genome scan in chickens for quantitative trait loci affecting carcass traits", *Poult.Sci.*, vol. 78, no. 8, pp. 1091-1099.

Visscher, P. M., Thompson, R., & Haley, C. S. 1996, "Confidence Intervals in QTL Mapping by Bootstrapping", *Genetics*, vol. 143, no. 2, pp. 1013-1020.

Visscher, P. M. 2006, "A note on the asymptotic distribution of likelihood ratio tests to test variance components", *Twin Res.Hum.Genet.*, vol. 9, no. 4, pp. 490-495.

Weller, J. I. 1986, "Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers", *Biometrics*, vol. 42, no. 3, pp. 627-640.

Weller, J. I., Kashi, Y., & Soller, M. 1990, "Power of Daughter and Granddaughter Designs for Determining Linkage Between Marker Loci and Quantitative Trait Loci in Dairy Cattle", *Journal of Dairy Science* , vol. 73, no. 9, pp. 2525-2537.

Williams, J. T. & Blangero, J. 1999, "Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples", *Genet.Epidemiol.*, vol. 16, no. 2, pp. 113-134.

Wolf, J. B., Cheverud, J. M., Roseman, C., & Hager, R. 2008, "Genome-wide analysis reveals a complex pattern of genomic imprinting in mice", *PLoS.Genet.*, vol. 4, no. 6, p. e1000091.

Wong, G. K. S., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z. J., Meng, Q. S., Zhou, J., Li, D. W., Zhang, J. J., Ni, P. X., Li, S. G., Ran, L. H., Li, H., Zhang, J. G., Li, R. Q., Li, S. T., Zheng, H. K., Lin, W., Li, G. Y., Wang, X. L., Zhao, W. M., Li, J., Ye, C., Dai, M. T., Ruan, J., Zhou, Y., Li, Y. Z., He, X. M., Zhang, Y. Z., Wang, J., Huang, X. G., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G. Q., Dong, L., Lan, F. D., Sun, Y. Q., Zhang, Z. P., Yang, Z., Yu, Y. P., Huang, Y. Q., He, D. D., Xi, Y., Wei, D., Qi, Q. H., Li, W. J., Shi, J. P., Wang, M. H., Xie, F., Wang, J. J., Zhang, X. W., Wang, P., Zhao, Y. Q., Li, N., Yang, N., Dong, W., Hu, S. N., Zeng, C. Q., Zheng, W. M., Hao, B. L., Hillier, L. W., Yang, S. P., Warren, W. C., Wilson, R. K., Brandstrom, M., Ellegren, H., Crooijmans, R. P. M. A., van der Poel, J. J., Bovenhuis, H., Groenen, M. A. M., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J. R., Rogers, S., Walker, B. A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S. J., Zhou, H. J., Hocking, P. M., Morrice, D., De Koning, D. J., Law, A., Bartley, N., Burt, D. W., Hunt, H., Cheng, H. H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M. T., Andersson, B., Webber, C., Ponting, C. P., Overton, I. M., Boardman, P. E., Tang, H. Z., Hubbard, S. J., Wilson, S. A., Yu, J., Wang, J., & Yang, H. M. 2004, "A genetic variation map for chicken with 2.8 million single- nucleotide polymorphisms", *Nature*, vol. 432, no. 7018, pp. 717-722.

Xiao, J., Li, J., Yuan, L., & Tanksley, S. D. 1995, "Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers", *Genetics*, vol. 140, no. 2, pp. 745-754.

Xu, S. & Atchley, W. R. 1995, "A Random Model Approach to Interval Mapping of Quantitative Trait Loci", *Genetics*, vol. 141, no. 3, pp. 1189-1197.

Xu, S. H. & Atchley, W. R. 1995, "A Random Model Approach to Interval Mapping of Quantitative Trait Loci", *Genetics*, vol. 141, no. 3, pp. 1189-1197.

Yokomine, T., Kuroiwa, A., Tanaka, K., Tsudzuki, M., Matsuda, Y., & Sasaki, H. 2001, "Sequence polymorphisms, allelic expression status and chromosome locations of the chicken IGF2 and MPR1 genes", *Cytogenet.Cell Genet.*, vol. 93, no. 1-2, pp. 109-113.

Zeegers, M., Rijsdijk, F., & Sham, P. 2004, "Adjusting for covariates in variance components QTL linkage analysis", *Behav.Genet.*, vol. 34, no. 2, pp. 127-133.

Zeng, Z. B. 2005, "QTL mapping and the genetic basis of adaptation: recent developments", *Genetica*, vol. 123, no. 1-2, pp. 25-37.

Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister-Genskow, M., Witte, L. A., Grignola, F. E., Uimari, P., Thaller, G., & Bishop, M. D. 1998, "Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree", *Genetics*, vol. 149, no. 4, pp. 1959-1973.

Zhang, Z. M., Zhao, M. J., Ding, H. P., Rong, T. Z., & Pan, G. T. 2006, "QTL mapping analysis of plant height and ear height of maize (Zea mays L.)", *Genetika.*, vol. 42, no. 3, pp. 391-396.

Zhou, X., Chen, W., Swartz, M. D., Lu, Y., Yu, R., Amos, C. I., Wu, C. C., & Shete, S. 2007, "Joint linkage and imprinting analyses of GAW15 rheumatoid arthritis and gene expression data", *BMC.Proc.*, vol. 1 Suppl 1, p. S53.