

**Genetic Mapping of Quantitative Trait Loci
for Body Weight on the X-chromosome in Mice**

Kellie Anne Rance
Bsc Agr (Hons) 1993

Ph.D.
The University of Edinburgh
1996



DECLARATION

I declare that this thesis has been composed by me.
Specific contributions of others are acknowledged

Kellie Anne Rance

November 1996

Contents

CONTENTS	I
LIST OF TABLES	IV
LIST OF FIGURES	VI
ACKNOWLEDGEMENTS	IX
PUBLICATIONS	X
ABSTRACT	XI
CHAPTER 1.....	1
GENERAL INTRODUCTION.....	1
CHAPTER 2.....	4
LITERATURE REVIEW	4
2.1 INTRODUCTION	4
2.2 METHODS OF MAPPING QUANTITATIVE TRAIT LOCI	5
2.2.1 <i>Introduction</i>	5
2.2.2 <i>Methods for Mapping QTL</i>	5
2.2.3 <i>Interpretation of mapping results; Thresholds and Confidence Intervals.</i>	13
2.2.4 <i>Experimental Power</i>	17
2.2.5 <i>Non-inbred Populations.</i>	22
2.2.6 <i>A quantitative assessment of QTL mapping.</i>	23
2.3 SELECTION LINE HISTORY	29
2.3.1 <i>Introduction.</i>	29
2.3.2 <i>The Replicate Lines and Preliminary results.</i>	29
2.3.3 <i>Physiological Basis of Response.</i>	31
2.3.4 <i>Investigation of the Genetic Basis of Response.</i>	35
2.4 THE X-CHROMOSOME IN QTL MAPPING.....	38
2.4.1 <i>Introduction.</i>	38
2.4.2 <i>X-chromosome inactivation and selection response</i>	38
2.4.3 <i>Variability of the X-chromosome</i>	40
2.4.4 <i>X-chromosome Mapping Experiments.</i>	41
CHAPTER 3.....	43
MATERIALS AND METHODS	43
3.1 DNA EXTRACTION	44
3.2 PCR AMPLIFICATION PROTOCOL	48
3.3 MOUSE HUSBANDRY.....	50
3.4 ESTIMATION OF CARCASS FAT PERCENTAGE	51
3.5 STATISTICAL ANALYSIS	52

CHAPTER 4.....	55
SINGLE MARKER ANALYSIS OF A RECIPROCAL F_2 POPULATION TO MAP THE PUTATIVE QTL FOR BODY WEIGHT ON THE X-CHROMOSOME.....	55
4.1 INTRODUCTION.....	55
4.2 MATERIALS AND METHODS	56
4.2.1 <i>The F_2 Population mating structure and measurements taken</i>	56
4.2.2 <i>Genotyping of Animals</i>	56
4.2.3 <i>Analysis</i>	57
4.3 RESULTS	58
4.3.1 <i>Estimation of sex linked body weight effects in the F_1 generation</i>	58
4.3.2 <i>Rates of polymorphism on the X-chromosome</i>	59
4.3.3 <i>Single marker analysis of F_2 data set</i>	60
4.4 DISCUSSION	65
4.4.1 <i>The reciprocal F_1 generation and sex linked effects</i>	65
4.4.2 <i>Polymorphism on the X-chromosome</i>	65
4.4.3 <i>Estimated marker associated QTL effects in the F_2 population</i>	66
4.4.5 <i>General discussion of results</i>	67
CHAPTER 5.....	69
ADAPTATION OF AUTOSOMAL QTL ANALYSIS TO ENABLE ANALYSIS OF X-LINKED DATA	69
5.1 INTRODUCTION.....	69
5.2 MATERIALS AND METHODS	70
5.2.1 <i>Development of an X-linked Multiple Regression Analysis</i>	70
5.2.2 <i>Autosomal multiple regression analysis</i>	74
5.2.3 <i>Simulated data sets</i>	75
5.2.4 <i>Thresholds and confidence intervals for QTL detection</i>	76
5.3 RESULTS	78
5.3.1 <i>The effects of population size on QTL detection</i>	78
5.3.2 <i>QTL detection and the effect of marker spacing</i>	81
5.3.3 <i>The effect of QTL position in relation to marker position on the power of QTL detection</i>	82
5.3.4 <i>Comparison of X-Linked and autosomal data analysis</i>	84
5.3.5 <i>Investigation of different modes of inheritance</i>	85
5.3.5 <i>Analysis of two QTL</i>	87
5.4 DISCUSSION	90
5.4.1 <i>Effects of Population Size and Marker Spacing on QTL detection</i>	90
5.4.2 <i>X-linked data and analysis: specific properties compared to autosomal data and analysis</i> .	94
5.4.3 <i>Performance of the two QTL Model</i>	98
5.4.4 <i>General discussion and extension of method for non-inbred populations.</i>	99
CHAPTER 6.....	102
MAPPING QUANTITATIVE TRAIT LOCI FOR BODY WEIGHT ON THE X-CHROMOSOME USING THE X-LINKED MULTIPLE REGRESSION METHOD	102
6.1 INTRODUCTION.....	102
6.2 MATERIALS AND METHODS	103
6.2.1 <i>The F_2X Population</i>	103
6.2.2 <i>Genotyping the F_2X population</i>	105
6.2.3 <i>Data Analysis</i>	106
6.3 RESULTS	110
6.3.1 <i>Sex linked effects in the reciprocal F_1 generation</i>	110
6.3.2 <i>Polymorphic markers in the F_2X pedigree</i>	111

6.3.3 Description of the F_2X data set.....	113
6.3.4 QTL mapping using the X-linked regression analysis.....	116
6.4 DISCUSSION	130
6.4.1 The reciprocal F_1 generation and sex-linked effects	130
6.4.2 The distribution of polymorphic markers in the F_2X	131
6.4.3 Discussion of Phenotypic Measurements	131
6.4.4 Detection of X-linked QTL influencing body weight.....	134
6.4.5 Interpretation of results of an X-linked QTL for 10 week weight	140
6.4.6 Summary of Results	145
CHAPTER 7.....	146
THE PROGRESSIVE BACKCROSS POPULATION.....	146
7.1 INTRODUCTION.....	146
7.2 MATERIALS AND METHODS	148
7.2.1 DNA extraction and PCR amplification protocol.....	148
7.2.2 The mating scheme for BC populations and summary of the data collected	148
7.2.3 Preliminary analysis of the Backcross generations	153
7.2.4 Analysis of the Backcross population using all pedigree information.....	155
7.3 RESULTS	159
7.3.1 Changes in the distribution of traits measured over the generations of backcrossing.....	159
7.3.2 The preliminary analysis of the Backcross population	160
7.3.3 Gibbs sampling of the Backcross population using all information through the pedigree to estimate QTL position(s) and effects.....	164
7.4 DISCUSSION	168
7.4.1 Changes in the distribution of the trait measurements through the BC generations.....	168
7.4.2 Preliminary analysis of marker data within the BC generations.....	169
7.4.3 Analysis of the BC population using Gibbs sampling.....	171
CHAPTER 8.....	174
GENERAL DISCUSSION	174
8.1 GENERAL DISCUSSION OF THE ESTIMATED QTL POSITIONS AND EFFECTS	174
8.1.1 The estimated QTL position.....	176
8.1.2 The QTL effect associated with the X-chromosome	178
8.1.3 Conclusions from the QTL mapping on the X-chromosome.....	180
8.2 THE X-LINKED QTL FOR BODY WEIGHT AND CANDIDATE LOCI	181
8.3 FUTURE WORK.....	184
8.4 GENERAL DISCUSSION OF QTL MAPPING IN ANIMAL BREEDING.....	187
REFERENCES.....	189

List of Tables

CHAPTER 2.....	4
LITERATURE REVIEW	4
<u>Table 2.1</u> The fraction of simulations in which a significant LOD score was obtained (from van Ooijen (1992)).....	24
<u>Table 2.2</u> The mean and standard deviation (shown in parenthesis) of the length of the confidence interval for QTL position. The LOD drop off of 1 represents a 95% CI and the LOD drop off of 2 represents the 99% CI. Estimates of the fraction of CI with contain the simulated QTL are also shown (from van Ooijen, 1992).	26
<u>Table 2.3</u> Changes in Genetic Variances in the P lines with selection over generations (Beniwal <i>et al.</i> , 1992a). Estimates of V_A ($\times 10^{-4}$ on the log scale)	34
<u>Table 2.4</u> Mean body weights at 10 weeks of reciprocal F_1 generations and contemporaneous selection lines. Dams first in crossing nomenclature (from Hastings & Veerkamp, 1993). Mean number in each class=33.	36
CHAPTER 3.....	43
MATERIALS AND METHODS	43
CHAPTER 4.....	55
SINGLE MARKER ANALYSIS OF A RECIPROCAL F_2 POPULATION TO MAP THE PUTATIVE QTL FOR BODY WEIGHT ON THE X-CHROMOSOME.....	55
<u>Table 4.1</u> Mean body weight (g) at 6 and 10 weeks of reciprocal halves of F_1 P6 line crosses. Dam shown first in crossing nomenclature	58
<u>Table 4.2.</u> Mean marker associated body weight (g) at 10 weeks, and number of animals within each group (shown in <i>italics</i>). Standard errors of means are shown in parenthesis. The map positions were taken from the Chromosome Committee (Mouse Genome Informatics, 1996).	61
CHAPTER 5.....	69
ADAPTATION OF AUTOSOMAL QTL ANALYSIS TO ENABLE ANALYSIS OF X-LINKED DATA	69
<u>Table 5.1</u> Expectations of mean genotypic effects of an X-linked QTL for all possible marker genotypes in a reciprocal F_2 population.	73
<u>Table 5.2.</u> Expectations for the mean genotypic effects of an autosomal QTL for all possible flanking marker genotypes in an F_2 population.	75
<u>Table 5.3.</u> Mean results of parameter estimates over 50 replicate simulations with varying population size. The QTL was simulated at 35cM with additive effect of $a=\pm 0.25$ in males and females, with no dominance deviation. Mean standard errors of estimates are shown in parenthesis.	80
<u>Table 5.4.</u> Mean results of parameter estimates over 50 replicate simulations, with varying marker density. Mean standard errors shown in parenthesis.	82
<u>Table 5.5.</u> The X-linked QTL genotypes found in an F_2 population, with mean realised QTL effects for additive, dominant and recessive QTL.	86
<u>Table 5.6.</u> Mean results of parameter estimates over 50 replicate simulations, with simulations set up to investigate the power of X-linked analysis. Mean standard errors shown in parenthesis.	87
<u>Table 5.7.</u> The mean power and genotyping requirements for simulations generated in this study.	92

CHAPTER 6.....	102
MAPPING QUANTITATIVE TRAIT LOCI FOR BODY WEIGHT ON THE X- CHROMOSOME USING THE X-LINKED MULTIPLE REGRESSION METHOD	102
<u>Table 6.1</u> Summary of the mating structure of the F_2X population, and the distribution of the animals born per litter.	104
<u>Table 6.2</u> Mean body weight (g) at 6 weeks of individuals in the reciprocal halves of F1 line crosses. The dam is shown first in crossing nomenclature	110
<u>Table 6.3</u> The distribution of individuals with full records (genotypic and phenotypic data) in the F_2X population for segregating <i>DXMit</i> marker loci. Ticks denote litters which were fully informative at the marker loci (the three fully informative marker loci in all litters are not shown).	112
<u>Table 6.4</u> Description of phenotypic data obtained in the F_2X generation. The estimates shown are the least square means for body weights (g), body weight gains (g), and estimated fat percentage.	114
<u>Table 6.5</u> The residual phenotypic correlations of traits measured and derived traits in the F_2X population.	115
<u>Table 6.6</u> The map positions of markers estimated from the F_2X population using CRIMAP, and taken from the Mouse Genome Informatics (1996). Estimated PCR product sizes are given for the marker alleles predominant in the high or low line.	117
<u>Table 6.7</u> Summary of QTL effects (g) for the body weights obtained using X-linked multiple regression method. Standard errors of estimates are shown in parenthesis.	119
<u>Table 6.8</u> The estimated QTL effects (see Table 6.7) for the three body weight measurements expressed as percentages of mean body weight (see Table 6.4).	120
<u>Table 6.9</u> Summary of the QTL effects (g) for the body weight gains obtained by the X-linked multiple regression method. The estimates of QTL effects expressed as percentages of mean body weight gains (see Table 6.4). Standard errors of the estimates are shown in parenthesis.	125
<u>Table 6.10</u> Summary of marker associated effects (g) at the marker <i>DXMit50</i> (the marker estimated to be in close linkage with the QTL for body weight) for all body weights and body weight gains. The standard errors of the estimates are shown in parenthesis.	126
CHAPTER 7.....	146
THE PROGRESSIVE BACKCROSS POPULATION.....	146
<u>Table 7.1</u> A summary of the numbers of successful matings and individuals weaned in each generation of the backcross population. The traits recorded are also shown.	153
<u>Table 7.2</u> Changes in the means and variances of 6 and 10 week body weights over the BC generations (corrected for the fixed effects and covariates in model 1 section 7.2.3, with the exception of the fixed effect of marker genotype). The mean litter sizes are shown for each BC generation.	160
<u>Table 7.3</u> The mean effect (g) for 10 week weight associated with segments of the X-chromosome in males and females from generations 3, 4 and 5 of the progressive backcross population (standard errors shown in parenthesis). The ticks represent markers which can be traced unequivocally to the high selection line, and the remaining markers on the X-chromosome can be traced to the low inbred line.	163
<u>Table 7.4</u> The estimated QTL effects associated with 6 week and 10 week weight obtained from the Gibbs sampling analysis (using the model outlined in section 7.2.4). Standard errors of estimates shown in parenthesis.	166
CHAPTER 8.....	174
GENERAL DISCUSSION	174
<u>Table 8.1</u> Summary of estimated X-chromosome effects and estimated X-linked QTL effects in males, and the contributions to mean body weight expressed as a percentage of mean male body weight. The estimated QTL effects are shown in bold.	179

List of Figures

I

CHAPTER 2.....	4
LITERATURE REVIEW	4
Figure 2.1 Lines established from the G-lines	30
Figure 2.2 Representation of the formation of the P6 selection lines.....	32
Figure 2.3 Selection response of the P replicate lines (selected for fat free body mass), and the P6 lines selected for body weight (Mbaga, 1996).....	33
CHAPTER 3.....	43
MATERIALS AND METHODS	43
Figure 3.1 Validation of the method of obtaining genomic DNA samples using ear clips taken from individuals at weaning. Genotypes were obtained for marker DXMit50 using the DNA from high and low line mice. H/L ⁺ show the positive controls, H/L ^{Sp} are spleen sample DNA controls for the two animals used, and H/L ^e show the ear clip samples in the order taken.....	47
CHAPTER 4.....	55
SINGLE MARKER ANALYSIS OF A RECIPROCAL F₂ POPULATION TO MAP THE PUTATIVE QTL FOR BODY WEIGHT ON THE X-CHROMOSOME.....	55
Figure 4.1 An example of a Gel photo showing the eight high and low line females from generation 51 of selection genotyped at the marker loci DXMit25. The gel shows females from the low line are all homozygous for the same allele (approximately 184 base pairs (bp)), the high line individuals H ² and H ⁸ are homozygous for the 170bp allele, and the remaining high line females are heterozygous.....	60
Figure 4.2 Illustration of the mean marker associated QTL effects at 10 week weight for markers along the X-chromosome. The additive effects represent half the difference between the homozygote females, and half the difference between the hemizygous males. Map positions were taken from Chromosome Committee (Mouse Genome Informatics, 1996). The error bars represent the standard errors associated with the estimates.....	63
CHAPTER 5.....	69
ADAPTATION OF AUTOSOMAL QTL ANALYSIS TO ENABLE ANALYSIS OF X-LINKED DATA	69
Figure 5.1 The effect of varying population size on the mean test statistic of 50 replicate simulations. The QTL was simulated at 35cM with additive effect $a=\pm 0.25$ in males and females, no dominance deviation was simulated.....	79
Figure 5.2 The effect of marker spacing on the mean TS obtained for 50 replicate simulations of population size of 350 with QTL of additive effect of $a=\pm 0.25$ phenotypic standard deviations, no dominance deviation was simulated.....	81
Figure 5.3 The effect of simulated QTL position in relation to marker loci positions on the mean test statistic obtained over 50 replicate simulations. The population size simulated was 350 animals, with marker spacing 20cM, and QTL effects of $a=\pm 0.25$ in males and females, and no dominance deviation was simulated.....	83
Figure 5.4 Comparison of the mean likelihood profile of X-linked and Autosomal data explaining equal phenotypic variance in the F ₂ . The QTL was simulated at 35cM in a population size of 350 animals. Results presented are the mean values over 50 replicate simulations.....	84
Figure 5.5 The effect of QTL action on the mean test statistic obtained from X-linked regression analysis. Comparison of additive, dominant and recessive QTL inheritance.....	87
Figure 5.6 Likelihood profiles obtained from analysing two simulated QTL using models fitting one QTL (Fig. a) and two QTL (Fig. b). The QTL were simulated at 30cM and 90cM both with additive effect of $a=\pm 0.5$ phenotypic standard deviations, on a 120cM chromosome with dominance deviation. The results presented are of one simulation.....	89

Figure 5.7 Illustration of the contribution of male and female F_2 individuals to the total power to detect X-linked QTL.....	94
Figure 5.8 Comparison of the mean likelihood profile of X-linked and Autosomal QTL. The QTL were simulated at 35cM in a population size of 350 animals explaining equal phenotypic variance in the F_2 population. The X-linked data set (a) had the QTL simulated with equal additive effect in males and females, and the data set (b) had the QTL simulated with equal variance explained by the QTL in the males and females. No dominance deviation was simulated.	96
CHAPTER 6.....	102
MAPPING QUANTITATIVE TRAIT LOCI FOR BODY WEIGHT ON THE X-CHROMOSOME USING THE X-LINKED MULTIPLE REGRESSION METHOD	102
Figure 6.1 The likelihood profiles for body weights at (a) 3 week and (b) 6 week, obtained using the X-linked multiple regression method (Chapter 5).....	118
Figure 6.1c The likelihood profiles for 10 week weights.	119
Figure 6.2 The distribution of the residual trait value (g) for 10 week weight in the three female, and two male marker classes of the marker DXMit50. (a) F_2 females, and (b) F_2 males.....	121
Figure 6.3 The likelihood profiles for the analysis of body weight gains between growth measurements taken at, (a) 3-6 weeks, (b) 3-10 weeks, and (c) 6-10 weeks, using the X-linked multiple regression method.	123
Figure 6.4 Analysis of body weight gain from 6-10 weeks using the model testing two QTL vs. one QTL.....	125
Figure 6.5 The distribution of residual carcass fat percentage (estimated from freeze dried body weight) in the marker classes of the marker DXMit50, (a) Females, and (b) Males.....	128
Figure 6.6 The Likelihood profile for estimated carcass fat content (estimated using freeze dried body weight in both males and females). This analysis did not reach the significance threshold of TS 9.7 (P<0.10).	129
CHAPTER 7.....	146
THE PROGRESSIVE BACKCROSS POPULATION.....	146
Figure 7.1 The mating scheme for the BC population, showing the progressive backcrossing of selected BC females to P6 inbred low line males.....	149
Figure 7.2 An example of the maintenance of a high line X-chromosome segment on the inbred low line background, showing a BC_n female mated to an inbred low line male, and some examples of the possible BC_{n+1} progeny	151
Figure 7.3 Illustration of the objective of the backcross experiment, to establish seven backcross lines with high line X-chromosome segments maintained by eight X-linked Mit markers. The bold lines represent regions of the X-chromosome inherited from the high selection line on an inbred low line background. The marker map positions shown were those estimated in the F_2X population (Chapter 6).	152
Figure 7.4 The marker associated QTL effects at 10 weeks in males and females in the backcross generations 3, 4 and 5. The male effects show the difference between the two hemizygous marker genotypes and the female effects show the difference between the homozygous low and heterozygous marker genotypes. The error bars represent the standard errors associated with the estimates.....	161
Figure 7.5 The estimated marginal posterior distribution of the QTL position. Analysis was carried out using all pedigree information by Gibbs sampling. The nearest flanking markers are DXMit50 and DXMit25 at 19.8cM and 33.7cM respectively.....	165
Figure 7.6 The estimated marginal posterior distributions for the QTL effects associated with 10 week weight. The estimated effect in males represents the difference between the two hemizygous genotypes (2a), and in females the effect shows the difference between the homozygous low genotype and heterozygous genotype (a+d).	166
CHAPTER 8.....	174
GENERAL DISCUSSION	174
Figure 8.1 Illustration of the QTL positions and 95% CI estimated for 6 and 10 week weight data in the F_2X population and progressive BC population, (a) progressive BC population 6 week weight, (b) progressive BC population 10 week weight, (c) F_2X population 6 week weight, and (d) F_2X population 10 week weight	177

Acknowledgements

I thank Peter Keightley and Bill Hill for supervising my project. Peter, thanks for your help and guidance in finding the direction the research should take and for reading through numerous drafts. Bill, thank you for your help and concise comments during the course of my studies and particularly during writing up, where your input has been both invaluable and encouraging. Thanks also to my third supervisor, Alan Archibald of the Roslin Institute for his help and good humour (and for being a great dancing partner in Guelph, 1994). I would also like to thank Sara Knott and Simon Heath for several useful discussions on some of the aspects of the data analysis. I acknowledge Simon Heath specifically for analysing the data described in Chapter 7 by Gibbs sampling, and for his inexhaustible help and encouragement during the course of my studies. Many thanks to Linda May, Fiona Oliver, Pat Pignatelli and Paul Rooney for their help in the lab and the mouse house.

On a more personal note I would like to thank all my friends and family who have supported and motivated me over the last three years. All my friends in the department, Hon for his help and encouragement, Philippe and Lutz for being entertaining office mates of the last year, but most of all the other students who have made life more colourful and helped me into financial ruin! Thanks to Simon H, Said, Eimear, Simon K, Jens, Ricardo, Angus, Andrew, Tony, Mimi, Dominic E, Jennie, Magnus, Grant, Huw, and everyone else!! A special thanks must go to Vicki, Tim, Dominic N, Fiona, Duncan, Emy and especially Mary, who have continued to support, reassure, and most of all distract me when everything was not going to plan. I thank my parents and brother from the bottom of my heart for their love and never ending encouragement. Finally I would like to thank Edward for his proof reading, for always being the optimist, but most of all, for seeing the best in me while I have been at my worst.

Publications

K. A. Rance, I. M. Hastings, W. G. Hill, and P. D. Keightley (1994). Mapping of putative QTL influencing body weight on the X-chromosome in Mice. *Proc. 5th. World Cong. Genet. Appl. Livest. Prod. Guelph*: Vol 21 pp 268-271

J. B. Alfred, K. A. Rance, B. A. Taylor, S. J. Phillips, C. M. Abbott, and I. J. Jackson (1996). Mapping in the region of Danforth's short tail and the isolation of tail length modifiers. *Genome Research* (submitted)

Abstract

Evidence of a large sex linked additive effect accounting for 25% (approximately 5g) of the divergence in weight between mouse lines selected for body weight has been described previously. A study was undertaken to map putative X-linked quantitative trait loci (QTL), with the objective of isolating the position(s) and number of QTL, and thereby provide information regarding the genetic mechanisms controlling growth and body weight.

A preliminary experiment was carried out to map the QTL on the X-chromosome. An F₂ generation was genotyped at 7 microsatellite loci found to be polymorphic in the selection lines. The results of single marker analysis showed significant linkage between X-linked markers and body weight, but DNA was not available for all individuals in the pedigree, so fully informative loci in the F₂ generation could not be determined. To enable analysis of fully informative marker loci, the F₂ experiment was repeated retaining DNA samples from all individuals within the pedigree. The F₂ population was generated from a reciprocal F₁ population between an inbred low line (derived from the low selection line) and the high selection line.

To enable statistical analysis of an F₂ population genotyped at markers on the X-chromosome, an analytical technique was developed based on the multiple regression method of Haley and Knott. The validity of the analysis was tested using simulated data, and the method was developed further to enable analysis of non-inbred data.

Using parental genotypes to determine fully informative marker loci in the F₂ generation, the phenotypic and genotypic data were analysed using the X-linked multiple regression method for non-inbred data. The analysis of data on 10 week weight indicated a single QTL of large effect ($\pm$ s.e.) situated at about 23 ± 2 cM from the proximal end of the chromosome, the likelihood curve showing a single well defined peak. The estimates for additive genotypic effects in males and females (half

the difference between the homozygote females and hemizygote males) were both approximately 2.6g (i.e. 17-20% of the total body weight in males and females at 10 weeks). Dominance effects in the females were found not to be significant in the analysis. To evaluate the contribution of X-linked QTL affecting carcass fat percentage, estimates of this trait were analysed, but no significant QTL were detected.

In addition to the F₂ population, a progressive backcross experiment was established with the objective of backcrossing regions of the high selection line X-chromosome onto the inbred low line background. This experimental design allows specific regions of the X-chromosome to be examined on an inbred genetic background, thus removing the contribution of QTL on autosomes and other regions of the X-chromosome and enabling the QTL position(s) and estimates to be resolved with greater accuracy. Data were collected over 5 generations of backcrossing. Estimation of the effects associated with the backcrossed X-chromosome regions in generations 3-5 support the results of the F₂ population analysed by X-linked multiple regression analysis, indicating the QTL is associated with the same X-linked markers. The estimates of the total magnitude of the QTL effect at 10 weeks in the backcross population were approximately 4g in both sexes (approximately 18 to 19% of the total body weight). Analysis of the backcross population using Gibbs sampling provided convincing evidence that the X-linked QTL was situated in the marker interval from approximately 20cM to 34cM, with the estimated position ($\pm$ s.e.) of 25.4 ± 2.8 cM, and no evidence for QTL elsewhere on the chromosome.

Analysis of the F₂ population by X-linked multiple regression analysis and the progressive backcross population by Gibbs sampling enable us to conclude the presence of a major QTL for body weight positioned at 26.4 ± 1.2 cM on the X-chromosome, which increases body weight by approximately 18% at 10 weeks. The estimated effect of the major X-linked QTL appears to explain nearly all the X-linked effect.

CHAPTER 1

General Introduction

Most economic traits in animal production are of a quantitative nature. Selection for these traits, bringing about animal improvement, has been performed effectively for many years with little knowledge of individual gene actions and interactions which contribute to the phenotypic variation. The approach utilising quantitative genetic variation has been effective as a result of the relative insensitivity of the methodology used to failures in simple models, such as the infinitesimal model of many loci of small effect. However, identification of individual quantitative trait loci (QTL) will provide the potential for improvements in the methodologies used in animal breeding.

The discovery of an abundance of molecular genetic markers, for example restriction fragment length polymorphisms and simple sequence length repeats or microsatellites, has provided the opportunity to resolve quantitative genetic variation into individual loci. The basic methodology for detection of QTL is the identification of statistical associations between phenotypic trait values and marker genotypes in a segregating population, such as an F_2 or backcross population derived from two inbred lines or strains differing substantially for the trait under investigation. Mapping QTL will allow new understanding of the number and effects of genes influencing complex traits. It will eventually provide fundamental knowledge of individual gene actions and interactions of QTL, and this will allow a more realistic model of phenotypic variation to be established, resulting in the development of more effective methods for prediction of breeding values and selection procedures. Identification of QTL linked to markers may provide the possibility of utilising marker assisted selection (MAS), where selection can be carried out directly on QTL (via linkage to markers) to bring about breed improvement. MAS will be of particular importance in traits that are

expensive to measure, expressed late in development, or not expressed in the sex under selection (e.g. milk traits in dairy cattle). In addition, it may be possible to use markers to introgress genes of economic importance from one breed to another (e.g. fecundity genes from the Chinese Meishan to Large White or Landrace pigs). The ultimate goal for QTL mapping is the possibility of positional cloning of genes, which will allow the molecular causes of genetic variation to be investigated and may ultimately enable improved alleles to be produced by direct molecular intervention.

Unfortunately many of these techniques are not economically viable in many commercial species, as a result of high costs to establish and maintain experimental populations. However, laboratory species, particularly mice, provide excellent models to investigate the genetic factors influencing quantitative variation.

As outlined above, mapping QTL has greater significance than simply to increase the basic knowledge of number, location and effects of genes associated with quantitative variation. In addition to the benefits to animal breeders there may also be substantial benefits to genetic research in humans. Investigation of quantitative traits such as obesity, growth and reproductive performance is difficult in humans, so the use of well-defined animal models and comparative gene mapping may indicate possible chromosomal regions of human disease loci in the human genome.

Evidence of a large additive X-linked effect which accounted for approximately 25% of the total divergence has been observed in lines of mice lines divergently selected on body weight. The purpose of this study was to map QTL influencing body weight on the X-chromosome in these mouse lines. A review of literature will be given to show an overview of the methodology available for QTL mapping, a brief description of the mouse lines used in the study, and finally a short review of the X-chromosome. Following this, the three experiments carried out to map X-linked QTL will be described; a preliminary F_2 population, followed by an F_2 population set up with the objective of mapping X-linked QTL, and finally a progressive backcross population. In addition to describing the experiments, analytical procedures adapted for the

analysis of X-linked data will also be shown. Finally a general discussion of the findings will be given.

CHAPTER 2

Literature Review

2.1 INTRODUCTION

The chapter is divided into three major sections: a review of methods involved in mapping of quantitative trait loci, the selection lines used in this study, and finally a brief review of literature on the X-chromosome.

2.2 METHODS OF MAPPING QUANTITATIVE TRAIT LOCI

2.2.1 Introduction

The conflict between the Mendelian theory of qualitative inheritance and the observation that many traits exhibit continuous variation was eventually resolved with the concept of quantitative inheritance; where a quantitative trait is controlled by multiple genes of small effect. Pioneering experiments by Sax (1923) showed that linkage between markers and quantitative trait loci (QTL) could be detected. This and further experiments (Thoday, 1961; Shrimpton & Robertson, 1988a,b) were hampered by a poor distribution through the genome of the morphological markers used to detect linkage to QTL. Today, with a near inexhaustible number of highly polymorphic markers throughout the genome, such as simple sequence length polymorphisms (SSLP or microsatellites) and restriction fragment length polymorphisms (RFLP), it is now possible to map QTL to specific regions of the genome (e.g. Paterson *et al.*, 1988; Andersson *et al.*, 1994; Cheverud *et al.*, 1996).

The objective of this section is to review the literature regarding mapping of QTL in experimental populations, covering statistical methods for mapping QTL, data interpretation, and experimental design.

2.2.2 Methods for Mapping QTL

The basic starting point for mapping QTL in experimental populations is a cross between two inbred lines. Crosses between genetically divergent inbred lines provide the most powerful means of detecting QTL by linkage (Knott & Haley, 1992). The assumption is made that QTL and markers are fixed at different alleles in the parental

lines, so that all F_1 individuals are heterozygous at all marker loci and QTL.

Segregating populations are scored for both the trait(s) of interest and for a number of genetic markers. Linkage between QTL and markers is observed in segregating populations, typically in an F_2 intercross ($F_1 \times F_1$) or backcross (BC_1) ($F_1 \times \text{Parental}$). Crosses between inbred lines provide a bench mark for assessing methods for investigation of power to detect QTL and accuracy of QTL estimates. In this section the basic methods for mapping QTL using single marker analysis and interval mapping by Maximum Likelihood and Multiple Regression will be discussed.

The traditional method for detecting QTL, marker linkage, was simply to compare the phenotypic means for different marker classes (Sax, 1923), assuming the QTL is in complete linkage with the marker locus. The phenotypic effect of the different marker-QTL classes can be inferred, and tested for significance (significantly different from zero) by a simple t-test. Although this method is extremely simple, it has a number of problems (from Lander & Botstein, 1989):

1. If the QTL does not occur in complete linkage with the marker locus its phenotypic effect may be underestimated by a factor which depends on the recombination fraction. If the recombination fraction between marker and QTL is r , then the inferred phenotypic effect of the QTL will be biased downwards by a factor of $(1-2r)$ in a BC population and $(1-3r/2)$ in an F_2 population.
2. If the marker and QTL are not in complete linkage, the variance explained by the marker will decrease by a factor of $(1-2r)^2$ and $(1-3r/2)^2$ in a BC population and F_2 population respectively. The number of progeny required to detect a QTL of given effect will increase by the reciprocal of the variance explained.
3. The single marker analysis does not give the likely position of a QTL. The analysis is unable to distinguish between tight linkage to a small effect, or loose linkage to a large effect.

Lander & Botstein (1989) suggested the shortcomings of single marker analysis result from the fact that marker loci were analysed one at a time. To remedy the difficulties

in these methodologies, they proposed the use of ‘marker intervals’ to estimate QTL positions and effects.

Following the initial development of ‘interval mapping’ using Maximum Likelihood (ML) by Lander & Botstein (1989), a number of additional methods have been proposed. The majority of these methods fall broadly into ML methodologies (Lander & Botstein, 1989; Luo & Kearsey, 1992), and Multiple Regression methods (Haley & Knott, 1992; Martinez & Curnow, 1992; Moreno-Gonzalez, 1993; Whittaker *et al.*, 1996). Both ML and regression methods assume normal distribution of data (when hypothesis testing), and no segregation distortion.

As mentioned previously, interval mapping was initially proposed by Lander & Botstein (1989) for a BC population, and was adapted for an F₂ population by Luo & Kearsey (1992). In interval mapping, instead of considering a single marker-QTL association, a chromosome interval is considered, flanked by two markers with a putative QTL within the segment:

$$\begin{array}{c} A-----Q-----B \\ | \leftarrow \quad r \quad \rightarrow | \end{array}$$

where A and B are the flanking markers, Q is the QTL, and the recombination distance between A and B is r and known. Under the null hypothesis of no QTL present within the interval, maximum likelihood estimations (MLEs) for QTL effects can be obtained at given positions within the interval. Using the estimates of QTL effects at putative QTL positions, a likelihood ratio test statistic can be obtained for the QTL at a given position within the interval. The estimate for QTL position is obtained by the highest likelihood ratio test statistic.

Lander & Botstein (1989) utilise LOD scores to interpret the likelihood ratio test statistic (the test statistic commonly used in human genetics),

$$\text{LOD} = \log_{10}(L_1/L_0),$$

where L_1 is the overall ML where QTL effects are estimated in the model, and L_0 is the overall ML where the QTL effects estimated are restricted to zero.

The maximum likelihood methods using flanking markers have been shown by simulation to provide good estimates of parameters, and allow detection of QTL in segregating populations derived from inbred lines (Lander & Botstein, 1989; Luo & Kearsey, 1992; Knapp, 1991; Paterson *et al.*, 1991). Because iterative numerical methods are required to maximise the likelihoods of estimates, ML methods for QTL mapping are computationally demanding and become increasingly intractable as the model becomes more complex. As a result of this, methods utilising regression have been developed (Haley & Knott, 1992; Martinez & Curnow, 1992; Moreno-Gonzalez, 1993). The basic methodology as described by Haley & Knott (1992) utilises regression of phenotypic values onto the mean (m) and coefficients of additive (a) and dominance (d) terms. The coefficients of a and d are expressed in terms of recombination fractions between flanking markers at putative QTL positions (e.g. 1cM intervals between flanking markers). Numerical values of the coefficients of a and d at each of the putative QTL positions can be calculated for every marker genotype combination. Multiple regression is carried out at all putative QTL positions within the marker interval, regressing the phenotypic value onto the mean (mean of the homozygote parents) and the numerical values of the coefficients of a and d (Haley & Knott, 1992). The most likely position of the QTL is obtained at the lowest value of the residual sum of squares (RSS) (Martinez & Curnow, 1992; Haley & Knott, 1992), and estimates of a and d can be obtained at this position.

This method can be compared to the ML method by using a test statistic comparable to the likelihood ratio test statistic:

$$\text{test statistic} = n \log_e (\text{RSS}_{\text{reduced}} / \text{RSS}_{\text{full}}),$$

where n is the number of individuals in the analysis, $\text{RSS}_{\text{reduced}}$ is the RSS of the regression with no QTL parameters fitted, and RSS_{full} is the RSS with the QTL parameters fitted. This test statistic is asymptotically distributed as χ^2 with degrees of freedom (d.f.) equal to d.f. full model - d.f. reduced model (e.g. for an F_2 population 2d.f. for a and d). The likelihood ratio test statistic can be converted to LOD scores by division by $2(\log_e 10)$.

The multiple regression methods of Haley & Knott (1992) and Martinez & Curnow (1992) were simplified by Whittaker *et al.* (1996) who showed that in F_2 and backcross populations, with a QTL with additive effect, the regression methods described above are exactly equivalent to regression of phenotype on marker type. The method of Whittaker *et al.* (1996) does not use an iterative numerical optimisation for QTL position and effects, as in the methods of Haley & Knott (1992) and Martinez & Curnow (1992) where a number of putative points are screened in sequence to determine the most likely QTL position. The method regresses phenotype on each pair of flanking markers in turn, selecting from the pairs of markers (in which both markers have regression coefficients of the same sign) the pair which give the smallest RSS. These markers provide the most likely interval for the QTL position. The QTL position and effect in an F_2 generation are determined by solving the equation,

$$\alpha^2 = [\beta_j + (1-2r) \beta_{j+1}] [\beta_{j+1} + (1-2r) \beta_j] / (1-2r),$$

where β_j and β_{j+1} are the coefficients of the linear regression of phenotype on the flanking markers j and $j+1$, and r is the recombination fraction between the flanking

markers. The QTL position depends on the ratio of β_j and β_{j+1} , both of which must have the same sign as α .

The ML and multiple regression methods outlined above have been shown (in simulation studies) to be effective in mapping QTL in segregating populations. The multiple regression methods (Haley & Knott, 1992; Martinez & Curnow, 1992; Moreno-Gonzalez, 1993) using marker intervals have provided a valuable alternative to the ML methods (Lander & Botstein, 1989; Luo & Kearsey, 1992). The regression methods are easier to implement using existing analytical packages to estimate QTL positions and effects (e.g. Genstat), which makes these methods more accessible for general use. The method of Whittaker *et al.* (1996) provides a quick method of determining marker intervals of interest. This method may have a limited application for the analysis of experimental data as it does not provide a likelihood surface for all areas searched, and the computation time required for a genome wide scan using the method of Haley & Knott (1992) is not problematic. The method of Whittaker *et al.* (1996) does, however, provide a quick and effective alternative to the ML and multiple regression methods, and this may be of particular use in simulation studies.

Models assuming greater than one QTL on a chromosome

Work carried out to compare ML and regression methods has shown that both methods yield similar results when mapping a single QTL. The correlation of test statistics between ML and regression methods never fell below 0.96 in the simulation study described by Haley & Knott (1992). This implies that nearly all the useful information about the QTL is contained in the marker class means. However, there are still problems with ML interval mapping and regression methods, most notably the inability to distinguish between one or more QTL, particularly between one QTL within an interval and two QTL in flanking intervals. Investigations have been carried out into both ML and regression methods under QTL models assuming two or more QTL (Lander & Botstein, 1989; Haley & Knott, 1992; Martinez & Curnow, 1992; Knapp, 1991; Zeng, 1994; Jansen, 1994).

Lander & Botstein (1989) suggested that when a likelihood curve shows evidence of two QTL, each of the QTL should, in turn, be fixed at their estimated positions. The likelihoods are then plotted against chromosome position for each of the fixed QTL, and a second high peak indicates the presence of a second QTL. However, Martinez & Curnow (1992) suggest this procedure is incorrect, as the estimated positions and sizes of effects are then not independent. In the methods of Haley & Knott (1992) and Martinez & Curnow (1992), two QTL models can be fitted. These methods allow a search over all possible pairs of values of the two positions and two sizes of effects. Both authors showed that two QTL could be distinguished from one and no QTL, in cases where the QTL were simulated in marker intervals separated by one marker interval with no QTL, and where there was sufficient recombination to break down linkage between the two QTL. Martinez & Curnow (1992) proposed the optimum procedure was always to search for two QTL, as in simulation studies; even when there are none or one QTL present, the procedure still gives good results. This strategy would prevent spurious identification of a single 'ghost' QTL, where two QTL are separated by a single interval. When testing for a single QTL the model produces a 'ghost' QTL in the interval flanked by the QTL in adjacent intervals (Martinez & Curnow, 1992). However, it should be noted that when using this strategy an increased threshold for QTL detection would need to be applied, as the number of degrees of freedom for the test statistic would be doubled for the test of two QTL vs. no QTL (see section 2.2.3).

The regression methods outlined above are effective in distinguishing between two or one QTL, where recombination is sufficient to break down the association between the two QTL (Haley & Knott, 1992; Martinez & Curnow, 1992). However, the computation involved with these models is not feasible when the number of QTL is large. To specifically address the difficulties encountered when mapping more than one QTL, methods have been developed to allow efficient mapping of multiple QTL (Jansen, 1993, 1996; Zeng, 1994). Both authors use information from markers other than the flanking markers of the interval being mapped, fitting the additional markers as 'cofactors' in the analysis to account for the effects of possible QTL elsewhere on

the genome. These markers eliminate background genetic noise (QTL on other chromosomes), and neutralise the effects of linked QTL, resulting in an increase in power and reduction in interference due to linked QTL.

The cofactors (additional markers) are effective in removing residual genetic variance only when linked to QTL. If we presume that only few of these markers are linked to QTL, it follows that many of the cofactors will tend to be redundant. Fitting redundant cofactors in the model can reduce the power of the test used (e.g. likelihood ratio test statistic for one QTL vs. no QTL), and increase sampling variance of estimates, particularly when the sample size is small (Zeng, 1994). Jansen (1993) addressed the problem of fitting redundant cofactors by selection of cofactors. To allow for change in likelihoods of the models with different parameter numbers, Jansen (1993) proposed a genetic model that leads to the highest log-likelihood (L), minus a penalty for the number of parameters fitted in the model (k) using

$$\text{Akaike's Information Criterion (AIC)} = -2(L-k).$$

A difference between AICs for two models greater than 2 is considered significant (Sakamoto *et al.*, 1986; cited in Jansen, 1993). Further work investigating these methods indicated that type I and type II errors were lower than for interval mapping using the likelihood ratio test statistic under the null hypothesis that no QTL is present in the marker interval (Jansen, 1994).

In a method further developed by Jansen (1996) a multiple QTL model (MQM) was proposed, utilising Monte Carlo expectation maximisation (EM) using the Gibbs sampler to generate Monte Carlo samples. This method also uses cofactors, but minimises the loss of degrees of freedom by cofactor selection, as outlined above, and also increases power due to the reduction in unexplained variance and interference due to other QTL. In addition to these advantages, the use of the Gibbs sampler allows estimation of the 'real' genotypes, and corrects for missing data, both resulting in an increase in power (Jansen, 1996).

The methods described in section 2.2.2 give an overview of the methods available for mapping QTL. When mapping QTL, workers must be aware of the methods available, and the relative attractiveness of the methods for the study of interest. As discussed earlier, the method described by Whittaker *et al.* (1996) is of particular use for simulation studies, but may not be the most appropriate method for experimental data. The multiple regression methods (Haley & Knott, 1992; Martinez & Curnow, 1992) are simple to implement, enabling more complex models to be fitted, and can be used in conjunction with cofactors to remove genetic noise due to QTL elsewhere on the genome. The regression methods are therefore easier to use than the ML methods of Lander & Botstein (1989) and Luo & Kearsey (1992), and computationally less demanding. Where QTL effects are small, simulation studies have indicated the MQM methods described by Jansen (1996) using Monte Carlo EM with cofactors may be the most effective, as this method has been shown to provide an increase in power.

2.2.3 Interpretation of mapping results; Thresholds and Confidence Intervals.

As discussed in the previous section, methods for interval mapping of QTL can be broadly divided into ML and regression methods. Extension of the one QTL model allows effective mapping of two QTL (Haley & Knott, 1992; Martinez & Curnow, 1992), and multiple QTL using cofactors (Jansen, 1993,1996; Jansen & Stam, 1994; Zeng, 1994; Jiang & Zeng, 1995). A problem common to all these methods is the difficulty in determining appropriate significance thresholds for test statistics (usually LOD scores or likelihood ratio test statistics) and confidence intervals for QTL positions, so that results from QTL mapping can be interpreted correctly. This section will discuss interpretation of results, outlining calculation of thresholds for QTL detection and confidence intervals (CI) for QTL positions.

The appropriate level at which to set the significance threshold when testing for the presence of QTL in many intervals is problematic. The objective is to produce a

threshold minimising type I errors where QTL are indicated but none are present, for an acceptable level of type II errors where QTL are present but are not detected. Lander & Botstein (1989) proposed the threshold for QTL detection was dependent on the size of the genome and marker density. Further development of the formula to calculate the significance threshold for QTL detection was carried out by Lander & Kruglyak (1995). They consider a normally distributed statistic, the Z-score, where the number of regions in which the statistic Z exceeds the pointwise significance level (T) has a Poisson distribution with mean:

$$\mu(T) = [C+2\rho GT^2] \alpha(T),$$

and the genome wide significance level is $\alpha(T) = 1 - e^{-\mu(T)}$. The quantities in the equation are defined:

$\alpha(T)$ = the pointwise significance threshold T ,

C = the number of chromosomes,

G = the genome length in Morgans,

ρ = the constant measuring the total crossing over rate between the genotypes being compared, indicating how rapidly the linkage statistic exceeds a specified threshold T by chance over the genome wide scan (Lander & Kruglyak, 1995). The results can be applied to LOD scores by replacing $[C+2\rho GT^2]$ with $[C+2\rho GX]$ in the equation above, where $X=(2\log_e 10)$. The pointwise significance level is determined from χ^2 tables with d.f. equal to the d.f. for the model with the QTL fitted (d.f. full), minus the d.f. of the model not fitting the QTL (d.f. reduced) (e.g. for an F_2 population, 2 d.f.). Lander & Kruglyak (1995) used this formula to obtain guidelines for thresholds for QTL mapping over a range of different experimental designs, and gave LOD thresholds for suggestive linkage ($P<0.10$) and significant linkage ($P<0.05$).

Other authors also suggest using χ^2 tables, with d.f. equal to (d.f. full - d.f. reduced) + 1 d.f. for the estimation of QTL position (Jiang & Zeng, 1995; Martinez & Curnow, 1992), e.g. for an F_2 population, 2+1 degrees of freedom. Jiang & Zeng (1995) stresses however, that this threshold was only a very rough approximation and the correct determination of the appropriate threshold was a complex statistical issue.

Jansen (1996) pointed out that the use of the χ^2 test statistic was valid only for single positions or intervals. Andersson *et al.* (1994) suggested that as a result of QTL mapping by ML (Lander & Botstein, 1989; Luo & Kearsey, 1992) and multiple regression methods (Haley & Knott, 1992; Martinez & Curnow, 1992) where sequential methods are employed to carry out multiple tests, the correct threshold cannot be easily derived theoretically. These authors proposed a test using actual marker data where 5000 replicates of phenotypic measurements are simulated with no QTL segregating. From these replicates the highest test statistics obtained were used to produce 5%, 1% and 0.1% probability of detecting QTL where no QTL was segregating. A similar method suggested by Churchill & Doerge (1994) uses a permutation test where the phenotypic data are permuted with respect to marker data a large number of times to sample from the distribution of test statistics under the null hypothesis of no phenotype-genotype associations. They suggest that this method is valid when used in conjunction with ML or regression based test statistics for any distribution of quantitative data.

The thresholds for QTL detection are very important in the prevention of type I and type II errors. The χ^2 test statistic should be used only as a preliminary estimate or a threshold or for use in a simulation study. The permutation test of Churchill & Doerge (1994) and the simulation method described by Andersson *et al.* (1994) appear to be the most appropriate methods for the analysis of experimental data. The only disadvantage of these methods are if the marker data is not available for all chromosomes, then it is not possible to consider a threshold for a genome wide scan. Lander & Schork (1994) stress that the threshold for a genome wide scan should always be applied to any new hypothesis, even when the search is over a small subset of the genome. This may appear to be a very restrictive statement, but if results are reported very clearly reflecting the thresholds used, the Lander & Schork (1994) statement need not apply, and thresholds for a single chromosome can still be used. An example of the reporting of a QTL detected with a significance threshold calculated for a single chromosome, is that the author should stress that the QTL is

significant in the analysis of the given chromosome, and no claims regarding significance for the entire genome can be made.

Another important issue in the correct interpretation of results from QTL mapping is the formulation of a confidence interval for the QTL position. Lander & Botstein (1989) proposed a one-LOD drop support interval (also referred to as the LOD-drop off method), where the CI for the QTL position is indicated by a drop of 1 LOD score from the maximum, which corresponds to a 95% CI. The use of $4 \times$ standard error (s.e.) of the estimated position of the QTL as a confidence interval has also been suggested (Darvasi *et al.*, 1993) and would represent approximately the length of a 95% CI. Comparison of $4 \times$ s.e. with empirical CI obtained by simulation showed the CIs obtained by both methods were generally quite close (Darvasi *et al.*, 1993). In addition to the one-LOD drop and $4 \times$ s.e. approximations of confidence intervals, bootstrapping has also been used to estimate CIs (Visscher *et al.*, 1996). In brief, the bootstrapping method discussed by Visscher *et al.* (1996) samples from the data set (phenotypic and genotypic records) with replacement for n bootstrap samples. After n bootstrap samples, the results of QTL analysis of the bootstrap samples are ordered, and the bottom and top 5 and 2.5 percentiles are taken to produce 90% and 95% CI. Comparison of bootstrapping and LOD-drop off methods to the empirical CI by Visscher *et al.* (1996) showed the CI obtained from bootstrapping were more stringent than estimated by LOD-drop off (which were biased downwards leading to CI which did not always contain the QTL). The bootstrap method is more stringent because it uses information from the likelihood curve for the entire chromosome to estimate the CI for the QTL position, whereas the LOD-drop off method simply uses information from the highest peak, often ignoring information from secondary peaks. It should be noted that in the study by Visscher *et al.* (1996) the LOD-drop off used was equivalent to a drop in LOD score of 0.83, as this was equivalent to $P < 0.05$ for a χ^2 with 1 degree of freedom. It is possible that if the one-LOD drop was used, the CI estimated by this method may have been closer to the empirical CIs. The downward biases of the CI determined by the LOD-drop off method were greater in populations where a QTL explains a small amount of the variation and for dense marker maps

(Mangin *et al.*, 1994). van Ooijen (1992) and Visscher *et al.* (1996) showed that type I errors occurring with the LOD-drop off method were greatly reduced with increased population size and heritability.

As described above, the LOD drop method and bootstrapping methods have been shown by simulation to produce good approximations to the empirical 95% CI. When QTL mapping power is high (i.e. large population and/or QTL of large effect) the LOD-drop off method provides a good approximation to the empirical CI, and therefore this method would be effective when a single well defined peak in the likelihood curve with a high maximum test statistic is obtained from the QTL analysis. Estimating the CI using the LOD-drop off method is very simple, and avoids the computation time required to estimate CI by bootstrapping. However, when a multimodal likelihood curve is obtained from the QTL analysis, the bootstrapping method would be the most appropriate method to use, as the LOD-drop off method is more likely to result in type I errors. It should be noted that whatever method is used, the workers must be aware of the potential inaccuracy of the estimated CI, as neither methods give completely accurate 95% CI.

2.2.4 Experimental Power

The use of interval mapping rather than single marker analysis with ML has been shown to increase the power of QTL mapping experiments, e.g. Backcross or F_2 intercross designs (Lander & Botstein, 1989; Knott & Haley, 1992), particularly at wide marker spacing (Knott & Haley, 1992). Comparisons of ML vs. regression methods have been shown to produce similar power of detection of QTL (Haley & Knott, 1992). In addition to the relative power of statistical methods used to map QTL, the power of QTL mapping can be affected by the experimental design, particularly segregating population type, population size, marker spacing, and the size of QTL effect.

The power to detect marker-QTL linkage is primarily determined by the QTL effect (Lander & Botstein, 1989; Darvasi *et al.*, 1993), as the QTL effect cannot be changed within a population under study. By varying the experimental design the power to map a QTL of a particular size may be increased, but ultimately the size of the QTL effect is the first factor which limits the power of a QTL mapping study. The effect of experimental design on the power to detect marker-QTL linkage will be discussed in this section.

The relative power of QTL detection in different experimental populations has been investigated (Beckmann & Soller, 1983; Lander & Botstein, 1989; Moreno-Gonzalez, 1993; Luo, 1993). Lander & Botstein (1989) suggested that the use of an F_2 inter-cross was about twice as powerful at detecting purely additive QTL as a BC population using interval mapping, as it provides information at twice as many meioses, resulting in 50-60% as many progeny required compared to the BC population. A simulation study conducted by van Ooijen (1992), however, showed the power to detect a QTL of entirely additive effect was slightly less for an F_2 than for a first generation backcross population (BC_1) (see Table 2.1 section 2.2.6). van Ooijen suggested that this was due to the fact that three mixing components have to be estimated in the F_2 population vs. two in a BC_1 . The author stressed that in the simulation study undertaken (simulating QTL explaining equal variance in the population), an additive QTL that explains a specified fraction of the total variance in the BC_1 will explain more than that fraction in the F_2 . The additive genetic variance of a QTL in an F_2 is twice that variance in a BC population, so van Ooijen (1992) suggested that this fact may have led to the conclusions of Lander & Botstein (1989) that an F_2 was twice as powerful for detection of additive QTL as a BC population. However, the two-fold genetic variance in the F_2 holds for QTL elsewhere on the genome as well. Therefore, the total variance due to all QTL also increase, while the non-genetic variance remains the same (van Ooijen, 1992). Simulation studies carried out by the same author showed the fraction of the variance explained by a single QTL in an F_2 will be less than twice the fraction in a BC_1 , reducing to even the same

fraction depending on the number of QTL segregating and the proportion of their genetic effects.

Simulation studies have indicated that BC_1 populations provide more power for QTL detection where there is no dominance. However, if we use the example of a backcross between the lines AA and BB carrying a QTL with partly (or entirely) dominant effect, one of the reciprocal backcrosses (e.g. AA×AB) will be more effective and the other backcross (e.g. BB×AB) less effective for mapping QTL compared to a BC population with a QTL of entirely additive effect. Mapping entirely recessive or dominant QTL in a BC_1 population may result in the QTL remaining undetected regardless of the size of the QTL effect or population size, (this will depend on the direction of the backcross investigated). Therefore for many experimental investigations F_2 populations are preferred. The F_2 populations enable both additive and dominance deviations to be estimated, whereas in the BC_1 it is not possible to estimate the dominance QTL parameter. If the QTL effect is known, e.g. estimated to be additive by a segregation analysis, a BC_1 population may be the preferred experimental design as it provides greater power for mapping QTL of a given size compared to an F_2 population.

Population size and marker density have a major effect on the power to detect QTL. Darvasi *et al.* (1993) showed that increasing population size increased power of QTL detection, and reduced the CI of the QTL position and the s.e. of estimates. It was noted that population size had a greater effect on the CI than marker spacing. Increasing the number of markers can reduce the confidence interval, but only to a given limit which is determined by the population size at a given gene effect (Darvasi *et al.*, 1993). These findings led to further work to establish optimum marker spacing. Darvasi & Soller (1994) suggest that over a range of experimental designs, wide marker spacings were optimum for initial studies of marker-QTL linkage (e.g. 45cM where population maintenance is 20 times the cost of genotyping at 1 marker). They suggest that if budget limits genotyping, power can be increased most efficiently by an increase in the population size.

Mapping QTL in a large population with wide marker spacings does provide greater flexibility in the experimental design. In a small population genotyped at a large number of markers, if linkage to a QTL is detected, there is no scope to improve the estimate of QTL position, resolve the CI, and increase the likelihood for the QTL by adding more markers to the map. These objectives can be only achieved by increasing the population size. Using a large population size in the initial experiment, as well as providing greater flexibility for the experiment design, will also enable a number of specific experimental designs to be implemented which can increase the efficiency of QTL mapping. Selective genotyping (Lander & Botstein, 1989; Darvasi & Soller, 1992), replicated progenies (Soller & Beckmann, 1990), sequential sampling (Motro & Soller, 1993), and pooled genotyping (Wang & Paterson, 1994) can all increase efficiency to detect QTL by reducing genotyping requirements and therefore leaving funds open to increase population size and therefore increase power.

Selective genotyping can decrease the number of individuals genotyped for a given power (Darvasi & Soller, 1992). The method utilises the fact that much of the information from QTL mapping comes from the extreme genotypes (Lander & Botstein, 1989; Darvasi & Soller, 1992). Using the example of a population of 500 animals with an additive QTL of ± 0.125 phenotypic standard deviations (S.D.), genotyping only half the animals (25% for each tail of the normal distribution), will result in a drop in power of detecting marker-QTL linkage from 0.80 to 0.77 (Darvasi & Soller, 1992).

The use of replicated progenies outlined by Soller & Beckmann (1990) is of particular interest for QTL mapping in plants. Replicated progenies are obtained for F_2 plants or F_3 families for selfed F_2 plants. The replicated progenies can be scored for phenotypic traits (reducing error variance for trait measurements and increasing power to detect QTL) and can also be scored for more than one phenotypic trait and/or over a range of environments. This will enable mapping of more than one trait, and allow genotype environmental interactions to be assessed. The F_2 individuals are the only individuals

to be genotyped at marker loci, therefore limiting the costs of genotyping for large number of phenotypic measurements.

Motro & Soller (1993) discussed the use of sequential sampling. Where individuals are genotyped in sub-groups of the population and using the obtained genotypes, marker-QTL linkage is assessed by simple comparison of means (e.g. for a cross between lines carrying marker genotypes AA and aa , comparison of AA and Aa in the backcross and AA and aa in the F_2). This method can substantially reduce the number of individuals genotyped to detect marker-QTL linkage. In combination with selective genotyping Motro & Soller (1993) suggest the number of markers scored may be reduced by eight fold to detect marker-QTL linkage, compared to the classical genotyping of fixed sample size from unselected offspring.

The use of pooled genotypes relies on the assumption that markers linked to QTL in a base population will be in linkage disequilibrium in a segregating population. If DNA pools of individuals in the extremes of the population are taken, it is assumed that markers in linkage disequilibrium with QTL will be at extremes in these pools, while unlinked markers will be randomly segregating (Wang & Paterson, 1994). Genotyping these pools gives a fast method of detecting marker-QTL linkage. This method is useful for initial scanning of the genome to detect QTL-marker linkage. Further mapping is required to obtain QTL parameter estimates, as no information regarding the size or action of the QTL effect is gained from mapping using pooled DNA.

As described above, the power of QTL mapping is affected by the population type, population size, marker spacing, and the size of QTL effect. Darvasi & Soller (1994) propose that, in both BC_1 and F_2 populations, wide marker spacings were optimum for initial studies of marker-QTL linkage, leaving funds (when limited) open to increase population size. A combination of the experimental designs described above could be used to maximise the power and reduce genotyping costs. In a large segregating population (BC_1 or F_2) derived from inbred lines, pooled genotyping could be used to detect QTL of large effect, and then followed by selective

genotyping used to isolate the QTL. To further resolve the QTL position, additional animals in the population could be genotyped. By repeating this procedure using the information at mapped QTL when producing the DNA pools (e.g. fitting marker genotypes for the markers closely linked to the mapped QTL as fixed effects), further QTL could be detected via pooled genotyping. It is clear that experimental design can have a substantial effect on the power and cost of mapping QTL, and should therefore be given due consideration when mapping QTL.

2.2.5 Non-inbred Populations.

In previous sections, mapping of QTL has been described utilising inbred lines where all marker loci are fully informative in a segregating population. While for many laboratory animals it is feasible to produce inbred populations of the lines of interest, unfortunately for commercial species this is not widely practicable. As a result, QTL mapping methods have been developed to map QTL in non-inbred populations. Haley *et al.* (1994) developed a method for mapping QTL in F_2 experimental populations derived from non-inbred lines (this method may also be adapted for other segregating populations). The method assumes that although there may be segregation at marker loci in the two lines, the QTL of interest are fixed at different alleles in the base population. The technique is based on the inbred multiple regression method (Haley & Knott, 1992). Using marker genotypes of parental lines, F_1 , and F_2 individuals the probabilities of the line origins of markers can be derived (where markers can be traced to either of the two foundation lines, weighted by the probability of each origin, see Haley *et al.* (1994)), some markers may be completely uninformative. Once the possible line origin combinations have been derived, the probabilities of each of the four line origin combinations for a QTL at a given position in an F_2 individual can be calculated conditional upon the possible line origin combinations of the marker and on the recombination fractions between the flanking markers, and the markers and QTL (Haley *et al.*, 1994). This method has been shown to be effective in mapping QTL in non-inbred populations (Andersson *et al.*, 1994),

and, like the method utilising inbred lines (Haley & Knott, 1992), it is relatively simple to apply and allows the exploration of more complex models to be undertaken (Haley *et al.*, 1994).

2.2.6 A quantitative assessment of QTL mapping

In previous sections the methodologies for mapping QTL in inbred populations and non-inbred experimental populations have been presented, and formulation of thresholds for QTL detection and confidence intervals for their positions have been outlined. The objective of this section is to give an indication of the size of QTL effects which can be detected and the CIs for their positions using the methodologies described previously.

Limits to the size of QTL effect which can be detected

van Ooijen (1992) compared the accuracy of mapping QTL in simulated data sets from a first generation backcross population (BC_1) and an F_2 population derived from two inbred lines, when data were analysed by ML interval mapping (Lander & Botstein, 1989; Luo & Kearsey, 1992), using the null hypothesis of no QTL in the marker interval. The simulations were generated with three population sizes and three sizes of QTL effects, using 500 replicate simulations. The significance thresholds for QTL detection were obtained for the BC_1 and F_2 populations by 16000 replicate simulations using the approximating equation of Lander & Botstein (1989).

Table 2.1 The fraction of simulations in which a significant LOD score was obtained (from van Ooijen (1992)).

N	σ^2_{exp}			F ₂		
	BC					
	0.01	0.05	0.10	0.01	0.05	0.10
100	0.01	0.11	0.41	0.02	0.06	0.31
200	0.02	0.41	0.87	0.02	0.29	0.79
400	0.07	0.84	1.00	0.05	0.76	1.00

N = population size, σ^2_{exp} = fraction of total background variance explained by the QTL.

The simulation results (Table 2.1) obtained by van Ooijen (1992) suggest that QTL of small additive effect (explaining approximately 1% of the total phenotypic variance) were very unlikely to be detected in populations of size 400 and less. The results suggest a population size of at least 200 individuals was required to map QTL, unless the QTL is very large, explaining greater than 10% of the total phenotypic variance. The van Ooijen (1992) simulation study indicates a population size of 400 individuals is required to reliably detect QTL explaining approximately 5% of the total phenotypic variance.

Experimental results suggest QTL explaining less than 5% of the total phenotypic variance can be mapped in similar population sizes to those outlined above (i.e. approximately 400 animals). In a study by Cheverud *et al.* (1996) using a population of 535 mice derived from two inbred lines which were genotyped at 75 microsatellite loci covering all the autosomes, QTL for body weight were detected on 16 of the 19 autosomes. Using a significance threshold calculated by the method previously described by Andersson *et al.* (1994) (see section 2.2.3), QTL were detected explaining $\geq 1.7\%$ ($P < 0.05$) of the total phenotypic variance. It should be noted that in the simulation study by van Ooijen (1992) and the experimental study by Cheverud *et*

al. (1996) two methods were used to obtain a significance threshold, and the results are therefore not strictly comparable.

Another experimental example of QTL mapping is the study described by Andersson *et al.* (1994) using non-inbred pig populations. When mapping QTL in a non-inbred population, the QTL detection rate will be a function of the level of uninformative marker loci within litters, as well as the population size, number of markers scored, and size of the QTL effect. The results of QTL mapping in an F_2 population derived from two non-inbred but genetically divergent pig lines, the European wild boar and the Large White, were presented by Andersson *et al.* (1994). Using the analysis method developed by Haley *et al.* (1994) (see section 2.2.5), QTL were mapped in 200 F_2 individuals genotyped at 105 genetic markers covering 75% of the porcine genome. Utilising the same simulation method to obtain a significance threshold for QTL detection as used by Cheverud *et al.* (1996), QTL were detected ($P < 0.05$) explaining 7.5% of the phenotypic variance of growth from birth to 30kg body weight. Other QTL were detected for the traits measured, but all explained a greater percentage of the total phenotypic variance.

Estimated confidence intervals for QTL positions

In addition to the detection of QTL, an important consideration for interpretation of the results of QTL mapping is the CI for the QTL position. As discussed in section 2.2.4, confidence intervals are dependant on population size, QTL effect, and marker density.

Work by van Ooijen (1992) investigated the influence of population size and QTL effect in F_2 and BC_1 populations when the marker density was 5cM. The results obtained showed that CI were smaller at a given simulated QTL effect as population size increased. Investigation of the influence of QTL size showed larger QTL effects produced smaller CI for the QTL positions. Estimation of the percentage of CI which contain the simulated QTL indicated the estimated CI were more accurate with larger population sizes and larger QTL effects. Comparison of the F_2 and BC_1 populations

showed the CI estimated in the F₂ populations were narrower than the contemporary BC₁ populations, this occurs as the F₂ populations provide information at twice as many meioses. In the small population sizes the reduced confidence interval in F₂ populations occurs at the expense of accuracy of the CI. The simulation results of van Ooijen (1992) indicate small CI are attainable only in large populations with QTL of large effect.

Table 2.2 The mean and standard deviation (shown in parenthesis) of the length of the confidence interval for QTL position. The LOD drop off of 1 represents a 95% CI and the LOD drop off of 2 represents the 99% CI. Estimates of the fraction of CI with contain the simulated QTL are also shown (from van Ooijen, 1992).

N	LOD drop off	σ^2_{exp}							
		BC				F ₂			
		0.05		0.10		0.05		0.10	
		CI	%QTL	CI	%QTL	CI	%QTL	CI	%QTL
200	1	17 (8)	78	15 (6)	90	16 (6)	73	14 (6)	85
	2	37 (15)	94	29 (13)	99	33 (13)	92	28 (12)	98
400	1	15 (7)	84	11 (4)	95	14 (6)	84	11 (4)	94
	2	31 (15)	96	18 (8)	99	27 (13)	96	19 (8)	99

N = population size, σ^2_{exp} = fraction of total variance explained by the QTL, CI = confidence interval, %QTL = percent of simulated QTL which fall within the estimated CI

The findings of van Ooijen (1992) were supported by the results of Darvasi *et al.* (1993) where the authors demonstrated by simulation that even for a QTL of large effect (≈ 0.5 phenotypic standard deviations, i.e. a QTL explaining approximately 12.5% of the total phenotypic variance) in a backcross population of 1000 animals scored at infinite marker loci, the empirical 95% confidence interval obtained was 11cM. The simulation results therefore suggest that using the simple F₂ or BC₁ experimental design interval mapping is not able to refine QTL position to less than

approximately 10cM. However, experimental data suggest that the CI for QTL of larger effect can be refined further.

The results obtained by Horvat & Medrano (1995) suggest that 99% CI (calculated by a drop in LOD score of 2) obtained in an experimental population can be less than 10cM. In the study by Horvat & Medrano (1995), the recessive *hg* gene which increases body weight by 30-50% was mapped in a population of 403 F₂ individuals, producing a CI of 3.7cM in the females and 5.4cM in the males using a sex averaged map density of approximately 1.7cM on chromosome 10. However, the small CI obtained in this study is likely to be a function of the large contribution of the *hg* gene to the total phenotypic variance.

In summary, the results of simulation studies suggest that QTL can be mapped with a significance threshold of $P < 0.05$ in large population sizes (e.g. 400 individuals) explaining approximately 5% of phenotypic variance (van Ooijen, 1992), and larger population sizes will enable QTL of smaller effect to be detected. When a QTL explaining approximately 12.5% of the total phenotypic variance is simulated in a population size of 1000 individuals genotyped at infinite marker interval mapping is unable to isolate the QTL to a CI of less than 10cM (Darvasi *et al.*, 1993).

Experimental data, however, do indicate that QTL of smaller effect can be mapped in large populations explaining as little of 1.7% of the total phenotypic variance (Cheverud *et al.*, 1996), and CI for QTL of very large effect (i.e. greater than 30% of total phenotypic variance) can be isolated to a chromosomal region of less than 5cM (Horvat & Medrano, 1995). To date analytical techniques and experimental designs commonly employed in mapping QTL have been unable to map QTL with great precision, producing narrow CI.

In this review, the analytical methods and experimental designs outlined enable preliminary mapping of QTL. For finer resolution mapping of QTL, special experimental designs must be employed. These designs largely rely on the production of congenic lines for the QTL of interest. An example of this approach is the method

of Shrimpton & Robertson (1988a,b) used to map bristle number in *Drosophila melanogaster*. This approach was based on the analysis of several congenic lines recombinant within small targeted regions. The Shrimpton & Robertson (1988a,b) experimental design is more powerful than simple F₂ or backcross populations, as the congenic lines are assumed to be genetically identical in regions unlinked to the congenic region of interest. This results in an elimination of genetic noise due to unlinked QTL and enables high resolution mapping of QTL.

2.3 SELECTION LINE HISTORY

2.3.1 Introduction.

Growth and body composition are of major economic importance in animal production. Improvement of these traits to meet market requirements are two of the most important problems that animal breeders face. In order to make improvements in these areas, it is helpful to have detailed knowledge of the genetic determination and interrelationships of these traits. Ideally these investigations would be carried out in the species of interest, but unfortunately in many agriculturally important species this is not always practical. Long generation intervals, high mature age, and high maintenance costs of populations, are the major reasons why investigations directly on commercial species are not carried out. However, laboratory animals, especially mice, provide excellent models to investigate genetic determination of growth and body composition and their interrelationships. Their short generation interval and low maintenance costs (compared to commercial species) enable a wide range of investigations to be carried out. The results of these investigations should be relevant to commercial mammalian species.

2.3.2 The Replicate Lines and Preliminary results.

Sharp *et al.* (1984) reported results on selection responses of mouse lines selected on growth, body composition and food intake. The objective of the study was to investigate mouse lines selected from the same base population but physiologically distinct in growth rate, food intake or fatness, and involved the establishment of a new strain, the 'G' strain. Two inbred lines, JU, and CBA, were crossed to form an F₁ population. The F₁ was crossed to an outbred strain, CFLP, which was obtained from Carnworth Laboratory in 1976 (Sharp *et al.*, 1984). One generation of random mating followed, and the next generation was designated as generation 0 of selection. Three

selection criteria were established, A-appetite, F-fat, and P-protein. There were replicates within each selection criterion to enable selection response to be distinguished from genetic drift. High, Low and Control lines were maintained in each replicate (Figure 2.1). This section will concentrate on the P-lines, as these lines were used in this study, the F-lines and A-lines will be reviewed in brief for comparison.

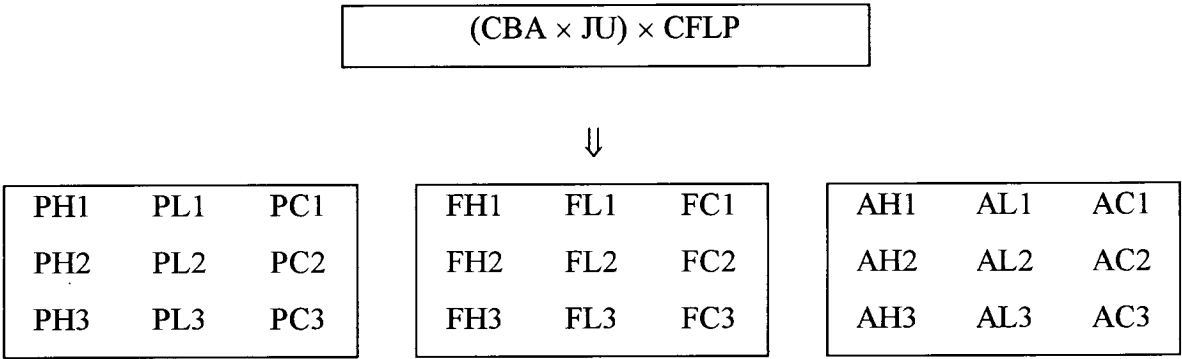


Figure 2.1 Lines established from the G-lines

The selection criteria were (from Sharp *et al.*, 1984):

- P - Protein lines: selected on body weight corrected for fat percentage.

$$\text{body weight (g)} - (8 \times \text{gonadal fat pad (g)})$$

(8 × as gonadal fat pads represent approximately one eighth of the total carcass fat in 10 week old males). This index is highly correlated to fat free mass (r=0.94). The males were mated at 8 weeks and then killed at 10 weeks and measurements taken.

Offspring were retained from selected males. Two males were measured in the control lines.

- A - Appetite lines: selected for individual 4-6 week food intake adjusted for 4 week weight.

- F - Fat lines: selected on the ratio of gonadal fat pad weight to body weight in 10 week males.

The growth curves at generation 14 presented by Bishop & Hill (1985) showed that at 4 weeks the P lines show much larger divergences in body weight compared to the A

and F-lines, and this divergence extends over the entire growth period. The selection response obtained for the three replicate P selection lines are shown in Figure 2.1. The mean body weight in the P-lines at generation 20 were 38.1g, 27.5g, and 23.1g for the high, control, and low lines respectively (Hastings & Hill, 1989).

The investigation of correlated responses at generation 20 of selection (Hastings & Hill, 1989) showed no significant change in carcass fat percentage in the A and P lines in males dissected at 10 weeks. Significant changes were found in the F line, with increasing ratio in the high line compared to decrease in the low line. Feed conversion efficiency (FCE) measured by Bishop & Hill (1985) showed no differences in FCE during the growth period between the A line high and low selection direction. The P high selection lines had higher efficiency than the P low and control lines. Observations on the F lines show greater efficiency in the fatter high line selection direction than the leaner low line.

2.3.3 Physiological Basis of Response.

At generation 20 of selection, replicates within the P and F selection lines were crossed to form the P6 and F6 lines (Figure 2.2). Crossing of the 3 replicate lines within high and low selection direction was intended to restore much of the genetic variability lost through random drift over the first 20 generations of selection. There was no line replication, and control lines were not established. At this stage the A-line was discontinued. This section will concentrate on the results of analysis of the P and P6 selection lines used in this study.



Figure 2.2 Representation of the formation of the P6 selection lines.

Analysis of data at generation 20 of selection indicated that the only selection criterion which led to a substantial change in body composition was that used in the F-line (Hastings & Hill, 1989). Selection for the ratio of gonadal fat pat to body weight produced a further divergent response in the carcass fat percentage, where the changes in body weight in these lines could be solely attributed to differing amount of carcass fat. Body composition analysis of the P lines showed that selection for fat free body weight resulted in no significant change in body composition.

The selection responses of the P line replicates and P6 lines are shown in Figure 2.3 (from Mbaga, 1996). The mean body weights at generation 50 of selection were 49.3g and 17.3g in the high and low lines respectively.

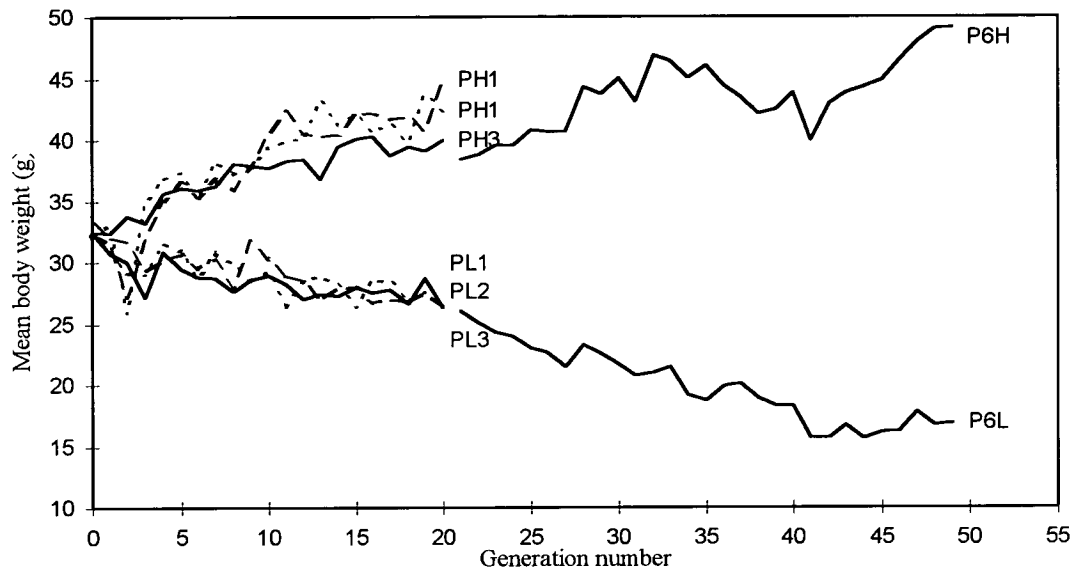


Figure 2.3 Selection response of the P replicate lines (selected for fat free body mass), and the P6 lines selected for body weight (Mbaga, 1996)

The relationship between water and fat proportion was also investigated by Hastings & Hill (1989). Lipids are strongly hydrophobic, therefore proportion of carcass fat can be estimated from carcass water content. Using results from carcass composition analysis and freeze dried carcasses weights, a linear regression was used to allow of carcass fat to be estimated from freeze dried carcass weights:

$$Y = a + bX,$$

where : Y = fat weight/wet weight

X = dry weight/ wet weight.

Least square estimates were $a = -0.302 \pm 0.009$ and $b = 1.13 \pm 0.003$ and the correlation between the estimated carcass fat and the fat content determined by carcass composition analysis was 0.98 (Hastings & Hill, 1989). In addition it was found that the effects of selection for estimated lean mass in the P-line could not be distinguished from selection for body mass.

As a result of these findings, selection criteria in the newly formed P6 and F6 lines were:

F6: fat proportion estimated using $Y = -0.302 + 1.13 X$

(see above for Y and X) on 14 week old males,

P6: Body weight of males and females at 10 weeks.

These new selection criteria in the P6 line were cheaper, easier and less time consuming to measure, and allowed selection to be practised in both males and females (Hastings & Hill, 1989).

Analysis of data to generation 20 in the P-lines by Beniwal *et al.* (1992a) showed an overall heritability of 0.52 ± 0.02 ; and this estimate was consistent across the replicates. Beniwal *et al.* (1992a) also investigated changes in genetic parameters during the course of selection. In order to investigate how variances in the selection lines change with time, the data were split into different periods of selection e.g. generation 0-7, 8-20 etc.. The results showed that in both the high and low lines there was a substantial and significant drop in genetic variances in later generations (Table 2.3).

Table 2.3 Changes in Genetic Variances in the P lines with selection over generations (Beniwal *et al.*, 1992a). Estimates of V_A ($\times 10^{-4}$ on the log scale)

Generations	High line V_A	Low line V_A
0-4	71	48
5-7	51	28
8-14	42	20
15-20	12	15

The drop in genetic variance could not be ascribed to the Bulmer effect (where in a selected population gametic phase disequilibrium between loci of opposing effect results in a negative covariance of gene effects (Falconer & Mackay, 1996), or inbreeding as these are catered for in the REML analysis with the animal model

(Beniwal *et al.*, 1992a). This suggests a departure from the infinitesimal model, such that all genes influencing predicted lean mass do not have a sufficiently small effect that the changes in gene frequency and consequent changes in variances are negligible.

2.3.4 Investigation of the Genetic Basis of Response.

During the course of the P-lines and P6 line selection, a number of experiments have been set up to determine the nature of genetic response (Hastings, 1990; Hastings & Veerkamp, 1993; Veerkamp *et al.*, 1993). Reciprocal F_1 populations were set up at generation 31 (Hastings, 1990), and generation 38 (Hastings & Veerkamp, 1993) of the P6 lines, and at generation 44 of the P-lines (only the first 20 generations were selected) (Hastings & Veerkamp, 1993) (see Table 2.4). There were sex differences in the mean body weight in the F_1 generation, where the weight of males were biased to that of the female parent (Hastings, 1990; Hastings & Veerkamp, 1993). If this was a maternal effect both sexes would be affected in the same way, but F_1 females show body weights midway between the parental lines for 10 week weight in both reciprocal halves.

A difference between the two sexes is that females inherit one maternal X-chromosome and one paternal X-chromosome, whereas males receive only a maternal X-chromosome. Initial inspection of the data suggested the X-chromosome may have a significant effect on body weight (Hastings, 1990). Further investigation was carried out to confirm an X-linked effect. A backcross population was set up at generation 31, to test for evidence of cytoplasmic or mitochondrial inheritance, and Y-linked effects. The hypothesis that cytoplasmic or mitochondrial inheritance had a significant effect on mean body weight was tested by comparing the mean body weights of females, which inherited either the high line or low line cytotype. For example in the F_1 generation the type 1 F_1 females with high line dam and low line sire will have a high line cytotype, and type 2 F_1 females with low line dam and high line sire will inherit a low line cytotype. If there were significant cytoplasmic or mitochondrial

linked effects the type 1 F₁ females would be significantly heavier than the type 2 females. The data did not indicate any evidence for cytoplasmic or mitochondrial inheritance (Hastings, 1990). The Y-linked effects were tested by comparing the mean body weight of backcross males. Backcross males with the same dam genotype, but with sire carrying a Y chromosome inherited from either high or low line, are assumed to be genetically identical for the autosomes and the X-chromosome, but carry a Y-chromosome inherited from the high (H) or low (L) selection lines. For example males with H dams, and F₁ sires (with H dam and L sire) have high line X-chromosome and low line Y chromosome, and the males with H dams, and F₁ sires (with L dam and H sire) have high line X and Y chromosome. The mean body weights of these two F₁ groups indicated no significant Y-linked effects (Hastings, 1990).

Table 2.4 Mean body weights at 10 weeks of reciprocal F₁ generations and contemporaneous selection lines. Dams first in crossing nomenclature (from Hastings & Veerkamp, 1993). Mean number in each class=33.

Origin	H×H	H×L	L×H	L×L	Mean s.e.
Replicate 6					
Gen.31 female	42.9	28.6	29.9	19.4	0.6
male	50.4	36.6	31.1	22.7	0.6
Gen.38 female	39.6	26.3	26.4	16.7	0.5
male	44.7	33.4	28.9	20.2	0.6
Rep. 1 female	31.2	25.5	26.1	18.8	0.5
male	37.6	36.1	29.5	21.4	0.5
Rep. 2 female	26.1	25.8	23.2	21.7	0.5
male	31.4	33.8	26.0	26.4	0.6
Rep. 3 female	31.7	27.4	26.8	19.0	0.6
male	41.9	37.1	29.1	24.0	0.7

The presence of the X-linked effect in all 3 replicates of the P-lines (Table 2.4) suggests the effect did not arise from a single spontaneous mutation during selection,

but was present in the base population. The size of the X-linked effect can be estimated from the mean body weights of the F₁ males and females:

$$(H \times L \text{ males} - H \times L \text{ females}) - (L \times H \text{ males} - L \times H \text{ females}),$$

where H×L are individuals with high line dams and low line sires, and L×H are individuals with low line dams and high line sires. The magnitude of the effect appears similar in all lines and replicates, ranging from approximately 4.6g in the P6 line at generation 38 of selection, up to 7.6g in replicate 3 of the P line. This suggests the QTL was fixed at different alleles in the high and low direction of each of the P replicate lines by generation 20, where selection ceased (Hastings & Veerkamp, 1993). Beniwal *et al.* (1992a) showed rapid decrease in genetic variance over the initial selection in the P-lines (Table 2.5), which is consistent with rapid fixation of a major gene in the population. Maximum likelihood segregation analysis of the backcross population set up at generation 31 carried out by Veerkamp *et al.* (1993) showed no evidence for autosomal major gene in the P6 lines, but significant improvement in the likelihood when a sex linked monogenic effect was fitted in the model. Despite strong evidence for a monogenic effect on the X-chromosome caution should be taken before concluding a monogenic effect on the X-chromosome. Firstly, the rapid reduction in genetic variance would also be consistent with the fixation of a small number of sex linked genes. Selection over the first 20 generations was carried out on males only, therefore X-linked alleles in a hemizygous state would be strongly selected, as their contribution to response would be proportionally larger than autosomal alleles (Charlesworth *et al.*, 1987). Secondly, in the segregation analysis recombination on the X-chromosome was possible only in the F₁ females, so recombination may not break down the association between linked polygenes, and these linked polygenes may stay together and act as a single major gene (Veerkamp *et al.*, 1993). Clearly we are unable to ascertain whether the X-linked effect is due to one or more loci.

2.4 THE X-CHROMOSOME IN QTL MAPPING

2.4.1 Introduction.

To date the majority of experiments carried out mapping QTL in mammals have concentrated on the autosomes. The project described here is specifically concerned with mapping a putative QTL for body weight on the X-chromosome (as reviewed in section 2.3.4).

In this section the specific properties of the X-chromosome which must be considered when mapping QTL on the X-chromosome will be reviewed, and these include dosage compensation, selection responses for X-linked loci, polymorphism rates of markers, and QTL mapping on the X-chromosome.

2.4.2 X-chromosome inactivation and selection response

During the course of evolution the Y-chromosome has become the male determining chromosome in mammals. The Y-chromosome is maintained in a hemizygous state with the X-chromosome in males, and in females the X-chromosome is maintained in a homogametic state. The hemizygous existence of sex linked genes in males should be hazardous, since monosomy of autosomes is lethal or causes severe phenotypic abnormalities in mammals (Ohno, 1967; Lyon, 1992). However, X-chromosome inactivation is the buffer mechanism evolved to offset the genetic disparity between the males and females (Ohno, 1967; Lyon, 1992).

In mammals, random inactivation of the X-chromosome in the early stages of development of the female ensures equal gene dosage of X-linked genes in males and females. The inactivated X-chromosome assumes a compact, condensed structure, replicates its DNA asynchronously, and is not transcribed to messenger RNA (Lyon,

1996). Inactivation occurs early in development. Observations have shown X-inactivation in the mouse embryo in the trophoblast cells at about 4 days gestation (Lyon, 1992). Once an X-chromosome is inactivated, it remains inactivated in all descendant cells (Ohno, 1967), leading to a situation where females are in effect mosaics with respect to functional X-chromosomes. The process of inactivation spreads from a region called the inactivation centre. Penny *et al.* (1996) provided conclusive evidence that the *X-ist* gene (X-inactive specific transcript) forms an essential part of the inactivation which spreads along the chromosome.

The hemizygous state of the X-chromosome in males plays a special role in selection response from X-linked loci. The X-chromosome contains a disproportionately large share of genes (per unit of genetic length) compared to the autosomes (Charlesworth *et al.*, 1987). Partially recessive favourable X-linked loci undergoing selection in males and females (or males alone) will show a greater response to selection than similar autosomal genes (Griffing, 1965; Charlesworth *et al.*, 1987). A greater response is predicted in X-linked genes than autosomes as a result of partial shielding of favourable recessives when in a heterozygous state, while for X-linked genes, complete exposure in the hemizygous male leads to greater response to selection (Charlesworth *et al.*, 1987). As a result of the greater selection response possible with selection for X-linked loci, rapid fixation of selected alleles in the population is likely to occur. Investigation of the mouse lines used in this study (the P-lines) showed an X-linked effect which accounted for approximately 25 % of the total selection response (section 2.3.4). Where X-linked loci are of large effect, the fixation of selected alleles would result in observed changes in genetic variance. As reviewed in section 2.3.3, during the course of selection of the P-lines, analysis showed a rapid drop in the genetic variance during the course of selection, and the fixation of the X-linked loci of the large effect in the population may have contributed significantly to the change in genetic variance.

2.4.3 Variability of the X-chromosome

The X-chromosome has been shown to be more stable than the autosomes, with lower mutation rates (Ohno, 1967). Ohno (1967) postulated that the X-chromosome has been conserved *in toto* in the evolution of mammals, which is supported by similarities in size, banding patterns, and the conservation of X-linked gene sequences between mammals (Ohno, 1967; Brockdorff *et al.*, 1991; Blair & Boyd., 1994).

Less variation at restriction sites on the X-chromosome compared to the autosomes has been observed (Hofker *et al.*, 1986; Mullins *et al.*, 1988). Analysis by Hofker *et al.* (1986) showed the frequency of X-chromosome restriction fragment length polymorphisms (RFLPs), per base pair investigated, was three times lower than on the autosomes, and this difference was highly significant ($P < 0.001$). Dietrich *et al.* (1994; 1996) also observed a clear deficit in the number of microsatellite markers isolated on the X-chromosome (assuming random distribution of markers throughout the genome) and quantified the under-representation in markers to be 57%. Similar results were found by Dib *et al.* (1996) in the human genome (75% of expected polymorphic sites). Data indicated that the marker deficit observed was primarily due to a shortage of microsatellite loci rather than lower rates of polymorphic sites (Dietrich *et al.*, 1996).

It was suggested by Hofker *et al.* (1986) that the hemizygous state of the X-chromosome in males may result in a reduction in the number of new mutations (which may result in fewer genomic marker sites, as well as lower rates of polymorphism at marker sites). The lower mutation rates may be a result of:

1. The absence of X-chromosome recombination in males, resulting in a reduction in one possible source of new mutations relative to the autosomes.
2. Spermatocytes acquire more mutations than oocytes because of greater number of cell divisions in gamete maturation of the sperm. As males are hemizygous for the X-

chromosomes, and this implies that 1/3 of X-chromosome are replicated via spermatogenesis and 2/3 replicated by the lower mutation rate oogenesis.

The factors influencing the mutation rate differentially in males and females may lead to the presence of fewer microsatellite loci as well as lower rates of polymorphism at microsatellite sites on the X-chromosomes.

The under-representation of polymorphic sites on the X-chromosome, and lower rates of polymorphism at the marker loci are problematic when mapping QTL in the X-chromosome, particularly in fine resolution mapping as this relies on a dense marker map. The lower number of markers available poses the first problem as the map density of markers on the X-chromosome is lower than the autosomes, and this is further exacerbated by the lower rates of polymorphism at marker sites. A consequence of these observations is that a larger number of marker loci must be investigated to identify polymorphic X-linked marker loci compared to autosomal marker loci.

2.4.4 X-chromosome Mapping Experiments

Mapping quantitative trait loci on the X-chromosomes is achieved by the same methods used to map loci on the autosomes: by the association of phenotypic traits with regions of the chromosome in a segregating population. Due to the hemizygous state of the X-chromosome in males, a number of special considerations must be examined when mapping X-linked loci.

Using crosses between two inbred lines, examination of the heterogametic sex (which will be considered as the male) provides a simple method to detect a large effect of the sex chromosomes on quantitative traits. The reciprocal F_1 crosses between inbred lines are genetically identical except for the sex chromosomes, so examination of the F_1 heterogametic sex provides a useful test for the presence of loci of interest on the X-chromosomes (Hastings, 1990; Hagger & Stranzinger, 1992). Mapping

experiments normally utilise segregating populations (F_2 or BC_1) but for X-linked loci segregation does not occur in the heterogametic sex. Considering an F_2 cross, recombination will occur in only F_1 females, therefore a recombinant X-chromosomes from the dam only is passed on to the F_2 progeny. Mullins *et al.* (1988) discussed the merit of using males only in analysis as unequivocal segregation of X-linked loci could be followed in the hemizygous animals. A result of the hemizygous state of the X-chromosome in males is that males provide greater power to detect QTL compared to females. If the QTL is assumed to have an equal effect in both males and females, the percentage of the total phenotypic variance explained by the QTL is greater in males than in females. The power to detect QTL is strongly influenced by the percentage variance explained by the QTL (Lander & Botstein, 1989; Darvasi *et al.*, 1993).

The additional power to detect QTL in males is of obvious benefit when attempting to map an X-linked loci, and experimental design should attempt to exploit the greater power to detect QTL in the males. Unfortunately, the advantages in mapping X-linked QTL in practise are partially off-set by poor distribution of marker loci on the X-chromosome (Dietrich *et al.*, 1994, 1996; Hofker *et al.*, 1986; Dib *et al.*, 1996) and lower rates of polymorphism at isolated marker loci (Dietrich *et al.*, 1996).

CHAPTER 3

Materials and Methods

3.1 DNA Extraction

Extraction of genomic DNA from spleen tissue

DNA extraction from spleen tissue samples was carried out on frozen tissue fragments of approximately 2mm³. The spleen fragment was homogenised in 0.5ml spleen buffer (1.0M NaCl, 10mM Tris pH 8.0, 25mM EDTA pH 8.0, 0.5% SDS) with 200µg ml⁻¹ proteinase K in a 1.5 ml snap top tube using a custom made homogeniser, complementing its shape. This mixture was incubated at 37°C overnight with gentle mixing. Care was taken to keep all implements clean between samples to avoid contamination.

Following incubation, 100µg ml⁻¹ RNAaseA (DNAase free) was added to the mixture and incubated for a further 2 hours at 37°C. The final stage of digestion was the addition of 200µg ml⁻¹ proteinase K, and incubation over night.

Following the digestion stages, where the majority of protein is removed by digestion with proteinase K, phenol-chloroform extraction was carried out to remove the remaining protein impurities. The digested sample was transferred to a 2ml screw top ependorff tube containing 0.5ml saturated phenol. This mixture was inverted (shaken) for 20 minutes. The mixture was allowed to settle and 0.5ml chloro/isoamylalcohol added (24:1 chloroform:isoamylalcohol) the isoamylalcohol inhibits any remaining RNAase activity (Sambrook *et al.*, 1989). The mixture was then centrifuged at 3600 rpm for 10 minutes. The top 0.5ml aqueous phase was removed into a new screw top tube, and the organic phase and interface were discarded. A further phenol:chloroform extraction was carried out as outlined above, using 50:50 phenol:chloroform, total volume 1.25ml. Purification of the sample was carried out using chloroform extraction, which removes any lingering traces of phenol from the nucleic acid purification. Two extractions (0.8ml chloroform) rotating for 10 minutes and spinning for 5 minutes were carried out, before precipitation of the DNA. The final supernatant was added to 1ml 100% ethanol, and the mixture was shaken

vigorously to precipitate the DNA. The precipitate was spun at 3600 rpm for 20 minutes to form a pellet. The liquid was removed from the tube, and the pellet washed in 70% ethanol and allowed to dry. The dry DNA pellet was re-suspended in 300µl 1× TE (Tris-EDTA buffer) pH 8.0. The DNA samples were diluted by a standard ratio: 20µl DNA sample + 200µl double distilled filtered H₂O (dd H₂O), which yields approximately 50ng µl⁻¹ DNA.

Genomic DNA extraction from ear clip samples

Ear clips were taken at weaning for identification purposes as part of the standard procedure. Genomic DNA extraction from ear clip tissue samples precludes the need for an additional procedure to obtain DNA and therefore reduces potential stress to the animals (tail clipping is the standard method to obtain tissue samples for DNA extraction from living animals, and it requires animals to receive a general anaesthetic).

At weaning two ear clips were taken from each animal using the standard genetics mouse house ear clipper. To avoid contamination between ear clips, all traces of hair and tissue were brushed from the ear clippers. The two ear clips, approximately 2mm in diameter, were placed in a 1.5ml snap top tube and stored at -20°C prior to extraction. The ear clips were digested in 200µl tail buffer (0.3M NaAc pH 9.0, 10mM Tris pH 7.9, 1mM EDTA pH 7.0, 1% SDS) with 200µg ml⁻¹ proteinase K. The mixture was incubated at 37°C overnight with gentle mixing. Following digestion the sample was first extracted in 200µl saturated phenol for 15 minutes, and then centrifuged at 3600rpm for 10 minutes. Subsequent to the phenol extraction stage the top layer was extracted in 100µl phenol and 100µl chloroform. The final extraction stage was a single chloroform extraction carried out in order to remove traces of phenol from the sample; where 200µl of chloroform was added to the aqueous layer, mixed for 5 minutes, then centrifuged for 5 minutes. The aqueous layer was removed and added to 500µl of 100% ethanol. The DNA was precipitated as described in the previous section, washed in 70% ethanol, and re-suspended in 30µl 1.0×TE.

To investigate the possibility that contamination could occur between samples when the ear clips were taken, the protocol was tested using the following procedure. Two culled mice (one from the high selection line, and one from the low selection line) were used to obtain two ear clips successively from high and low line individuals. The ear clips were taken in a recorded order so that if contamination was present the point at which it occurred could be identified. A total of five sets of ear clips were obtained. Genomic DNA was extracted from the ear clips using the protocol described above, and then genotyped at the marker *DXMit50*, (Protocol 1, section 3.2). The genotypes obtained are shown in Figure 3.1, and indicate that there was insufficient contamination between samples to cause incorrect genotypes to be recorded.

Streamlined Genomic DNA extraction from ear clip samples

Streamlining of the protocol used to extract DNA from ear clips was carried out in order to reduce the materials used and decrease the time required to extract DNA for PCR. Following the digestion stage as outlined above, the majority of protein is removed by digestion with Proteinase K. The digested sample was simply added to 500µl 100% ethanol and shaken vigorously to precipitate the DNA. The precipitated DNA sample was looped out using a 1µl disposable plastic loop (Merck, UK), and left to dry in a 1.5 ml screw top ependorff tube. When dry, the sample was re-suspended in 40µl 0.1×TE. All DNA samples extracted from ear clips were diluted by a standard ratio of 15µl DNA sample + 30µl double distilled filtered H₂O, which yields sufficient DNA for PCR amplification

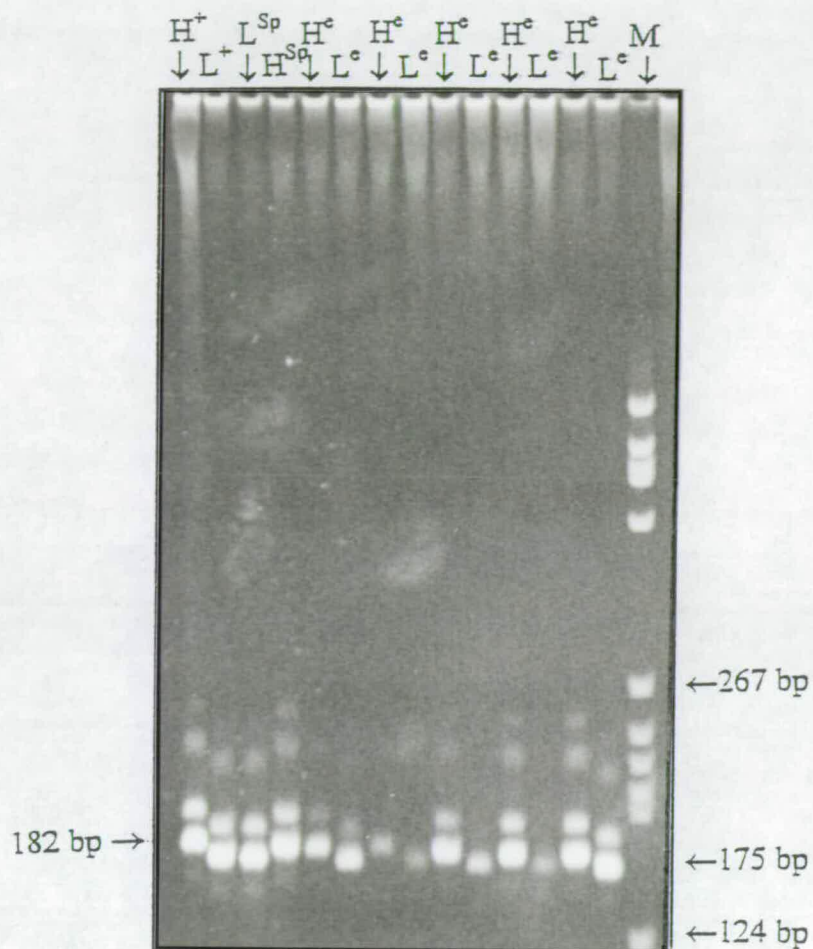


Figure 3.1 Validation of the method of obtaining genomic DNA samples using ear clips taken from individuals at weaning. Genotypes were obtained for marker DXMit50 using the DNA from high and low line mice. H/L⁺ show the positive controls, H/L^{Sp} are spleen sample DNA controls for the two animals used, and H/L^c show the ear clip samples in the order taken.

3.2 PCR amplification protocol

Protocol 1.

Individuals were genotyped at simple sequence length polymorphism (SSLP or microsatellite) markers using the polymerase chain reaction based on the protocol described by Dietrich *et al.* (1992). Genomic DNA was amplified at microsatellites using the following reaction mixture:

- 1µl genomic DNA
- 0.8µl forward (20pM) primer
- 0.8µl reverse (20pM) primer
- 2.0µl dNTP stock (made from 1 part of 100mM dATP, dCTP, dGTP, dTTP and double distilled filtered H₂O, diluted 1:10 with dd H₂O)
- 2.0µl 10x PCR buffer (Promega, UK)
- variable MgCl₂, typically 1.25mM final concentration (25mM stock)
- dd H₂O to make total reaction volume 20µl
- 0.125µl Taq Polymerase (0.625u) per reaction

Reactions were overlaid with 20µl light mineral oil and amplified using an OmniGene PCR machine (Hybaid, UK). The PCR thermocycling protocol was:

1×	94°C for 3 min	55°C 1min	72°C 30sec
32×	94°C for 30sec	55°C 1min	72°C 1min
1×	94°C for 30sec	55°C 2min	72°C 5min

Following PCR amplification 8µl Ficoll/Bromophenol blue buffer (17% Ficoll, 0.1M Tris, 0.1M EDTA, 0.1% Bromophenol Blue) was added. This acts as a 'stop' buffer to prevent any further reaction, and as a loading dye. Products were run on 20cm long vertical running polyacrylamide gels for 3-4 hours at 155 Volts.

Gels were typically 6% polyacrylamide:

6ml 40% acrylamide/bis-acrylamide

4ml 10×TBE

30ml H₂O

130μl TEMED

130μl 10% ammonium persulphate

Gels were visualised after ethidium bromide staining. The gels were stained for 25mins in 500 ml 50μl l⁻¹ ethidium bromide (stock 10mg ml⁻¹), visualised under ultra violet light, and photographed using professional black and white instant film 8.5×10.8cm (Polaroid, UK).

Considerable streamlining of procedures was carried out to allow more efficient and economic genotyping of individuals. Many procedures have been adapted to utilise an 8 channel pipette (Dynatech Laboratories, Ltd). Genomic DNA samples were stored in 96 well sample boxes (Beckman, UK), and PCR reactions were amplified in flexible 96 well microtitre plates (Hybaid, UK). Loading of gels also utilised the 8 channel pipette, where custom made gel combs were designed so that tip array corresponds to alternate teeth in the gel comb.

Protocol 2.

Additional modifications were made to Protocol 1 PCR reaction protocol. These modifications further reduced the time required to set up each reaction, in addition to enabling substantial reductions in the reagents used. These modifications have resulted in more economical genotyping of the animals.

Changes to the standard PCR reaction protocol were based on the protocols reported by Routman & Cheverud (1994). The 55°C annealing temperature previously used in the thermocycling protocol was reduced to 50°C. This was carried out in conjunction with a 50% reduction in the reaction mixture volume, i.e. the total reaction volume of 20μl used in Protocol 1 was reduced to 10μl. Ficoll buffer was no longer added to the reaction mix after the PCR reaction, instead 1/5 volume loading dye (60% sucrose,

1.0mM cresol red) was added to the reaction mixture prior to PCR amplification. A typical reaction mix was:

- 0.5µl genomic DNA
- 0.4µl forward primer
- 0.4µl reverse primer
- 1.0µl stock dNTP (as above)
- 1.0µl 10x PCR buffer (Promega, UK)
- 2.0µl loading dye
- variable MgCl₂ Typically 1.25mM final concentration (25mM stock)
- dd H₂O to make total reaction volume 10µl
- 0.0625µl Taq Polymerase (0.625u) per reaction

The addition of the loading dye prior to the PCR reaction reduces pipetting effort, and seems to facilitate weak PCR reactions (Routman & Cheverud, 1994). Products were run on 6% polyacrylamide gels at 200 volts for 1.5 to 2 hours, and visualised as described above.

All genotypes were scored on at least two separate occasions to minimise incorrect genotypes. Pedigree information (where available) also provided a means to detect genotypes which were incorrect. Due to the mode of X-linked inheritance a number of genotypes were not possible in the F₂X generation, e.g. heterozygous genotypes for male individuals. Where discrepancies were identified, further investigations were carried out and, where necessary, PCR reactions were repeated.

3.3 Mouse Husbandry

All mating pairs were set up using individuals of a minimum age of 6 weeks. During the breeding period all mating pairs were maintained on Rat and Mouse No. 3 Breeding Diet (Special Diets Service, UK). At birth all litters were adjusted to a maximum litter size of 12 pups per litter. The records taken during the mating period

were: the date of birth, the number of pups born, and the adjusted number of pups. All mating pairs were maintained for up to four parities, and the parity number was recorded for all litters.

All animals were weaned at 3 weeks, when records taken were the total number of pups, and the number of males and females weaned. At weaning, individuals were given ear marker identification, separated into single sex cages, and maintained in groups of up to 10 individuals on Rat and Mouse No. 1 Modified Maintenance Diet (Special Diets Service, UK). At weaning, litter mates of the same sex were kept together. If animals were small at weaning, all animals in the cage were given access to wet mashed diet on the cage floor.

3.4 Estimation of carcass fat percentage

Carcass fat percentage was estimated in all individuals using the dry matter content of the carcass (section 2.3.3). Partially thawed animals were prepared for freeze drying by scoring the skin surface (including the tail), and making deep cuts to the rib cage and skull (this facilitated complete drying of the carcass). The prepared animals were placed in groups of 2-3 in foil containers with identification and re-frozen. Completely frozen animals in batches of approximately 50 animals were dried under high vacuum for 72 hours. Fat percentage was estimated according to Hastings & Hill (1989) where:

$$\text{freeze dried fat percentage} = [(\text{freeze dried weight/dead weight}) \times b - a] \times 100,$$

where $a=0.302$ and $b=1.13$. This relationship was determined in the F-lines at generation 20 of selection; for further information see section 2.3.3.



3.5 Statistical Analysis

Sex linked effects in the reciprocal F₁ generation

Body weights were recorded for the reciprocal F₁ populations described in Chapters 4 and 6. The mean body weight of males and females in the reciprocal halves of the F₁ generation were compared to test for the sex linked effect previously observed at generation 31 and 38 of selection in the P6 lines (Hastings, 1990; Hastings & Veerkamp, 1993; section 2.3.4). Using Restricted Maximum Likelihood (REML) in Genstat 5.3 (Genstat 5 Committee, 1993), body weights recorded in the reciprocal F₁ cross were analysed fitting sex nested in dam line genotype (i.e. dams from either low or high line) using the model below:

$$Y_{ijkl} = \mu + SO_{ij} + D_k + \beta \cdot W_k + e_{ijkl},$$

where:

Y_{ijkl} = the observations on the l th individual (of the reciprocal F₁ generations described in Chapter 4 and 6)

μ = overall mean

SO_{ij} = fixed effect of the i th sex nested within the j th line origin of the dam

D_k = random common environmental effect associated with the k th dam

β = the linear regression coefficient of Y on the number weaned (W_k) from the litter of the k th dam

e_{ijkl} = the residual error associated with the l th individual

Model Evaluation

A number of fixed and environmental effects were recorded during data collection as part of the routine procedure in the maintenance of the mouse populations investigated in this study (for examples of the effects recorded, see section 3.3). To determine the importance of individual terms in the models for different phenotypic traits a likelihood ratio statistic was calculated by the REML directive in Genstat 5.3

(Genstat 5 Committee, 1993). The statistic used tests a fixed model against a nested sub-model. This test provides log likelihoods calculated from the same projected data set, using the same random model, but with designated fixed effects constrained to zero. The difference in log likelihoods therefore gives a likelihood ratio test statistic: $-2(L_1 - L_0)$, where L_1 is the overall likelihood of the full model, and L_0 is the likelihood for the nested sub model with effects restricted to zero. The null hypothesis for this test is that the effects restricted to zero are not significant in the model ($P > 0.05$). This test statistic has an asymptotic chi squared distribution with degrees of freedom equal to the number of degrees of freedom in the fixed model term constrained to be zero in the sub-model (Genstat 5 Committee, 1993).

Using the example of the model testing for sex linked effects in the reciprocal F_1 generation (described above), but assuming data on first and second parity litters were available. To determine the importance of parity in the model, the effect associated with the fixed effect of parity would be constrained to be zero in the nested sub model:

Sub model:
$$Y_{ijklm} = \mu + SO_{ij} + D_{km} + \beta \cdot W_{km} + e_{ijklm},$$

Full fixed model
$$Y_{ijklm} = \mu + SO_{ij} + D_{km} + \beta \cdot W_{km} + P_m + e_{ijklm},$$

where:

Y_{ijklm} = the observations on the l th individual of the reciprocal F_1 generation

μ = overall mean

SO_{ij} = fixed effect of the i th sex nested within the j th line origin of the dam

D_k = random common environmental effect associated with m th parity litter of the k th dam

β = the linear regression coefficient of Y on the number weaned (W_{mk}) from the m th parity litter of litter of the k th dam

P_m = the fixed effect of the m th parity

e_{ijklm} = the residual error associated with the l th individual

The number of degrees of freedom of the test statistic would be one (as first and second parity litters were recorded). The importance of the term in the model would be determined by the likelihood ratio test statistic: $-2(L_1-L_0)$, such that if the test statistic was greater than 3.84 (χ^2 $P<0.05$, with 1d.f.) the fixed effect of parity would be significant. If the test statistic was less than 3.84 the fixed effect of parity would be dropped from the model.

The analytical methods described above will be referred to in later chapters. In addition to the methods described in this section, additional statistical analysis will be described in greater depth in later chapters.

CHAPTER 4

Single marker analysis of a reciprocal F₂ population to map the putative QTL for body weight on the X-chromosome

4.1 INTRODUCTION

As reviewed previously (section 2.3.4), investigation of the basis of response to selection in mouse lines divergently selected for body weight suggested a large contribution to the total divergence by the X-chromosome. An experiment was undertaken to confirm the existence of the putative X-linked QTL using molecular genetic techniques. Phenotypic records and DNA samples were available from an experiment set up to investigate the basis of response by segregation analysis in the P6 mouse lines. The experiment was not set up with the purpose of mapping the putative X-linked QTL discussed in section 2.3.4. Using the material and data available, the results of mapping the putative X-linked QTL will be presented.

4.2 MATERIALS AND METHODS

4.2.1 The F₂ Population mating structure and measurements taken

A reciprocal F₂ generation had been set up using foundation animals from generation 45 of selection of the P6 lines (by Dr. I. M. Hastings). The P6 line animals were used to form a reciprocal F₁ generation with eight F₁ families used in each reciprocal half. From this F₁ generation a reciprocal F₂ generation was established representing all possible combinations

HL × HL

HL × LH

LH × LH

LH × HL,

where HL = F₁ individuals with high selection line dams and low selection line sires and LH = F₁ individuals from the reciprocal F₁ matings. F₁ parents were selected in order to represent each F₁ family in equal frequency. The F₂ generation comprised of 195 animals (100 females, 95 males) from 20 families (first parity litters only). Records of 6 and 10 week weights were taken on all individuals in the pedigree. Spleen tissue samples were taken from all F₂ individuals. DNA samples were not available from the P6 high and low line parents and the F₁ parents.

4.2.2 Genotyping of Animals

Genomic DNA was extracted from spleen tissue fragments from each of the F₂ individuals using the protocol described in section 3.1. DNA extracted from eight high and low line females from generation 51 of selection were used to identify polymorphic markers (samples were not available from parental individuals in generation 45 of selection). All individuals were genotyped at microsatellite markers (using Protocol 1, section 3.2) to detect variation between the high and low lines at

molecular markers. All F₂ individuals were genotyped at the polymorphic markers isolated (see Table 4.2) using Protocol 1 of section 3.2.

4.2.3 Analysis

F₁ Phenotypic Data Analysis

Records of 6 and 10 week weights were analysed to test for the sex linked effect previously observed in the P and P6 selection lines (section 2.2.4). Analysis was carried out using the model described for the reciprocal F₁ data in section 3.6. The model included only effects which were significant (P<0.05).

F₂ Single Marker Analysis

The importance of individual terms in the model for analysis of F₂ body weight data was ascertained using the method described in section 3.6 for model evaluation. The final model used for analysis of 6 and 10 week weights was:

$$Y_{ijkl} = \mu + MS_{ij} + D_k + \beta.W_k + e_{ijkl},$$

where:

Y_{ijkl} = the phenotype of 6 and 10 week weights on the l th individual

μ = overall mean

MS_i = fixed effect of the i th marker genotype within the j th sex

D_j = random common environmental effect associated with the k th dam

β = the linear regression coefficient of Y on the number weaned (W_k) from the litter of the k th dam

e_{ijkl} = the residual error associated with the l th individual

Analysis of F₂ data to determine the association between the QTL and markers was carried out using REML in Genstat 5.3 (Genstat 5 Committee, 1993).

4.3 RESULTS

4.3.1 Estimation of sex linked body weight effects in the F₁ generation

The mean body weights of the males and females in the two reciprocal halves of the F₁ generation were compared to investigate if the sex linked QTL effect described previously (section 2.3.4) could be identified. Contemporaneous high and low line individuals were not available for comparison. Analysis of the 6 week weight data showed there were differences between the mean body weights of males and females in both reciprocal halves of the F₁ generation, significantly so in the high line dam reciprocal half ($P < 0.05$ using a Student t-test). The males were 5.6g heavier than females in the high line dam reciprocal half, and 2.5g heavier in the low line dam reciprocal half. Using the relationship described in section 2.3.4, the size of the sex linked effect can be estimated from the mean body weights of the males and females:

$$(H \times L: A - Aa) - (L \times H: a - Aa),$$

where H×L represents the F₁ reciprocal half (dam shown first in crossing nomenclature), Aa are the F₁ females, and a and A are the F₁ males. The estimated sex linked effect was therefore 3.1g at 6 weeks.

Table 4.1 Mean body weight (g) at 6 and 10 weeks of reciprocal halves of F1 P6 line crosses. Dam shown first in crossing nomenclature

Weight	Sex	L×H	s.e.	No.	H×L	s.e.	No.
6 week	female	21.9	1.69	20	20.3	1.58	25
	male	24.4	1.70	19	25.9	1.59	26
10 week	female	26.4	1.02	20	24.6	0.94	25
	male	29.6	1.03	19	34.1	0.93	26

s.e. = standard error of mean, No.= number within class.

The estimated mean body weights at 10 weeks show there were significant differences in body weight between the males and females within the two reciprocal halves of the F_1 generation ($P < 0.05$), where the mean body weights of males were 3.2g and 9.5g greater than females in the low line and high line dam reciprocal halves respectively. Using the relationship described above the estimated sex linked effect was approximately 6.3g for 10 week weight. The estimated sex linked effects at 6 and 10 weeks are of a similar magnitude to those shown previously (Hastings, 1990; Hastings & Veerkamp, 1993).

4.3.2 Rates of polymorphism on the X-chromosome

An example of products obtained from PCR amplification of the microsatellite loci (*DXMit25*) are shown in Figure 4.1. The results show the genotyping of DNA obtained from the eight high and low line females of generation 51 of the P6 selection line.

Initial genotyping of the test females from the high and low lines (section 4.2.2) at 42 microsatellite marker loci showed only 12 loci to be polymorphic (29%), indicating low rates of polymorphism at the X-linked marker loci tested. Of the 12 polymorphic loci isolated none were found to be fixed at different alleles in the high and low lines. From these 12 polymorphic loci, seven markers produced consistently scoreable PCR products (see Table 4.2). These markers were used to genotype all F_2 individuals. The remaining five informative markers could not be consistently scored.

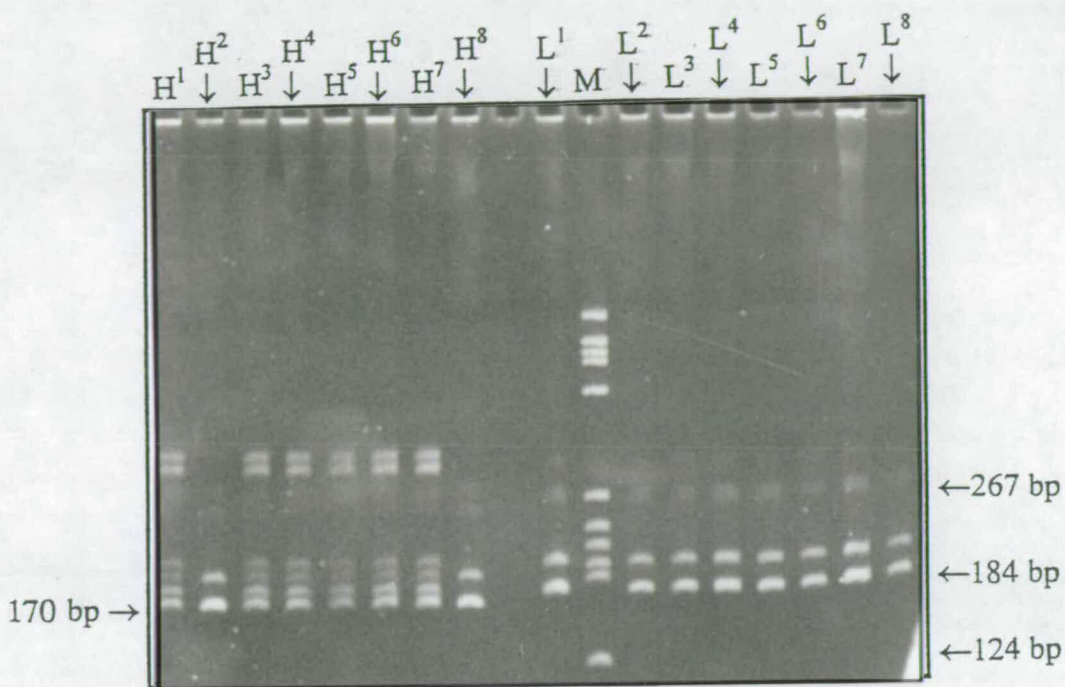


Figure 4.1 An example of a Gel photo showing the eight high and low line females from generation 51 of selection genotyped at the marker loci DXMit25. The gel shows females from the low line are all homozygous for the same allele (approximately 184 base pairs (bp)), the high line individuals H² and H⁸ are homozygous for the 170bp allele, and the remaining high line females are heterozygous.

4.3.3 Single marker analysis of F₂ data set.

The mean body weights ($\pm$ S.D.) in the F₂ generation were 25.29g $\pm$ 5.74 and 21.99g $\pm$ 4.97 at 6 weeks, and 31.24g $\pm$ 7.15 and 27.11g $\pm$ 5.70 at 10 weeks in males and females respectively. These two traits were highly correlated, having a residual phenotypic correlation of $r=0.86$ (trait values corrected for sex and number weaned per litter).

Table 4.2. Mean marker associated body weight (g) at 10 weeks, and number of animals within each group (shown in italics). Standard errors of means are shown in parenthesis. The map positions were taken from the Chromosome Committee (Mouse Genome Informatics, 1996).

Marker	Map Position	Males		Females		
		♂ <i>a</i>	<i>A</i>	<i>aa</i>	<i>Aa</i>	<i>AA</i>
<i>DXMit55</i>	1.5cM	30.8 (0.6) <i>61</i>	31.9 (0.7) <i>39</i>	26.6 (0.8) <i>30</i>	28.2 (1.3) <i>53</i>	27.2 (0.6) <i>12</i>
<i>DXMit50</i>	11.0cM	29.6 (0.6) <i>53</i>	33.2 (0.6) <i>47</i>	25.8 (1.1) <i>14</i>	27.9 (0.7) <i>49</i>	26.9 (0.6) <i>32</i>
<i>DXMit25</i>	27.8cM	29.0 (0.7) <i>66</i>	32.4 (0.6) <i>34</i>	26.5 (0.7) <i>46</i>	27.7 (0.7) <i>40</i>	26.9 (1.3) <i>9</i>
<i>DXMit16</i>	36.0cM	27.7 (1.0) <i>15</i>	31.9 (0.5) <i>85</i>	26.4 (3.7) <i>2</i>	27.3 (0.5) <i>13</i>	26.4 (1.1) <i>80</i>
<i>DXMit64</i>	36.5cM	29.6 (0.9) <i>22</i>	31.7 (0.5) <i>78</i>	26.7 (2.0) <i>4</i>	27.2 (0.5) <i>19</i>	26.9 (0.9) <i>72</i>
<i>DXMit79</i>	44.5cM	30.8 (1.0) <i>15</i>	31.3 (0.5) <i>85</i>	26.2 (1.3) <i>13</i>	27.1 (0.6) <i>24</i>	27.8 (0.9) <i>58</i>
<i>DXMit38</i>	48.0cM	30.2 (0.6) <i>69</i>	33.7 (0.7) <i>31</i>	26.7 (0.7) <i>13</i>	28.9 (1.1) <i>37</i>	27.1 (0.7) <i>45</i>

♂ *aa* and *AA* are the homozygote marker genotypes in females predominant in the low and high line respectively, and *a* and *A* are corresponding the hemizygote marker genotypes in males.

The mean 10 week body weight of males and females associated with the markers genotyped on the X-chromosome are shown in Table 4.2. All high line marker genotypes (marker alleles predominant in the high line test females) have positive marker associated effects. In the males, the marker associated mean body weights at low line marker alleles range between 27.7g at marker *DXMit16* and 30.8g at markers *DXMit55* and *DXMit79*. For the high line marker alleles the mean body weight ranges between 31.3g at marker *DXMit79* and 33.7g at marker *DXMit38*.

One of the assumptions for the single marker analysis is that the marker and QTL are fixed for different alleles in the base population, i.e. the F₁ female parents are heterozygous at the markers and QTL. As the parental DNA is not available for the F₂ population, it is not possible to determine if the F₁ females were all heterozygous for marker alleles. However, if all F₁ females were heterozygous at all marker alleles, in the reciprocal F₂ generation the ratio of marker alleles would not deviate significantly from a 1:1 ratio. The marker allele frequencies observed in the F₂ population (with the exception of marker *DXMit50*) suggest that the marker allele frequencies were significantly different ($P < 0.05$) from the 1:1 segregation ratio expected in a cross between lines where markers are fixed at different alleles in the base population (using a χ^2 test with 1 d.f.).

In the females, the homozygote marker associated QTL effects at 10 week weight are 25.8-26.7g at low line marker alleles, and 26.4-27.8g at high line marker alleles. The mean body weights associated with all heterozygous marker genotypes (with the exception of marker *DXMit79*) show higher mean body weight than mean body weight associated with high line marker genotypes. If we assume a single QTL on the X-chromosome, this relationship suggests over-dominance of the marker associated QTL effect.

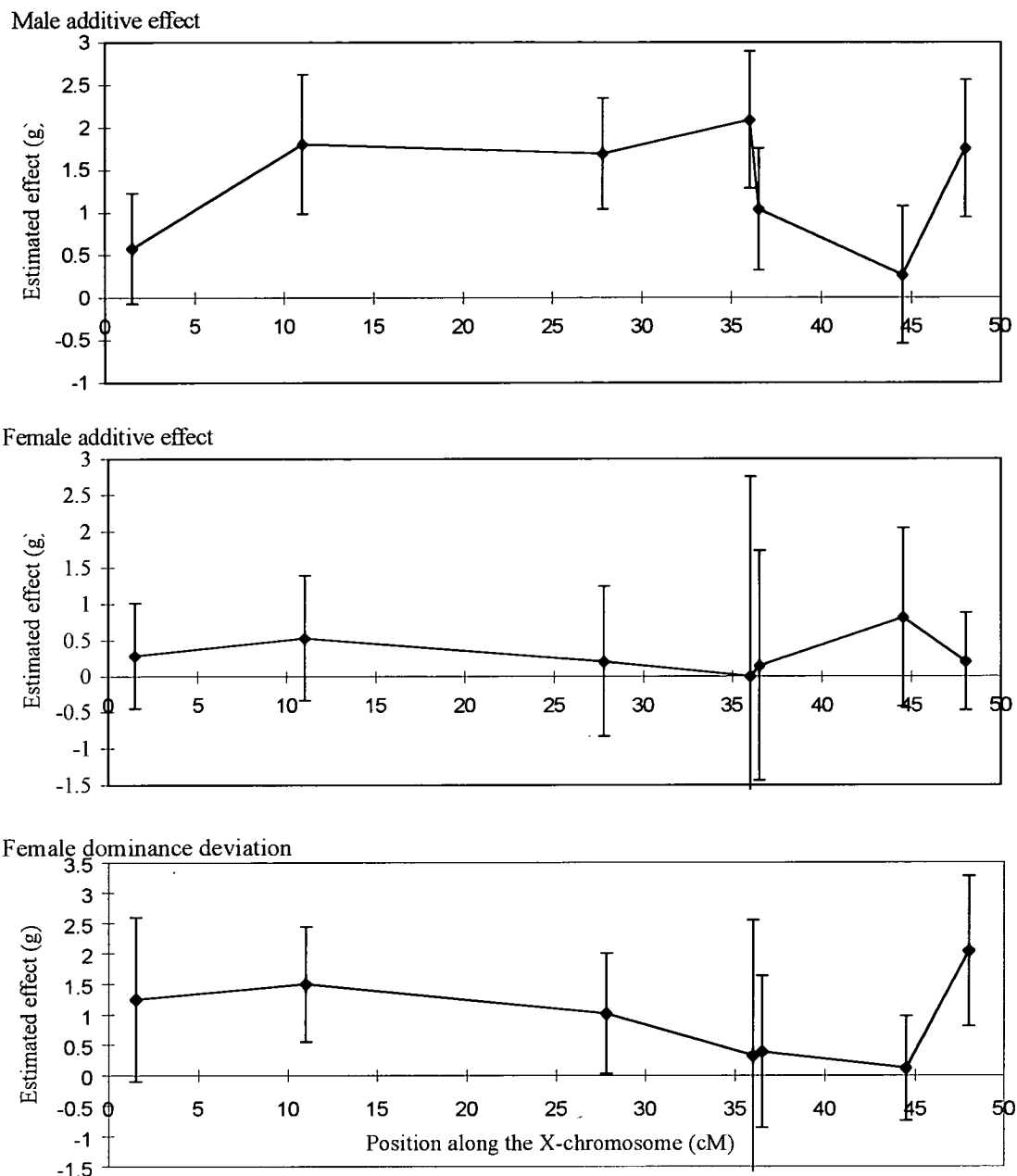


Figure 4.2. Illustration of the mean marker associated QTL effects at 10 week weight for markers along the X-chromosome. The additive effects represent half the difference between the homozygote females, and half the difference between the hemizygous males. Map positions were taken from Chromosome Committee (Mouse Genome Informatics, 1996). The error bars represent the standard errors associated with the estimates.

The results presented in Figure 4.2 show the mean marker associated QTL effects for 10 week body weight at different positions on the X-chromosome. No significant marker associated QTL effects were detected in the females ($P>0.05$). The estimates of additive marker associated QTL effects were low, ranging from 0.8g at marker *DXMit79* to less than 0.1g at markers *DXMit16* and *DXMit79*, and these estimates were associated with high standard errors of 0.7-2.8. The dominance deviations were positive at all markers, but like the additive estimates, none of the estimates were significantly different from zero. In the males significant marker associated effects were detected at all marker loci with the exception of markers *DXMit55*, *DXMit64* and *DXMit79*. The magnitude of the effects at markers *DXMit50* and *DXMit25* at the proximal end of the X-chromosome were approximately 1.8g. A similar size effect was associated with the marker *DXMit38* at the distal end of the X-chromosome. The highest effect was associated with marker *DXMit16* at 36cM with an effect of 2.1g.

Similar relationships as outlined above were observed in the analysis of 6 week weight data using the model described in section 4.2.3 (data not shown). In general, smaller estimated marker associated QTL effects were seen at 6 week weights than 10 week weights. The standard errors associated with estimates were also lower compared to standard errors associated with 10 week weight estimates. None of the markers which did not show significant effects in the analysis of 10 week weight showed significant associated effects when analysed with 6 week weights.

4.4 DISCUSSION

4.4.1 The reciprocal F₁ generation and sex linked effects

The 6 and 10 week weight data from the F₁ generation were analysed to determine if the sex linked effect observed previously (see section 2.3.4) could be isolated in this F₁ population. The mean body weights of males were heavier than females in both reciprocal halves of the F₁ generation at 6 and 10 weeks, but the difference was larger with high line dams. Using the relationship described (section 4.3.1) the estimated sex linked effect was 3.1g at 6 weeks and 6.3g at 10 weeks. Both estimates of the sex linked effects are of a similar magnitude to those estimated previously at 6 and 10 weeks (Hastings, 1990; Hastings & Veerkamp, 1993). Therefore, as discussed earlier, mean body weights of males and females from the reciprocal F₁ generation in this study support the hypothesis proposed by Hastings (1990) of a major effect for body weight being present on the X-chromosome.

When the X-linked effects are expressed as a mean of the F₁ male body weight (as the sex linked effect is expressed only in males in the F₁ generation) the percentage contribution is approximately 12% at 6 weeks and 20% at 10 weeks.

4.4.2 Polymorphism on the X-chromosome

The initial genotyping of high and low line test females presented in section 4.3.2 indicate low rates of polymorphism at microsatellite loci on the X-chromosome. The genotypes indicate only 29% of marker loci investigated were polymorphic, and these results support previous observations in mice (Dietrich *et al.*, 1994, 1996). Dietrich *et al.* (1996) presented results showing the average pairwise polymorphism rates among *Mus musculus* laboratory strains investigated were approximately 50%, but polymorphism on the X-chromosome was significantly lower at only 33%. A consequence of low rates of polymorphism at marker loci is that a greater number of

marker loci need to be investigated to isolate polymorphic marker loci for QTL mapping. In addition to the lower rates of polymorphism at isolated marker loci, there is also an under-representation of isolated marker loci on the X-chromosome in mice (Dietrich *et al.*, 1994, 1996) and man (Dib *et al.*, 1996; Hofker *et al.*, 1986). The poor distribution of markers on the X-chromosome exacerbates problems of low rates of polymorphism at isolated marker loci. Isolation of informative marker loci on the X-chromosome is therefore considerably harder than on the autosomes. Hofker *et al.* (1986) suggested that lower mutation rates in the hemizygous male are responsible for the poor distribution of marker loci on the X-chromosome (see section 2.4.4).

4.4.3 Estimated marker associated QTL effects in the F₂ population

The mean body weights associated with the marker genotypes were high for all markers predominant in the high selection line compared to the low selection line marker genotypes. This relationship was observed in both males and females. The F₂ data show significant additive effects associated with all markers in the F₂ males at 10 weeks with the exception of *DXMit55*, *DXMit64* and *DXMit79*. In females the additive and dominance effects were positive but not significantly different from zero. The trend in females showed the additive effects were considerably less than those estimated in males, although the difference was not significant ($P > 0.05$).

Expressing the marker associated effects at marker *DXMit50* (this marker was chosen as its segregation ratio did not deviate significantly from a 1:1 ratio and it was associated comparatively high effects in both males and females) as a percentage of the total mean body weight will give an indication of the total contribution of the marker associated effects. The marker associated effects (the difference between the homozygote female genotypes and hemizygote male genotypes) explains approximately 11% and 4% of the mean body weights at 10 weeks in males and females respectively. The estimated percentage contribution of the marker associated effects at *DXMit50* are considerably less than the total X-linked effect estimated in the F₁ generation and estimated previously (Hastings, 1990; Hastings & Veerkamp,

1993). However, this may be as expected due to the inability to determine informative markers in the F_2 generation.

Analysis of data using single marker analysis assumes that the marker and putative QTL are in complete linkage. To establish marker-QTL linkage, it must be possible to trace the origin of markers unequivocally to either the high or low line, which is not possible in this data set. In an F_2 generation derived from a cross between two lines when marker loci are segregating within lines (as in this population), informative markers can be determined by genotyping all animals in the F_2 pedigree, i.e. parents of the F_1 and F_2 generations. Tissue samples for DNA extraction were not available from parents of the F_1 and F_2 generations to enable fully informative loci within litters to be determined.

Regardless of the inability to determine fully informative marker loci, the marker associated effects still indicate a QTL present on the X-chromosome. Large significant effects were associated with 4 of the 7 markers genotyped on the X-chromosome in males. If there was no QTL segregating in the F_2 generation we would not expect to observe such large (and significant) marker associated effects.

4.4.5 General discussion of results

The results obtained from the F_1 analysis support previous evidence of a putative QTL for body weight on the X-chromosome. The analysis of F_2 data show support for this hypothesis, but, full interpretation of the results of F_2 data set is not easily achieved. Analysis of this data could be carried out effectively by Gibbs sampling, where estimates are conditioned on all possible marker genotypes weighted by their conditional probabilities. The Gibbs sampling would enable full analysis of the data, but the method is complicated, and computationally demanding. In addition to the problems associated with analysing the F_2 data set, the power of this data set to map QTL will be limited by the density of the marker map and the level of informative marker data. The map density could be increased by isolating additional markers

segregating in the F_2 population, but since full interpretation of the results of this F_2 generation is difficult without complex analysis, it was decided that the most appropriate expenditure of additional resources would be to repeat the F_2 experiment. In the repeated F_2 population (Chapter 6, referred to as the F_2X population), DNA samples were available for all individuals in the pedigree, parental lines and F_1 , allowing fully informative marker loci to be determined. The F_2X population also enabled the contribution of carcass fat percentage to the X-linked QTL for body weight to be investigated.

CHAPTER 5

Adaptation of autosomal QTL analysis to enable analysis of X-linked data

5.1 INTRODUCTION

The results of the analysis of reciprocal F_1 generations (Hastings, 1990; Hastings & Veerkamp, 1993; Chapter 4) show a strong indication of an X-linked QTL for body weight in the divergently selected P6 mouse lines. In addition to the findings of Hastings (1990) and Hastings & Veerkamp (1993), analyses of an F_2 generation of the P6 mouse lines using single markers (Chapter 4) also indicate X-linked QTL for body weight.

As discussed in section 2.2.2, QTL analysis in segregating populations using single markers, although simple to implement, has a number of shortcomings leading to the inability to accurately estimate QTL effects and position. To address these problems, an analysis using marker intervals to estimate QTL position and effects was developed (Lander & Botstein, 1989).

To date, statistical methods for mapping QTL in segregating populations have been developed almost exclusively for autosomal data (section 2.2.2). Due to the X-chromosome being hemizygous in males, a number of special points must be considered when mapping X-linked QTL. This Chapter will outline the analysis developed specifically for X-linked data, and provide validation of the method using simulated data.

5.2 MATERIALS AND METHODS

5.2.1 Development of an X-linked Multiple Regression Analysis

An X-linked multiple regression analysis was developed based on the autosomal multiple regression method of Haley & Knott (1992). Initial development and testing considered a cross between two inbred lines; the method was further developed for a cross between two non-inbred lines, which will be discussed later in this chapter. The basic method considers a reciprocal cross between two inbred lines, with a QTL (Q) positioned between two co-dominant flanking markers (A and B). The two inbred lines were assumed to carry different alleles at all three loci,

e.g. Dam: $A_2A_2Q_2Q_2B_2B_2$ $\times$ Sire: $A_1Q_1B_1$,

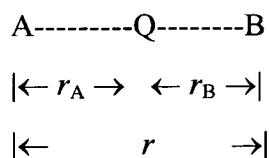
and Dam: $A_1A_1Q_1Q_1B_1B_1$ $\times$ Sire: $A_2Q_2B_2$.

These crosses produce 5 possible QTL genotypes in the F_2 , three female genotypes Q_2Q_2 , Q_1Q_2 , and Q_1Q_1 and two male hemizygous genotypes Q_2 and Q_1 . The phenotypes of the 5 QTL classes were assumed to have the same variance and to be normally distributed or could be transformed to be so. The mean genotypic effects in the F_2 (with additive terms estimated in males and females separately) were set as,

$$\begin{array}{ll} m+a_f \text{ for } Q_1Q_1 & m+a_m \text{ for } Q_1 \\ m+d \text{ for } Q_2Q_1 & \\ m-a_f \text{ for } Q_2Q_2 & m-a_m \text{ for } Q_2, \end{array}$$

where m was the mid-parent (mean of the homozygote females and hemizygote males), a_f is the additive deviation in females (half the difference between the homozygotes), a_m is the additive deviation in males (half the difference between the hemizygotes), and d is the dominance deviation in females only.

The recombination distance between A and Q was r_A and Q and B was r_B .



The recombination distances between the two markers was r , which was assumed to be known, or could be calculated from marker data prior to analysis. For all analyses it was assumed there was no crossover interference, thus:

$$r = r_A + r_B - 2r_A r_B.$$

Haldane's (1919) mapping function was used to convert distances (x) in Morgans to recombination fractions, $r = 0.5(1 - e^{-2x})$.

The expected means of the F_2 individuals for each flanking marker genotype combination can be derived in terms of recombination distances between markers and putative QTL genotypes. The calculations of expected marker genotype means for the putative QTL were considered for the male and female parent separately, in order to account for the hemizygous state of the X-chromosome. The following example shows the derivation of the expected mean in terms of QTL genotype for a female F_2 individual with homozygote flanking marker genotype $A_1 A_1 B_1 B_1$.

The homozygote female flanking markers genotype:	$A_1 A_1 B_1 B_1$,
the possible female parent gametes were:	$A_1 Q_1 B_1 \quad A_1 Q_2 B_1$,
these occurred with expected frequency:	$(1-r_A)(1-r_B)/2 \quad r_A r_B/2$.

The male parent gamete $A_1 Q_1 B_1$ has expected frequency of 0.5 as the only other possible combination within the F_2 population is $A_2 Q_2 B_2$. The female homozygotes of marker genotype $A_1 A_1 B_1 B_1$ have an expected frequency of $(1-r)/4$ in the F_2 population.

The expected frequencies of the 3 possible QTL genotypes with marker genotype $A_1A_1B_1B_1$ are therefore:

$$Q_1Q_1 \quad (1-r_A)(1-r_B)/4$$

$$Q_1Q_2 \quad r_Ar_B/4$$

$$Q_2Q_2 \quad 0.$$

Summing over the QTL genotypes, and scaling for expected frequencies of marker genotypes, the expected mean performance of an F_2 female of homozygous marker genotype $A_1A_1 B_1B_1$ is:

$$m + a (1-r_A)(1-r_B)/(1-r) + d r_Ar_B/(1-r),$$

where a is the coefficient of additive genetic deviation and d is the coefficient of dominance genetic deviation. The same procedure can be followed to derive expected mean performance for male F_2 individuals. For example the male hemizygous marker genotype A_1B_1 has expected frequency of $(1-r)/2$ in the F_2 population. The expected frequencies of the two possible QTL genotypes in the F_2 generation are:

$$Q_1 \quad (1-r_A)(1-r_B)/2$$

$$Q_2 \quad r_Ar_B/2.$$

As with females, summing over QTL genotypes and scaling for the expected marker frequencies, the expected mean performance of an F_2 male with hemizygote marker genotype A_1B_1 is:

$$m + a [(1-r_A)(1-r_B)-(r_Ar_B)]/(1-r).$$

($d=0$ as the X-chromosome is a hemizygous in males, therefore no dominance deviation can arise). For the expected mean performance of males and females the same mean (m) was used, as it is assumed that the fixed effect of sex fitted in the analysis removes variance due to sex (although this is not necessary if males and females are analysed separately). The male parent is not considered when estimating mean performance in the F_2 males, as there is no contribution of the X-chromosome from the male parent. The coefficients of a_m , a_f and d in terms of recombination fractions for each of the 12 possible flanking marker combinations (4 males, and 8 females) on the X-chromosome in an F_2 population are shown in Table 5.1.

Table 5.1 Expectations of mean genotypic effects of an X-linked QTL for all possible marker genotypes in a reciprocal F2 population.

Sex	Marker genotypes	Coefficients of:	
		α (additive genetic deviation)	d (dominance genetic deviation)
Male	A ₁ B ₁	$[(1-r_A)(1-r_B)-(r_Ar_B)]/(1-r)$	-
	A ₁ B ₂	$[(1-r_A)r_B-r_A(1-r_B)]/r$	-
	A ₂ B ₁	$-[(1-r_A)r_B-r_A(1-r_B)]/r$	-
	A ₂ B ₂	$-[(1-r_A)(1-r_B)-(r_Ar_B)]/(1-r)$	-
Female (A ₁ B ₁ male parent)	A ₁ A ₁ B ₁ B ₁	$[(1-r_A)(1-r_B)]/(1-r)$	$[r_Ar_B]/(1-r)$
	A ₁ A ₁ B ₁ B ₂	$[(1-r_A)r_B]/r$	$[r_A(1-r_B)]/r$
	A ₁ A ₂ B ₁ B ₁	$[r_A(1-r_B)]/r$	$[(1-r_A)r_B]/r$
	A ₁ A ₂ B ₁ B ₂	$[r_Ar_B]/(1-r)$	$[(1-r_A)(1-r_B)]/(1-r)$
Female (A ₂ B ₂ male parent)	A ₁ A ₂ B ₁ B ₂	$-[r_Ar_B]/(1-r)$	$[(1-r_A)(1-r_B)]/(1-r)$
	A ₁ A ₂ B ₂ B ₂	$-[r_A(1-r_B)]/r$	$[(1-r_A)r_B]/r$
	A ₂ A ₂ B ₁ B ₂	$-[(1-r_A)r_B]/r$	$[r_A(1-r_B)]/r$
	A ₂ A ₂ B ₂ B ₂	$-[(1-r_A)(1-r_B)]/(1-r)$	$[r_Ar_B]/(1-r)$

The numerical values of the expectations of α_f , α_m and d were calculated for each flanking marker genotype using programs written in 'C'. For a given marker interval, a regression of the phenotypic value onto the numerical values of additive and dominance terms were carried out at putative QTL positions (e.g. 1cM intervals all along the chromosome). The regression calculations were made using Genstat 5.3 (Genstat 5 Committee, 1993). The estimates of additive and dominance genetic deviations were obtained at each putative position. A test statistic was used to obtain a likelihood curve for the QTL position along the chromosome.

The test statistic (TS) used has been shown to provide a very close approximation to the likelihood ratio test (Haley & Knott, 1992),

$$TS = n \log_e (RSS_{\text{reduced}} / RSS_{\text{full}}),$$

where n is the number of individuals in the analysis, RSS_{reduced} is the residual sum of squares (RSS) of the regression with no QTL parameters fitted, and RSS_{full} is the RSS with the QTL parameters fitted (Haley & Knott, 1992). This test statistic is asymptotically distributed as χ^2 with degrees of freedom (d.f.) equal to the d.f. full model minus d.f. reduced model. In the model described above, fitting the numerical values of the coefficients of a_m , a_f , and d , the number of d.f. is 3. The null hypothesis used in the analysis was that there was no QTL or variance explained by the marker interval tested. The likelihood ratio test can be converted to LOD scores by division by $2(\log_e 10)$.

A model for analysis of two QTL

As shown by Haley & Knott (1992) the model to test for the presence of one QTL, against the null hypothesis of no QTL present on a chromosome, can be extended to test for two QTL (with the null hypotheses of one QTL or no QTL). The two QTL model simply searches the chromosomes in two dimensions, testing every possible combination of positions for the two QTL. The significance of the two QTL model was determined from the change in the test statistic for two QTL vs. one or no QTL.

5.2.2 Autosomal multiple regression analysis

To allow evaluation of the comparative power of X-linked and autosomal data and methods of analysis, programmes were also written to calculate the expectations in terms of a and d genetic deviations for autosomal linked marker data. The coefficients of additive and dominance genetic deviations in terms of recombination fraction for each of the nine possible marker combinations are shown in Table 5.2 taken directly from Haley & Knott (1992).

Table 5.2. Expectations for the mean genotypic effects of an autosomal QTL for all possible flanking marker genotypes in an F_2 population.

Marker genotype	Coefficients of :	
	a (additive genetic deviation)	d (dominance genetic deviation)
$A_1A_1B_1B_1$	$[(1-r_A)^2(1-r_B)^2 - r_A^2 r_B^2]/(1-r)^2$	$[2r_B(1-r_B) r_B(1-r_B)]/(1-r)^2$
$A_1A_1 B_1B_2$	$[(1-r_A)^2 r_B(1-r_B) - r_A^2 r_B(1-r_B)]/r(1-r)$	$[r_A(1-r_A)(1-r_B)^2 + r_A(1-r_A) r_B^2]/r(1-r)$
$A_1A_1 B_2B_2$	$[(1-r_A)^2 r_B^2 - r_A^2(1-r_B)^2]/r^2$	$[2r_A(1-r_A) r_B(1-r_B)]/r^2$
$A_1A_2 B_1B_1$	$[r_A(1-r_A)(1-r_B)^2 - r_A(1-r_A) r_B^2]/r(1-r)$	$[(1-r_A)^2 r_B(1-r_B) + r_A^2 r_B(1-r_B)]/r(1-r)$
$A_1A_2 B_1B_2$	0	$[r_A^2 r_B^2 + r_A^2(1-r_B)^2 + (1-r_A)^2 r_B^2 + (1-r_A)^2(1-r_B)^2]/[r^2 + (1-r)^2]$
$A_1A_2 B_2B_2$	$[r_A(1-r_A)r_B^2 - r_A(1-r_A)(1-r_B)^2]/r(1-r)$	$[(1-r_A)^2 r_B(1-r_B) + r_A^2 r_B(1-r_B)]/r(1-r)$
$A_2A_2 B_1B_1$	$[r_A^2(1-r_B)^2 - (1-r_A)^2 r_B^2]/r^2$	$[2r_A(1-r_B) r_B(1-r_B)]/r^2$
$A_2A_2 B_1B_2$	$[r_A^2 r_B(1-r_A) - (1-r_A)^2 r_B(1-r_B)]/r(1-r)$	$[r_A(1-r_A)(1-r_B)^2 + r_A(1-r_A) r_B^2]/r(1-r)$
$A_2A_2 B_2B_2$	$[r_A^2 r_B^2 - (1-r_A)^2(1-r_B)^2]/(1-r)^2$	$2r_A(1-r_A) r_B(1-r_B)]/(1-r)^2$

The estimates of a and d , and calculation of the test statistic were carried out using the same methods as outlined in the previous section for X-linked analysis.

5.2.3 Simulated data sets

A number of different simulated data sets were generated to determine the effectiveness of the analytical methods described in section 5.2.1 and 5.2.2. The simulation programs were written in 'C', using the basic simulation protocol:

n F_1 parents were simulated with m marker loci with known map distances in Morgans. QTL(s) were simulated at xcM from the first marker on the chromosome, such that parents of the F_1 carried QTL alleles which differed by y phenotypic standard deviation units in the base population. The phenotypic trait values were generated given the individuals' QTL genotype and a random error term representing environmental (non-genetic) noise, such that the phenotypic measurements generated were normally distributed. The additive and dominance terms in the females were generated independently from the male additive terms. A reciprocal F_2 generation, e.g. Dam: $A_2A_2Q_2Q_2B_2B_2 \times$ Sire: $A_1Q_1B_1$, and Dam: $A_1A_1Q_1Q_1B_1B_1 \times$ Sire: $A_2Q_2B_2$, was derived from the simulated F_1 population, with recombination events in the F_1 females being generated at random with no crossover interference. The litter size simulated in the F_2 population was 10, and each sex was represented equally. In the simulations all parameters could be varied readily.

Simulated populations were also generated with autosomal linked QTL. These simulations were the same as the X-linked simulations in all respects, with the exception that additive and dominance terms for males and females were not generated independently and recombination events were generated in the male and female F_1 parents.

5.2.4 Thresholds and confidence intervals for QTL detection

As discussed in section 2.2.3, the calculation of thresholds for QTL detection is a complicated statistical issue. For the purpose of this study, the threshold used for QTL detection was set at a TS of 14, which is approximately equal to a LOD score of 3 (commonly used by human geneticists (Lander & Botstein, 1989)). The validity of the threshold used for QTL detection will be discussed in greater depth in section 5.4.2.

The confidence intervals for QTL detection were calculated using the one-LOD drop method (Lander & Botstein, 1989). This method estimates the confidence interval by a drop of 1 LOD score from the maximum LOD score obtained, which is approximately equal to a 95% CI for QTL position (Lander & Botstein, 1989). The one-LOD drop method is equivalent to a drop in test statistic of 4.6. Ideally, in a simulation study, empirical CI would be calculated. It was decided that as a result of the low number of replicated simulations analysed (50 replicates), calculation of empirical confidence intervals would be inappropriate and results would be misleading due to the small number of replicate simulations.

5.3 RESULTS

This section will present the results of simulated data analysed using the X-linked regression method (section 3.2.1) and the autosomal regression method (Haley & Knott, 1992). Unless otherwise stated, the results presented show the mean values of 50 replicated simulations. The simulations were restricted to 50 replicates, due to the computational time requirements of Genstat 5.3 (Genstat Committee, 1993).

The objective of the simulation studies carried out in this chapter was to check the validity of the X-linked multiple regression method in estimation of QTL position and effects, and explore the specific properties of the X-linked regression method compared to the autosomal regression method of Haley & Knott (1992).

5.3.1 The effects of population size on QTL detection

As outlined in section 2.2.4, population size affects the power of QTL detection, where (given that there is a QTL present), power represents the probability of detecting the QTL. Simulations were generated to show the effect of varying the population size on the power to detect a QTL with an additive effect of $a = \pm 0.25$ phenotypic standard deviations in both males and females, and no dominance deviation. A QTL of this size explains 4.7% of the phenotypic variance in an F_2 population using the relationship:

$$\text{X-linked } F_2 \text{ phenotypic Variation} = \frac{a_f^2}{4} + \frac{a_m^2}{2},$$

where $a_f^2/4$ is the additive variance contribution from the females, and $a_m^2/2$ is the additive variance contribution from the males. Simulations were carried out for population sizes of 200, 350, and 500 animals. An 80cM chromosome was simulated with marker spacing of 10cM, with the first marker at 0.0cM. The QTL was simulated at 35cM from the first marker.

Figure 5.1 shows the mean likelihood curve for QTL detection for the three population sizes simulated. The mean test statistic obtained from the simulation with 200 animals failed to reach the significance threshold of 14, as the 50 replicate simulations produced a maximum mean TS of 13.0. Using the population sizes 350 and 500, results of analysis show maximum mean TS of 17.1 and 25.7 respectively. As expected, the population size of 500 F_2 individuals shows a marked improvement in QTL detection over population sizes of 200 and 350.

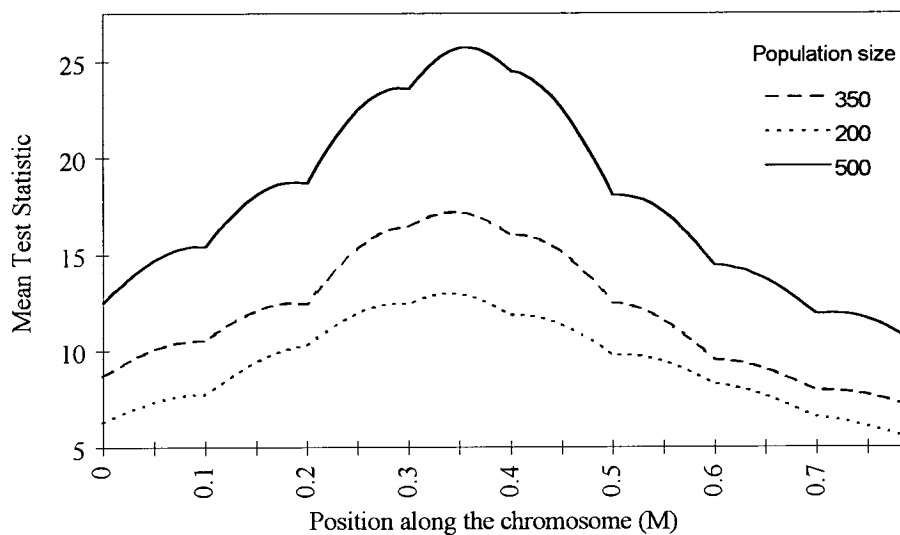


Figure 5.1. The effect of varying population size on the mean test statistic of 50 replicate simulations. The QTL was simulated at 35cM with additive effect $a=\pm 0.25$ in males and females, no dominance deviation was simulated.

Table 5.3. Mean results of parameter estimates over 50 replicate simulations with varying population size. The QTL was simulated at 35cM with additive effect of $\alpha=\pm 0.25$ in males and females, with no dominance deviation. Mean standard errors of estimates are shown in parenthesis.

Population size	Mean QTL position	Estimated 95% CI	QTL estimates:		
			a_m	a_f	d
200	34.0 cM	47 cM	0.27 (0.11)	0.26 (0.14)	-0.05 (0.21)
350	34.5 cM	35 cM	0.25 (0.08)	0.24 (0.11)	0.02 (0.16)
500	35.5 cM	24 cM	0.26 (0.06)	0.26 (0.09)	0.03 (0.13)

Estimated 95% CI represents the length of the CI for QTL position.

The parameter estimates obtained from simulations are shown in Table 5.3. The results show that the QTL position estimates were close to the simulated QTL position for all population sizes. The estimates of QTL size were also good over all population size simulations. The mean estimated QTL effects were not significantly different from the simulated QTL effects of $\alpha=\pm 0.25$ and $d=0$, and the standard errors associated with the estimated QTL effects are lower for males than females in all population sizes. Increasing the population size resulted in smaller estimated confidence intervals and lower standard errors associated with estimates. Using the threshold for QTL detection of a test statistic of 14, a population size of 200 enabled QTL to be detected in 44% of replicates, in a population size of 350, QTL were detected in 70% of replicates, and with a population size 500, QTL were detected in 92% of replicates. Investigation of the percentage of simulations in which QTL were detected (i.e. reaching the significant threshold of TS 14) shows clearly that increasing population size results in an increase in power to detect QTL.

5.3.2 QTL detection and the effect of marker spacing

The effect of marker spacing was investigated in a population size of 350, with a QTL of additive effect $a = \pm 0.25$ in males and females. The marker spacings investigated were 10cM, 20cM and 40cM with the first marker simulated at 0.0cM (i.e. for marker spacing 40cM one marker at either end of the chromosome and one in the middle). The QTL was simulated at 35cM from the first marker on the 80cM chromosome.

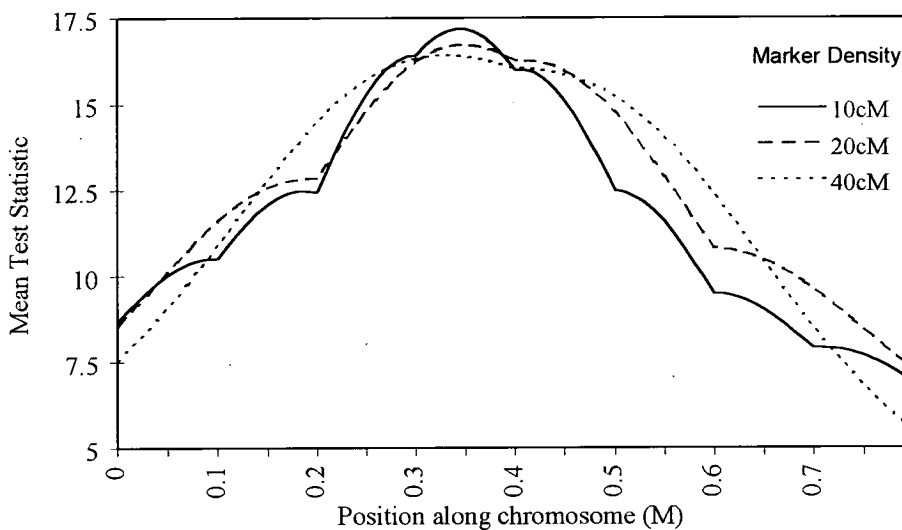


Figure 5.2. The effect of marker spacing on the mean TS obtained for 50 replicate simulations of population size of 350 with QTL of additive effect of $a = \pm 0.25$ phenotypic standard deviations, no dominance deviation was simulated.

The effect of varying marker densities on QTL detection are presented in Figure 5.2 and Table 5.4. The mean likelihood curves show that the mean TS and standard errors of estimates do not vary greatly over the three marker spacings for a QTL of $a = \pm 0.25$ phenotypic standard deviations. Examination of the percentage of replicate simulations analysed which reached the threshold for QTL detection shows at the marker density of 40cM, 60% of replicates reach the significance threshold of TS 14. At the marker spacings of 20cM and 10cM, QTL were detected in 66% and 70% of

replicates respectively. The mean estimated QTL position at the marker spacings of 10cM, 20cM and 40cM were close to the simulated QTL positions. The estimated confidence intervals for QTL positions rose dramatically with decreased marker density, the CIs estimated were 30cM, 43cM and 51cM for the marker densities 10cM, 20cM and 40cM respectively.

Table 5.4. Mean results of parameter estimates over 50 replicate simulations, with varying marker density. Mean standard errors shown in parenthesis.

Marker density	Mean QTL position	Estimated 95% CI	QTL estimates:		
			a_m	a_f	d
10cM	34.5 cM	30 cM	0.25 (0.08)	0.24 (0.11)	0.02 (0.16)
20cM	36.0 cM	43 cM	0.25 (0.08)	0.27 (0.11)	0.02 (0.16)
40cM	33.5 cM	51 cM	0.24 (0.08)	0.25 (0.11)	0.03 (0.16)

The results obtained from the analysis of different marker densities show little increase in the maximum TS with the increased marker density. The low variability in the maximum TS may have been influenced by the simulated QTL positions. In all simulations the QTL were at 35cM, and this position was only 5cM proximal to the nearest marker at 40cM. To investigate the effect of the vicinity of simulated QTL to the nearest marker, simulations were set up with different QTL positions.

5.3.3 The effect of QTL position in relation to marker position on the power of QTL detection

To investigate the effect of the simulated QTL position in relation to the marker position, an 80cM chromosome was simulated with 20cM marker spacing (the first marker was simulated at 0.0cM). A QTL of additive effect $\alpha=\pm 0.25$ phenotypic standard deviations was simulated at 30cM, 32.5cM, 35cM and 37.5cM (no dominance deviation was simulated).

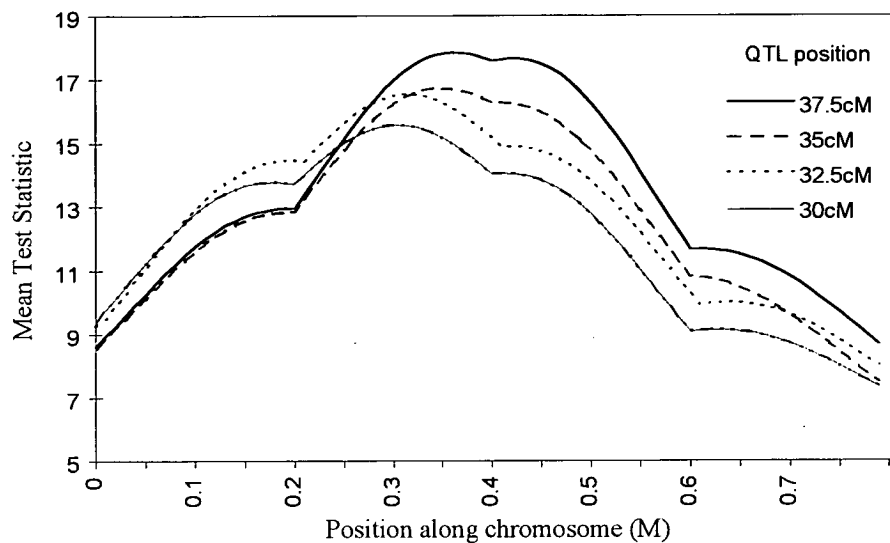


Figure 5.3. The effect of simulated QTL position in relation to marker loci positions on the mean test statistic obtained over 50 replicate simulations. The population size simulated was 350 animals, with marker spacing 20cM, and QTL effects of $a=\pm 0.25$ in males and females, and no dominance deviation was simulated.

Comparison of the mean likelihood curves show that the mean test statistic at the most likely QTL position increases as the QTL is simulated in closer linkage with the nearest marker (it should be noted that the QTL search was carried out at 1cM intervals). The mean estimated QTL positions were 37cM, 36cM, 32cM and 31cM for the simulated QTL positions of 37.5cM, 35cM, 32.5cM, and 30cM respectively. The increase in mean TS was associated with a modest increase in power. The percentage of simulations in which QTL analysis reached the significance threshold of TS 14, were 70% at simulated QTL position 37.5cM and 62% at simulated QTL position 30.0cM. The mean estimated QTL effects were not significantly different from those simulated at all simulated QTL positions.

5.3.4 Comparison of X-Linked and autosomal data analysis

An initial comparison of X-linked and autosomal data was carried out using simulated additive effects of equal size in both X-linked and autosomal data sets, and dominance deviations were set to zero. The X-linked and autosomal multiple regression methods were used as described in sections 5.2.1 and 5.2.2. It was noted that an additive effect of $a = \pm 0.25$ explained 4.7% of the phenotypic variance for an X-linked QTL, and 3.1% of the phenotypic variance in an F_2 generation with an autosomal linked QTL. To account for this difference in the comparison, 50 replicate simulations of 350 animals with an X-linked additive effect of $a = \pm 0.25$ and an autosomal additive effect of $a = \pm 0.306$ were simulated. These simulated QTL effects explain approximately 4.7% of the variance (averaged over the sexes) in both X-linked and autosomal data sets. These simulations were considered to be a more valid comparison of the power of X-linked and autosomal linked QTL data and analysis. The replicate simulations were made with marker spacings of 10cM (the first marker at 0.0cM) with the QTL simulated on an 80cM chromosome 35cM from the first marker.

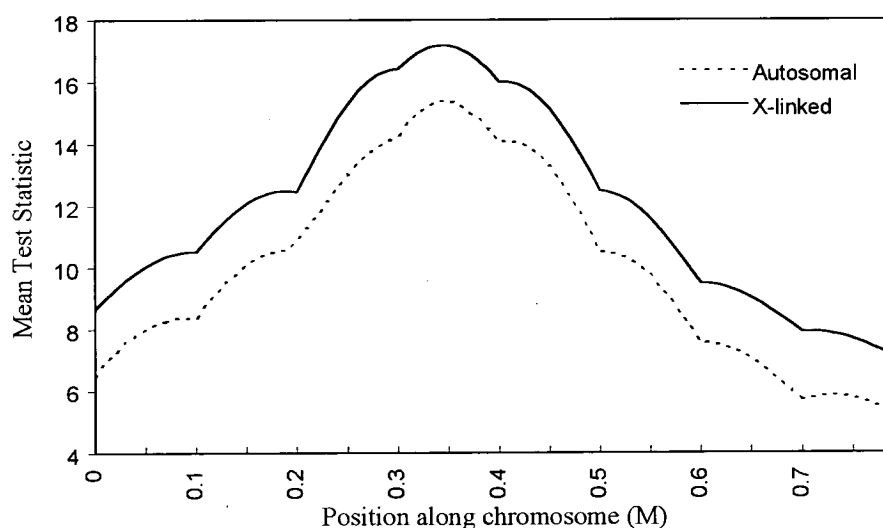


Figure 5.4 Comparison of the mean likelihood profile of X-linked and Autosomal data explaining equal phenotypic variance in the F_2 . The QTL was simulated at 35cM in a

population size of 350 animals. Results presented are the mean values over 50 replicate simulations.

Comparison of analyses of X-linked data and autosomal data with QTL simulated explaining 4.7% of the phenotypic variance in the F_2 population, show that X-linked data give higher mean test statistic over the 50 replicate analysis than the autosomal data (Figure 5.4). Investigation of the power of X-linked data compared to autosomal data showed the X-linked data to be more powerful for QTL detection than autosomal data with a simulated QTL explaining the same percentage of phenotypic variation in the F_2 population. Using the significant threshold of TS 14, analysis of X-linked data detected QTL in 70% of the replicates, whereas in autosomal data QTL were detected in 64% of replicates. Both X-linked and autosomal analyses provided estimates of QTL effects and positions close to the simulated QTL parameters (not significantly different from those simulated).

5.3.5 Investigation of different modes of inheritance

To investigate the relative power of X-linked QTL with additive, dominant, and recessive inheritance, simulations were set up to determine the relative power for these modes of inheritance. Fifty replicate simulations were analysed with a QTL simulated with effect of $a = \pm 0.25$ phenotypic standard deviations in a population size of 350 animals, and marker density of 10cM (the first marker simulated at 0.0cM). The QTL were simulated with additive, complete dominance, or complete recessive inheritance. Table 5.5 shows the expected means of the QTL effects in males and females.

Table 5.5. The X-linked QTL genotypes found in an F_2 population, with mean realised QTL effects for additive, dominant and recessive QTL.

Sex	QTL genotype	Mean effect		
		Additive ($d=0$)	Dominant ($d=+a$)	Recessive ($d=-a$)
Female	Q_1Q_1	$m + a$	$m + a$	$m + a$
	Q_2Q_1 and Q_1Q_2	$m + 0$	$m + a$	$m - a$
	Q_2Q_2	$m - a$	$m - a$	$m - a$
Male	Q_1	$m + a$	$m + a$	$m + a$
	Q_2	$m - a$	$m - a$	$m - a$

The results presented in Figure 5.5 and Table 5.6 show the relative performance for the three modes of inheritance. Mean parameter estimates presented in Table 5.6 show that all estimates were close to those simulated. The differences between mean estimated and simulated QTL effects were not significant, and the estimated QTL positions were also close to the simulated QTL position (35cM). The mean likelihood curves presented in Figure 5.5 illustrate that the results obtained from analysis of the dominant and recessive QTL were very similar, with both curves showing higher mean TS than in the additive QTL analysis. The relative power of QTL detection in population with dominant and recessive QTL was greater than those with additive QTL, with 78% of the replicates with dominant or recessive QTL reaching the significant threshold of TS 14, compared to 70% of the replicates simulated with additive QTL.

Table 5.6. Mean results of parameter estimates over 50 replicate simulations, with simulations set up to investigate the power of X-linked analysis. Mean standard errors shown in parenthesis.

Marker density	Mean QTL position	QTL estimates:		
		a_m	a_f	d
Additive	34.5 cM	0.25 (0.08)	0.24 (0.11)	0.02 (0.16)
Dominant	36cM	0.26 (0.08)	0.25 (0.11)	0.21 (0.16)
Recessive	36cM	0.25 (0.08)	0.25 (0.11)	-0.26 (0.16)

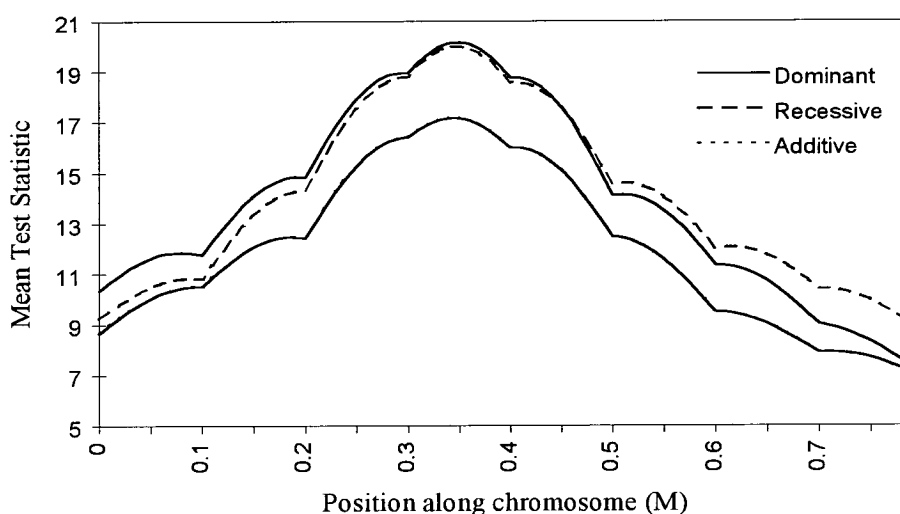


Figure 5.5 The effect of QTL action on the mean test statistic obtained from X-linked regression analysis. Comparison of additive, dominant and recessive QTL inheritance.

5.3.5 Analysis of two QTL

To investigate the performance of the model for analysis of two QTL, a single simulation of a 120cM chromosome with 20cM marker spacing was made (the first marker being simulated at 0.0cM). Two QTL were simulated at 30cM and 90cM both with an effect of $\alpha=\pm 0.5$ phenotypic standard deviations. The two QTL were

simulated such that the parental lines carried the QTL loci in association (i.e. one inbred line carried both decreasing alleles, and the other carried both increasing alleles). The likelihood curve obtained from the analysis using the model fitting one QTL (Figure 5.6a) shows multiple peaks with no clear position for the QTL. The model fitting two QTL is shown in Figure 5.6b; the likelihood surface enables the two simulated QTL to be distinguished. In both models the null hypothesis of no QTL on the chromosome was used. The estimated QTL positions were the same as the positions simulated, and the estimated QTL effects obtained at the most likely QTL positions were not significantly different from the simulated QTL effects. Estimates obtained were:

30cM	$a_m = 0.56 (0.08)$	$a_f = 0.68 (0.16)$	$d = 0.13 (0.18)$
90cM	$a_m = 0.67 (0.09)$	$a_f = 0.22 (0.17)$	$d = 0.18 (0.18)$

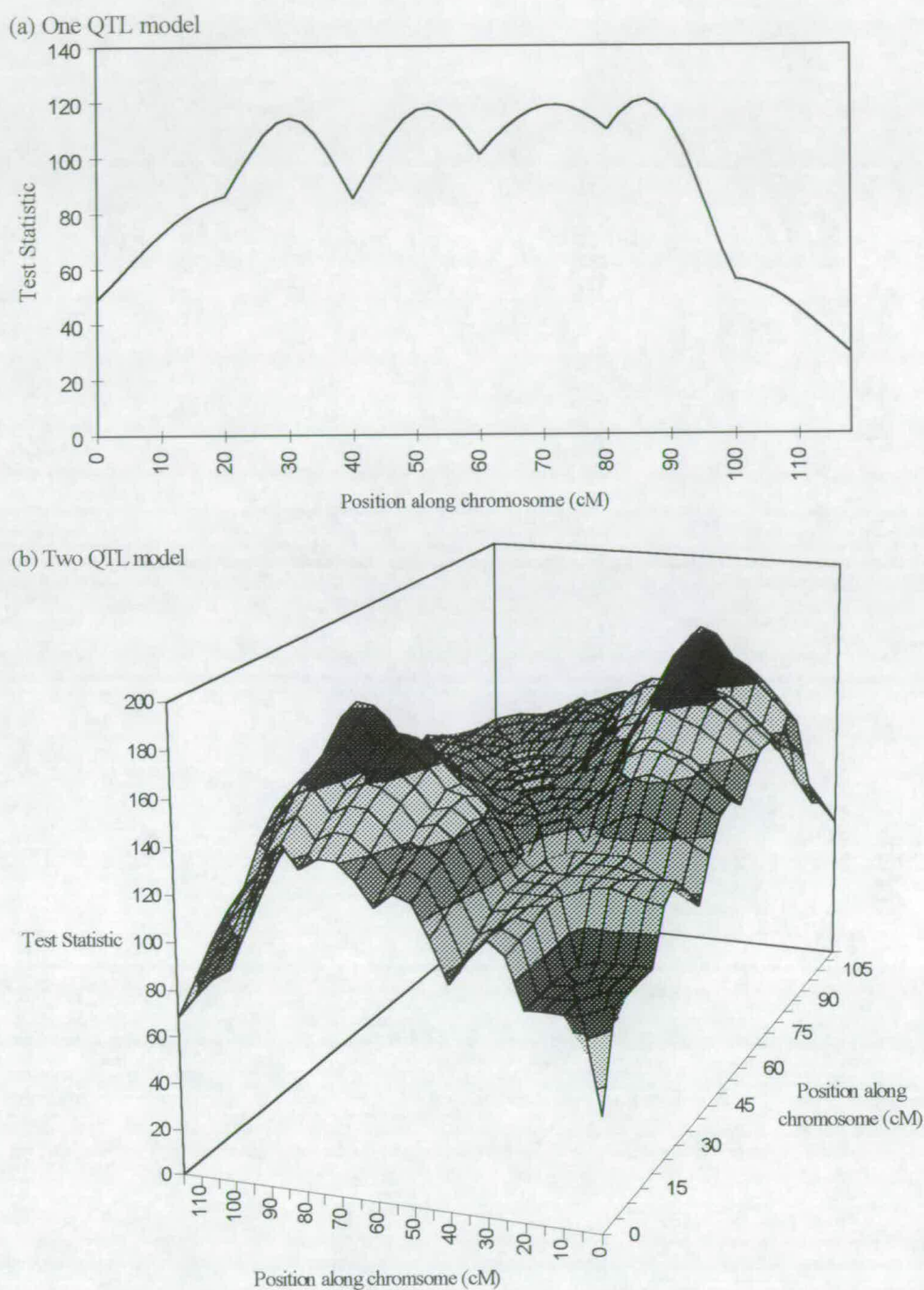


Figure 5.6. Likelihood profiles obtained from analysing two simulated QTL using models fitting one QTL (Fig. a) and two QTL (Fig. b). The QTL were simulated at 30cM and 90cM both with additive effect of $a=\pm 0.5$ phenotypic standard deviations, on a 120cM chromosome with dominance deviation. The results presented are of one simulation.

5.4 DISCUSSION

5.4.1 Effects of Population Size and Marker Spacing on QTL detection

The effect of population size on the ability to detect QTL was clearly illustrated by the results of simulations outlined in section 5.3.1. Greater population sizes give greater power to detect QTL; a higher percentage of significant QTL detection over replicates, and show a higher mean TS. The larger populations give better estimates for QTL parameters (lower s.e.) and lower estimated confidence intervals for QTL position.

Simulation studies by van Ooijen (1992) showed that for an autosomal QTL explaining 5% of the total variance, a population size of at least 400 was needed to result in reliable QTL detection ($P < 0.20$). For the purpose of this study, QTL of similar contribution to the total phenotypic variance were simulated, explaining only 4.7% of the phenotypic variance of the F_2 generation. Over the 50 replicates a population size of 500 enabled QTL to be detected in 92% of simulations. With the population size of 200, the detection rate was only 44%. These results support observations that population size has a dramatic effect on QTL detection (Lander & Botstein, 1989; van Ooijen, 1992; Darvasi *et al.*, 1993; Darvasi & Soller, 1994). Large population sizes are required to reliably detect QTL with small percentage contribution to total variance in the F_2 generation. For a QTL explaining 4.7% of the phenotypic variation, a large population size (e.g. 350-500 individuals) would be required to reliably detect the QTL. QTL with larger effect can be detected reliably in small population sizes, as a result of the greater contribution of the QTL to the total F_2 phenotypic variance (this was not illustrated in this study).

The results obtained from simulations with different marker densities showed that the mean TS was not dramatically affected by different marker densities, with the maximum TS varying between 19 and 16 over all marker densities. However, the

power of QTL detection was affected, with only a 60% detection rate at 40cM marker density, rising to 68% and 70% at 20cM and 10cM respectively. Simulations with a 40cM marker density showed greater variation in maximum TS (data not shown), resulting in a high mean TS, but poor overall QTL detection rate.

The QTL position was accurately estimated at 40cM, 20cM and 10cM map densities. As stated in the results section, for all simulated map densities the QTL was simulated 5cM proximal to the nearest marker at 40cM. The results of the simulations outlined in section 5.3.3, where QTL were simulated at 2.5cM, 5cM, 7.5cM and 10cM proximal to the nearest marker at 40cM (with marker density of 20cM), showed the power of the QTL analysis was increased as the QTL was simulated in tighter linkage with the nearest marker. These results indicate that the relative power of different marker densities was influenced by the position of the QTL with respect to the nearest marker. This suggests that the power of QTL detection at the three marker densities investigated would be reduced to a greater degree at the wider marker spacing if the QTL were simulated in the middle of the marker intervals at all marker densities investigated.

Darvasi & Soller (1994) suggested that increasing population size will increase power for QTL detection more efficiently compared to increasing marker densities in most experiments. As a result of this finding, they suggest the use of wide marker spacings for initial studies to detect marker QTL linkage. These authors proposed that the optimal marker spacing, in economic terms, is a function of the ratio (c) between cost of genotyping an individual at a single marker, and the cost of raising an individual for an F_2 design. Darvasi & Soller (1994) suggest that for mouse QTL mapping, the cost of rearing and trait scoring an individual is 100 times the cost of scoring an individual marker ($c=0.01$). Simulation studies showed that for a backcross population with a QTL of 0.5 S.D units, with a type I error rate of 0.05 the optimal marker spacing for $c=0.01$ was approximately 30cM for a population size of 700 individuals. The results obtained in this study also indicate that genotyping animals at wide marker spacing would be effective at determining linkage between QTL and regions of the genome

when the QTL is of large effect. When linkage was detected (or when suggestive linkage was detected (e.g. $P < 0.20$), more markers could then be genotyped to increase the TS and narrow the CI for QTL position.

The simulation studies carried out in this chapter provide evidence to support the findings of Darvasi & Soller (1994). Table 5.5 shows the total number of genotypes required per simulated data set, and the power of different population sizes and marker densities simulated in this study (power representing the total number of simulated data sets analysed which reached the significance threshold of 14, expressed as a percentage of the total simulations analysed).

Table 5.7. The mean power and genotyping requirements for simulations generated in this study.

Population size	Marker density	Power (%)	Number of genotypes
200	10cM	44	1800
350	10cM	70	3150
500	10cM	92	4500
350	20cM	68	1750
350	40cM	60	1050

Comparison of the simulated population sizes of 350 show that the increase in power between marker densities of 40cM and 20cM was 8%, and this required an additional 700 genotypes. Between the marker densities of 20cM and 10cM the increase in power was 2%, which require an additional 1400 genotypes. Examination of the results obtained for varying the population size showed that an increase in power of 26% and 22% was obtained between the population sizes of 200 and 350, and 350 and 500 respectively, and the corresponding change in number of genotypes was 1350 in both cases. Although we did not observe the changes in QTL detection rate at different marker densities for population sizes of 200 and 500 F_2 individuals, the results suggest that the increase in power for different the population sizes was

dramatic for the corresponding changes in genotyping requirements, when compared to the relationship observed in simulated data sets with different marker densities. The conclusions drawn from this study regarding the relationships between marker densities and population size, and the question of the most effective means to maximise power, rely on the assumption that the relationship between power and number of markers investigated follow the same trends within population size.

The findings of Darvasi & Soller (1994), and to a limited extent the findings from this study, suggest the use of low marker densities in large populations is most effective for the initial scanning of the genome for marker-QTL detection. When significant marker-QTL linkage is detected additional makers can be added to refine the QTL position estimate. Clearly, the marker spacing used initially will be dependent on population size and the size of the QTL effect.

The results of simulations to compare additive, dominant and recessive QTL action are as expected. Table 5.5 shows the QTL classes in males and females and the QTL effect associated with each genotype for the additive, dominant and recessive QTL. If we consider a QTL of additive effect, little information will come from the female heterozygote QTL classes (Q_2Q_1 and Q_1Q_2) as the estimate of $d=0$ and will not contribute to the overall genetic variance. Therefore much of the information in the females will come from the homozygous QTL genotypes Q_1Q_1 and Q_2Q_2 . In males both QTL classes contribute to the overall power, regardless of the QTL action. The greater power of the recessive and dominant QTL occurs due to additional information from the heterozygote females, where $d= \pm a$ (increasing the overall genetic variance explained by females as the dominance genetic variance is greater than zero). The extra information results in the increased (and equal) power for the dominant and recessive QTL detection.

5.4.2 X-linked data and analysis: specific properties compared to autosomal data and analysis

The results in which simulated QTL explain equal F_2 phenotypic variance in X-linked and autosomal data showed that the analysis of X-linked data is more powerful for the detection of QTL, producing a higher percentage of QTL detected over replicates and a higher mean TS. Figure 5.6 shows the results of 50 replicates of analysis using simulated X-linked data with a QTL with additive effect of $a=\pm 1.0$ phenotypic standard deviations at 35cM. The population size in the 50 replicate simulations was 350 animals with a marker density of 10cM (the first marker simulated at 0.0cM). The top curve shows results of combined analysis of F_2 males and females, and the bottom curves show the contribution of the sexes separately.

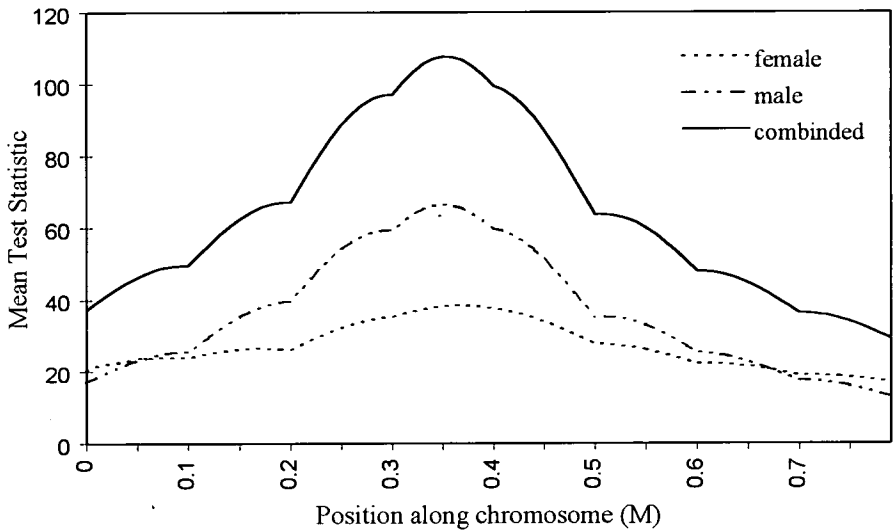


Figure 5.7 Illustration of the contribution of male and female F_2 individuals to the total power to detect X-linked QTL.

The results presented in Figure 5.7 illustrates that when a QTL of equal effect in males and females is simulated, the males contribute greater power to QTL detection. There are two factors which may result in the difference in power observed between the sexes:

1. When a QTL is simulated with equal additive effect in males and females, the variance explained by the QTL is greater in males than females.
2. The number of QTL classes, and therefore mixing components that have to be estimated, are different between the sexes. In males there are only two QTL classes, in females there are three. It should be noted that within litter there are two female classes as only heterozygotes and homozygotes genotypes can occur in the F_2 generation.

To explore the contribution of the two specific components of X-linked QTL data outlined above, a simulation study was set up with QTL explaining equal variance in the males and females. A QTL was simulated explaining 4.7% of the phenotypic variance in males and females ($a_m = \pm 0.217$ and $a_f = \pm 0.306$). The QTL was simulated at 35cM with marker density of 10cM in a population size of 350 animals, and 50 replicates were analysed. The results obtained were compared to simulated population carrying QTL explaining equal phenotypic variance, either simulated for autosomal QTL or X-linked QTL with sex averaged additive effect (see section 5.3.4).

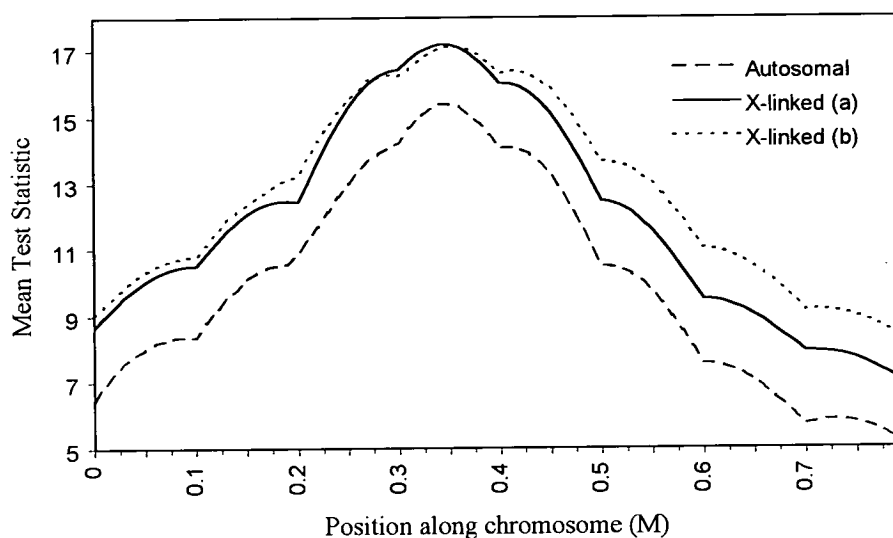


Figure 5.8 Comparison of the mean likelihood profile of X-linked and Autosomal QTL. The QTL were simulated at 35cM in a population size of 350 animals explaining equal phenotypic variance in the F_2 population. The X-linked data set (a) had the QTL simulated with equal additive effect in males and females, and the data set (b) had the QTL simulated with equal variance explained by the QTL in the males and females. No dominance deviation was simulated.

The mean likelihood curves presented in Figure 5.8 show that the two X-linked data sets, X-linked (a) with simulated QTL explaining 4.7% variance averaged over sexes, and X-linked (b) with simulated QTL explaining 4.7% variance in males and females produce equal maximum TS. This indicates that the data provide equal power regardless of the distribution of the variance explained between the sexes. The results obtained from these simulations suggest that the additional power of the X-linked data analysis is likely to be a result of the number of mixing components which are estimated in X-linked analysis. This relationship has been used by van Ooijen (1992) to explain the greater power for QTL detection observed in first generation backcross populations compared to F_2 populations of equal size with QTL explaining equal variance (see Table 2.1). Analysis of simulated data with X-linked QTL explaining equal variance in males and females, where the contribution of the males and females

were separated, showed both sexes contribute equally to the power to the analysis (data not shown).

The standard errors of estimated additive effects are lower for males compared to females in all simulations. This occurs as a result of the greater information in males available to estimated additive effects compared to females (see Table 5.5).

Thresholds for autosomal and X-linked data sets

It may be argued that the use of a common threshold for autosomal and X-linked data sets was not valid. The first argument suggests the use of a higher threshold for X-linked QTL detection, as the X-linked method uses a greater number of degrees of freedom compared to the autosomal analysis. The threshold for X-linked analysis uses 4 degrees of freedom (3 for the QTL effects and 1 for the QTL position) compared to 3 d.f. in the autosomal linked QTL (where male and female expectations of mean genotypic effects are not distinguished in the analysis). This argument suggests a higher threshold should be used in X-linked QTL analysis. Simulation studies comparing separate estimation of parameters in males and females for a QTL of additive effect and joint estimation, do not show a significant difference in TS obtained, where the likelihood curves for the two analysis were indistinguishable (data not shown). The only advantage in analysing male and female additive effects separately was the ability to estimate different additive effect in males and females. Therefore, analysis of these parameters separately (using an additional d.f.) was not essential. Estimating the pooled additive effect for males and females would result in an increase in power due to a reduction of the number of d.f. estimated in the threshold used.

A converse argument can be made to set a lower threshold for QTL detection on the X-chromosome. When analysing autosomal data, the threshold for QTL detection considers a genome wide scan for QTL, regardless of the number of chromosomes mapped (19 autosomes in the mouse). The threshold for a genome wide scan is used as a result of there being no prior information to predict the location of the QTL.

When mapping QTL on the X-chromosome, information from phenotypic observations can provide evidence for X-linked QTL (Hastings, 1990; Hagger & Stranzinger, 1992), and this leads to a situation where the threshold used should only consider one chromosome. Lander & Kruglyak (1995) suggest that a reduction in TS threshold may be countenanced where strong prior evidence exists to restrict the search to one chromosome.

As previously discussed, the formulation of thresholds for QTL detection is a complex issue, and beyond the scope for full consideration in this project. The use of a common QTL threshold for both data sets almost certainly provides a conservative threshold for X-linked QTL, where there is joint estimation of additive parameters.

5.4.3 Performance of the two QTL Model

The X-linked multiple regression analysis for mapping QTL was effective in distinguishing between two QTL at the large QTL effects simulated. This agrees with results obtained using two QTL model in the autosomal multiple regression method (Haley & Knott, 1992). The results shown in Figure 5.6b, indicate that when QTL were simulated at 30cM and 90cM, it was possible to estimate QTL positions and effects using the two QTL model (simulations with QTL positioned with closer map positions were not investigated). Simulation studies by Haley & Knott (1992) were used to investigate if the properties of the two QTL model, where two QTL were simulated in association in inbred lines (i.e. both increasing QTL alleles carried by one line and both decreasing alleles carried by the other inbred line) and two QTL were simulated in dispersion in the inbred lines (where each line carried one increasing and one decreasing QTL allele). The results showed that QTL mapping in these data sets using the one QTL model resulted in the mean TS being inflated in the association model and deflated in the dispersion model, and it was concluded that this was due to covariances between the linked QTL. Analysis using the two QTL model showed a marked improvement in fit of the model compared to the one QTL model when QTL were simulated 50cM apart. The estimates of the QTL parameters obtained were very

good for the two simulated QTL in association and dispersion. Further simulations showed QTL which were simulated 20cM apart were difficult to separate, and consequently the parameters were poorly estimated. The simulation study indicated that the two QTL model works effectively at separating two QTL on a chromosome when there is sufficient recombination between the two loci to break down the linkage. Using the results obtained by Haley & Knott (1992) when closely linked QTL are analysed, the multiple regression method will not be effective at estimating positions and estimates.

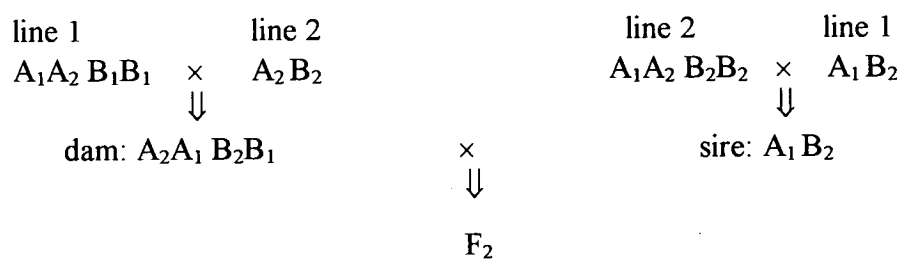
5.4.4 General discussion and extension of method for non-inbred populations.

The results of simulations shown in this chapter have illustrated that the X-linked regression analysis works as expected in the parameter set investigated, providing mean estimates of QTL effects and positions close to those simulated. The simulated data sets analysed indicated that the power of analysis varies with the population size and marker density in the same way as autosomal data and analysis, although the X-linked data have been shown to provide greater power to detect QTL than autosomal linked QTL data explaining an equal percentage of the total phenotypic variance in the F_2 generation.

The analytical method described in this chapter is effective at estimating QTL effects and position for simulated populations derived from a cross between two inbred lines. However, the marker frequencies presented in Chapter 4 indicate that the mouse lines investigated in this study show segregation at marker loci within line. As a result of this finding, the analytical method described above was adapted to enable analysis of an F_2 population where there was considerable segregation at marker loci within lines.

Consider the following example of a reciprocal F_1 population derived from a cross between two lines where two marker alleles were segregating at marker loci within lines. The two lines differ considerably for the quantitative trait under investigation,

and it was assumed that the two lines were fixed for different QTL alleles. All individuals in the pedigree were genotyped at flanking markers (A and B):



The above example shows one possible combination of marker alleles in the pedigree of an F_2 family derived from a reciprocal F_1 mating. In the cross illustrated there was segregation at marker alleles in the two lines (type 1 marker alleles were predominant in line 1, and type 2 marker alleles were predominant in line 2). The dam has inherited type 2 marker alleles from its sire, and type 1 marker alleles from its dam (the probability of type 2 marker allele being inherited from its dam was equal to the probability of the inheritance of type 1 marker allele). This results in the dam being fully informative at both markers A and B, where the origin of each marker can be traced unequivocally back to line 1 or line 2. Examination of marker loci inheritance shows that the sire has inherited both type 1 and type 2 marker alleles from its dam. Despite the inheritance of the marker allele A_1 from its dam, the sire was also fully informative at both loci. The sire is fully informative as males inherit only their X-chromosome from their dam. Therefore, all progeny in the reciprocal F_2 generation are fully informative at both marker loci.

As a consequence of the mode of inheritance of the X-chromosome, the analysis of a non-inbred F_2 population was carried out using a similar method to the X-linked multiple regression analysis for inbred populations. Informative marker loci in the F_2 population were determined by the F_1 dam marker genotypes. For a given marker interval, informative flanking markers were determined, and coefficients of a_m , a_f and d calculated for the interval. When individuals were not informative for markers at the extreme ends of the chromosome, a single marker analysis was used to provide information at all positions on the chromosome to be used in all individuals. Hence,

information, either by single marker analysis (at chromosome ends) or by flanking marker analysis, is available at all positions on the X-chromosome (between the extreme X-linked markers) for all individuals in the F_2 generation.

The non-inbred regression analysis effectively combines all information present to give an estimate of QTL position and effects. In non-inbred data sets, F_2 individuals effectively have marker data at different marker spacings; as markers become uninformative, flanking marker spacings become wider. The power to detect QTL and accuracy of estimated positions, as for data from inbred populations, will be dependent on the population size, QTL effect, and marker spacing. The effect of marker spacing on power, accuracy of QTL estimates, and position will be a function not only of the number of loci genotyped in the population, but also the level of information at markers scored, i.e. how much uninformative data was present.

CHAPTER 6

Mapping quantitative trait loci for body weight on the X-chromosome using the X-linked multiple regression method

6.1 INTRODUCTION

The results presented in Chapter 4 show strong support for the hypothesis of a QTL, of major effect on body weight, positioned on the X-chromosome in the P6 mouse selection lines. Parental DNA was not available for the F_2 population described in Chapter 4, which meant that fully informative marker loci could not be determined to allow QTL mapping analysis. To allow fully informative marker data to be established, a further F_2 experiment was set up with the objective of mapping the putative X-linked QTL of large effect. This population is referred to as the F_2X to allow the two F_2 populations to be distinguished.

The F_2X experiment will be presented in this chapter. The analysis of the F_2X data set is carried out using the non-inbred X-linked multiple regression method (Chapter 5), with the objective of estimating the number of X-linked QTL of large effect (see section 2.2.6), the size of their effects and their position(s).

6.2 MATERIALS AND METHODS

6.2.1 The F₂X Population

A reciprocal F₁ population was set up between the P6 high selection line at generation 52 of selection and the P6 inbred low line at generation 7 of inbreeding (established at generation 45 of selection). The aim of utilising an inbred low line was to reduce segregation at markers within the base population lines. An inbred high line was not available, due to difficulties encountered with generating such an inbred line.

Fourteen F₁ matings were set up, with a total of seven families in each reciprocal half (female high (H) × male low (L), and female L × male H). Due to the poor reproductive performance in the F₁ matings, the F₁ individuals used as parents of the F₂ generation were sourced from a total of only 7 families (5 families from low inbred females and 2 from high selection females). A reciprocal F₂ population was bred comprising of 18 families (for the mating structure see Table 6.1). Both F₁ reciprocal halves were represented equally in the F₂ generation. Of the 18 F₂ matings, 17 were successful and were maintained for up to three parities, yielding a total of 340 F₂X individuals surviving to 10 weeks. A summary of the litter data is shown in Table 6.1. All mating pairs and weaned animals were maintained as outlined in section 3.3. Spleen samples were taken from all individuals in the pedigree for use as a source of DNA for marker analysis.

Table 6.1 Summary of the mating structure of the F_2X population, and the distribution of the animals born per litter.

Litter	♣ Dam	♣ Sire	Numbers born		
			Parity 1	Parity 2	Parity 3
1	HL 1.1	LH 6.1	9	11	12
2	HL 1.2	LH 4.1	10	12	13
3	HL 1.3	LH 3.1	8	4	8
4	HL 1.4	LH 7.1	9	8	-
5	LH 3.1	HL 1.1	4	7	7
6	LH 4.1	HL 1.2	8	8	9
7	LH 3.2	HL 1.3	9	5	12
8	LH 6.1	HL 1.4	10	6	10
9	LH 6.2	HL 1.5	9	8	-
10	HL 2.1	LH 7.2	11	14	9
11	HL 2.2	LH 3.2	12	9	-
12	HL 2.3	LH 5.1	7	7	9
13	LH 5.1	HL 1.5	11	10	-
14	LH 3.3	HL 2.1	10	11	13
15	LH 4.2	HL 2.1	8	14	-
17	HL 2.4	LH 5.1	11	11	-
18	LH 6.3	HL 1.3	8	-	-

♣ HL= high selection line dam and inbred low line sire, LH = inbred low line dam and high selection line sire. The identification numbers show the F_1 mating pair and then the individual number (within sex), e.g. 1.2 is individual 2 from the F_1 litter 2.

Records of 6 week weight were taken on all F_1 individuals. The body weight measurements recorded in the F_2X generation were 3 week weight, 6 week weight, and 10 week weight. At 10 weeks of age all F_2X generation animals were killed by cervical dislocation, spleens removed, and placed in labelled screw top tubes and snap frozen in liquid nitrogen. The gonadal fat pads (GFP) were removed from all F_2X

males, weighed and then replaced in the body cavity. These were used to produce estimates of carcass fat percentage (section 2.3.2) according to,

$$\text{GFP Fat percentage} = [(\text{GFP/dead weight}) \times 8] \times 100.$$

All F_2X animals were weighed after removal of the spleen and frozen individually in polythene bags with identification cards. To produce a second estimate of the carcass fat percentage, all individuals were freeze dried, and the freeze dried weights were used to estimate carcass fat percentage as outlined in section 3.4. Full records (genotypic and phenotypic data) were available on a total of 334 animals.

6.2.2 Genotyping the F_2X population

Genomic DNA was extracted from spleen samples taken from all individuals in the pedigree by phenol:chloroform extraction as described in section 3.1. All individuals in the pedigree were genotyped using Protocol 2 (section 3.2) at the seven microsatellite loci found to be polymorphic in the F_2 population described in Chapter 4. In addition to these microsatellite loci, additional polymorphic marker sites in the F_2X population were identified using the high selection line and inbred low line parents of the F_1 generation. A total of 61 marker loci were investigated (including markers previously investigated in Chapter 4) yielding six additional polymorphic marker loci. The thirteen microsatellite loci found to be segregating in the F_2X population and their estimated PCR product sizes are listed in Table 6.6. To allow fully informative loci within litters to be identified, all individuals in the pedigree were genotyped at the polymorphic marker loci. All genotypes were scored on two separate occasions to minimise incorrect genotypes as described in section 3.2.

6.2.3 Data Analysis

Sex linked effects in the reciprocal F₁ generation

The mean body weight of males and females in the reciprocal halves of the F₁ generation were compared to establish if the sex linked effect observed initially in the P6 selection line by Hastings (1990) could be detected in the F₁ generation of the F₂X population. Data on 6 week body weight of the F₁X individuals was analysed using the model described in section 3.5 to test for the sex linked body weight effect (see sections 2.3.4 and 4.3.1).

Description of trait data

There were a number of fixed and environmental effects recorded during the collection of the data on the F₂X population. To determine the importance in all models of individual terms and interaction terms for the phenotypic traits measured, the likelihood ratio statistic was used (see model evaluation, section 3.5). The following model was obtained using the method referenced above,

$$Y_{ijkl} = \mu + S_i + D_{jk} + \beta.(B \text{ or } W)_{jk} + e_{ijkl}, \quad (1)$$

where:

Y_{ijkl} = the observations on the l th individual

μ = overall mean

S_i = fixed effect of the i th sex

D_{jk} = random common environmental effect associated with the k th parity litter nested within the j th dam

β = the linear regression coefficient of Y on the number born (B_{jk}) or weaned (W_{jk}) from the k th parity litter of the j th dam

e_{ijkl} = the residual error associated with the l th individual

When regressing the trait observations on the number born or weaned, the number born was significant ($P < 0.05$) for the analysis of 3 week weight, and for body weight gain from 3-6 weeks and 3-10 weeks. The number weaned per litter was found to be significant ($P < 0.05$) in the analysis of estimated fat percentage, and the remaining body weight models.

To investigate whether the freeze drying batch (see section 3.4) had a significant effect on the freeze dried carcass weight, a fixed effect of batch was fitted in the model analysing freeze dried weights. Batch was not found to be significant in the analysis and was dropped from the model. The fixed effect of parity was not significant ($P > 0.05$) in the analysis of the phenotypic measurements described above.

To illustrate the distribution of the data, residual phenotypic trait values (trait values corrected for all fixed effects and covariates in the model described above) were obtained using Genstat 5.3 (Genstat 5 Committee, 1993). Estimation of residual phenotypic correlations were carried out using Minitab (Ryan, Joiner & Ryan, 1985).

Regression analysis to estimate QTL effects

Evidence from the analysis of the F_2 population described in Chapter 4 showed an association between markers on the X-chromosome and body weight at 6 and 10 weeks. To estimate the number of X-linked QTL of large effect, their position(s), and the size of their effects, marker data and phenotypic data from the F_2X population were analysed using the non-inbred X-linked multiple regression method (Chapter 5). The test statistic (TS) used in the analysis was:

$$TS = n \log_e (RSS_{\text{reduced}} / RSS_{\text{full}}),$$

where n is the number of individuals in the analysis, RSS_{reduced} is the residual sum of squared (RSS) of the regression with no QTL parameters fitted, and RSS_{full} is the RSS with the QTL parameters fitted (Haley & Knott, 1992). The reduced models used in the analysis were those outlined above. The full models had the additional fitted effects of the numerical values of the coefficients of a_m , a_f and d calculated using

the terms given in Table 5.1 (section 5.2.1). The marker map positions used in the regression analysis of the X-linked data were estimated from F_2X marker data, using CRIMAP.

The null hypothesis (H_0) used in the analysis was that there were no QTL or variance for the trait analysed on the X-chromosome. Previous analyses (section 2.3.4 and Chapter 4) have indicated the X-chromosome carried loci with large effect for body weight, which implies the H_0 used may be inappropriate. The most applicable alternative would be to set the null hypothesis that the variance explained by the X-chromosome was due to a large number of loci of small effect. This hypothesis could be tested by fitting all marker on the chromosome as fixed effects in the reduced model, i.e. testing for a single QTL of large effect vs. variance explained by multiple loci on the X-chromosome. This model cannot be fitted for the F_2X data set, as non-informative marker loci result in too much non-informative (missing) data.

The significance threshold for QTL detection was obtained using the method previously described by Andersson *et al.* (1994) (see section 2.2.3). An empirical threshold was estimated from 1000 replicated analyses of simulated trait data in which no QTL was segregating (the trait data were simulated from a normal distribution with mean of zero and variance of one) and the pedigree and marker genotypes from the F_2X population. By ordering the maximum TS obtained from each replicate analysis and taking the 99, 95 and 90 percentiles, the thresholds for $P<0.01$, $P<0.05$ and $P<0.10$ are obtained under the null hypothesis of no QTL segregating in the population. The estimated empirical thresholds for the one QTL model were: TS 15.5 for $P<0.01$, 11.4 for $P<0.05$ and 9.7 for $P<0.10$. The analysis was limited to 1000 replicates due to the computation time required by Genstat 3.5 (Genstat 5 Committee, 1993).

The confidence intervals for the QTL position were estimated using the one-LOD drop method (Lander & Botstein, 1989), where the 95% confidence interval for QTL position is obtained by a drop in LOD score of one from the maximum LOD score

detected. The one-LOD drop method is equivalent to a drop in the test statistic used in this study of 4.6. An estimated 99% confidence interval can be obtained by a drop in LOD score of 2 (a drop in the test statistic of 9.2).

The choice of the null hypothesis tested, the threshold used to determine significant QTL detection, and the method used to calculate the confidence intervals for the QTL positions will be discussed in greater depth later in this chapter.

For the analysis using the X-linked multiple regression method, the data must be normally distributed when the hypothesis is tested. Residuals obtained from fitting the fixed effects and covariates in the model outlined above were tested for fit to a normal distribution in Genstat 5.3 (Genstat 5 Committee, 1993), but no significant deviation from normality was detected.

6.3 RESULTS

6.3.1 Sex linked effects in the reciprocal F₁ generation

Mean body weights of males and females in the reciprocal halves of the F₁ generation were compared to establish if the significant sex-linked effect observed in previous generations (section 2.3.4 and Chapter 4) could be found in the F₁ generation of the F₂X population (Table 6.2).

Table 6.2 Mean body weight (g) at 6 weeks of individuals in the reciprocal halves of F1 line crosses. The dam is shown first in crossing nomenclature

Sex	L × H	s.e.	No.	H × L	s.e.	No.
female	16.7	1.4	12	27.2	1.9	8
male	20.0	1.2	19	33.7	2.0	7

s.e. = standard error of estimate.

There are significant differences in body weights between the males and females within the two reciprocal halves of the F₁ population ($P < 0.05$ using student t-test). In the low line dam reciprocal half, the estimated mean body weight in males was 3.3g greater than females, and in the high line dam reciprocal half the difference in mean body weight between the sexes was 6.5g. Previous data (Hastings, 1990; Hastings & Veerkamp, 1993; Chapter 4) have indicated that there is a large X-linked effect. The X-linked effect can be estimated using the mean body weights of the heterozygous females and hemizygous males in the two reciprocal halves of the F₁ generation,

$$(H \times L: A - Aa) - (L \times H: a - Aa),$$

where H×L are individuals from the high line dam reciprocal half, L×H are individuals from the low line dam reciprocal half, Aa are the F₁ females, and a and A are the F₁ males. Using this relationship the estimated X-linked effect was 3.2g at 6 week weight.

The progeny from the high line dams have significantly higher body weights than the contemporaneous individuals with low line dams, suggesting a large maternal effect influencing the body weight at 6 weeks.

6.3.2 Polymorphic markers in the F_2X pedigree

To identify polymorphic markers in the F_2X population a total of 61 microsatellite markers were tested (including those investigated in the analysis of the F_2 generation described in Chapter 4). Of the markers investigated, thirteen were found to be polymorphic and to produce consistently scoreable PCR products. Of these thirteen markers, three were identified to be fully informative in the F_2X population (*DXMit187*, *DXMit50*, and *DXMit13*) where marker alleles could be traced unequivocally to the high selection line or the inbred low line founder parents in all F_2X individuals.

All animals (with full records) in the F_2X pedigree were genotyped at all thirteen polymorphic markers. Table 6.3 summarises the distribution of fully informative markers in the F_2X population. The percentage information at each marker is given (determined by the percentage of individuals in the F_2X generation where marker information could be traced unequivocally to the base population), with the exception of the fully informative marker loci *DXMit187*, *DXMit50*, and *DXMit13*. The results show that the percentage of fully informative individuals in the F_2X generation ranges from approximately 16% to 54%, with the exception of the three fully informative loci. At 11 of the 13 markers genotyped in the F_2X population, there were only two alleles segregating in the F_2X generation. At markers *DXMit55* and *DXMit187* there were three alleles segregating. The pedigree information at marker *DXMit187* (fully informative in the F_2X population) showed unequivocally one allele present in the high selection line, and two different alleles present in the inbred low line. The marker loci *DXMit121* and *DXMit31* were both fully informative in the same litters, and this may be because the two loci were closely linked (Table 6.6).

Table 6.3 The distribution of individuals with full records (genotypic and phenotypic data) in the F_2X population for segregating *DXMit* marker loci. Ticks denote litters which were fully informative at the marker loci (the three fully informative marker loci in all litters are not shown).

The F_2X generation				<i>DXMit</i> marker loci								
Litter	Sire X-chr.	No. females	No. males	55	46	25	62	16	64	79	38	31 & 121
1	L	8	14		✓	✓	✓			✓	✓	
2	L	13	13		✓	✓	✓			✓	✓	
3	L	10	7		✓	✓	✓	✓			✓	✓
4	L	1	4		✓	✓	✓	✓			✓	✓
5	H	6	8		✓				✓			
6	H	11	11		✓				✓			
7	H	14	14		✓				✓			
8	H	11	9	✓		✓	✓	✓				
9	H	7	8	✓		✓		✓				
10	L	15	13	✓					✓	✓	✓	
11	L	13	8	✓					✓			
12	L	12	3									
13	H	19	11	✓		✓	✓	✓			✓	✓
14	H	14	6		✓				✓	✓	✓	
15	H	9	16		✓				✓			
17	L	11	7									
18	H	3	5					✓				
Number of informative individuals (total 334)				84	179	135	120	95	158	96	148	52
Percentage of informative individuals				25.1	53.6	40.4	35.9	28.4	47.3	28.7	44.3	15.6

Sire X-chr = line origin of the sire X-chromosome, high (H) or low (L) line

6.3.3 Description of the F₂X data set.

Mortality and Litter Effects

A total of 402 animals were born in the F₂X generation. After adjustment to a maximum litter size of 12, there were 395 pups from a total of 43 litters weaned. The mean litter size at birth ($\pm$ S.D.) was 9.4 ± 2.4 , and the mean number of offspring weaned was 8.9 ± 2.2 .

The percentage mortality varied over the growth periods measured. From birth to weaning at 3 weeks, mortality was 3.0%, from weaning to 6 weeks it was 9.4%, and from 6 to 10 weeks the mortality was only 0.6%. The mortality rates support the observation of a high level of illness in the population (particularly diarrhoea post weaning), it should be noted that normal mortality rates are negligible (Rooney, pers.com. 1996). A two sample t-test was used to investigate if body weight at weaning was a significant factor in the post weaning mortality, i.e. if smaller animals were significantly more vulnerable to disease than their contemporaries. Comparison of mean body weight ($\pm$ s.e.) at 3 weeks of those not surviving the post weaning period ($9.6\text{g} \pm 0.3$) to the group surviving to 6 weeks ($10.1\text{g} \pm 0.1$), showed the mean body weights in the group not surviving to 6 weeks were 0.5g lower than the surviving group, but this difference was not significant ($P=0.11$).

Marker data were used to further investigate the contribution of the X-linked QTL for body weight and its association with the rates of mortality observed. If the X-linked QTL for body weight had a significant effect on the mortality in the F₂X generation, the segregation ratio of markers linked to the QTL may deviate significantly from the expected 1:1 ratio. Allele frequencies in the F₂X population were investigated at the three fully informative marker loci. Using a χ^2 test (1 d.f.) there were no significant deviations from the expected 1:1 ratio of the high and low line alleles (all male data and female homozygote genotype data) at the marker loci examined (*DXMit187* 121:133, *DXMit50* 124:136, and *DXMit13* 132:122).

Phenotypic data

Full records, genotypic and phenotypic data, were available on a total of 334 animals in the F_2X generation. This section will describe the phenotypic data obtained for these animals.

Table 6.4 Description of phenotypic data obtained in the F_2X generation. The estimates shown are the least square means for body weights (g), body weight gains (g), and estimated fat percentage.

Trait	Male ($n = 157$)			Female ($n = 177$)		
	mean	standard deviation	CV (%)	mean	standard deviation	CV (%)
3 week weight	10.36	1.57	15.2	10.11	1.51	14.9
6 week weight	25.09	3.44	13.7	21.03	2.81	13.4
10 week weight	31.54	4.00	12.7	25.89	3.19	12.3
gain 3 to 6 weeks	14.58	2.58	17.7	10.87	1.96	18.0
gain 3 to 10 weeks	21.06	3.17	15.1	15.87	2.85	17.0
gain 6 to 10 weeks	6.52	2.21	33.9	4.96	1.89	38.1
Fat % at 10 weeks	12.4	4.07	32.8	8.2	2.86	34.9
GFP % at 10 weeks	13.3	4.40	33.1			

CV = coefficient of variation. Fat % calculated using freeze dried weight with the formula in section 3.4. GFP % fat percentage in males estimated from gonadal fat pads (GFP) using the formula in section 6.2.1. n = number of animals in sample.

Table 6.4 shows that the mean male body weights at all ages were higher than the female body weights, and significantly so at 6 and 10 weeks ($P < 0.05$). The coefficients of variation (CV) associated with body weight at 3 weeks were high, indicating a high variance associated with the mean of this trait. Within both sexes the body weight gains over the middle growth period (3-6 weeks) were greater than in late growth period (6-10 weeks). The CV associated with body weight gain for the 6-10 week period were approximately 34% and 38% in males and females respectively.

The estimates of the total carcass fat percentage shown in Table 6.4 indicate the mean carcass fat percentage estimated from GFP was slightly higher than estimated from the freeze dried weights. High CV (approximately 33-35%) were associated with all estimates of carcass fat percentage. The correlation between fat percentage calculated using freeze dried weight in males (Fat%) and fat percentage calculated using gondal fat pad weight was $r=0.89$, indicating that both methods provide similar estimates of carcass fat percentage. The estimate of mean carcass fat percentage at 10 weeks calculated from freeze dried weights was higher in the males than the females.

Table 6.5 The residual phenotypic correlations of traits measured and derived traits in the F₂X population.

Traits	Traits:					
	3 wk wt	6 wk wt	10 wk wt	gn 3-6	gn 3-10	gn 6-10
6 wk wt	0.676					
10 wk wt	0.575	0.842				
gn 3-6 wk	0.268	0.883	0.739			
gn 3-10 wk	0.210	0.684	0.914	0.766		
gn 6-10 wk	0.013	0.007	0.541	-0.002	0.640	
Fat % at 10 wk	-0.053	-0.119	-0.025	-0.106	0.001	0.146

wt=weight, wk=week measured, gn=gain between weeks measured, Fat% estimated using freeze dried weights.

To investigate the degree to which the different traits measured were correlated, residual phenotypic correlations were calculated for all traits (residuals obtained from fitting the fixed effects and covariates in the model described in section 6.2.3). The residual phenotypic correlations presented in Table 6.5 show negative correlations between estimated carcass fat percentage at 10 weeks and body weight measurements at all ages. When body weight measurements were divided into growth periods, the correlation was negative between residual fat percentage and residual body weight

gain between 3 and 6 weeks, but became positive for late growth between 6 and 10 weeks.

6.3.4 QTL mapping using the X-linked regression analysis

In this section the results of analysis using the non-inbred X-linked multiple regression method (Chapter 5) carried out to estimate QTL position and effects will be presented. The map positions for the markers genotyped in the F_2X population, estimated using CRIMAP, are shown in Table 6.6. The table also shows published map positions obtained from the Mouse Genome Informatics (1996). The marker positions estimated from the F_2X data show the same order as the Mouse Genome Database (MGD) and the Chromosome Committee, and produce a total map length of a similar size to the published maps. The published map distances show a high degree of variation for estimated map positions at markers. These difference occur as a result of sampling error associated with the independent studies, and also the number of recombination events observed in the different mouse strains investigated. The map distances used in the regression analysis were those obtained from the F_2X population using CRIMAP, as the map distances estimated from the marker data were estimated in the same mouse strain. The estimated PCR product sizes of the alleles predominant in the high or low line are also listed.

Table 6.6 The map positions of markers estimated from the F_2X population using CRIMAP, and taken from the Mouse Genome Informatics (1996). Estimated PCR product sizes are given for the marker alleles predominant in the high or low line.

Marker	✦PCR product size (base pairs)		F_2X	MGD	Chr.Comm	MIT
	High	Low				
<i>DXMit55</i>	142 159	123	0.0	1.0	1.5	0.0
<i>DXMit187</i>	132	116 120	12.2	5.0	5.0	5.0
<i>DXMit50</i>	182	175	19.8	12.5	11.0	22.5
<i>DXMit46</i>	151	146	30.4	20.0	24.5	29.6
<i>DXMit25</i>	184	170	33.7	28.0	27.8	32.0
<i>DXMit62</i>	127	138	34.6	30.0	34.5	34.2
<i>DXMit16</i>	119	98	41.9	36.0	36.0	38.7
<i>DXMit64</i>	125	139	48.9	38.0	36.5	41.0
<i>DXMit79</i>	155	145	52.6	44.0	44.5	47.9
<i>DXMit38</i>	135	144	56.0	58.0	48.0	43.7
<i>DXMit13</i>	212	206	60.9	60.0	48.0	54.7
<i>DXMit121</i>	153	142	71.4	66.0	60.0	61.4
<i>DXMit31</i>	124	111	72.0	70.0	70.0	70.5

MGD = Mouse Genome Database, MIT = Massachusetts Institute of Technology, Chr.Comm = Chromosome Committee.

✦ The PCR product sizes were estimated from marker data.

QTL for body weight on the X-chromosome

To estimate linkage between marker loci and QTL for body weight at 3, 6 and 10 weeks, data were analysed using the X-linked multiple regression method and the results are shown in Figure 6.1a,b,c. The likelihood curves for all of the body weight measurements indicate the QTL detected were highly significant ($P < 0.01$), reaching

maximum test statistics (TS) of 36.1, 86.2 and 112.4 for 3, 6 and 10 week weight data respectively. The maximum TS were found at 23cM for all body weight measurements, with all the likelihood curves showing a smooth single peak (see Figure 6.1a,b,c). Using the equivalent of a one LOD-drop off to obtain a confidence interval for QTL position (a drop of 4.6 in the TS), the confidence intervals (CI) obtained were 13cM, 10cM, and 8cM for 3 week, 6 week and 10 week weight respectively. A summary of the estimates are shown in Table 6.7.

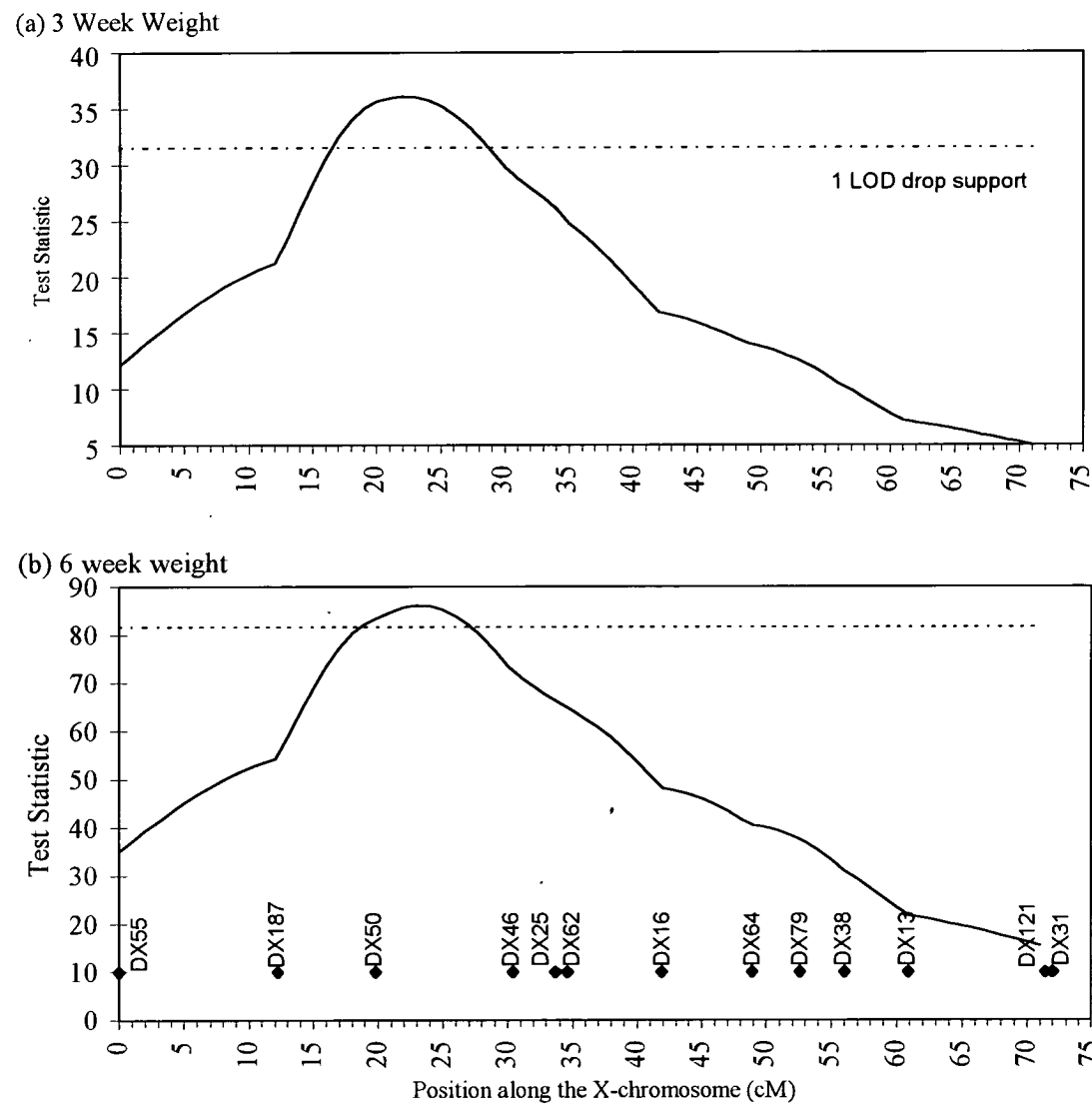


Figure 6.1 The likelihood profiles for body weights at (a) 3 week and (b) 6 week, obtained using the X-linked multiple regression method (Chapter 5).

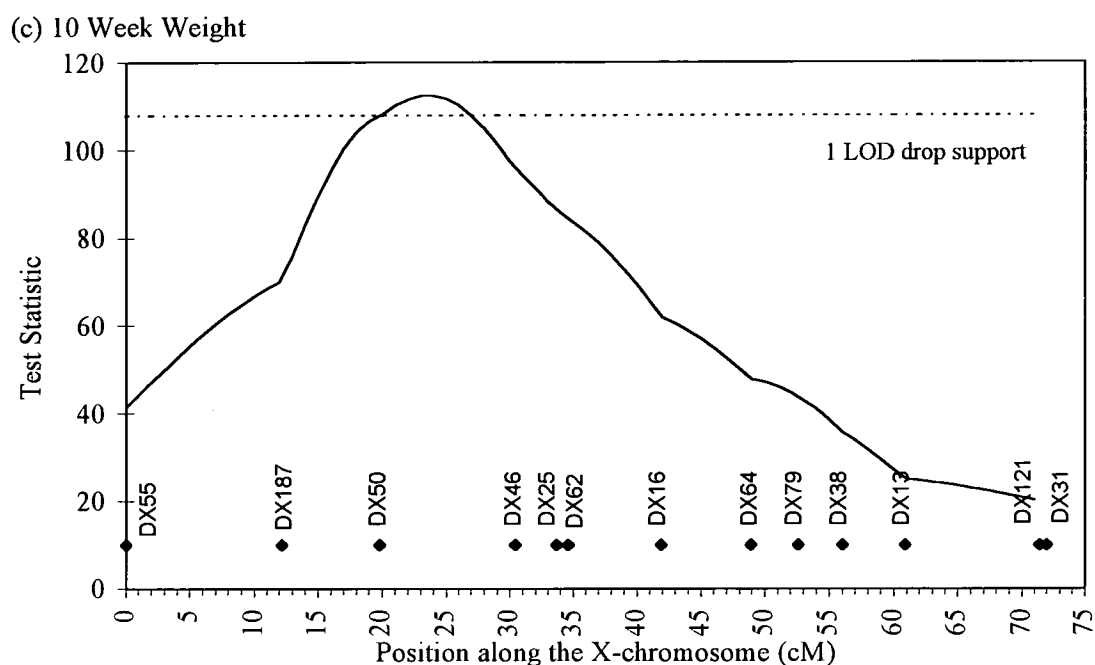


Figure 6.1c The likelihood profiles for 10 week weights.

Table 6.7 Summary of QTL effects (g) for the body weights obtained using X-linked multiple regression method. Standard errors of estimates are shown in parenthesis.

Body weights	Max TS	Male	Female		♦Percent F ₂ variance
		additive	additive	dominance	
3 week	36.1	0.69 (0.12)	0.39 (0.22)	0.10 (0.27)	7.7
6 week	86.2	2.00 (0.24)	1.83 (0.42)	0.38 (0.53)	14.2
10 week	112.4	2.63 (0.27)	2.62 (0.48)	0.37 (0.60)	27.9

♦ The percentage F₂ variance explained by the QTL was calculated at the most likely QTL position as the reduction in the residual variance of the F₂ population by the inclusion of the QTL effects in the model.

The estimated additive effects in males (half the difference between the hemizygotes) were greater than the additive effect in the females (half the difference between the

homozygotes) at 3 and 6 weeks. The difference between the male and female additive effects were not significant ($P < 0.05$ using the student t-test) for all body weights measured. In males and females the estimated additive effects for 10 week weight were both 2.6g, which corresponds to a total difference of 5.2g between the two homozygote female genotypes and the two hemizygote male genotypes. Smaller standard errors were associated with the male estimates of additive effects compared to the female estimates of additive effects, due to the greater information regarding the additive QTL effect coming from the hemizygous males (section 5.4.2). The dominance deviations estimated in females were positive (i.e. increasing body weight) but not significantly different from zero for the body weight traits.

Table 6.8 The estimated QTL effects (see Table 6.7) for the three body weight measurements expressed as percentages of mean body weight (see Table 6.4).

Age at measurement	Male		Female		
	mean weight (g)	additive %	mean weight (g)	additive %	dominance %
3 week	10.4	6.7	10.1	3.9	1.0
6 week	25.1	7.8	21.0	8.7	1.8
10 week	31.5	8.3	25.9	10.1	1.4

The estimated QTL effects expressed as a percentage of mean body weights (Table 6.8) give an indication of the percentage contribution of the QTL at the three body weights measured. The percentage effect of the QTL in males ranges between approximately 7% and 8% of the corresponding mean body weight, increasing slightly with age. In females, the percentage of the body weight explained by the QTL were higher than those obtained in the males, at 6 and 10 weeks. At 3 weeks the estimate of the additive QTL effect in females was much lower than in males, explaining approximately 4% of the mean body weight.

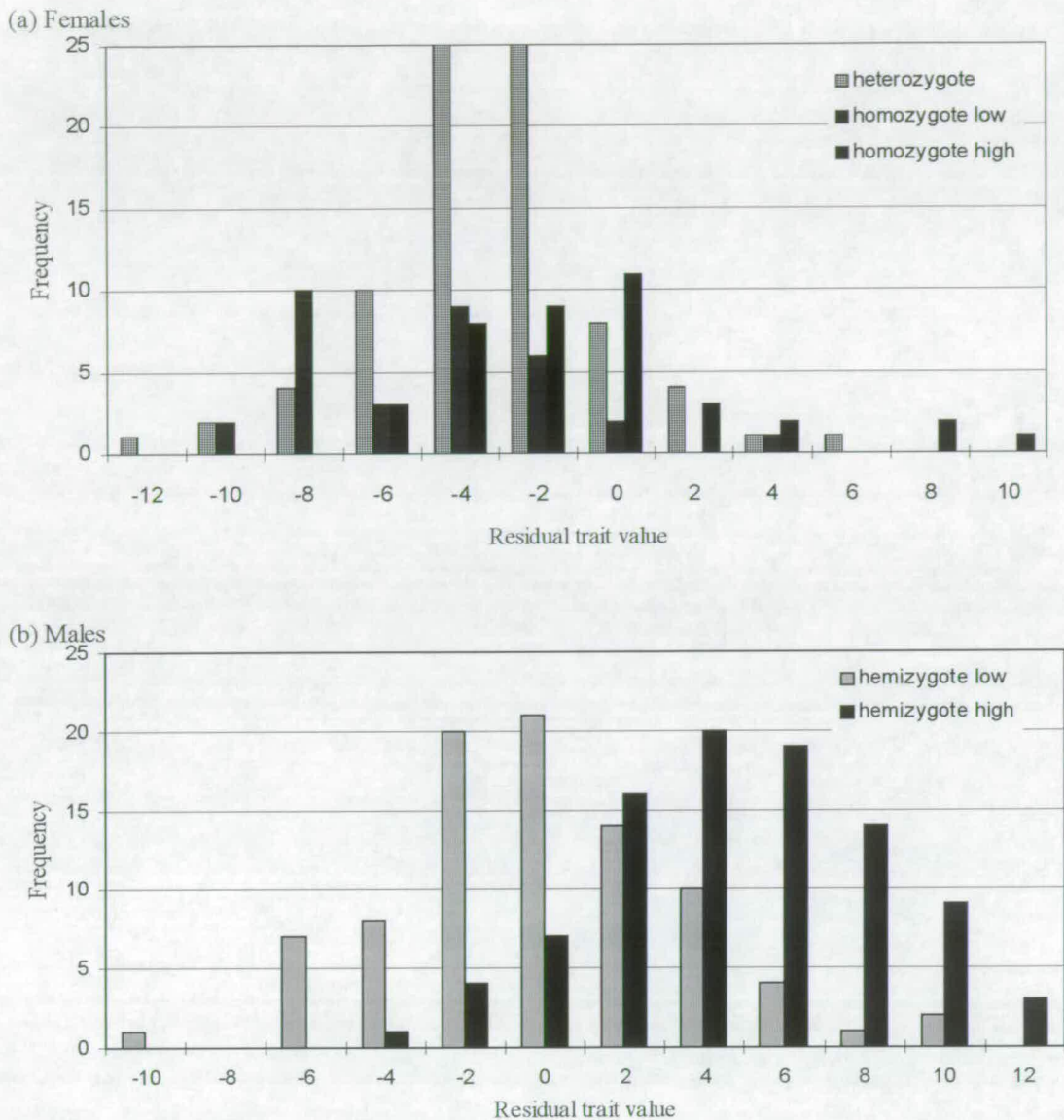


Figure 6.2 The distribution of the residual trait value (g) for 10 week weight in the three female, and two male marker classes of the marker *DXMit50*. (a) F_2 females, and (b) F_2 males.

The results from the QTL analysis of body weights at all ages using the X-linked multiple regression method indicate that the marker *DXMit50* was estimated to be in close linkage with the QTL for body weight. Since the marker *DXMit50* was fully informative in the F_2X population, the marker genotypes can be traced unequivocally

to the high or low line in all individuals. The distribution of residual 10 week weight within the marker classes of *DXMit50* is illustrated in Figure 6.2. In the female data set (Figure 6.2a), the heterozygotes have intermediate marker associated QTL effects compared to the two homozygote genotypes. This clearly illustrates the predominantly additive action of the marker associated QTL effect. For the two marker classes found in males, the distribution of the marker associated QTL effect shows a difference in the mean (Figure 6.2b). Assuming complete linkage between marker *DXMit50* and the QTL, the two QTL genotypes cannot be distinguished visually by the residual phenotypic measurements alone, as the sum of the two distributions illustrated in Figure 6.2b give a unimodal distribution (figure not shown). The marker associated QTL effects at *DXMit50* for 10 week weight were estimated by single marker analysis using the model described in section 6.2.3 with the additional fixed effect of marker genotype analysed within sex (Table 6.10). The estimated effects ($\pm$ s.e.) indicate the additive effect in males ($2.43\text{g} \pm 0.52$) were greater than in females ($2.24\text{g} \pm 0.77$), although the difference was not significant ($P < 0.05$); again the dominance deviations in females were not significantly different from zero.

The X-linked QTL and body weight gain

To determine if the X-linked QTL for body weight influenced body weight predominantly in one growth period compared to the other growth periods measured, body weight gains were analysed using the X-linked regression method. The results of the analysis of body weight gain from 3-6 weeks, 3-10 weeks and 6-10 weeks using the non-inbred X-linked multiple regression method are shown in Figure 6.3a, b, and c respectively. The likelihood curves for 3-6 week and 3-10 week weight gain are both smooth with a single peak at 21cM and 23cM respectively. The peak test statistics found for the analysis were 52.8 for the 3-6 week body weight gain, and 76.7 for the 3-10 week weight gain. Using a drop in TS of 4.6, the CIs obtained for the QTL positions were 9cM and 11cM for 3-6 week, and 3-10 week body weight gains respectively.

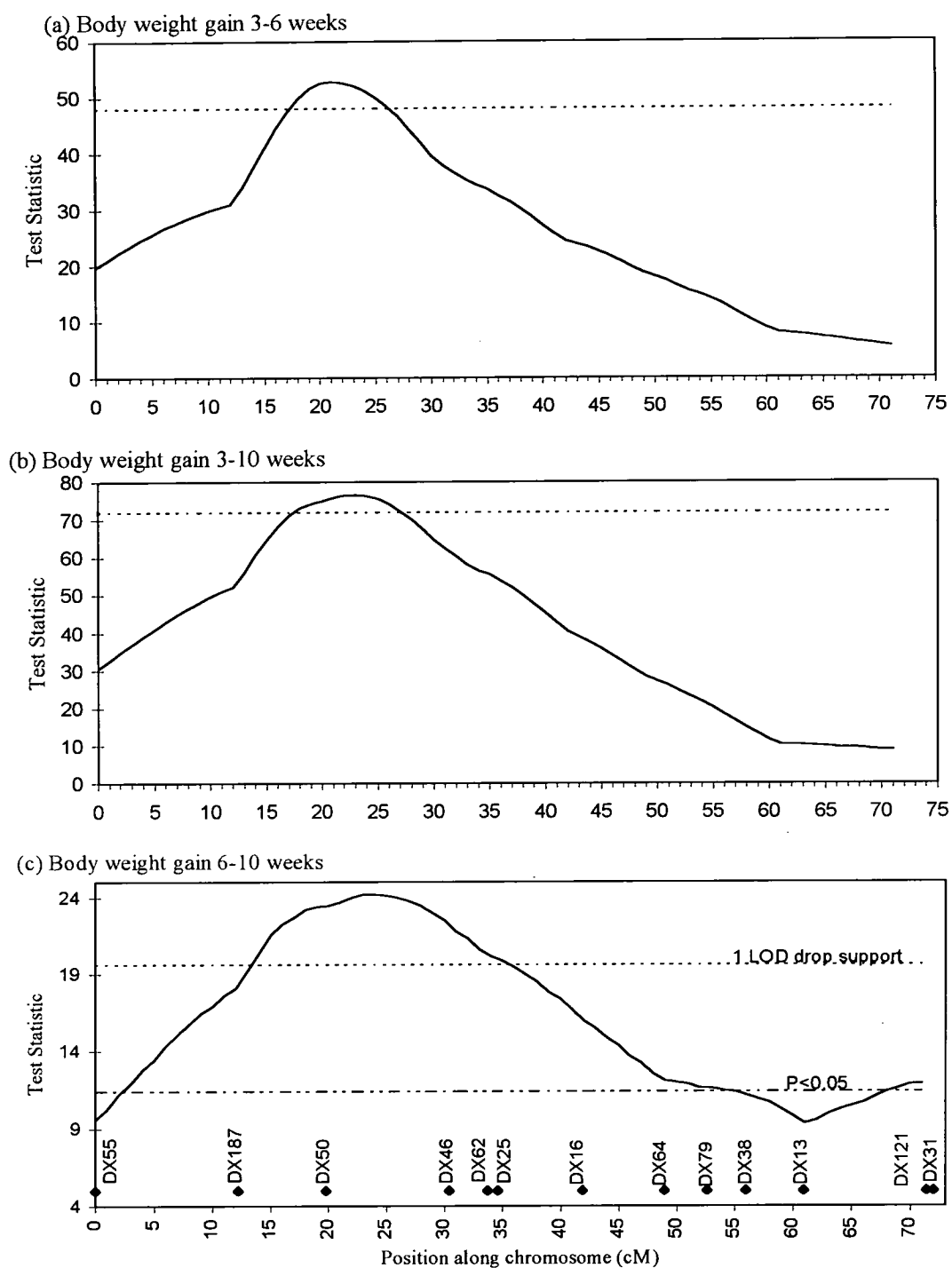


Figure 6.3 The likelihood profiles for the analysis of body weight gains between growth measurements taken at, (a) 3-6 weeks, (b) 3-10 weeks, and (c) 6-10 weeks, using the X-linked multiple regression method.

In the QTL analysis of body weight gain between 6-10 weeks the maximum TS reached was 24.1 at 23cM, although a rise in TS was seen towards the distal end of the X-chromosome. The second peak reached the threshold of $P < 0.05$, suggesting the presence of a second QTL on the X-chromosome. An analysis of the data fitting a two QTL vs. one QTL model did not show a significant change in the TS at a position other than 23cM, using the estimated empirical threshold of TS 19.4 ($P < 0.05$) and TS 16.8 ($P < 0.10$). This threshold has 4 degrees of freedom (d.f.) (8 d.f. for the full model [6 for the QTL effects for the two QTL, plus 2 for the two QTL positions] minus 4 d.f. for the reduced model fitting one QTL). The estimated empirical thresholds for the two QTL model vs. one QTL model were obtained by 500 replicated analyses as described in sections 6.2.3. To reduce computation time, the numbers of replicates were limited to 500 and the two dimensional chromosome search (see section 5.3.5) was carried out at 5cM intervals. The likelihood surface obtained from the QTL analysis using a model fitting two QTL vs. one QTL is shown in Figure 6.4. The QTL effects associated with the two maxima were:

23cM Maximum TS = 22.9 $a_m = 0.66$ (0.17) $a_f = 0.82$ (0.31) $d = 0.55$ (0.41)

71cM Maximum TS = 13.0 $a_m = 0.13$ (0.19) $a_f = -0.19$ (0.34) $d = -1.51$ (0.89)

The QTL effects at both loci have high standard errors. Only the estimates of the additive effects at the QTL positioned at 23cM were significantly different from zero ($P < 0.05$ using student t-test).

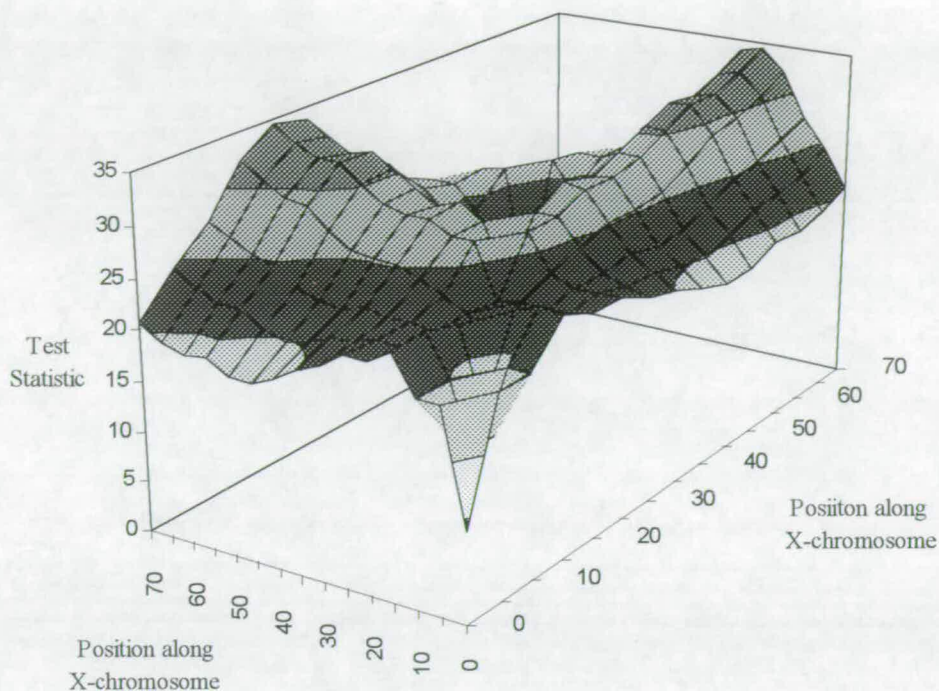


Figure 6.4 Analysis of body weight gain from 6-10 weeks using the model testing two QTL vs. one QTL.

Table 6.9 Summary of the QTL effects (g) for the body weight gains obtained by the X-linked multiple regression method. The estimates of QTL effects expressed as percentages of mean body weight gains (see Table 6.4). Standard errors of the estimates are shown in parenthesis.

Gains	Male		Female			
	additive	% of mean	additive	% of mean	dominance	% of mean
3-6 week	1.59 (0.24)	10.9	1.20 (0.44)	9.3	0.54 (0.54)	4.2
3-10 week	2.25 (0.29)	10.6	2.25 (0.52)	12.0	0.52 (0.65)	2.8
6-10 week	0.72 (0.18)	11.4	0.74 (0.31)	12.8	0.04 (0.39)	0.7

A summary of the estimates of QTL parameters for body weight gains using the one QTL model are shown in Table 6.9. For the growth intervals of 3-6 weeks and 3-10 weeks there was, as already stated, a significant association with QTL positioned at 21cM and 23cM respectively. The estimates of the additive QTL effects in males and females at all body weight gains measured were not significantly different ($P < 0.05$), both showing very similar estimates. Again, the dominance deviations in females were not significantly different from zero in the analysis. The estimated QTL effects expressed as a percentage of the mean body weight gains (Table 6.9) indicate the additive effects of approximately 11% in males. In females, the percentage effects of the QTL are between 9% and 13% for all body weight gains.

Table 6.10 Summary of marker associated effects (g) at the marker *DXMit50* (the marker estimated to be in close linkage with the QTL for body weight) for all body weights and body weight gains. The standard errors of the estimates are shown in parenthesis.

Trait	Male		Female			
	additive		additive		dominance	
3 week weight	0.64	(0.33)	0.18	(0.48)	0.06	(0.34)
6 week weight	1.89	(0.48)	1.43	(0.65)	-0.07	(0.32)
10 week weight	2.43	(0.52)	2.24	(0.77)	0.17	(0.29)
3-6 week gain	1.28	(0.33)	1.19	(0.52)	0.17	(0.26)
3-10 week gain	1.77	(0.38)	2.00	(0.44)	0.03	(0.22)
6-10 week gain	0.56	(0.44)	0.84	(0.71)	0.21	(0.34)

Analysis of the body weight and body weight gain data by the non-inbred multiple regression method indicate that the marker *DXMit50* (at 19.8cM) is closely linked to the QTL for body weight on the X-chromosome (estimated at approximately 23cM). Table 6.10 shows the estimated marker associated QTL effects obtained using the single marker analysis using the model described in section 6.2.3 with the additional fixed effect of marker genotype at marker *DXMit50* (males and females were analysed

separately). The estimates for all traits show similar trends to those obtained by QTL mapping analysis, but for all traits the estimated effects were lower (with the exception of the estimated marker associated QTL effect in females for 6-10 week weight gain). This is expected as the single marker analysis assumes complete linkage between the marker and QTL. Therefore, recombination between the marker and QTL can break down the association and result in an underestimation of the QTL effect (associated with an increase in the standard errors of the estimates). It should be noted that the single marker analysis of the 3 week weight and 6-10 week body weight gain data did not indicate the marker associated QTL effects were significantly different from zero ($P>0.05$).

Linkage between fat percentage and the X-linked QTL for body weight

In addition to the body weight measurements taken on the F_2X generation, estimates of carcass fat percentage were also calculated. Previous analysis has indicated the marker *DXMit50* was closely linked to the X-linked QTL for body weight. To investigate the contribution of the QTL for body weight to fat percentage at 10 weeks, the distribution of residual fat percentage (corrected for the fixed effects and covariates in the model described in section 6.2.3) within the *DXMit50* marker classes was plotted (Figure 6.5). The distribution of residual fat percentage does not indicate an association between the X-linked QTL for body weight (at approximately 23cM) and fat deposition. This observation is supported by single marker analysis carried out using the model described in section 6.2.3 plus the additional fixed effect of marker genotype, where the estimated marker associated QTL effects ($\pm$ s.e.) were: $a_m = -0.55 \pm 0.69$, $a_f = -0.45 \pm 1.05$, and $d = -0.46 \pm 0.78$. The estimated marker associated QTL effects do not indicate linkage between the marker *DXMit50* on the X-chromosome and a QTL influencing fat percentage.

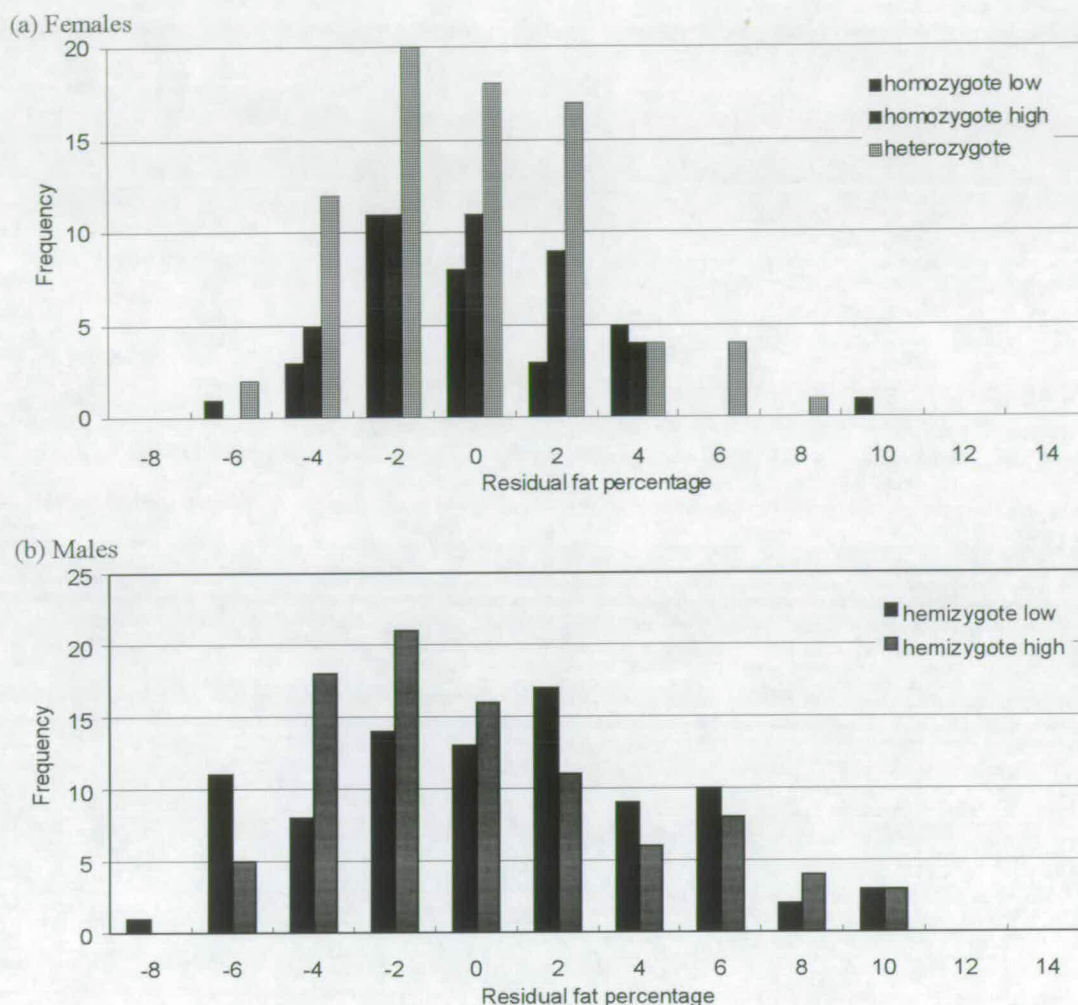


Figure 6.5 The distribution of residual carcass fat percentage (estimated from freeze dried body weight) in the marker classes of the marker *DXMit50*, (a) Females, and (b) Males.

To test for linkage between QTL influencing fat deposition and markers elsewhere on the X-chromosome, the predicted carcass fat percentage was analysed using the non-inbred X-linked regression analysis (Chapter 5). The maximum TS reached in the QTL analysis of fat percentage estimated from gonadal fat pad weight in males was TS 5.8. In the QTL analysis of fat percentage from freeze dried body weights (estimated in males and females) the maximum TS 5.1 was reached (Figure 6.5). This indicates that fat percentage does not show a significant association with a QTL on the X-chromosome ($P > 0.10$).

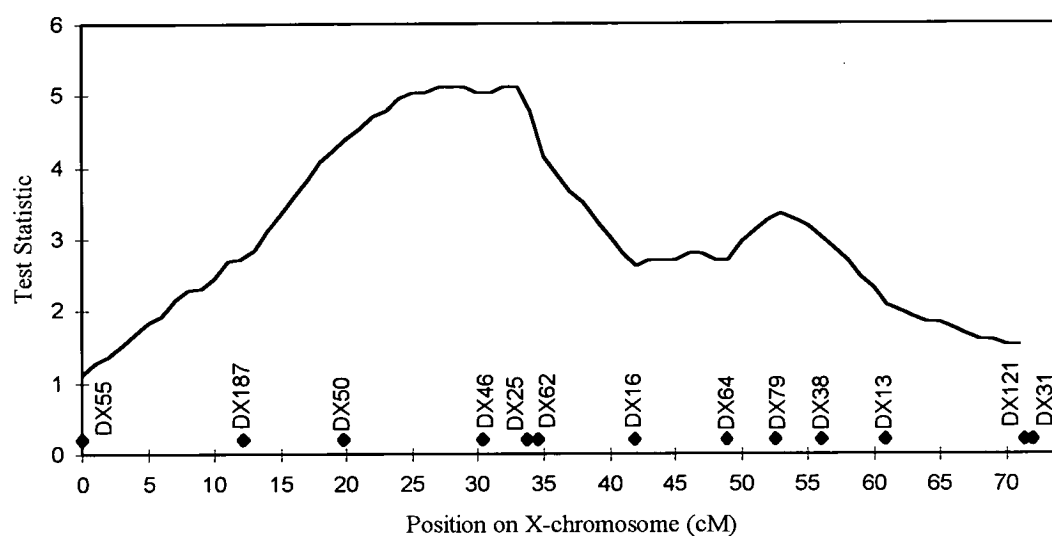


Figure 6.6 The Likelihood profile for estimated carcass fat content (estimated using freeze dried body weight in both males and females). This analysis did not reach the significance threshold of TS 9.7 ($P < 0.10$).

6.4 DISCUSSION

6.4.1 The reciprocal F_1 generation and sex-linked effects

The F_1 generation for the F_2X population was analysed (using the model shown in section 3.5) to establish if the sex linked effect observed previously (section 2.3.4 and 4.3.1) could be found. The mean body weights of males were significantly heavier than females within each reciprocal half ($P < 0.05$). The difference in mean body weight between the sexes in the low line dam reciprocal half was 3.3g and was 6.5g in the high line dam reciprocal half, and therefore the estimated sex linked effect is 3.2g at 6 weeks. The estimated X-linked effect is a similar magnitude to those estimated previously at 6 weeks (Hastings, 1990; Chapter 4), and show support for the significant sex linked effect associated with the X-chromosome.

The mean body weights of the males and females in the F_1 generation of the F_2X population show individuals with the high line dams have a significant advantage for body weight at 6 weeks compared to the F_1 reciprocal half with the low line dams. The results from this study suggest there is a strong influence of the maternal environment on the mean body weight at 6 weeks. The strong maternal effects observed in this population may be due to the small number of successful matings in the F_1 generation (seven litters in total). All individuals in the high line dam reciprocal half were sourced from a total of two families. The low number of litters representing the high line dam reciprocal half will result in a large sampling error associated with the results obtained.

6.4.2 The distribution of polymorphic markers in the F₂X

The results obtained from genotyping the eight test females from the high and low line at generation 52 outlined in Chapter 4, showed that 29% of markers genotyped were polymorphic. The investigation discussed in this chapter showed that the rate of polymorphism in the P6 lines (section 6.3.1) were also very low, 13 of the 61 markers investigated were polymorphic in the F₂X pedigree producing consistently scoreable PCR products (a further 4 of the 61 markers investigated were polymorphic but were not consistently scoreable in the F₂X population). Once again, the percentage of polymorphic sites observed suggests a low rate of polymorphism on the X-chromosome compared to those on the mouse autosomes. As discussed in sections 2.4.4 and 4.4.2, the under representation of informative marker loci are a result of low rates of polymorphism at isolated marker loci (Dietrich *et al.*, 1996). The low number of informative marker loci caused by depressed rates of polymorphism at X-linked marker loci are exacerbated by the under representation of marker loci on the X-chromosome (Dietrich *et al.*, 1994, 1996). It should be noted that low rates of polymorphism observed for X-linked marker loci investigated in the P6 lines may also be due to poor detection of polymorphic loci. The method used to detect polymorphism at marker loci (section 3.2) is able only to reliably detect difference in PCR products of approximately 4 base pairs. However, it is unlikely that this fully explains the low rates of polymorphism observed on the X-chromosome.

6.4.3 Discussion of Phenotypic Measurements

Litter effects

The mean litter size at birth, weaning and subsequent weighings (6 and 10 weeks) suggest a high level of mortality in the F₂X generation, particularly in the post weaning period, where mortality was nearly 10% of the total number weaned. The calculated mortality rate supported the observation of a high level of illness (particularly diarrhoea post weaning) in the population.

At weaning, animals are taken from their mothers, given ear-clip identification and mixed with individuals (of the same sex) from other litters. In addition to the physical stresses at weaning, animals are also given a change of diet (normal mouse facility practice at this time, see section 3.3). The particularly high level of disease observed in the post weaning period is likely to have occurred as a result of a high level of disease challenge present, exacerbated by the stress of weaning and diet change. Observations during the maintenance of the population showed that there was a higher level of mortality in the first parity litters, with illness affecting entire litters, suggesting there may have been a higher level of disease challenge during this period.

Animals which did not survive the post-weaning period were lighter at weaning when compared to the surviving animals, although the difference was not significant ($P=0.11$). Weight data do not support a hypothesis that body weight was a significant factor in the mortality of individuals in the F_2X population in the post weaning period. The analysis of the fully informative marker data at *DXMit50* at 19.8cM (estimated to be closely linked to the QTL for body weight) using a χ^2 test with 1 d.f. did not indicate a significant deviation from the expected 1:1 ratio of alleles in the F_2X generation. Therefore the marker data do not suggest that the X-linked QTL for body weight was a significant factor in the mortality of animals. Therefore, we can assume that the QTL genotype of individuals was not a significant factor in post weaning mortality.

Phenotypic Measurements

To illustrate the distribution of the phenotypic traits measured in the F_2X population, the means and standard deviations are shown in Table 6.3. The weights recorded at 3, 6 and 10 weeks were used to calculate body weight gain during the growth periods of 3-6 week, 3-10 week and 6-10 weeks weight. The body weight measurements show rapid growth in the period of 3-6 week weight and declining growth over the period of 6-10 weeks, which agrees with the normal mammalian sigmoid growth curve, where early gains are slow, growth is rapid in the mid-growth period, and slows again as animals approach maturity.

The residual phenotypic correlations (phenotypic correlations of traits corrected for fixed effects and covariates in the model) of all body weight measurements and derived body weight gains are shown in Table 6.5. Carcass fat percentage show a negative correlation with early weight gain (3-6 weeks) and a positive correlation with the late weight gain (6-10 weeks). This is consistent with the basic knowledge of mammalian growth, where rapid lean growth is seen during the early stages of the growth curve, and as the animal approaches maturity lean growth begins to plateau corresponding with an increase in fat deposition.

As expected, the residual phenotypic correlations between different body weight measurements diminishes through the growth curve, i.e. mean weight at 3 weeks has a higher correlation with weight at 6 weeks than at 10 weeks. The changing correlations are largely a consequence of the reduction in the environmental maternal effect during the course of growth. Early growth is strongly influenced by the mothering ability of the dam, e.g. milk yield and nest building. During later growth (post weaning) the influence of the dam is gradually eroded, as the growth potential of the individual is expressed. In addition to the dam effects on the early growth, the genetic contribution of the individual must be also considered. Cheverud *et al.* (1996) suggested that the low correlation between early and late growth was a result of QTL affecting growth at different stage were distinct; some loci affecting early growth only and some loci affect only late growth. The results of linkage analysis carried out on the population of mice which were genotyped at markers on all the autosomes, and weighed at one week intervals, suggested that QTL for age-specific weights were 7 QTL at one week, up to 17 QTL at 10 weeks. The early growth effects tending to taper off by 6 weeks, and the later growth genes first appear at 3-6 weeks (Cheverud *et al.* 1996). It can be concluded that both dam effects and genetic effects of the individual, (e.g. age specific QTL effecting growth), have a strong influence on early growth.

6.4.4 Detection of X-linked QTL influencing body weight

Genotyping all individuals in the F_2X pedigree enables the fully informative marker loci within litters to be determined (as outlined in section 5.4.4). The marker data obtained for the F_2X pedigree show that three markers were fully informative, and the remaining markers range from approximately 16% to 54% informative. Single marker analysis of the F_2X pedigree data may give an indication of the position and effect of the X-linked QTL (indicated by the highest estimates of QTL effects, and the lowest standard errors associated with the estimates). However, this method is unable to circumvent the problems resulting from the varying information content at the markers along the chromosome. The single marker analysis would produce sampling error associated with different markers, as the animals used to provide estimates of QTL effects at marker loci along the chromosome vary for each marker. Analysis by the non-inbred X-linked multiple regression method (Chapter 5) enables information from all individuals (at each putative QTL position) to be used to obtain estimates for QTL effects and position.

The map positions used in the QTL analysis of the F_2X population were those obtained by two point and multipoint linkage analysis using CRIMAP. The same marker order was estimated in the F_2X population as the order published by the Mouse Genome Database and the Chromosome Committee (both obtained from the Mouse Genome Informatics (1996)). The Massachusetts Institute of Technology map order, however, shows the markers *DXMit79* and *DXMit38* to be in reverse order to the other maps referenced (The Encyclopaedia of Mouse Genome, 1996). As mentioned in the results section, disparity between the marker positions obtained in the maps referenced occurs as a result of sampling error associated with the independent studies, and also the different amounts of recombination observed in the mouse strains used in the different studies. There are no obvious arguments to suggest the use of one particular published map rather than another. The best map to use is clearly the one most related to the strain of mouse used in the study. Hence, the map

estimated from the F_2X data using CRIMAP was used in the QTL analysis. Using the estimated marker positions from other studies in the analysis does not change the marker interval in which the QTL is positioned, but does change the estimated position of the QTL as the map position of the closest linked markers vary (data not shown).

The QTL for body weight on the X-chromosome

The main objective of the investigation described in this chapter was to estimate the number of X-linked QTL of large effect, the size of their effects, and their positions. The likelihood curves obtained from the QTL analysis show substantial evidence for a QTL at approximately 23cM (approximately 3cM distal to marker *DXMit50*) affecting growth at all the body weights measured (3, 6 and 10 week) showing repeatability of the evidence for QTL position. The marker *DXMit50* was fully informative in the F_2X population, and markers distal to this marker had varying information content (28% to 53%), until the next fully informative marker *DXMit13* at 60.9cM (see Tables 6.2 and 6.5). It may be hypothesised that the information content of markers in the interval *DXMit50* to *DXMit13* may influence the estimated QTL position. If we assume the QTL is positioned in the interval *DXMit50* to *DXMit13* the fully informative marker information varies for different litters. The varying marker information for individuals from different litters results in different marker densities, and potentially the QTL positioned different distances from the nearest marker. The results from simulation studies (Chapter 5), where QTL were simulated at varying distances from the nearest marker, suggest that the vicinity of the QTL to the nearest marker does not influence the estimated QTL position. Additional simulation studies indicate that marker density does not affect the accuracy of estimated QTL position. Therefore, there is no evidence that the level of heterozygosity at markers in the F_1 generation of the F_2X population influences the estimated QTL position.

The effect of the X-linked QTL explains approximately 7.7% of the phenotypic variance at 3 weeks, increasing to 14.2% at 6 weeks and 27.9% at 10 weeks. The

estimated percentage of the F_2 variance explained by the QTL appears to increase with age. As discussed previously, early growth is strongly influenced by maternal environmental effects, where much of the variation in body weight in early growth is due to the dam, resulting in only up to 70% of the variation being due to the direct effects of genes carried by the pups themselves (Riska *et al.*, 1984). This relationship may result in an underestimation of the percentage of phenotypic variance explained by the QTL, particularly at early body weight measurements.

The estimate of the additive QTL effect at 3 and 6 weeks was slightly higher in males than females, although the difference was not significant. The same estimated additive effect of 2.6g was obtained for 10 week weight in both males and females. The dominance deviations at 3, 6 and 10 weeks were not significantly different from zero suggesting the X-linked QTL has a completely additive action. The additive QTL effects expressed as a percentage of body weight range from approximately 7% to 8% in males at all three body weight measurements. In the females, the percentage additive effects were slightly higher than the estimates in males at 6 weeks and 10 weeks (9% and 10% in females compared to 8% in males). Whereas, at 3 weeks the additive effect expressed as a percentage of the mean body weight in females was approximately 4%, which does not agree with the results obtained in males. The effect of sex was fitted in the model used in the X-linked regression analysis, therefore differences between the body weight of the males and females due to autosomal acting QTL were removed (and the same overall mean for males and females was fitted in the QTL analysis). The estimated additive effects associated with the X-linked QTL suggest there is a sex difference in the action of the QTL in early growth. The additive effects at 3 weeks indicate the QTL has a greater contribution for mean body weight at 3 weeks in males compared to females. We are unable to determine whether the estimate for female additive effect at 3 weeks is a spurious result, or a property of the X-linked QTL, without repeating the experiment.

The influence of the X-linked QTL on specific growth periods

Further investigation of the period at which the X-linked QTL affects growth was carried out with the examination of body weight gain between measurements. The likelihood curves from the QTL analysis of the body weight gain between 3-6 week, 3-10 week, and 6-10 week suggest the QTL to be position at 21cM, 23cM and 23cM on the chromosome respectively, again showing strong support for the X-linked QTL being positioned at approximately 23cM.

The likelihood curves obtained for all body weight gains show the QTL effects were highly significant ($P < 0.01$). As shown in Figure 6.3c, the TS obtained for 6-10 week body weight gain did not show a single peak, but showed a peak at 23cM, and an elevated TS towards the distal end of the X-chromosome (the second peak reaching the $P < 0.05$ threshold). Caution must be taken when interpreting these results, as the elevation in the TS in the analysis fitting one QTL does not occur until after the marker *DXMit13* (which was fully informative in the F_2X population, see Figure 6.3c). As shown the Table 6.3, the markers (*DXMit31* and *DXMit121*) more distal to marker *DXMit13* were both only 15.6% informative, and therefore the remaining 84.4% of individuals contribute information to the QTL position and effect by single marker analysis. Analysis using single markers contributes no information to the QTL position, as it is unable to distinguish between closely linked QTL of small effect, and loosely linked QTL of large effect. Another shortcoming of this method is the inability to determine if the QTL is positioned more distal or proximal to the marker analysed. It is therefore likely that the elevated TS seen towards the distal end of the X-chromosome is due to sampling error associated with the 15.6% of animals informative at markers *DXMit31* and *DXMit121*. Analysis of the 6-10 week body weight gain data with the X-linked multiple regression method fitting two QTL vs. one QTL did show a significant change in the TS ($P > 0.10$) at the peak at 23cM, but not elsewhere on the X-chromosome. It can be concluded that data for 6-10 week body weight gain do not support the hypothesis of a second QTL on the X-chromosome.

In the study by Cheverud *et al.* (1996) carried out to examine the effects of QTL on age specific body weights in mice, body weights at one week intervals (1-10 weeks of age) and genotypes at 75 microsatellite markers spanning 19 autosomes were obtained on 535 mice. The results show QTL mapped to 16 of the 19 autosomes using an estimated empirical threshold of $P < 0.10$ (obtained using the method previously described by Andersson *et al.* (1994) see section 2.2.3). Linkage analysis indicated that many of the QTL had significant effect for early (1-3 week) and middle (3-6 week) growth periods or the middle and late (6-10 weeks) growth periods, but not for postnatal growth as a whole (Cheverud *et al.* 1996). Comparison of the mapping results for body weight gains obtained in this study with the results of Cheverud *et al.* (1996) may suggest that the QTL positioned at 23cM on the X-chromosome affects growth in the middle and late growth periods. Early growth from 1-3 weeks cannot be commented on, as one week weight was not measured in this study, although the QTL effects were significant for 3 week weight data.

In all QTL analysis of body weight measurements and body weight gains there was no indication of significant dominance effects for the QTL. This supports previous findings by Veerkamp *et al.* (1993), where ML segregation analysis suggested an X-linked effect with additive action. Consideration of the state of the X-chromosome in males and females may suggest we may not expect to observe a dominance deviation. In males the X-chromosome is hemizygous, and therefore no dominance deviation can be observed. Whereas in females the X-chromosome undergoes random inactivation early in development, leading to a hemizygous mosaic for the active X-chromosome (section 2.4.2). If the product of the X-linked QTL was active only within the cell of origin, then like the males, a dominance deviation would not be observed in the females (if we assume equal frequency of the two active X-chromosomes in the females). If the QTL product had systemic action on growth, then it would be possible to observe dominance deviation in females. No firm conclusions can be made regarding the action of the QTL product until the product is determined. The results obtained from analysis of phenotypic data by Veerkamp *et al.* (1993) and QTL

analysis presented in this chapter do not suggest any dominance deviation in the action of the X-linked QTL for body weight.

The influence of the X-linked QTL on fat deposition

Body weight data analysed by the non-inbred X-linked multiple regression method (Chapter 5) indicated a QTL of large effect on body weight was positioned on the X-chromosome at approximately 23cM. Data on carcass fat percentage was analysed to investigate the possible contribution of the QTL for body weight to fat deposition. Single marker analysis of the marker *DXMit50* (closely linked to the X-linked QTL for body weight) was undertaken to estimate a possible correlated effect associated with fat percentage. The estimated marker associated effects at *DXMit50* did not indicate the X-linked QTL at approximately 23cM influenced the deposition of carcass fat, as the estimated effects were not significantly different from zero ($P>0.05$). These results do not allow us to reject the hypothesis that there is no X-linked QTL influencing carcass fat percentage. Firstly, the single marker analysis is not as powerful as the X-linked regression method. For example, the data for 3 week weight and body weight gain between 6-10 weeks did not indicate significant marker associated QTL effects at the marker *DXMit50*, but the regression analysis showed significant TS ($P<0.01$) at the marker *DXMit50*, with a peak at 23cM. Secondly there may be a QTL influencing fat percentage elsewhere on the X-chromosome. To investigate the possibility of QTL influencing fat deposition elsewhere on the X-chromosome, the estimated carcass fat percentage was analysed using the non-inbred X-linked multiple regression method. The likelihood curve for both estimates of carcass fat percentage (from GFP and freeze dried body weights) showed there was no significant effect associated with the X-chromosome, and therefore suggesting that carcass fat percentage is unchanged by the action of X-linked QTL.

In summary, the initial examination of the result of QTL analysis of data for body weights and body weight gains between weighings, suggest that the QTL for body weight on the X-chromosome has an additive mode of action. The QTL also appears to affect growth throughout the growth curve, as the contribution of the X-linked

QTL for body weight was significant for 3 week weight data, and significant for all body weight gains. The estimated QTL effects at 10 weeks was approximately 5g difference between the homozygote female and hemizygote male genotypes, and this is approximately equal to the estimated X-linked effect (Hastings, 1990; Hastings & Veerkamp, 1993), suggesting the QTL at approximately 23cM explains the entire X-linked effect in the P6 lines.

The QTL mapping analysis for body weight measurements and body weight gains between measurements do not indicate that the QTL has a significantly greater contribution over any specific growth period to warrant examination of this period in greater detail. Data on 10 week weight encompasses all growth periods measured, consequently the results obtained from the QTL analysis of body weight at 10 weeks will be used in the remaining discussion.

6.4.5 Interpretation of results of an X-linked QTL for 10 week weight

To enable correct interpretation of results obtained from the analysis using the non-inbred X-linked multiple regression method, the null hypothesis for QTL detection, the significance thresholds, and the confidence intervals for QTL position must be considered in depth.

The null hypothesis for QTL detection

As reviewed in section 2.2.3 the formulation of an appropriate threshold for QTL detection is problematic. In the majority of studies regarding the formulation of thresholds, the null hypothesis used was that no genetic variance was explained by the marker segment under consideration. However, Visscher & Haley (1996) suggest a more appropriate and realistic null hypothesis should be based on testing the hypothesis that there is one, or a limited number, of QTL of relatively large effect in the region, versus the hypothesis that there are many loci of small effect. As described in section 2.3.4 and Chapter 4, the previous analysis of the P6 lines had shown the X-

chromosome has a significant effect on body weight in the P6 lines, and therefore, the null hypothesis used in this study (no QTL or variance explained by the marker interval) may not be the most appropriate. A simulation study by Visscher & Haley (1996) indicated that the incorrect null hypothesis (e.g. no variance explained by the marker interval) will result in an over estimation of the significance level (i.e. an inflated TS), compared to the null hypothesis assuming variance is explained by the chromosome, and can therefore result in type I errors. As explained in section 6.2.3, a null hypothesis assuming the variance explained by the X-chromosome was due to many loci of small effect could not be fitted due to the nature of the data set (i.e. the level of non-informative or missing marker genotypes). However, it is extremely unlikely that the threshold obtained for 10 week weight data would be reduced sufficiently by fitting a null hypothesis that the variance explained by the chromosome was due to a large number of loci of small effect, such that the TS was reduced to less than the $P < 0.01$ threshold.

Thresholds for QTL detection

As described in section 6.2.3, the empirical thresholds for QTL detection were estimated by the method previously described by Andersson *et al.* (1994), where marker data were analysed over a large number of replicates with simulated phenotypic data generated from a normal distribution with no QTL present (replicates range from 500 by Cheverud *et al.* (1996) to 5000 by Andersson *et al.* (1994)). An alternative and arguably more appropriate method is the permutation test of Churchill & Doerge (1994) where the phenotypic data are randomly shuffled with respect to the genotypic data to break down the association between phenotype and genotype. Repeating this procedure over a large number of replicates (e.g. 1000), results in a test statistic under the null hypothesis of no association between phenotypic and genotypic data. Ordering the maximum TS from the replicates and taking the 95 and 99 percentiles produces the thresholds for $P < 0.05$ and $P < 0.01$. The advantage of this method to determine the threshold for QTL detection is that the threshold is generated under the null hypothesis specific to the data set used. The distribution of phenotypic measurements, marker densities, and number of linkage groups analysed

are all specific to the data set in question. As a result of the computation time required to estimate empirical thresholds using the permutation test of Churchill & Doerge (1994) for the separate traits recorded in this study, the empirical thresholds were estimated using the method previously described by Andersson *et al.* (1994).

The results obtained from analysis of all data sets using the X-linked multiple regression method show maximum TS were greater than the estimated empirical threshold for $P < 0.01$. The likelihood curve for 10 week weight data shows the TS along the entire X-chromosome surpasses the significance threshold of 15.5 ($P < 0.01$), indicating all markers show a significant association with the QTL. This occurs as a result of recombination events not being sufficient to break down the association between the QTL of large effect and all markers on the X-chromosome in all individuals. The thresholds for QTL detection are important to prevent type I and type II errors. The QTL mapping results presented in this chapter, however, provide convincing evidence of a QTL of large effect on the X-chromosome. The maximum TS of 112.4 reached for the analysis of 10 week weight data indicates the X-linked QTL was very highly significant.

From the shape of the likelihood curve using the model fitting one QTL in analysis of 10 week weight data, it was thought unlikely that analysis using the two QTL vs. a single QTL model would show a significant change in TS. The analysis using the two QTL model did not result in a significant change ($P > 0.10$) in the TS at any position other than the peak at 23cM. The results suggest a single QTL of large effect contributing to the X-linked body weight effect at 10 weeks, and inspection of the shape of the likelihood curves obtained for other body weight measurements do not indicate greater than one QTL. However, we are unable to eliminate the possibility that the X-linked effect is controlled by two (or more) closely linked QTL within the 12cM 99% confidence interval. The inability to distinguish between closely linked loci occurs as a result of recombination events which are insufficient to break down the association between closely linked loci. The lack of recombination is further exacerbated on the X-chromosome as recombination only occurs in the F_1 females.

Further experimentation would be needed to resolve the CI further and enable the number of loci to be determined without doubt.

Confidence intervals for QTL positions

Comparison of the estimated QTL positions for the three body weight measurements and three body weight gains suggest that the most likely position of the QTL is at approximately 23cM from the most proximal marker (approximately 3cM distal from marker *DXMit50* at 19.8cM). The 10 week weight data set suggest the estimated 95% confidence interval for this position using the equivalent to a one LOD-drop off support interval (a 4.6 drop in the TS) is 8cM, and the 99% CI (9.2 TS drop) is 12cM (Figure 6.1c).

The formulation of an effective confidence interval for QTL position is important for full interpretation of the results of QTL mapping. As outlined in section 2.2.3, the method of Visscher *et al.* (1996) using bootstrapping was very effective in the formulation of CI for QTL position in a simulation study, being over stringent compared to the empirical CI for QTL position. The effectiveness of bootstrapping occurs as a result of information from the entire likelihood curve being used to produce the CI. This is particularly useful when the likelihood curve is multimodal. The results of a simulation study by Visscher *et al.* (1996) showed the LOD-drop off method (equivalent to $P < 0.05$) for a population size of 200 individuals with simulated QTL with h^2 of 0.01-0.1 gave CI that contained the QTL in 92-93% of simulations over 1000 replicates, and the corresponding bootstrap samples gave CI that contained the simulated QTL in 96-99% of simulations. The mean width of the CI were greater when using the bootstrap method compared to the LOD-drop off method. It should be noted that Visscher *et al.* (1996) used a drop in LOD score of 0.83 (which corresponds to a 95% CI when the TS is distributed as a χ^2 with 1 degree of freedom), and it is possible that if a drop in LOD of one had been used the LOD-drop off method may have produced CI closer to the empirical 95% CI. Work by Visscher *et al.* (1996) and van Ooijen (1992) showed the LOD-drop off method was more

effective when the population size and h^2 are high, i.e. when the experimental power is high.

Like the formulation of the estimated empirical thresholds for QTL detection, the bootstrap method is computationally demanding. The advantages of the bootstrapping method over the LOD-drop off method decline when the experimental power is high, and a smooth likelihood curve is obtained with a single well defined peak. The contribution of the X-linked QTL at 10 weeks is very large, contributing in excess of 27% of the total phenotypic variation. The smooth shape of the single well defined peak obtained from the QTL analysis of the 10 week weight data suggests that the one LOD-drop off is appropriate to obtain a CI for QTL position. If the peak was less well defined, with greater than one peak, there may be a significant advantage in using the bootstrapping method to formulate a CI for QTL position. It should be noted that both the LOD-drop off method and the bootstrap method do not provide completely accurate 95% CI for QTL position. Therefore, whatever method is used, potential inaccuracies may occur in the estimated confidence interval.

6.4.6 Summary of Results

The main findings of the F_2X experiment set up with the objective of mapping QTL on the X-chromosome in the P6 lines were:

1. The body weight measurements at 3, 6 and 10 weeks and growth periods between weight measurements indicate a QTL for body weight positioned at 23cM on the X-chromosome.
2. QTL analysis of all the body weight measurements suggest that an X-linked QTL affects body weight throughout the growth curve.
3. Analysis of the data on 10 week weight shows the total additive effect of the X-linked QTL is 2.6g in males and females, i.e. the QTL results in a total effect of 5.2g between the homozygote females and the hemizygote males, explaining approximately 20% of the mean body weight. The estimated QTL effect at approximately 23cM appears to explain the entire X-linked effect in the P6 lines (section 2.3.4 and Chapter 4).
4. There was no significant evidence for dominance deviations in the females at this locus.
5. The confidence interval for the QTL position at 10 weeks is estimated to be 8cM for a 95% CI and 12cM for a 99% CI.
6. The estimated fat percentage data analysed by the X-linked multiple regression method show no indication of an X-linked effect associated with carcass fat percentage.

CHAPTER 7

The progressive backcross population

7.1 INTRODUCTION

The analyses of the F_2X population using the non-inbred X-linked multiple regression method (Chapters 5 and 6) indicate the presence of a major QTL for body weight at approximately $23\text{cM} \pm 4\text{cM}$ on the X-chromosome, with an estimated effect (the difference between the homozygous females and hemizygous males) of 5.2g on 10 week weight in both males and females. The F_2X population has provided valuable information on the size of the QTL effect and its position.

As outlined in section 2.2.6, simulation studies have indicated that F_2 and BC_1 experimental designs, often employed in QTL mapping experiments, are unable to map QTL with great precision, largely due to genetic noise resulting from QTL elsewhere on the genome, and to a limited extent (when large population sizes are used) the lack of recombination events. Analytical methods for mapping QTL have been developed to remove genetic noise due to linked QTL and QTL on other chromosomes, by fitting multiple markers as cofactors in the analysis (section 2.2.2). These methods require segregating populations to be genotyped at markers spanning the entire genome in order to remove all background genetic noise. An experimental alternative to this analytical procedure (to remove noise due to QTL elsewhere in the genome) is to progressively backcross regions of the genome of interest onto an

inbred background. This experimental design was used in the progressive backcross population set up to map the X-linked QTL for body weight.

The progressive backcross (BC) population was based on work carried out on *Drosophila* by Shrimpton and Robertson (1988a), where morphological markers were used to maintain specific regions of the third chromosome (C) in a progressive backcross population, producing several congenic lines. This experimental design is more powerful than a simple F₂ or BC₁ population, as the congenic lines can be assumed to be genetically identical in all regions of the genome unlinked to the congenic region of interest, thus removing noise due to unlinked QTL. The objective of the experiment described in this chapter was to backcross specific regions of the high line X-chromosome onto an inbred low line background. This experimental design results in the removal of body weight effects associated with the high line autosomes and the X-chromosome other than the maintained region.

The progressive backcross population experimental design will be presented in this chapter; the data set over generations of backcrossing will be described and finally the results of analysis of the progressive backcross population will be shown.

7.2 MATERIALS AND METHODS

7.2.1 DNA extraction and PCR amplification protocol

Genomic DNA was extracted from spleen samples or ear clip samples (section 3.1). As described previously (section 3.1), ear clips were taken as part of the standard procedure at weaning for identification purposes and therefore, DNA extraction from ear clip tissue samples precludes the need for additional procedures to obtain DNA. The DNA extracted from individuals in generation 3-5 of the progressive backcross used the streamlined procedure (section 3.1).

The standard PCR amplification Protocol 1 (section 3.2) was used to genotype animals from BC generations 1-3. Genotyping of animals in the BC generations 4 and 5 utilised the procedure described in Protocol 2 (section 3.2).

7.2.2 The mating scheme for BC populations and summary of the data collected

Females from generation 51 of selection of the P6 high line and males from the P6 low inbred line, at generation 6 of inbreeding (set up at generation 45 of selection), were mated to produce an F₁ generation. All F₁ females were mated to inbred low line males (generation 7 of inbreeding) to form generation 1 of the progressive backcross population (BC₁). Thereafter, all mating pairs and weaned animals were maintained as described in section 3.3. The numbers of litters and progeny generated are summarised in Table 7.1.

In the first generation of backcrossing (BC₁) ear clip samples were taken from all individuals at weaning for DNA extraction. The parents of the BC₁, and parents of the F₁ generation were killed by cervical dislocation, spleens were removed, placed in labelled screw top tubes and snap frozen in liquid nitrogen. Following DNA extraction

from all the parents and the BC₁ females, individuals were genotyped at 9 loci spanning the length of the X-chromosome (*DXMit55*, *DXMit50*, *DX2Mit25*, *DXMit16*, *DXMit64*, *DXMit79*, *DXMit38*, *DXMit121*, and *DXMit31*). The genotypes were used to ascertain fully informative marker loci in the BC₁ females, where marker alleles could be traced unequivocally to the high selection line or the inbred low line. Female BC₁ individuals were selected to be dams of the BC generation 2, where the markers genotyped were fully informative. Figure 7.1 summarises the mating scheme of the BC population.

The P6 inbred low line was maintained in parallel to the backcross population by full sib mating to provide sires for the progressive BC population. Ear clip samples were taken from all inbred low line males for DNA extraction. The inbred low line males were genotyped at all marker loci described above to enable selection of males carrying low line marker alleles only.

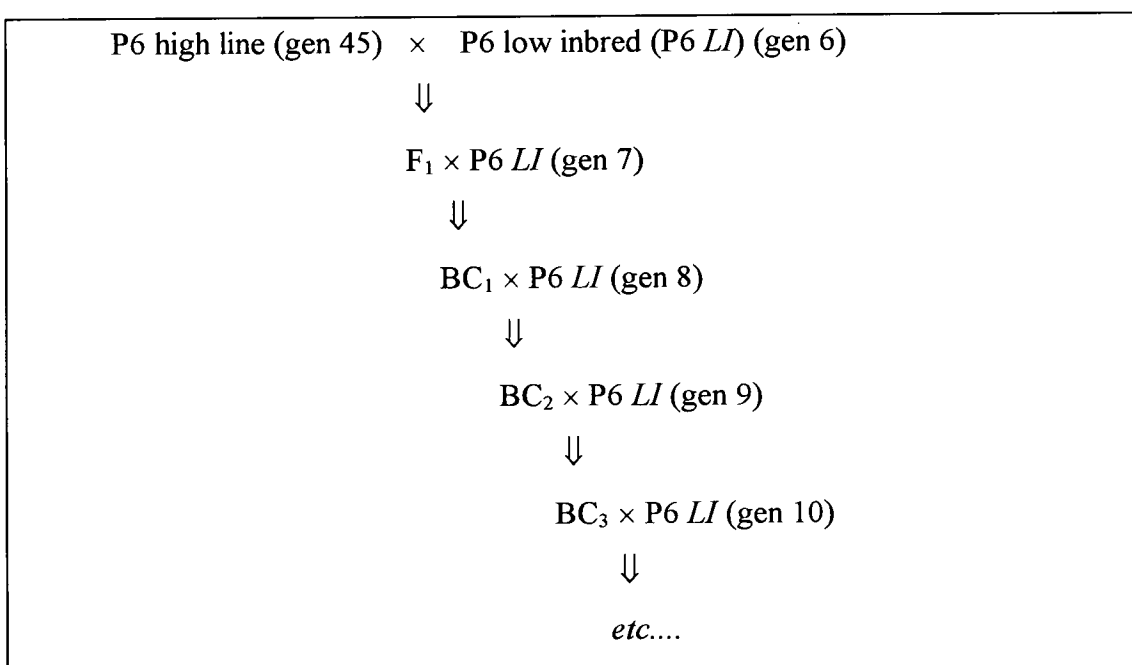
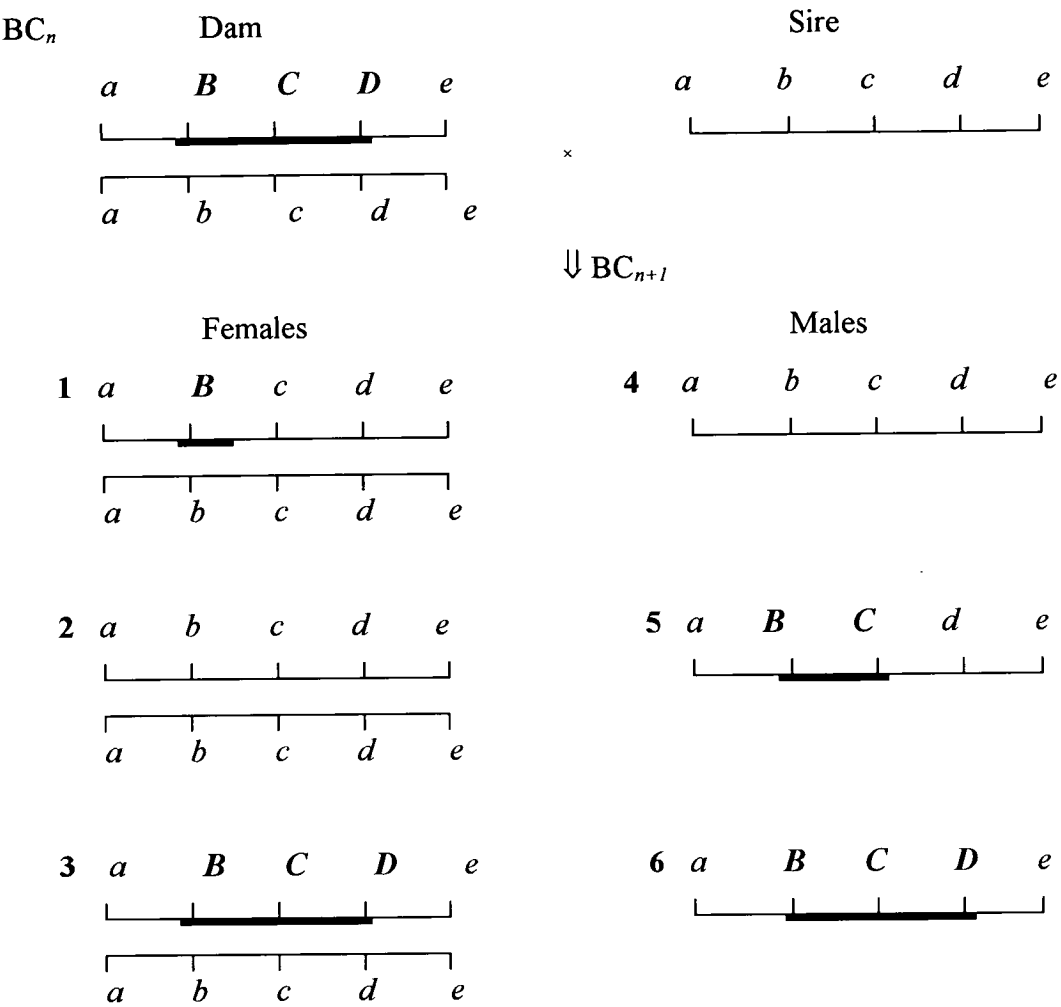


Figure 7.1 The mating scheme for the BC population, showing the progressive backcrossing of selected BC females to P6 inbred low line males.

The objective of the mating scheme was to maintain specific segments of the high line X-chromosome in a heterozygous state in the BC population dams (using markers spanning the length of the X-chromosome). An example of the conservation of an X-chromosome segment is shown in Figure 7.2. Using the example of the marker segment illustrated in Figure 7.2 (maintaining marker *B*, *C*, and *D* from the high selection line), and assuming no double crossovers between flanking markers, the chromosome segment between marker *B* and *D* would be inherited from the high selection line, and marker alleles *a* and *e* inherited from the inbred low line. The percentage high line X-chromosome between the markers *a* and *B*, and *D* and *e*, will tend to decline with each generation of backcrossing.



Marker alleles *B*, *C*, and *D* were inherited from the high selection line, and markers *a*, *b*, *c*, *d*, and *e* from the inbred low line. The female progeny number 3 would be retained for breeding in the next generation to maintain high line marker alleles *B*, *C*, and *D* on the inbred low line background.

Figure 7.2 An example of the maintenance of a high line X-chromosome segment on the inbred low line background, showing a BC_n female mated to an inbred low line male, and some examples of the possible BC_{n+1} progeny

The mating scheme was designed to establish a number of lines (illustrated in Figure 7.3) carrying overlapping segments of the X-chromosome inherited from the high

selection line, carried on the genetic background of the inbred low line. It should be noted that uniform segments were not obtained until later generations. As described above, all BC females were genotyped at 9 microsatellite loci on the X-chromosome. Unfortunately the marker loci *DXMit64* was found to be uninformative from generation BC₂ onwards, leaving 8 markers used to maintain segments of the X-chromosome in BC₂ onwards.

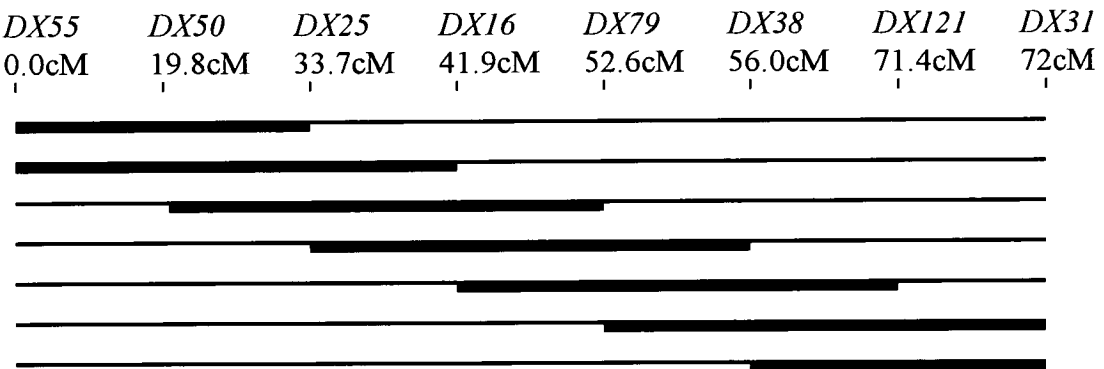


Figure 7.3 Illustration of the objective of the backcross experiment, to establish seven backcross lines with high line X-chromosome segments maintained by eight X-linked Mit markers. The bold lines represent regions of the X-chromosome inherited from the high selection line on an inbred low line background. The marker map positions shown were those estimated in the F₂X population (Chapter 6).

In BC generations 3, the marker *DXMit187* at 12.2cM on the X-chromosome was added to the panel of 8 markers used previously (Figure 7.3). This marker (*DXMit187*) was found to be fully informative in the BC pedigree, where all marker alleles could be traced unequivocally to the high or low line.

Table 7.1 A summary of the numbers of successful matings and individuals weaned in each generation of the backcross population. The traits recorded are also shown.

Generation	No. Litters	Males			Females		
		No. weaned	6 wk wt	10 wk wt	No. weaned	6 wk wt	10 wk wt
F ₁	5	19 †	✓		23	✓	
BC ₁	15	62 †	✓	✓	84	✓	✓
BC ₂	11	-	-	-	57	✓	
BC ₃	14	90	✓	✓	87	✓	✓
BC ₄	17	96	✓	✓	83	✓	✓
BC ₅	22	123	✓	✓	112	✓	✓

† DNA available, but no genotypes recorded

As summarised in Table 7.1, tissue samples were taken from all individuals in the BC pedigree, with the exception of males in BC generation 2 (these animals were killed at weaning). Data on 6 week weight was collected for all individuals weaned in the BC population, and body weight at 10 week weight was recorded for all individuals in the pedigree with the exception of F₁ individuals and females in the BC₂ generation.

7.2.3 Preliminary analysis of the Backcross generations

Single marker analysis

A single marker analysis was carried out to infer associations between the marker loci genotyped in the BC population and the 6 and 10 week weight data. Data were analysed for each generation separately and each marker was fitted in the model in turn.

The data were analysed by Genstat 5.3 (Genstat 5 Committee, 1993) using the model:

$$Y_{ijklm} = \mu + S_i + D_j + \beta \cdot W_{jk} + G_l + e_{ijklm}, \quad (1)$$

where:

Y_{ijklm} = the observations of 6 and 10 week weights on the m th individual

μ = overall mean

S_i = fixed effect of the i th sex

D_j = random common environmental effect associated with the j th dam

β = the linear regression coefficient of Y on the number weaned (W_{jk}) from the k th parity litter of the j th dam

G_l = fixed effect of the l th marker genotype ($n = 0$ or 1 in males and 0 or 2 in females, where 0 are low line genotypes, 1 is the male hemizygous high line genotype, and 2 is the female heterozygous genotype)

e_{ijklm} = the residual error associated with the m th individual

This model enables the effects associated a high line marker allele (in a hemizygous state in males, and heterozygous state in females) to be estimated, compared to the hemizygous or homozygous low line marker allele in males and females respectively. The estimates obtained represent $2a$ in males and $a+d$ in females.

Segment analysis

In addition to the single marker analysis outlined above, a segment analysis was carried out to estimate the effect associated with carrying a specific segment of the high line X-chromosome flanked by known marker loci (for examples of these 'segments' see Figure 7.3). The numbers of individuals carrying each segment were low within each generation, because uniform segments were not established until the final generation of the BC population. To increase the numbers of individuals within each segment group, information from generations 3, 4, and 5 of the BC population was combined. To attempt to account for changes in body weight observed over the generations analysed (see Table 7.2), a fixed effect of generation was fitted. The data

were analysed in Genstat 5.3 (Genstat 5 Committee, 1993) using the model outlined below,

$$Y_{ijklmn} = \mu + S_i + D_j + \beta \cdot W_{jk} + SEG_n + GN_l + e_{ijklmn}, \quad (2)$$

where:

Y_{ijklmn} = the observations of 6 or 10 week weights on the m th individual

μ = overall mean

S_i = fixed effect of the i th sex

D_j = random common environmental effect associated with j th dam

β = the linear regression coefficient of Y on the number weaned (W_{jk}) from the k th parity litter of the j th dam

SEG_n = fixed effect of the n th segment

GN_l = fixed effect of the l th generation of origin

e_{ijklmn} = the residual error associated with the m th individual

7.2.4 Analysis of the Backcross population using all pedigree information

The objective of analysing data from the BC population is to obtain an estimate for the X-linked QTL position and the size of effect associated with the QTL. This will provide independent estimates to those already obtained from the analysis of the F_2X population (Chapter 6). In the F_2X population X-linked QTL of small effect may have been masked by the QTL of large effect estimated at $23\text{cM} \pm 4\text{cM}$. The design of the BC experiment should allow QTL (of small effect) elsewhere on the X-chromosome to be mapped, providing estimates of QTL position and size of effect.

To enable all the information collected through the pedigree to be included in the analysis, the pedigree structure of the population must be accounted for. In the QTL analysis of an F_2 or a simple first generation backcross population, using ML or multiple regression methods (see section 2.2.2), we assume the parental genotypes are

known, i.e. we assume the founder lines to be fixed at different QTL alleles, and all the F_1 parents are heterozygous at the QTL. As a result of this assumption, we can consider all the F_2 and first generation backcross individuals independently. The pedigree structure of the progressive BC population means all animals cannot be considered to be independent (with the exception of animals within the first generation of backcrossing). The individuals within generations are dependant on common ancestors in previous generations (resulting in dependencies across generations), and individuals from different litters may have ancestors in common in the previous generation (leading to dependencies within generations).

The basic methodology for QTL mapping is to sum over all possible genotypes, at all loci consistent with the observed data within the population. For a simple F_2 or first generation backcross population this can be done one individual at a time (as all animals are independent). In more complex pedigree structures (e.g. the pedigree structure of the progressive backcross population) multiple individuals must be considered at once. Analysis of the progressive BC population using a Markov chain Monte Carlo (MCMC, see Metropolis *et al.*, 1953 and Hastings, 1970) sampling approach enables the genetic configuration of the BC pedigree to be accounted for. Gibbs sampling (a special case of MCMC, see Geman & Geman, 1988) was used to produce estimates of the QTL positions and effects. The estimates of QTL positions and effects were obtained from the joint posterior distribution. The joint posterior distribution for the parameter vector θ and the data Y is defined:

$$f(\theta, Y) = f(Y | \theta) \times f(\theta),$$

or joint posterior distribution = likelihood $\times$ prior.

Uniform priors were used for all QTL effects and QTL positions so that the posterior distributions are proportional to the likelihood functions. The full conditional posterior distribution of a given model parameter (the posterior distribution of the parameter conditional on all other parameters in the model) is then used to obtain an estimate for the parameter in question.

The model used in the analysis was:

$$Y_{ijklm} = \mu + \alpha + D_j + P_k + \beta_1 \cdot W_{jk} + QTL_{li} + \beta_2 \cdot H_j + e_{ijklm},$$

where:

Y_{ijklm} = the observations of 6 and 10 week weights on the m th individual

μ = overall mean

α = polygenic component (representing the breeding value)

D_j = random common environmental effect associated with the j th dam

P_k = the fixed effect of the k th parity litter

β_1 = the linear regression coefficient of Y on the number weaned (W_{jk}) from the k th parity litter of the j th dam

QTL_{li} = fixed effect of the l th genotype nested within the i th sex

β_2 = the linear regression coefficient of Y on the percentage of high line genotype (H_j) of individuals with the j th dam (this covariate is used to account for the changes in the means and variances of body weight due to the changing percentage of high line autosomal loci, see section 7.3.1)

e_{ijklm} = the residual error associated with the m th individual

Using the model described above, the Gibbs sampler is used to sample all unknown variables (e.g. the QTL effects, positions, and frequencies) from their full conditional distributions (conditional on all other parameters in the model). To illustrate the basic methodology of Gibbs sampling, consider the following example of three variables in a joint density $f(x, y, z)$, the marginal distributions of X , Y , Z are required with densities $f(x)$, $f(y)$, and $f(z)$ respectively. Let (x_0, y_0, z_0) represent an arbitrary set of starting values for the three random variables, then the Gibbs sampler proceeds as follows. The first sample is:

$$x_1 \text{ from } f(x | Y = y_0, Z = z_0),$$

where $f(x | Y = y_0, Z = z_0)$ is the conditional distribution of X given $Y = y_0$ and $Z = z_0$.

The second sample is:

$$y_1 \text{ from } f(y | X = x_1, Z = z_0),$$

note that we condition now on x_1 not x_0 . The third sample is:

$$z_1 \text{ from } f(z | X = x_1, Y = y_1), \quad .$$

now conditional on both x_1 and y_1 . This process of three sampling steps constitutes the first iteration of a Gibbs sampler. The second step of the Gibbs sampler will start:

$$x_2 \text{ from } f(x | Y = y_1, Z = z_1),$$

where $f(x | Y = y_1, Z = z_1)$ is the conditional distribution of X given $Y = y_1$, and $Z = z_1$, and so on. This process is repeated t times, where t is known as the length of the Gibbs sequence. In this study, the Gibbs sampler was run for $t = 40,000$ cycles.

The genotypes (i.e. at QTL and missing marker genotypes) for all individuals were sampled for each locus in turn conditional on the marker observations, the trait values, the model parameters, and the genotypes at other loci. In the sampling of genotypes of individuals, the genotypes were updated for all individuals at a given locus simultaneously (Kong, 1991). This method speeds up the mixing of the chain. Mixing is an intuitive term used to describe how well the Gibbs sampler is working. For the parameter estimates to be accurate, realisations from the Gibbs sampler must be taken from the stationary posterior distribution of each parameter, where mixing of the chain describes how efficiently the sampler is moving around the sample space. By speeding up the mixing of the chain, the number of cycles of the Gibbs sampler required, is reduced. The analysis described above was used to produce estimates of the posterior distributions of QTL positions and effects. The Gibbs sampling was carried out by Dr S. C. Heath using a method specifically modified to allow analysis of X-linked loci (a derivation of the methods described in Heath (1995)).

The confidence intervals for the QTL positions were obtained from the $4 \times$ standard error (s.e.) of the mean position as proposed by Darvasi *et al.* (1993).

7.3 RESULTS

7.3.1 Changes in the distribution of traits measured over the generations of backcrossing

To illustrate the distribution of the body weight traits measured in the BC population, means and variances were calculated for each generation (Table 7.2). The data were corrected for the fixed effects and covariates fitted in model 1 (section 7.2.3), with exception of the fixed effect of marker genotype. The estimates obtained show that the mean 6 week and 10 week body weights decline with each generation of backcrossing to the inbred low line, with the exception of the estimates obtained for BC₂ (in this generation the 6 week weights were available only on females). The reduction in mean body weight is associated with a reduction in the variance of the body weight measurement. This observation is expected, as the decrease in the mean and the variance of body weight occurs as the percentage contribution of high line autosomal linked QTL for body weight decline by an average of 50 percent with each generation of backcrossing; such that in the BC₅ generation, individuals carry on average 1.56% high line autosomal alleles.

The data presented in Table 7.2 also show that mean litter size declines with each generation of backcrossing to the inbred low line. Brien *et al.* (1984), Beniwal *et al.* (1993) and Mbaga (1996) have investigated the correlated responses in litter size which occurred as a result of divergent selection for fat free body mass in the P lines. The correlated responses showed the mean difference observed between the high and low selection lines were approximately 5.4 pups born per litter in the P6 lines at generation 50 of selection (Mbaga, 1996). As the contribution of the low line genes in the BC females increases, the mean litter size would be expected to decline.

Table 7.2 Changes in the means and variances of 6 and 10 week body weights over the BC generations (corrected for the fixed effects and covariates in model 1 section 7.2.3, with the exception of the fixed effect of marker genotype). The mean litter sizes are shown for each BC generation.

Generation	Litter size	6 week weight (g)		10 week weight (g)	
		Mean	Variance	Mean	Variance
F ₁	8.95	25.04	5.29	-	-
BC ₁	9.09	18.38	5.26	22.10	6.69
BC ₂	8.11	15.09	1.33	-	-
BC ₃	6.92	18.18	4.01	22.67	5.67
BC ₄	5.98	17.61	3.09	21.49	4.90
BC ₅	5.89	17.28	3.10	21.31	3.97

7.3.2 The preliminary analysis of the Backcross population

To investigate the QTL effects associated with the marker loci maintained in the BC population, a single marker analysis was carried out in each generation of the progressive backcross population. The results give an indication of the X-linked QTL position and the size of its effect. In the BC population, markers closely linked to the QTL would provide the highest estimated marker associated QTL effects. As recombinations break down linkage between the QTL and the markers along the chromosome, the estimated marker-QTL effects would decline and standard errors associated with the estimates would increase.

The results presented in Figure 7.4 show the changing marker associated QTL effects along the chromosome estimated in the males and females from generations 3, 4 and 5 of the BC population. The results are presented for the final three generations only, as the numbers of individuals with full information (trait and genotypic data) were limited in the initial generation of the BC population.

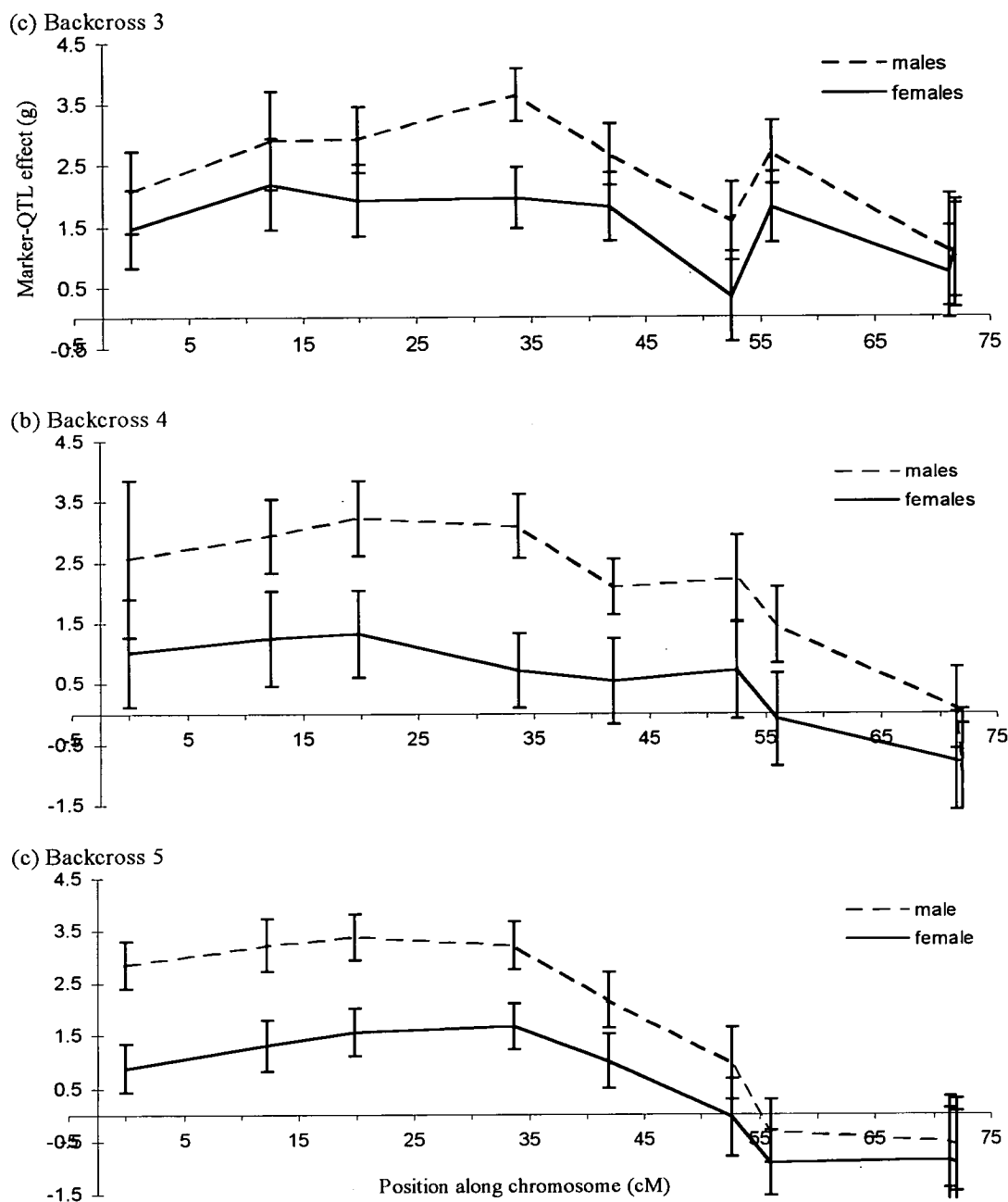


Figure 7.4 The marker associated QTL effects at 10 weeks in males and females in the backcross generations 3, 4 and 5. The male effects show the difference between the two hemizygous marker genotypes and the female effects show the difference between the homozygous low and heterozygous marker genotypes. The error bars represent the standard errors associated with the estimates.

From the single marker analysis in generations 3, 4, and 5 of the backcross population the highest marker associated QTL effects were associated with the markers at the proximal end of the X-chromosome. The highest estimates were found at markers *DXMit187*, *DXMit50* and *DXMit25*, at 12.2cM, 19.8cM and 33.7cM respectively. In all generations, the estimates in males were larger than in females. In the marker region *DXMit187* to *DXMit25* the estimated marker associated QTL effects in males were approximately twice the estimates in females, such that the marker associated QTL effects (g) at *DXMit50* were 2.92 (0.54) and 1.92 (0.58) in generation 3, 3.22 (0.62) and 1.32 (0.71) in generation 4, and 3.38 (0.44) and 1.56 (0.45) in generation 5 in the males and females respectively (standard errors of estimates shown in parenthesis).

In the BC₅ generation, the marker associated QTL effects were highest between *DXMit187* and *DXMit25*, and these markers were associated with the lowest standard errors. The estimated effects declined towards the distal end of the X-chromosome in BC₅, where estimates were negative in males and females but not significantly different from zero ($P > 0.05$ using a Student t-test).

The results of the segment analysis (model 2) for 10 week weight are shown in Table 7.3. The objective of this analysis was to estimate the effects associated with carrying specific segments of the high line X-chromosome on an inbred low line background. The segment estimates represent the difference between the two hemizygous genotypes in males ($2a$) and the difference between the homozygous low and heterozygous genotypes in females ($a+d$).

In males, the marker segments at the most distal end of the X-chromosome (between *DXMit16* and *DXMit31*) were associated with small or negative estimates (these estimates were not significantly different from zero $P > 0.05$). The estimated effects associated with the markers at the most proximal end of the X-chromosome (between *DXMit55* and *DXMit16*) were significantly different from zero ($P < 0.05$), and explained approximately 4g of the difference in mean body weight between the two hemizygote segment groups in males. The estimated effect associated with the marker

segment 7 (*DXMit55* to *DXMit187*) was not significantly different from zero. The same trends are shown in the females from the BC generations 3, 4, and 5, although the estimated effects were approximately half of the estimates in males. Similar trends can be seen in the estimates associated with segments at 6 weeks (data not shown). It should be noted that the origin of the segments of X-chromosome between high and low line markers (i.e. where there has been a recombination event) are not known, but with each generation of backcrossing the percentage of high line X-chromosome between these markers will tend to decline.

Table 7.3 The mean effect (g) for 10 week weight associated with segments of the X-chromosome in males and females from generations 3, 4 and 5 of the progressive backcross population (standard errors shown in parenthesis). The ticks represent markers which can be traced unequivocally to the high selection line, and the remaining markers on the X-chromosome can be traced to the low inbred line.

Estimated mean effects (g)		Mit Marker:								
males	females	<i>DX55</i>	<i>DX187</i>	<i>DX50</i>	<i>DX25</i>	<i>DX16</i>	<i>DX79</i>	<i>DX38</i>	<i>DX121</i>	<i>DX31</i>
0.48 (0.89)	-2.74 (2.00)								♠1 ✓	✓
-0.15 (0.96)	-1.05 (0.59)						2 ✓	✓	✓	
-0.38 (0.72)	-1.80 (0.91)					3 ✓	✓	✓		
4.26 (0.95)	2.03 (0.67)			4 ✓	✓	✓				
4.41 (0.75)	1.82 (0.59)	5 ✓	✓	✓	✓					
3.81 (0.57)	1.09 (0.42)	6 ✓	✓	✓						
1.11 (0.83)	-0.12 (0.77)	7 ✓	✓							
5.50 (0.84)	1.65 (0.78)	8 ✓	✓	✓	✓	✓	✓	✓		

♠ the segment label, enabling segments to be referred to easily, e.g. segment 1 refers to the marker segment *DXMit121* to *DXMit31*

The estimates obtained from the segment analysis suggest that the distal end of the X-chromosome (marker *DXMit16* to *DXMit31*) is not associated with a QTL for body weight, but there is a significant association between the proximal end of the X-chromosome and body weight at 6 and 10 weeks (between markers *DXMit50* to *DXMit25*).

7.3.3 Gibbs sampling of the Backcross population using all information through the pedigree to estimate QTL position(s) and effects

In order to use all information from the BC population, data were analysed using a Gibbs sampling method (section 7.2.4). The results presented in Figure 7.5 show the estimated marginal posterior distribution of the QTL position, using the prior that the QTL position was uniform for the whole X-chromosome, i.e. the probability of the QTL position was equal for all positions on the X-chromosome. The distribution was obtained from the frequency at which the QTL was estimated at each position along the chromosome. The plot shows only the region of the X-chromosome of 21cM to 32cM, as the probability of the QTL in all section outside this area were effectively zero. The mean estimated QTL positions were 25.4cM with an estimated 95% confidence interval of $\pm 2.8\text{cM}$ for 10 week weight, and $26.9\text{cM} \pm 3.0\text{cM}$ for 6 week weight. The estimated QTL positions provide strong evidence that the QTL is positioned within the marker interval *DXMit50* (19.8cM) to *DXMit25* (33.7cM). The shape of the of the curves presented in Figure 7.5 indicate the posterior distributions (and therefore the likelihoods) for QTL positions are irregular and jagged in shape. It should be noted that some of the irregularities may be due to Monte Carlo (sampling) error.

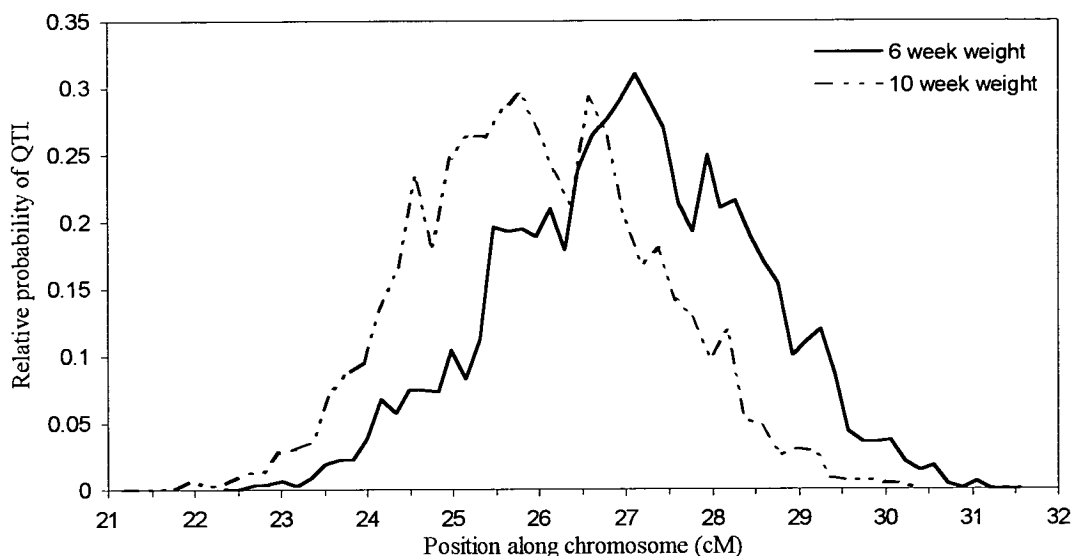


Figure 7.5 The estimated marginal posterior distribution of the QTL position. Analysis was carried out using all pedigree information by Gibbs sampling. The nearest flanking markers are *DXMit50* and *DXMit25* at 19.8cM and 33.7cM respectively.

The mean QTL effects estimated using the Gibbs sampling method are presented in Table 7.4 (when the QTL effects were sampled at the mean estimated QTL position). The estimated marginal posterior distribution for QTL effects (Figure 7.6) show that the sampled QTL effects are normally distributed. The mean QTL effects (Table 7.4) indicate close agreement between the estimated QTL effects in males and females. If we assume no dominance in the females, the estimated QTL effects (representing the difference between the hemizygous males and $2(Aa-aa)$ in females) were approximately 3g at 6 weeks and 4g at 10 weeks in both males and females. These estimated QTL effects were very close to those obtained using the segment analysis (Table 7.3)

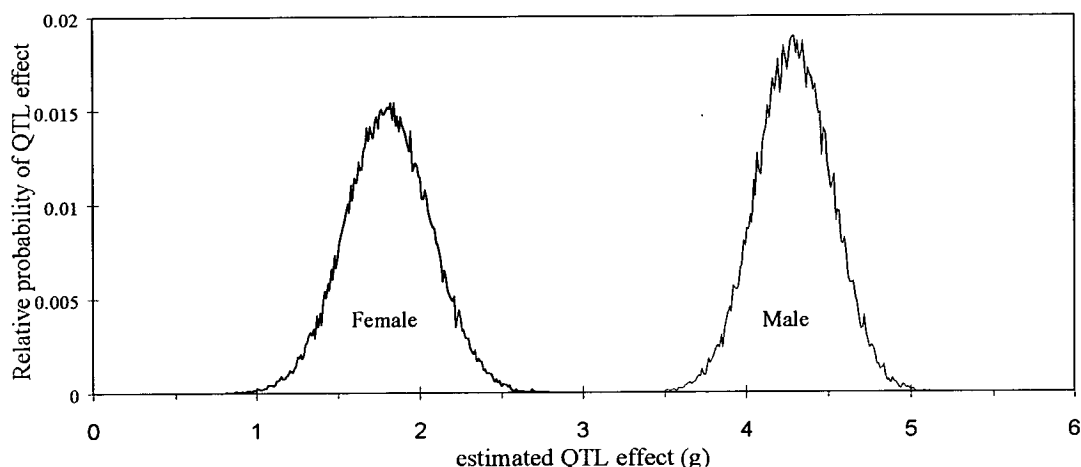


Figure 7.6 The estimated marginal posterior distributions for the QTL effects associated with 10 week weight. The estimated effect in males represents the difference between the two hemizygous genotypes ($2a$), and in females the effect shows the difference between the homozygous low genotype and heterozygous genotype ($a+d$).

Table 7.4 The estimated QTL effects associated with 6 week and 10 week weight obtained from the Gibbs sampling analysis (using the model outlined in section 7.2.4). Standard errors of estimates shown in parenthesis.

Body weight measurement	① Estimated QTL effects (g)		
	Males	Females	
	a	aA	⌕ $2(Aa-aa)$
6 week weight	3.39 (0.22)	1.44 (0.23)	2.88
10 week weight	4.29 (0.23)	1.81 (0.26)	3.62

① The estimated QTL effect are given as a deviation from the hemizygous low line genotype (a) in males, and homozygote low genotype (aa) in females.

⌕ $2(Aa-aa)$ represents the estimate of the total QTL effect (difference between the homozygote genotypes), assuming no dominance deviation.

Assuming no difference between the male genotype group ' a ' and the female genotype ' aa ', the difference in body weight due to sex in the population is approximately 3.12g and 4.33g at 6 weeks and 10 weeks respectively.

7.4 DISCUSSION

7.4.1 Changes in the distribution of the trait measurements through the BC generations

To illustrate the changes in the distribution of the body weights measured in each generation of the BC population, the means and variances were calculated for 6 and 10 week weights (corrected for the fixed effects and covariates in model 1 (section 7.2.3) with the exception of the fixed effect of marker genotype). Correction of the data for the fitted effects removes variance due to sex and common environmental effects associated with litter. The estimated means and variances show (with the exception of the estimates for females in BC₂) that with each generation of backcrossing the mean and variance of the body weight declines. The decrease in mean body weight is expected, as the contribution of high line autosomal QTL influencing body weight is reduced by approximately 50 % in each generation. As shown in section 7.3.1, the decline in mean body weight is associated with a reduction in variance. Again, this occurs as a result of the increased percentage of autosomal QTL from the inbred low line with each generation of backcrossing. The variance associated with inbred lines is lower than in outbreeding populations, due to a decrease in the level of heterozygosity as a result of inbreeding. In backcross generations 3, 4, and 5, the variance declines with respect to the mean 10 week weight, and this is illustrated by the coefficients of variation associated with the mean body weight. The estimated CV were 10.5%, 10.3% and 9.4% in BC generations 3, 4, and 5 respectively, which clearly shows a reduction in the variance associated with the mean body weights over the final generations of backcrossing.

As described above, during the progressive backcross the reduction in the contribution of high line autosomal genes will be approximately 50% in each generation, such that at generation 5 of the BC, the contribution of high line QTL will be less than 2%. The mean body weight of the inbred low line at 6 weeks (corrected

for the fixed effects and covariates in model 1 (section 7.2.3) with the exception of the fixed effect of marker genotype) of approximately 14.2g with a variance of 2.18, is considerably lower than the estimated body weight at generation 5 of the backcross population ($17.28\text{g} \pm 3.10$). The reduction in the mean body weight and particularly the variance of body weights may not be as rapid as expected, as segments of the high line X-chromosome were artificially maintained in each generation of backcrossing. This preserves the effects associated with the X-chromosome in the population, and consequently the decline in the mean and variance is retarded compared to the normal backcrossing situation. It should be noted that in the F_1 generation and the initial backcross generations there may be an increase in mean body weight due to heterosis, although the influence in the final generations will be negligible.

7.4.2 Preliminary analysis of marker data within the BC generations

To estimate the QTL effects associated with each marker maintained in the BC population, a single marker analysis was carried out in the BC generations 3, 4, and 5. The results presented are of the analysis of 10 week weight data. If we assume the QTL is additive (Veerkamp *et al.*, 1993; Chapter 6), the estimates in males (difference between the hemizygotes) would be approximately twice the estimates in females (the difference between the homozygote low and heterozygote genotypes). Comparison of the estimated marker associated QTL effects show that in all BC generations, as expected, the estimates in males are approximately twice the estimates in females.

The marker associated QTL effects along the chromosome indicate that in all generations the highest estimates are at the proximal end of the X-chromosome. Concentrating on the estimates obtained in the BC_5 generations, the highest marker associated QTL effects are linked to markers *DXMit187*, *DXMit50* and *DXMit25* at 12.2cM, 19.8cM and 33.7cM respectively. The difference between the hemizygous males and twice the difference between the heterozygous and homozygous low females, i.e. 2α (assuming no dominance in the females) were approximately 3g in

both sexes at the above mentioned markers. The marker associated estimates obtained in the BC₅ generation indicate the region between 12cM and 34cM is associated with the X-linked QTL. This suggests a wide region of the X-chromosome is associated with the QTL for body weight, as the estimated marker associated QTL effects do not show abrupt changes along the chromosome. The low variation in the estimated QTL effects along the chromosome are presumably due to covariances between the markers (and marker associated QTL effects) maintained in the BC population. The covariances occur as markers were maintained in pairs or groups of three or more in the backcrossing design. If single markers were maintained in the population, the estimates would probably show more abrupt changes in the estimated QTL effects.

The results obtained from the segment analysis (Table 7.3) provide a clearer estimate of the QTL position. Estimating the effects associated with the marker segments prevents the over estimation of marker associated QTL effects due to covariances between the markers (and marker associated effects) seen in the single marker analysis.

The estimated effects associated with segments along the chromosome, enable a single segment of interest to be isolated. As summarised in the results section, the estimates associated with segments at the most proximal end of the X-chromosome (markers *DXMit16* to *DXMit31*) were not significantly different from zero in both males and females. At the proximal end of the X-chromosome markers *DXMit55* to *DXMit16* (segments 4, 5, and 6), were associated with estimated effects of approximately 4g in males and between 1g and 2g in females. The effects associated with the marker segments 3 and 4, indicate markers *DXMit50* to *DXMit25* are associated with effects significantly different from zero, but the effects associated with the marker *DXMit16* (segment 3) were not significantly different from zero. The marker segments 5 and 7 also show the X-linked QTL is associated with the markers *DXMit50* and *DXMit25* but not the markers *DXMit55* and *DXMit187* (segment 7). This information leads to the hypothesis that the X-linked QTL is associated with the marker interval *DXMit50* to *DXMit25*, and that this segment of the X-chromosome

has a total effect ($2a$) of approximately 4g in males and females (assuming no dominance deviation in females). The results do not allow the conclusion that the X-linked QTL is in the interval *DXMit50* to *DXMit25*, as the points at which the recombination break points occurred in the intervals *DXMit187* to *DXMit50* and *DXMit25* to *DXMit16* are not known. To enable information from all individuals in the BC pedigree to be used to estimate the QTL position and effects, data were analysed using the Gibbs sampling based method (section 7.2.4) by Dr S. C. Heath.

7.4.3 Analysis of the BC population using Gibbs sampling

The BC population was analysed using Gibbs sampling (which takes account of the pedigree structure of the population), and this enabled the information collected from all generations of the BC population to be combined to estimate the QTL position and effects. The results presented in Figure 7.5 show the estimated marginal posterior distribution of the QTL position on the X-chromosome for 6 week and 10 week weight data. The frequency at which QTL were sampled outside the interval 21cM to 32cM were close to zero, and this provides convincing evidence that the QTL is positioned in the marker interval *DXMit50* to *DXMit25*, and suggests that there is no evidence for a second QTL elsewhere on the X-chromosome. The estimated positions of the QTL ($\pm 95\%$ CI) for 6 week weight and 10 week weight within this interval were $25.4\text{cM} \pm 2.8\text{cM}$ and $26.9\text{cM} \pm 3.0\text{cM}$ respectively. The positions estimated for the two body weight measurement show overlapping CI for the QTL position, with the estimated position obtained for 6 week weight being slightly more distal to the estimated position for 10 week weight. There is no obvious explanation why there is a shift in estimated QTL position for the 6 week and 10 week weight, except that the information on 10 week weight was incomplete in the BC pedigree. The Gibbs sampling method is able to account for the incomplete 10 week weight data, but in the process of sampling 10 week weight, the shift in estimated position may have resulted.

The estimated effects associated with the X-linked QTL at 6 and 10 weeks are presented in Table 7.4. As described in section 7.3.3, the QTL effects presented were those obtained from the posterior distribution of the QTL effects sampled at the maximum estimated QTL position. However, it should be noted that these estimates were not significantly different from those obtained from the full conditional distribution for the QTL effect, where the estimates from the two distributions differed by less than 0.02g. The QTL effect in males (the difference between the two hemizygous genotypes) were approximately 3.4g at 6 weeks, and 4.3g at 10 weeks. In females, assuming there is no dominance deviation, the estimated QTL effects ($2(Aa-aa)$) were 2.9g and 3.6g at 6 weeks and 10 weeks respectively. Using the mean standard error associated with the estimates shown in Table 7.4 of approximately 0.23, the estimated effects in females were significantly lower ($P < 0.05$ using student t-test) than the estimates in males. No firm conclusions can be drawn from this observation, as the estimated QTL effect in females assumes no dominance deviation. A small and non-significant dominance deviation could result in an underestimation (or overestimation) of the QTL effect calculated by the method described above. Despite the lower estimated QTL effects observed in females compared to males, the estimates do show the same trends in both sexes. However, when the estimated QTL effects in the two sexes are expressed as a percentage of the corresponding mean body weight (corrected for the fixed effects and covariates in model 2, with the exception of the fixed effect of segment), the QTL effects both explain 18% to 19% of the total mean body weights in both males and females.

In summary, the results of analysis of the BC population provide very strong evidence for a QTL of large effect (approximately 18% to 19% of the mean body weight at 6 and 10 weeks) is situated in the marker interval *DXMit50* to *DXMit25*, (19.8cM to 33.7cM). The estimated QTL effects are of equal magnitude to the estimated QTL effects presented in Chapter 6 (16% to 20% of the mean body weight) and appear to explain nearly all the X-linked effect in the P6 lines estimated previously (Hastings, 1990; Hastings & Veerkamp, 1993; Chapter 4). The analysis of data collected in the BC population using Gibbs sampling showed no evidence of additional QTL

positioned elsewhere on the X-chromosome outside the marker interval *DXMit50* to *DXMit25*. The estimated QTL position for the X-linked QTL at 10 weeks was approximately $25.4\text{cM} \pm 2.8\text{cM}$ (representing the 95% CI for the estimated QTL position).

CHAPTER 8

General Discussion

8.1 General discussion of the estimated QTL positions and effects

The objectives of this study were to provide molecular evidence for (one or more) X-linked QTL for body weight initially described by Hastings (1990) in the P6 selection lines. Following confirmation of the X-linked QTL, the objective of the project was to estimate the number of X-linked QTL of large effect, the size of their effects and their position(s).

Brief summary of the QTL mapping results

The preliminary study described in Chapter 4, (where body weight at 6 and 10 weeks was available for 195 F₂ animals, and genotypes were collected at seven microsatellite marker loci on the X-chromosome), provided strong evidence for a QTL of large effect on the X-chromosome. Unfortunately, parental DNA was not available to enable fully informative marker loci to be determined within litters; and as a result, data could not be easily analysed to enable QTL position and effect to be estimated. It was therefore decided that the F₂ experiment should be repeated with parental DNA retained to allow fully informative marker loci to be determined.

In addition to body weight measurements, estimates of carcass fat percentage were calculated so that the contribution of the X-linked QTL for body weight to fat deposition could be estimated. In the repeated F₂ experiment (the F₂X population) the 334 animals were genotyped at 13 marker loci spanning the length of the X-

chromosome (genotypes of the parents allowed fully informative marker loci to be determined).

Analysis of the marker and trait data by the non-inbred X-linked multiple regression method (Chapter 5) showed a single well defined peak in the likelihood curve (at approximately 23cM) for all body weights analysed. The multiple regression analysis of 10 week weight data provided very strong evidence for a single X-linked QTL ($\pm$ s.e.) at approximately 23 ± 2 cM in the marker interval *DXMit50* to *DXMit46* (at 19.8cM and 30.4cM receptively). The estimated QTL effects associated with the 10 week weight data (the difference between the two hemizygote genotypes in males and the two homozygote genotypes in females), were 5.2g (mean s.e. 0.38) in males and females. There was no significant dominance deviation estimated in females. The QTL analysis of the estimated carcass fat percentage data did not indicate an X-linked QTL influencing fat deposition.

In Chapter 7 the progressive backcross (BC) population was described. The objective of the BC population was to backcross specific regions of the P6 high selection line X-chromosome onto the P6 inbred low line background. This was achieved over five generations of backcrossing using nine microsatellite markers spanning the length of the X-chromosome. Analysis of this data set by REML in Genstat 5.3 (Genstat 5 Committee, 1993) using a segment analysis (section 7.2.3) and by Gibbs sampling (section 7.2.4) both provided convincing evidence for a single QTL situated in the marker interval *DXMit50* to *DXMit25* (19.8cM to 33.7cM). There was no evidence for X-linked QTL outside the marker interval *DXMit50* to *DXMit25*. The Gibbs sampling method estimated the QTL position ($\pm$ s.e.) at $25.4\text{cM} \pm 1.43\text{cM}$ with a total effect of approximately 4.3g (s.e. 0.23) at 10 weeks in males, the estimated effect in females was approximately 3.6g assuming no dominance deviation.

8.1.1 The estimated QTL position

The three experiments summarised above provide conclusive molecular evidence for an X-linked QTL for body weight. Evidence from the F_2X and progressive BC populations both indicate a single QTL of large effect. The estimated QTL positions obtained in the F_2X population and the progressive BC population were in good agreement. Figure 8.1 shows the estimated position obtained from the F_2X and the progressive backcross populations for 6 and 10 week weight. The estimated positions in the F_2X population are closer to marker *DXMit50* and have wider CI for the QTL position than those estimated in the progressive BC population. This is as expected, for the size of the progressive BC population was larger, and provided greater opportunity for recombination events. It should be noted that in the F_2X population marker data were available on the additional marker *DXMit46* (positioned at 30.4cM in the marker interval *DXMit50* to *DXMit25*), and this may have accounted for the difference in the estimated QTL positions in the two populations. Despite this fact, when the F_2X population data was analysed excluding the *DXMit46* data the likelihood curve indicated the same estimated QTL position (at 23cM) for 10 week weight.

The progressive BC population provided a powerful method to test for QTL elsewhere on the X-chromosome (as covariances in marker associated QTL effects are broken down by the experimental mating design). Analysis of the progressive BC population by Gibbs sampling (section 7.2.4) indicated no QTL (with effect large enough to detect) on the X-chromosome outside the marker interval *DXMit50* to *DXMit25*.

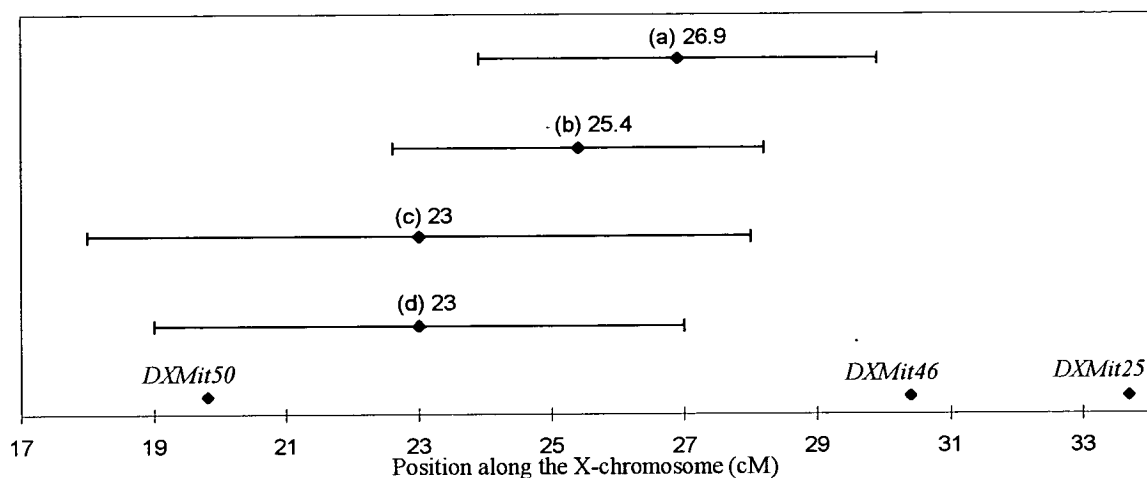


Figure 8.1 Illustration of the QTL positions and 95% CI estimated for 6 and 10 week weight data in the F₂X population and progressive BC population, (a) progressive BC population 6 week weight, (b) progressive BC population 10 week weight, (c) F₂X population 6 week weight, and (d) F₂X population 10 week weight

In summary, the evidence from the analysis of the F₂X population and the progressive BC population provide strong support for the hypothesis of a single X-linked QTL (or a number of very closely linked QTL) positioned in the marker interval *DXMit50* to *DXMit25*. Pooling the estimates for the QTL positions obtained in the two populations by scaling the estimates by the variance, the estimated QTL position ($\pm$ s.e.) was 24.6 ± 1.15 cM for 10 week weight and 25.8 ± 1.29 cM for 6 week weight. It should be noted that the QTL positions were estimated using two different methods in the two experiments, using multiple regression (Chapter 5) in the F₂X population and by Gibbs sampling assuming uniform priors for the QTL parameters in the progressive backcross population.

8.1.2 The QTL effect associated with the X-chromosome

Following the approximation of the QTL position, the size of the QTL effect (associated with the QTL between the markers *DXMit50* and *DXMit25*) must be considered. The estimated effects in males (i.e. the total difference between the hemizygote male genotypes) will be used to consider the size of the X-linked effect, as the additive effect can be estimated in females from the BC population only by assuming no dominance. It should be noted that the results presented in Chapter 6 provided strong evidence that the estimated QTL effect was of equal magnitude in males and females, and this supports previous investigations (Veerkamp *et al.*, 1993) where segregation analysis indicated that the X-chromosome was associated with an additive effect.

The X-linked QTL effects estimated in the F_2X population and the progressive BC population are summarised in Table 8.1 (the QTL estimates are shown in bold), where the effects in males are between 3.4g and 4.0g at 6 weeks and between 4.3g and 5.2g at 10 weeks (mean s.e. 0.2 - 0.3) in the BC population and F_2X population respectively. These estimates are in fairly close agreement. It should, however, be noted that the mean body weight of individuals from the two populations are very different (as a result of the different autosomal backgrounds). When the estimates are expressed as a percentage of the mean body weight, the QTL appears to have a very similar percentage contribution in both populations, which suggests the X-linked QTL has a multiplicative effect on mean body weight.

Table 8.1 Summary of estimated X-chromosome effects and estimated X-linked QTL effects in males, and the contributions to mean body weight expressed as a percentage of mean male body weight. The estimated QTL effects are shown in bold.

Body weight (g)	Source of data (method of estimation shown in parenthesis)	Mean body weight (g)	X-linked (QTL) effect (g)	
			estimate	percentage of mean body wt
6 week	② P6 lines Gen 31 (F ₁)	27.4	5.7	20.8
	Chapter 4 (F ₁)	25.2	3.1	12.3
	Chapter 6 (F ₁)	23.7	3.2	13.5
	Chapter 6 (Multiple reg.)	25.1	4.0	15.9
	Chapter 7 (Gibbs)	19.1	3.4	18.2
10 week	② P6 lines Gen 31 (F ₁)	33.2	6.8	20.5
	① P6 lines Gen 38 (F ₁)	30.9	4.6	14.9
	Chapter 4 (F ₁)	28.8	6.3	21.9
	Chapter 6 (Multiple reg.)	31.5	5.3	16.8
	Chapter 7 (Gibbs)	23.8	4.3	18.9

① from Hastings & Veerkamp (1993), ② from Hastings (1990)

F₁ = X-chromosome effect estimated in an F₁ population using the model described in section 3.5, Multiple reg. = estimated QTL effects obtained by non-inbred X-linked multiple regression method (Chapter 5), and Gibbs = estimated effect obtained by Gibbs sampling (see section 7.2.4)

As outlined above, the X-linked QTL appears to have an equal effect on mean body weight in both the F₂X and the progressive BC populations, but does the QTL account for the entire X-linked effect estimated from the F₁ populations ? To answer this question, the estimated QTL effects in males were compared to the X-linked effects in the F₁ generations (from Hastings & Veerkamp, 1993; Hastings, 1990 and Chapters 4 and 6), expressed as a percentage of the mean F₁ male body weight. From the estimated X-linked effects in the F₁ generations (Table 8.1), the estimated percentage effects at 6 weeks are between 12% and 21%, and at 10 weeks are

between 15% and 23% of the mean male body weight. The percentage influence of the QTL are in close agreement with the percentage contribution of the X-chromosome estimated from the F_1 males. This indicates that the single X-linked QTL appears to explain almost the total contribution of the X-chromosome, increasing mean body weight by approximately 18%.

Estimation of the X-linked effects in all three of the P-replicate lines by Hastings and Veerkamp (1993) provided convincing evidence that the X-linked effect was present in the base populations, rather than due to a mutations which occurred during the selection process. As described in section 2.3.2, the P-lines were derived from an F_1 cross between two inbred lines (JU and CBA) which was crossed to an outbred population (CFLP) from the Carnworth Laboratory. The two inbred lines JU and CBA are still available, and therefore if we assume the QTL originates for a single line, a reciprocal F_1 population made between these two inbred lines would provide a simple test for the presence of the X-linked QTL. Investigating if the X-linked QTL originated for the outbred populations is more problematic as, if the population is still available, it is unlikely to resemble the original CFLP due to random drift. However, there are a number of inbred lines derived from the Carnworth Lab outbred population, e.g. CFW and CFCW. These lines could be used to test for the X-linked effect if the inbred lines failed to show the X-linked effect.

8.1.3 Conclusions from the QTL mapping on the X-chromosome

In conclusion, the evidence obtained in this study indicates a single QTL of large effect (or a number of very closely linked QTL) situated in the marker interval *DXMit50* to *DXMit25* on the X-chromosome. The estimated QTL positions (in the studies described) indicate the estimated QTL position ($\pm$ s.e.) of 24.6 ± 1.15 cM for 10 week weight. The single QTL has an additive effect, showing no significant dominance deviation in females. The magnitude of the QTL effects are equal in both

sexes, explaining approximately 18% of the mean body weight at 10 weeks, and this approximately explains the entire effect associated with the X-chromosome.

8.2 The X-linked QTL for body weight and candidate loci

As discussed in the previous section, experimental results (Chapters 6 and 7) show convincing evidence for a single QTL of large effect on the X-chromosome. This section will compare the findings presented in this study, with other loci mapped on the X-chromosome.

In a study by Dragani *et al.* (1995) where two interspecific test-cross populations were made, HSB: female C3H/He $\times$ *M.spretus* and male C57BL/6J; and ASB: female A/J $\times$ *M.spretus* and male C57BL/6J, linkage analysis showed a strong indication that a region of the genome important in determining quantitative variation in body weight at 40 weeks was present on the X-chromosome. Using composite interval mapping (Zeng, 1994), two complementary QTL were isolated on the X-chromosome in the HSB test-cross (Bwt1 and Bwt2), together explaining 25.6% of the phenotypic variance. Linkage analysis of the ASB test-cross analysis suggested 3 QTL (Bwt1, Bwt2 and Bwt3) positioned on the X-chromosome explaining 24% of the phenotypic variance. The linkage analysis indicated the female groups (both ASB and HSB) did not show a significant association with putative QTL in the test-crosses. The Dragani *et al.* (1995) study provided evidence for X-linked QTL for body weight in males (Bwt1 and Bwt2) supported by two independent test-cross populations. The authors suggest that the QTL (Bwt1 and Bwt2) may have a recessive action (as there was no indication for the X-linked loci in the females).

Comparison of the results obtained by Dragani *et al.* (1995) (isolating two body weight QTL in common to the two test-crosses) and results obtained in this study (isolation of one loci of large effect) are unable to provide conclusive evidence that the QTL mapped in the two studies are the same loci. In the study by Dragani *et al.*

(1995) the results indicate the QTL is recessive, being only detectable in the hemizygous males. In the study presented here, the QTL has an additive action (Chapter 6), having equal effect in males and females at 10 weeks (Chapters 6 and 7). It should be noted that in the Dragani *et al.* (1995) study, data were analysed by composite interval mapping, there was no mention of tailoring the analysis for X-linked analysis, and the results obtained producing LOD scores between 3 and 7. As discussed in section 5.4.2, females do not provide as much power for X-linked QTL detection as males. The lack of evidence for the X-linked QTL in females may be a result of the lower contribution of females to the total power of X-linked QTL detection. Without access to the data, conclusions cannot be drawn to whether there was no significant indication of QTL in the females due to lower power, or the X-linked QTL did not contribute to the mean body weight in the test-cross females.

Comparison of the estimated QTL positions in the two studies show that the X-linked QTL mapped in this study (mapped to in the marker interval *DXMit50* to *DXMit25* on the X-chromosome) and the QTL Bwt1 mapped by Dragani *et al.* (1996) are both associated with the same chromosomal region. Markers *DXMit50* and *DXMit48* (closely linked to Bwt1) have the same estimated Mit map position on the X-chromosome (22.5cM). The estimated QTL effect at Bwt1 is approximately 17% of the mean body weight in the HSB mice and 14% in the ASB mice. Although the mode of action of the QTL in the two studies are different (additive in this study and recessive in the Dragani *et al.* (1996) study), the map positions indicate Bwt1 and the X-linked QTL in this study map to similar chromosomal regions. The origin of the body weight QTL in the same study is unlikely to be from common ancestors, as the X-linked QTL in the Dragani *et al.* (1996) study were derived from the *M.spretus* parental strains. It is clear that further investigation is needed before conclusion can be drawn to whether these QTL may be the same loci. The further investigations are likely to involve backcrosses between the *M.spretus* and P6 mouse lines, (as the QTL (Bwt1 and Bwt2) were observed in males from both ASB and HSB test-cross populations, and therefore it is likely these QTL originate from the common dam genotype, *M.spretus*). Molecular investigation of the region associated with the X-

linked QTL in the two studies, could determine if the X-linked QTL (from the two mouse lines) segregating on different genetic backgrounds resulted in the recessive and additive QTL phenotypes.

In addition to examining evidence for X-linked QTL influencing body weight isolated in other studies, candidate loci for the X-linked QTL for body weight were investigated. The Mouse Genome Database (1996) was used to search for candidate loci. The search was restricted to the region 10cM to 40cM (MGD map positions) using the following allele search types (the number of positive results are shown in parenthesis):

- Endocrine defects, hormones, growth and obesity (0)
- Enzymes (8)
- Receptors (4)
- Miscellaneous (2)

From the positive results obtained, only two of loci detected could influence growth and body weight:

30.0cM	G6pdx	Glucose-6-phosphate dehydrogenase-1 X-linked
34.7cM	Phk	Phosphorylase kinase.

G6pd is the first enzyme in the pentose phosphate pathway. The pentose phosphate pathway is a series of reactions with two purposes; the generation of NADPH for reductive biosynthesis and the formation of ribose-5-phosphate for synthesis of nucleotides (Stryer, 1981). The enzyme Phk is active in Glycogen metabolism (Stryer, 1981). However, the MGD map positions for these enzymes do not fall within the marker interval *DXMit50* to *DXMit25* (MGD positions 12.5cM and 28.0cM respectively). Therefore, if we assume the estimated QTL position is accurate, these loci are not suitable candidate loci. The Mouse Genome Informatics is constantly updated, therefore future investigation may provide suitable candidate loci. It may also be possible to isolate candidate loci by using the human/mouse genome comparative map. However, like the results presented above for the mouse genome search, no suitable loci have been isolated to date.

8.3 Future work

The work described in this thesis has been able to isolate the X-linked QTL of large effect to the marker interval *DXMit50* to *DXMit25*. The future goals for the X-linked QTL project are to map genetically the QTL to a high resolution, thus providing conclusive evidence there is a single QTL for body weight present on the X-chromosome. Fine resolution mapping will narrow the candidate region to a small genomic segment suitable for positional cloning. The ultimate objective would be to clone the X-linked gene, thus allowing molecular characterisation of the gene product, and enabling the role of the QTL in growth to be studied.

The first step in the cloning of the X-linked QTL would be the fine resolution mapping of the QTL. The congenic lines produced in the progressive BC experiment (Chapter 7) provide an excellent genetic resource for fine mapping. Evidence summarised in section 8.1 indicate the QTL is positioned in the interval *DXMit50* to *DXMit25*. The congenic line carrying the high line X-chromosome segment *DXMit50* to *DXMit25* on the inbred low line background will form the basis of the fine resolution mapping.

Fine mapping of the X-linked QTL would be based on animals recombinant in the interval *DXMit50* to *DXMit25* and a progeny test of these animals (as non-genetic noise can be reduced through progeny testing). The fine mapping would be substantially dependant on the density of available markers. Investigations have shown that there are approximately six additional informative Mit markers in the interval *DXMit50* to *DXMit25*. Clearly an additional source of polymorphic markers would be needed to narrow the interval to less than 0.1cM. Another source of markers are Single-Stranded Conformation Polymorphism (SSCP) and Random Amplified Polymorphic DNA (RAPD). The SSCP, provide a powerful method to detect single base pair differences between the high and low line X-chromosome; and where sequence tagged sites are not available in the region, (i.e. a gap in the physical map) RAPDs could be developed. It is hoped that these two forms of markers would

provide a large number of informative marker loci on the X-chromosome to enable a dense map to be produced.

The mating design used could be based on a backcross, using the females from the progressive BC population and males from the inbred low line. The progeny from this mating would be genotyped at the flanking markers to allow recombinants to be detected. The individuals recombinant for the region *DXMit50* to *DXMit25* (or possibly a further narrowed region) would be genotyped at additional markers to pinpoint the recombination break points. To allow the QTL to be isolated to within 0.1cM, approximately 1000 animals would be required.

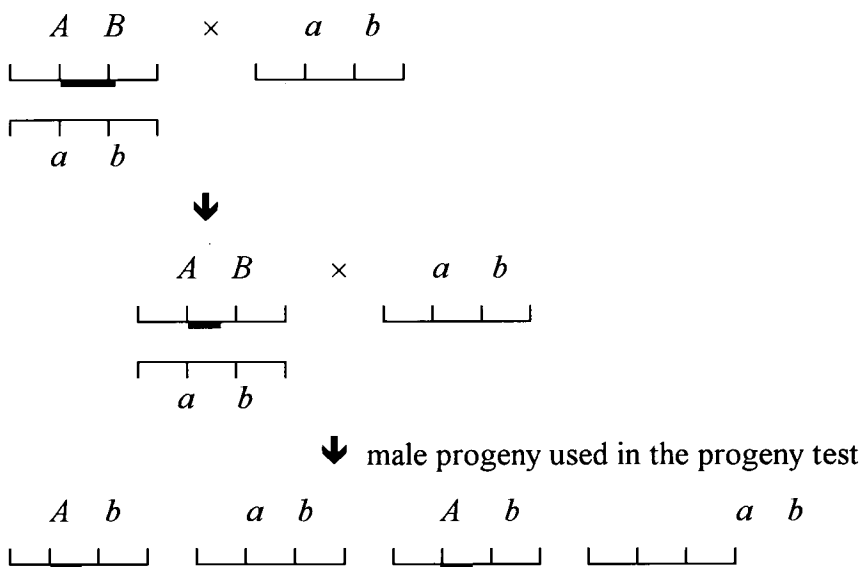


Figure 8.2 Illustration of the mating design of the population proposed used to fine map the X-linked QTL, and the progeny test of animals. The flanking markers genotyped are *A/a* and *B/b*, (*A* and *B* from the high line, and *a* and *b* from the low line). Female progeny carrying the *A* allele would be used to produce a large number of males for a progeny test.

If we assume preliminary analysis of the 1000 animals described above would enable the QTL to be isolated to within 5-10cM region, a subsequent progeny test would be

carried out for 50 to 100 families using the design illustrated in Figure 8.2.

Preliminary analysis of data collected during the course of the experiment is likely to enable the progeny test (i.e. one litter from each of the recombinant females) to be limited to considerably less than 50 families. The progeny test would allow other traits to be investigated, e.g. body weight at weekly intervals, muscle fibre density, the distribution of muscle on the carcass, and fat percentage (although prior evidence from Chapter 6 indicated there was no significant contribution of carcass fat percentage to the X-linked QTL for body weight).

Analysis of the data set generated from the large backcross population proposed will be dependant on whether it is possible to determine the QTL genotype of a given animal by a phenotypic measurement. To date it has not been possible to determine the QTL genotype by body weight measurements, but future investigation may allow this to be determined, e.g. by muscle fibre density, or the mass of specific muscle groups. If the QTL genotype of individuals remains unable to be determined, the analysis of data will be based on interval mapping (see section 2.2.2 and Chapter 5). If the QTL genotypes can be resolved, the experiment will be more powerful (alleviating the need for a progeny test) and the analysis will be based on simply isolating segregating marker loci in complete linkage with the QTL (i.e. by a simple linkage analysis).

Once the QTL is fine mapped to approximately 0.1cM, YAC(s) containing the QTL can be identified from the physical map.

Identifying the X-linked QTL and characterising the mode of action of the QTL will greatly increase knowledge regarding the control of the quantitative trait of body weight. In addition to the benefits which can be gained from greater understanding of the genetic control of body weight, there may be additional benefits to agriculture and human research (via the comparative mouse/human map project). Growth is of utmost importance in agricultural production systems, identification of a QTL for growth in commercial species would potentially have large economic benefits. Because the QTL for body weight mapped in this study is on the X-chromosome it is of particular

interest. As outlined in section 2.4.3, the genes on the X-chromosome are perhaps entirely conserved among mammals, and this provides the possibility of identification of loci influencing body weight in other mammalian species.

8.4 General Discussion of QTL mapping in Animal Breeding

This discussion will present a general overview of QTL mapping in the context of animal breeding. To date there have been a large number of experiments carried out to map QTL in the mouse (Collins *et al.*, 1993; West *et al.*, 1994; Das *et al.*, 1996, Taylor & Phillips, 1996, York *et al.*, 1996; Brockmann *et al.*, 1996; Cheverud *et al.*, 1996) and to a more limited extent in commercial livestock species (Clamp *et al.*, 1992; Andersson *et al.*, 1994; Larsen *et al.*, 1995; Geldermann *et al.*, 1996). These experiments have been successful in identifying areas of the genome associated with QTL and enabled statistical methods to be applied to experimental data.

The methodology for mapping QTL (e.g. Lander & Botstein, 1989; Haley & Knott, 1992) have been available for a number of years, and investigations into the relative power of different experimental designs have been extensive (Shrimpton & Robertson, 1988a,b; Lander & Botstein, 1989; Soller & Beckmann, 1990; Darvasi & Soller, 1992). Despite the wealth of literature published to allow the optimum power to be obtained from an experiment for a given level of expenditure, workers using mammalian species, still appear to implement blanket genotyping of the genome in moderate population sizes (i.e. only the investigation by Taylor & Phillips, 1996 mentioned above used a specific time/cost saving experimental design - pooled genotyping). The results obtained produce low maximum test statistics and large CI for the isolated QTL, and this may result in doubts regarding the repeatability of the results obtained. There is no doubt that the results published are valuable, particularly for driving the development of more effective statistical procedure. But while statistical methods are continuing to develop, the experimental designs used by

practical workers fail to implement the abundance of cost and time saving experimental designs.

Over the past few years there has been a number of papers published showing QTL for traits important for commercial production effectively mapped to areas of the genome (in mice mapping body weight QTL: Collins *et al.* (1993), Das *et al.* (1996), Brockmann *et al.* (1996), Cheverud *et al.* (1996), and mapping loci influencing carcass fat percentage West *et al.* (1994), Taylor & Phillips (1996), York *et al.* (1996)), but how is the work proceeding? In my opinion, if the knowledge regarding the mechanisms controlling quantitative traits is to advance, workers must go further than showing broad areas of the genome are associated with QTL. If the experimental designs described in section 2.2.4, e.g. pooled genotyping (Wang & Paterson, 1994), selective genotyping (Lander & Bostein, 1989; Darvasi & Soller, 1992) and sequential sampling (Motro & Soller, 1993) were implemented to reduce the laboratory cost, and if large initial populations were used to map QTL, it may be possible to resolve QTL to narrow regions of the genome. QTL should be mapped with high maximum TS (i.e. $P < 0.001$ for a genome wide scan) to regions of 10 - 15cM. These experiments should lead the way to attempt to clone QTL of large effect, i.e. approximately one phenotypic standard deviation (or more). These loci are likely to be few, but by concentrating on these loci, we will make a greater impact on increasing our knowledge of the mechanisms behind the control of quantitative traits than by producing many papers showing linkage to broad regions of the genome.

After more than five years of mapping QTL using abundant molecular markers, workers must now strive to resolve QTL to narrow regions of the genome using novel experimental designs (e.g. lines congenic for the regions of interest). Mapping QTL to a high resolution will be costly and time consuming, but with collaboration between institutes (i.e. using specialist knowledge for experimental design and for analysis of data), I believe this is attainable. Hopefully in the next five years of research the QTL described in this study and other such loci of large effect, will be mapped to a high resolution and possibly cloned!

References

- Andersson, L., C. S. Haley, H. Ellegren, S. A. Knott, M. Johansson, K. Andersson, L. Andersson-Eklund, I. Edfors-Lilja, M. Fredholm, I. Hansson, J. Håkansson, and K. Lundström (1994) Genetic mapping of Quantitative Trait Loci for growth and fatness in pigs. *Science*, **263**: 1771-1774.
- Beckmann, J. S. and M. Soller (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.*, **67**: 35-43
- Beniwal, B. K., I. M. Hastings, R. Thompson, and W. G. Hill (1992a) Estimation of changes in genetic parameters in selected lines of mice using REML with an animal model. 1. Lean mass. *Heredity*, **69**: 352-360
- Beniwal, B. K., I. M. Hastings, R. Thompson, and W. G. Hill (1992b) Estimation of changes in genetic parameters in selected lines of mice using REML with an animal model. 2. Body weight, body composition and litter size. *Heredity*, **69**: 361-371
- Bishop, S. C., and W. G. Hill (1985) Effects of selection on growth, body composition, and food intake in mice. III. Correlated responses: growth, body composition, food intake and efficiency and catabolism. *Genetical Research*, **46**: 57-74
- Blair, H. J., and Y. Boyd (1994) Positioning of 14 simple sequence repeat loci relative to markers on the gene-based map of the mouse X-chromosome. *Mouse Genome* (March) 127-128
- Brien, F. D., G. L. Sharp, W. G. Hill and A. Robertson (1984) Effects of selection on growth, body composition and food intake in mice. II. Correlated responses in reproduction. *Genetical Research*, **44**: 73-85

- Brockdorff, N., G. Kay, S. Smith, J. T. Keer, R. M. J. Hamvas, S. D. M. Brown, S. Rastan, (1991) High density molecular map of the central span of the mouse X-chromosome. *Genomics*, **10**: 17-22
- Brockmann, G, D.Timtchenko, P. Das, U. Renne, G. Freyer, S.Kuhla, F.Teuscher, J. Wolf, C. Kuhn, and M. Schwerin (1996) Detection of QTL for body-weight and body-fat content in mice using genetic-markers. *Journal Of Animal Breeding And Genetics-Zeitschrift Fur Tierzucht Und Zuchtungsbiologie*, **113**: 373-379
- Charlesworth, B., J. A. Coyne, and N. H. Barton (1987) The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, **130**: 113-146.
- Cheverud, J. M., E. J. Routman, F. A. M. Duarte, B. van Swinderen, K. Cothran, and C. Perel (1996) Quantitative trait loci for Murine growth. *Genetics*, **142**: 1305-1319
- Churchill, G. A., and R. W. Doerge (1994) Empirical threshold values for quantitative trait mapping. *Genetics*, **138**: 963-971
- Clamp, P.A., J.E. Beever, R.I. Fernando, D.G. McLaren, and L.B. Schook (1992) Detection of linkage between genetic-markers and genes that affect growth and carcass traits in pigs. *Jour. Anim. Sci.*, **70**: 2695-2706
- Collins, A.C, I.C.A. Martin, and B.W. Kirkpatrick (1993) Growth quantitative trait loci (QTL) on mouse chromosome-10 in a Quackenbush-Swiss x C57BL/6J backcross. *Mammalian Genome*, **4**: 454-458
- Darvasi, A., and M. Soller (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.*, **85**: 353-359.
- Darvasi, A., and M. Soller (1994) Optimum spacing of genetic markers for determining linkage between maker loci and quantitative trait loci. *Theor. Appl. Genet.*, **89**: 351-357
- Darvasi, A., A. Weinreb, V. Minke, J. I. Weller, and M. Soller (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics*, **134**: 943-951.

- Das, P., G. Brockmann, L. Meyer, U. Renne, G. Freyer, and M. Schwerin (1996) The effect of a restricted region of chromosome-11 on body- weight in mice under special consideration of the growth- hormone gene locus. *Archiv Fur Tierzucht-Archives Of Animal Breeding*, **39**: 185-194
- Dib, C., S. Fauré, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morissette and J. Weissenbach (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380**: 152-154.
- Dietrich, W., H. Katz, S. E. Lincoln, H. Shin, J. Friedman, N. C. Dracopoli, and E. S. Lander (1992) A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics*, **131**: 423-447.
- Dietrich, W. F., J. C. Miller, R. G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D. C. Joyce, M. Wessel, R. D. Dredge, A. Marquis, L. D. Stein, N. Goodman, D. C. Page, E. S. Lander (1994) A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics*, **7**: 220-225.
- Dietrich W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M. J. Daly, K. A. Ingalls, T. J. O'Connor, C. A. Evans, M. M. DeAngelis, D. M. Levinson, L. Kruglyak, N. Goodman, N. G. Copeland, N. A. Jenkins, T. L. Hawkins, L. Stein, D. C. Page, and E. S. Lander (1996) A comprehensive genetic map of the mouse genome. *Nature*, **380**: 149-152
- Dragani, T. A., Z. -B. Zeng, F. Canzian, M. Gariboldi, M. T. Ghilarducci, G. Manenti, and M. A. Pierotti (1995) Mapping of body weight loci on mouse chromosome X. *Mammalian Genome*, **6**: 778-781.
- Falconer, D. S. and T. F. C. Mackay (1996) Introduction to quantitative genetics, 4rd Edition, Longman Scientific and Technical.
- Geldermann, H., E. Muller, P. Beeckmann, C. Knorr, G. Yue, and G. Moser (1996) Mapping of quantitative-trait loci by means of marker genes in F₂ generations of Wild boar, Pietrain and Meishan pigs. *Journal Of Animal Breeding And Genetics-Zeitschrift Fur Tierzuchtung Und Zuchtungsbiologie*, **113**, 381-387

- Geman, S. and D. Geman (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721-741
- Genstat 5 Committee (1993) *Genstat 5 Release 3 Reference Manual*. Clarendon Press, Oxford.
- Griffing, B. (1965) Influence of sex on selection. *Australian Journal of Biological Science*, **18**: 1157-1170
- Hagger, C., and G. Stranzinger (1992) A method to detect large influences of specific sex chromosomes on quantitative traits. *Theor. Appl. Genet.*, **83**: 925-929.
- Haldane, J. B. S. (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. *J. Genet.*, **8**: 299-309
- Haley, C. S., and S. A. Knott (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**: 315-324.
- Haley, C. S., S. A. Knott and J. M. Elsen (1994) Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics*, **136**: 1195-1207.
- Hastings, I. M. (1990) An investigation of mouse lines selected on body weight for the presence of major genes. *Proc. 4th. World Cong. Genet. Appl. Livest. Prod. Edinburgh XIII*: 317-320
- Hastings, I. M., and W. G. Hill (1989) A note on the effect of different selection criteria on carcass composition in mice. *Anim. Prod.*, **48**: 229-233.
- Hastings, I. M., and W. G. Hill (1990) Analysis of lines of mice selected for fat content. 2. Correlated responses in the activities of enzymes involved in lipogenesis. *Genet. Res. Camb.*, **55**: 55-61
- Hastings, I. M., and R. F. Veerkamp (1993) The genetic basis of response in mouse lines divergently selected for body weight or fat content. 1. The relative contributions of autosomal and sex-linked genes. *Genet. Res. Camb.*, **62**: 169-175
- Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**: 97-109

- Heath S. C. (1995) *Inferences on the Genetic Control of Quantitative Traits from Selection Experiments*. PhD Thesis, University of Edinburgh
- Hofker, M. H., M. I. Skraastad, A. A. B. Bergen, M. C. Wapenaar, E. Bakker, A. Millington-Ward, G. J. B. van Ommen, and P. L. Pearson (1986) The X-chromosome shows less genetic variation at restriction sites than the autosomes. *Am. J. Hum. Genet.*, **39**: 438-451.
- Horvat, S., and J. F. Medrano (1995) Interval mapping of high growth (*hg*), a major locus that increases weight-gain in mice. *Genetics*, **139**: 1737-1748.
- Jansen, R. C. (1993) Interval mapping of multiple quantitative trait loci. *Genetics*, **135**: 205-211.
- Jansen, R. C. (1994) Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, **138**: 871-881.
- Jansen, R. C. (1996) A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics*, **142**: 305-311.
- Jansen, R. C. and P. Stam (1994) High resolution of quantitative traits into multiple loci via Interval mapping. *Genetics*, **136**: 1447-1455.
- Jiang, C., and Z-B. Zeng (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, **140**: 1111-1127
- Knapp, S. J. (1991) Using molecular marker to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progeny. *Theor. Appl. Genet.*, **81**: 333-338.
- Knott, S. A., and C. S. Haley (1992) Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. *Genet. Res., Camb.*, **60**: 139-151.
- Kong, A. (1991) Analysis of pedigree data using methods combining peeling and Gibbs sampling. *Computer Science and Statistics Proceedings of the 23rd symposium on the Interface*. 379-385
- Lander, E. S., and D. Botstein (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**: 185-199.
- Lander, E. S., and L. Kruglyak (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, **11**: 241-247.

- Lander, E. S., and N. J. Schork (1994) Genetic dissection of complex traits. *Science*, **265**: 2037-2048
- Larsen, N.J., H.Ellegren, P.B.Nielsen, and L.Andersson (1995) Genetic-variation at the growth-hormone locus in a wild pig intercross - test of association to phenotypic traits and linkage to the blood-group-d locus. *Theor. Appl. Genet.*, **91**: 1074-1077
- Luo, Z. W. (1993) The power of two experimental designs for detecting linkage between a marker locus and a locus affecting a quantitative character in a segregating population. *Genetics Selection Evolution*, **25**: 249-261.
- Luo, Z. W., and M. J. Kearsey (1992) Interval mapping of quantitative trait loci in an F_2 population. *Heredity*, **69**: 236-242.
- Lyon, M. F. (1992) Some milestones in the history of X-chromosome interaction. *Annu. Rev. Genet.*, **26**: 17-28.
- Lyon, M. F. (1996) Pinpointing the Centre. *Nature*, **379**: 116-117.
- Mangin B., B. Goffinet, and A. Rebai (1994) Constructing confidence intervals for QTL location. *Genetics*, **138**: 1301-1308.
- Martinez, O., and R. N. Curnow (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.*, **85**: 480-488.
- Mbaga, S. H. (1996) *Analysis and Inferences from long-term quantitative genetic selection experiments*. PhD Thesis, University of Edinburgh.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953) Equations of state calculations by fast computing machines. *J. Chem. Physics*, **21**: 1087-1091
- Moreno-Gonzalez, J. (1993) Efficiency of generations for estimating marker-associated QTL effects by multiple regression. *Genetics*, **135**: 223-231.
- Motro U., and M. Soller (1993) Sequential sampling in determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.*, **85**: 658-664.
- Mouse Genome Informatics (1996) *Mouse Genome Informatics Project, The Jackson Laboratory, Bar Harbor, Maine*. World Wide Web (URL:<http://www.informatics.jax.org>), July.

- Mullins, L. J., S. G. Grant, D. A. Stephenson, and V. M. Chapman (1988) Multilocus molecular mapping of the mouse X-chromosome. *Genomics*, **3**: 187-194
- Ohno, S., (1967) *Sex chromosomes and sex-linked genes*. New York, Springer.
- Ooijen, J. W. van (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.*, **84**: 803-811.
- Paterson, A. H., S. Damon, J. D. Hewitt, D. Zamir, H. D. Rabinowitch, S. E. Lincoln, E. S. Lander, and S. D. Tanksley (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics*, **127**:181-197.
- Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, and S. D. Tanksley (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, **335**: 721-726.
- Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan and N. Brockdorff (1996) Requirement for *Xist* in X-chromosome inactivation. *Nature*, **379**: 131-137
- Riska, B., W. Atchley and J. Rutledge (1984) A genetic analysis of targeted growth in mice. *Genetics*, **107**: 79-101
- Rooney, P. (1996) *personal communication*
- Routman, E. and J. Cheverud (1994) A rapid method of scoring simple sequence repeat polymorphisms with agarose-gel electrophoresis. *Mammalian Genome*, **5**: 187-188
- Ryan, B. F., B. L. Joiner and T. A. Ryan, Jr. (1985) *Minitab Handbook*, 2nd Edition. PWS-KENT publishing co., Boston.
- Sakamoto, Y., M. Ishiguro and G. Kitagawa (1986) *Akaike Information Criterion Statistics*. KTK Scientific Publishers, Tokyo; cited in Jansen, R. C. (1993) Interval mapping of multiple quantitative trait loci. *Genetics*, **135**: 205-211.
- Sambrook, J., E. F. Fritsch, and T. Maniatis (1989) *Molecular Cloning - A laboratory manual*. 2nd Edition. Cold Spring Harbor Laboratory Press.
- Sax, K. (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics*, **8**: 552-560.

- Sharp, G. L., W. G. Hill, and A. Robertson (1984) Effects of selection on growth, body composition and food intake in mice. 1. Responses in selected traits. *Genet. Res. Camb.*, **43**: 75-92.
- Shrimpton, A. E., and A. Robertson (1988a) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. I. Allocation of third chromosome sternopleural bristle effects to chromosome sections. *Genetics*, **118**: 437-443
- Shrimpton, A. E., and A. Robertson (1988b) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics*, **118**: 445-459
- Soller, M., and J. S. Beckmann (1990) Marker-based mapping of quantitative trait loci using replicated progenies. *Theor. Appl. Genet.*, **80**: 205-208.
- Soller, M., and J. S. Beckmann (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theor. Appl. Genet.*, **67**: 25-33.
- Stryer, I. (1981) *Biochemistry*, 2nd Edition. W. H. Freeman & Company, San Francisco.
- Taylor, B. A. and S. J. Phillips (1996) Detection of obesity qtls on mouse chromosome-1 and chromosome- 7 by selective DNA pooling. *Genomics*, **34**: 389-398
- Thoday, J. M., (1961) Location of polygenes. *Nature* **191**: 368-370.
- Veerkamp, R. F., C. S. Haley, S. A. Knott, and I. M. Hastings (1993) The genetic basis of response in mouse lines divergently selected for body weight or fat content. II. The contribution of genes with a large effect. *Genet. Res., Camb.*, **62**: 177-182
- Visscher, P. M., R. Thompson, and C. S. Haley (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics*, **143**: 1013-1020
- Visscher, P. M., and C. S. Haley (1996) Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. *Theor. Appl. Genet.*, (in press)
- Wang, G. -L. and A. H. Paterson (1994) Assessment of DNA pooling strategies for mapping of QTLs. *Theor. Appl. Genet.*, **88**: 355-361

- West, D.B., J.Goudeylefevre, B.York, and G.E. Truett (1994) Dietary obesity linked to genetic-loci on chromosome-9 and chromosome-15 in a polygenic mouse model. *Journal Of Clinical Investigation*, **94**: 1410-1416
- Whittaker, J. C., R. Thompson, and P. M. Visscher (1996) On the mapping of QTL by regression of phenotype on marker-type. *Heredity*, **77**: 23-32
- York, B., K. Lei, and D.B. West (1996) Sensitivity to dietary obesity linked to a locus on chromosome- 15 in a cast/C57BL/6J F₂ intercross. *Mammalian Genome*, **7**: 677-681
- Zeng, Z-B. (1994) Precision mapping of Quantitative Trait Loci. *Genetics*, **136**: 1457-1468